Article

Sociological Methods & Research
1-32

DOI: 10.1177/0049124119875961
journals.sagepub.com/home/smr
SAGE

Individual Components of Three Inequality Measures for Analyzing Shapes of Inequality

Tim Futing Liao[1]

Abstract

In common sociological research, income inequality is measured only at the aggregate level. The main purpose of this article is to demonstrate that there is more than meets the eye when inequality is indicated by a single measure. In this article, I introduce an alternative method that evaluates individuals' contributions to inequality as well as the between-group and within-group components of these individual contributions. I first highlight three common inequality measures, the Gini index and two generalized entropy measures—Theil's T and Theil's L indices—by presenting their individual components as a method for evaluating inequality. Five artificial data examples illustrate the use of these individual components first. An empirical analysis of the 2007 and 2017 Current Population Survey data then focuses on the differences in inequality revealed by such individual inequality components between the 2007 and 2017. The individual-level inequality measures can reveal patterns of inequality concealed by single measures at the aggregate level. In particular, the Gini individual measures differentiate cases better than the generalized entropy measures and tend to have smaller standard errors in a regression analysis.

[1] University of Illinois at Urbana–Champaign, IL, USA

Corresponding Author:
Tim Futing Liao, University of Illinois at Urbana–Champaign, 3210 Lincoln Hall, 702 S. Wright Street, Urbana, IL 61801, USA.
Email: tfliao@illinois.edu

Keywords

inequality, Gini index, Theil index, measurement, decomposition

It is a common practice to use an inequality measure such as the Gini index or one of the generalized entropy indices to indicate inequality and to regard a rise in such a measure as a corresponding rise in inequality. Such a usage can be misleading. As Osberg (2017) pointed out insightfully, rather different types or shapes of inequality can give rise to an identical Gini measure. The same can be said about other inequality measures. More generally, inequality may take on various shapes, especially when reflected by structural inequality defined by between-group differences and by within-group variations. This article presents a simple method for decomposing individual-level contributions to inequality that can reveal the underlying shapes and structures of inequality.

In this article, I define "shapes" and "structures" of inequality differently. I use the term "shape" to refer to various distributional forms of inequality, which may not necessarily relate to a social structural foundation. "Structures" of inequality, on the other hand, typically have a social structural base such as whether there is a correspondence between structures of inequality and class or occupational structure (Kenworthy 2007). Insofar as social structure is an important consideration, my application of the proposed method provides insight into structures of inequality.

What sets this article apart from the literature on inequality measurement? There has been a long tradition of decomposing inequality measures. For example, the decomposition of Theil's two indices by population or social groups has been available since Theil's (1967) groundbreaking work. Applications of Theil's decomposition to empirical data are too numerous to list. Attempts to decompose the Gini index by population or social groups have also be made by various scholars over the years (e.g., Dagum 1997; Modalsli 2017; Mussard, Terraza, and Seyte 2003; Yitzhaki 1994). While not decomposition per se, Ceriani and Verme (2015) showed that the Gini index can be viewed as a collection of individual contributions to the overall inequality. In this article, I integrate the two research traditions by presenting a method for the decomposition (by a person's membership in a group) of the individual components of the Gini as well as the two Theil indices, whose individual components have never been treated or analyzed.

In the pages to follow, I first present five artificial data situations to illustrate the limitations of relying on the overall inequality indices alone, before demonstrating a method for measuring inequality at the individual level, that is, the individual contribution and its between-group and within-group components of these inequality measures. I call such an individual contribution to the overall Gini “iGini” and an individual component of the Theil (either the T or the L version, a special case of the generalized entropy) “iTheil” measure. The iGini and iTheil components can be additional useful measures of inequality. Based on the difference between a person and the others in other social groups, the iGini between-group component is in principle consistent with Tilly’s (2003:37) concept of durable categorical inequality, “Durable categorical inequality refers to organized *differences* in advantages by gender, race, nationality, ethnicity, religion, community, and similar classification systems. It occurs when transactions across a categorical boundary (e.g., male-female) (a) regularly yield net advantages to *people* on one side of the boundary and also (b) reproduce the boundary” (emphasis added).

To illustrate the application of individual inequality components to real-world data, I also present an empirical analysis of the March 2007 and 2017 Current Population Survey Annual Social Economic Supplement (CPS ASEC) income data to gain an understanding of the changes in the shapes and structures of income inequality in the United States 10 years apart where the overall income inequality. Finally, a concluding section offers a summary of, and some further thoughts on, these individual-level inequality component measures and their usages in sociological research.

An Illustration With Some Simple Artificial Data

Table 1 presents five artificial data examples, each of which contains two groups of 10 cases each. I prepared the data by following the principle of Osberg’s (1981, 2017) “Adanac” data example, in which the rich and the poor were two discrete classes. The purpose of using the Adanac data example here is different from those of Osberg’s. In this article, I am interested in seeing how a single, overall inequality measure may conceal the kind of shapes and structures of inequality that can only be revealed by the components of individual contributions to inequality that I study here.

Here, the “rich” and the “poor” are defined as two relative categories, with the former referring to those with higher incomes as a group compared with the rest of the population while the latter referring to those with lower incomes as a group compared with the rest of the population. The values

Table 1. Five Artificial Data Examples With Gini = .40.

Data 1		Data 2		Data 3		Data 4		Data 5	
Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
2	2	2	2	3	3	5	5	10	10
2	2	2	2	3	3	5	5	10	10
2	2	2	12	3	3	5	5	10	10
2	2	2	12	3	3	5	5	10	10
2	2	2	12	3	49/3	5	5	10	10
2	2	2	12	3	49/3	5	5	10	10
12	12	2	12	3	49/3	5	30	10	10
12	12	2	12	3	49/3	5	30	10	10
12	12	2	12	3	49/3	5	30	10	90
12	12	2	12	3	49/3	5	30	10	90

Table 2. Summary Inequality Values of the Three Indices Using Table 1.

	Data 1	Data 2	Data 3	Data 4	Data 5
Gini	.4	.4	.4	.4	.4
Theil's T	.3348	.3348	.3389	.3819	.5108
Theil's L	.3819	.3819	.3389	.3348	.3681

reported may be regarded as income in \$10,000. In Aldanac_{60} or Data 1 , the poor population is set at 60 percent that share 20 percent of total income. Data 2 is a set of reordered Data 1 where the two groups no longer have equal distributions of the rich and the poor. In Aldanac_{70} or Data 3, the poor population is set at 70 percent that share 30 percent of total income. By the same logic, in Aldanac_{80} and Aldanac_{90}, the poor constitute 80 percent and 90 percent of the population that share 40 percent and 50 percent of total income, respectively.

Table 2 presents three common inequality measures, the Gini index and two generalized entropy indices when $\alpha = 1$ and when $\alpha = 0$, commonly known as Theil's T and Theil's L measures, respectively. All five data sets produced an identical Gini index of .4 even though they have quite different income distributional shapes. This, of course, is not a new finding: Osberg's (2017) figure 2 shows that one may draw an almost infinite number of shapes that occupy an identical area covered by the observed Lorenz curve. This explains why different shapes of inequality may produce identical Gini values.

Theil's T produces an identical value of .33 for Data 1 and Data 2, a slightly higher value of .34 for Data 3, and two higher values of .38 and .51 for Data 4 and Data5. Theil's L also gives an identical value of .38 for Data 1 and 2, lower values of .34 and .33 for Data 3 and Data 4, respectively, and a slightly increased value of .37 for Data 5.

Do these inequality measures really reflect the shapes and structures of inequality in these five data sets? To obtain a better sense of the shapes of inequality in the data, I show the five data sets in a series of treemaps. Treemaps provide a method for visualizing data by levels and categories (as tree branches) and can be used to compare different sets of data. These graphs, as shown below, will facilitate our understanding of the shapes of inequality in the five artificial data examples.

The size of a rectangle is determined by the proportional value of a data point relative to the others in the same data set. To make the graphs comparable across the data sets, I use the same spectral color scale across the five data sets. That is, the same colors indicate the same value ranges. This way, the income values are comparable within the data sets by the relative size of a rectangle and across them by the color of a rectangle. In addition, the two groups of cases are organized into two blocks of rectangles (or branches in these treemaps). The first two treemaps show that, despite the identical inequality index values, there exists a great deal of structural inequality, as displayed by the difference in the sizes of the branches, with a much higher degree of structural inequality in Data 2 than in Data 1 even though each of the three inequality indices records an identical value of overall inequality for the two data sets. Two general patterns of inequality are observed from Data 2 to Data 5. First, structural inequality decreases slightly from Data 2 to Data 3 to Data 4 to Data 5. On the other hand, the inequality between certain cases increases from Data 2 to Data 5, with the changing shades of color from light blue to azure blue to lime green to maroon red.

This compensation of between-group inequality by within-group inequality clearly is the reason that produces an identical Gini coefficient and relatively similar values for the other two indices for all five data situations. Because Theil's T is more sensitive to the upper end of a distribution (in this case, the two maroon-colored cases in Data 5), it produces a highest value there of the five data sets. On the other hand, Theil's L is more sensitive to the lower end of a distribution, hence having higher values of the first two data situations, and the two entropy measures mirror each other in terms of their low to high values. However, because none of the three inequality

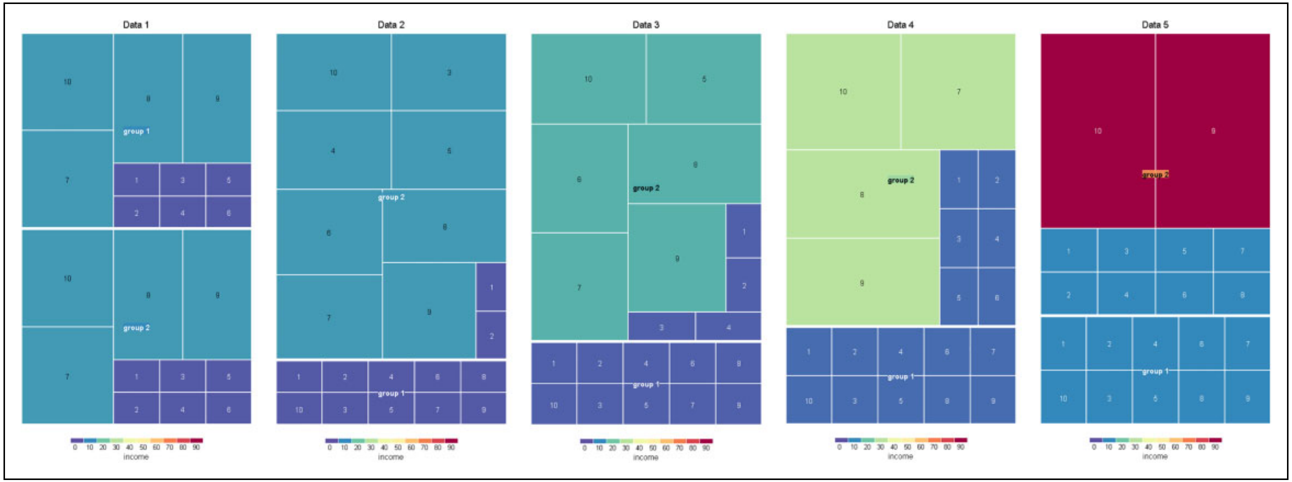

Figure 1. Structures of inequality portrayed by treemaps of the five data sets of Table 1.

measures are able to capture structural inequality in the five data sets, we postpone the discussion of how to evaluate structural inequality and more generally shapes of inequality until the individual components of these measures have been presented.

Individual Inequality Components

Individual Gini Components

Various computational formulas exist for implementing the Gini coefficient. Perhaps the most revealing one is that used by Dagum (1997, 1998) and Mussard et al. (2003):

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\bar{x}}, \tag{1}$$

where x_i and x_j represent the *i*th and the *j*th person's income (or, more generally, an attribute) in the overall sample or population of size n. The denominator can be replaced by $2n$ times the sum of x_i because $n \times \mu$ the sample mean equals the sum of x_i. Equation (1) states that G is a scaled measure of all possible pairwise contrasts or differences in absolute value, normed to lie in the range of [0, 1]. All other computational formulas in the literature produce the same results as equation (1). Equation (1) also implies that, by expressing what comes after the first summation sign as a single variable, G becomes the sum of individual contributions g_i or iGini, where $g_1 + g_2 + \cdots + g_n = G$:

$$g_i = \frac{\sum_{j=1}^{n}|x_i - x_j|}{2n^2\bar{x}}. \tag{2}$$

The iGini component or g_i, or simply g, can be interpreted as a person's scaled difference from all the other persons in the sample. Thus, the iGini component measures individual diversity and, as an individual measure of diversity, satisfies the six desirable properties of an inequality measure: continuity, additivity, linear homogeneity, translation invariance, symmetry, and anonymity (for further details, see Ceriani and Verme 2015).

Typically, the computation of equations (1) or (2) takes place in the order of the data, often arranged by case ID or i. Rearranging the data order of the pairwise computations in equations (1) or (2) in any way we choose does not change the mathematical results. We can, however, by rearranging the data in such a way that we separately compute the pairwise differences between members across different social groups and the pairwise differences between members within the same groups.

It follows that when the population is classified into K number of groups for $k = 1$ to K, for the kth group, person ith's iGini or g_{ik} can be regarded as having two components, one recording the difference between person i in group k and everyone else (person j) in a different group h for $h \neq k$ and the other between person i and everyone else (person j) in the same group k:

$$g_{ik} = \frac{\sum_{h=1}^{K}\sum_{j=1}^{n_h}|x_{ik} - x_{jh}|}{2n^2\bar{x}} + \frac{\sum_{j=1}^{n_k}|x_{ik} - x_{jk}|}{2n^2\bar{x}} = g_{bik} + g_{wik}, \quad (3)$$

where n_h indicates the size of group h and n_k is the size of group k where again $h \neq k$. In this formulation, the overall amount of inequality (G) consists of individual contributions (g_{ik}) expressed as the additive between-group and within-group components g_{bik} and g_{wik} for person i in group k. In other words, each case is contrasted against all members of the other groups and all members of the same group. For the overall sample, because we already use lowercase letter g (instead of G) to indicate individual-level measurement and because the group designation k is understood, we drop the subscripts i and k and obtain g_b and g_w, that is, iGini between and iGini within, respectively, for each individual in the overall sample or population.

We may simply sum over the sample space the respective components g_b and g_w to obtain the overall between and within components or, alternatively, we may derive these components separately as shown below if g_b and g_w are of no interest. In this case, we rearrange the overall sample of n individuals into K groups. Both n_h and n_k are defined as earlier. Without complication, the n^2 number of contrasts in equation (1) can be rearranged into two sets of contrasts or absolute differences between those in the same groups (i.e., group k) and those in different groups (i.e., groups h and k). Following this rearrangement, we obtain

$$G = \frac{\sum_{k=1}^{K}\sum_{h=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_h}|x_{ik} - x_{jh}|}{2n^2\bar{x}} + \frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}|x_{ik} - x_{jk}|}{2n^2\bar{x}}, \quad (4)$$

where the first term on the right-hand side defines the between-group component of the Gini coefficient and the second term, the within-group component. We call the first term G_b and the second term G_w, with $G = G_b + G_w$. The decomposition of equation (4) is also used by Modalsli (2017) for decomposing global inequality. We may instead sum the respective two terms in equation (3) over the sample space to obtain these two components.

Individual Theil Components

The Theil index of inequality, originally proposed by Theil (1967), is another widely applied inequality measure after the Gini coefficient. Unlike the Gini index, the two versions of the Theil index belong to the family of generalized entropy measures. When $\alpha = 1$, a parameter for the generalized entropy family regulating the weight given to distances between cases in different parts of a distribution, we have the typical Theil's measure or Theil's T, also known as Theil's first measure. When $\alpha = 0$, we have Theil's L, also known as Theil's second measure or the mean logarithm deviation.

The total amount of inequality measured by Theil's T or first measure is:

$$\mathrm{T}_T = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{\bar{x}}\ln\frac{x_i}{\bar{x}}, \tag{5}$$

where x_i stands for the income of individual i, $\bar{x}$ the overall mean income, and n the sample size. Note that Theil's T can be expressed first in terms of income share to population share ratio and then can be simplified to equation (5). Using the same notation, Theil's T total in equation (5) can be decomposed into a between-group and a within-group component (Akita 2003; Liao 2016a):

$$\mathrm{T}_T = \mathrm{T}_B + \mathrm{T}_W = \sum_{k=1}^{K} y_k \ln\frac{\bar{x}_k}{\bar{x}} + \sum_{k=1}^{K} y_k \sum_{n=1}^{n_k} y_{ik} \ln\frac{x_{ik}}{\bar{x}_k}, \tag{6}$$

Where y_k is the kth group's income share expressed as a proportion of the sample or population total income, $\bar{x}_k$ is the mean income of group k, y_{ik} represents the income share of the ith individual in the kth group, and x_{ik} stands for the ith individual's income in group k. Theil's L total or the second measure (L_T) can be expressed similarly as

$$\mathrm{L}_T = \mathrm{L}_B + \mathrm{L}_W = \sum_{k=1}^{K} N_k \ln\frac{\bar{x}}{\bar{x}_k} + \sum_{k=1}^{K} N_k \sum_{n=1}^{n_k} N_{ik} \ln\frac{\bar{x}_k}{x_{ik}}, \tag{7}$$

where N_k and N_{ik} are the kth group's group size proportion of the overall sample and the ith case's proportion of the kth group, respectively. For obtaining Theil's individual counterpart to the iGini measure, we simply extend the decomposition of equations (6) or (7) to the individual level without summing over the groups but summing over within group k for the within-group component. For the between-group component, we prorate it for each individual by its individual income share in group k. With such an extension, equation (6) for the individual component of $t_{\mathrm{T}i}$ becomes

$$t_{\mathrm{T}i} = t_{bi} + t_{wi} = y_{ik} \sum_{k=1}^{K} y_k \ln \frac{\bar{x}_k}{\bar{x}} + y_{ik} y_k \ln \frac{x_{ik}}{\bar{x}_k}. \tag{8}$$

Equation (7) can be similarly extended to obtain the individual components of $l_{\mathrm{T}i}$:

$$l_{\mathrm{T}i} = l_{bi} + l_{wi} = N_{ik} \sum_{k=1}^{K} N_k \ln \frac{\bar{x}}{\bar{x}_k} + N_{ik} N_k \ln \frac{\bar{x}_k}{x_{ik}} \ , \tag{9}$$

where N_{ik} is the proportion the ith case out of the kth group and N_k is the proportion the kth group out of the population, as previously defined.

Let us drop subscript i as we did for the iGini measures for the sake of brevity and use lowercase letters t and l for Theil's T and Theil's L measures at the individual level. Then, the iTheil components, that is, t_b, t_w, l_b, l_w, and the iGini measures of g_b and g_w can be computed with sampling weights, using the procedure described in Liao (2016b). Finally, if one wishes, one can sum up all these iGini and iTheil measures by group to obtain a group's contribution to the overall Gini/Theil value. However, a word or caution for using them is in order because the summed-up value can be a function of group size. To properly assess how much a group contributes to the overall inequality, we can compute a hypothetical overall measure, be it the Gini, Theil's T, or Theil's L, by dividing the sum of a kth group's iGini, iTheilT (i.e., the individual component of Theil's T), or iTheilL (i.e., the individual component of Theil's L) by its group size to obtain the group average iGini, iTheilT, or iTheilL and multiply it by the overall sample size.

$$\mathrm{T}_k = t_{\mathrm{T}k} \frac{n}{n_k}. \tag{10}$$

This hypothetical measure in equation (10) is expressed in terms of Theil's T for group k, where $t_{\mathrm{T}k}$ is the iTheilT measure for group k and n_k is the group size. It is hypothetical because the computation is not based on the observed population but on the assumption that the population of the same size as observed consisted entirely of members of this particular group. When compared with the original Gini, Theil's T, or Theil's L, the measure of equation (10) can reveal how much a group contributes to inequality for this hypothetical population. The word “hypothetical,” when it is applied to a measure, is used in the same way later in this article. The same operation can be performed on the within-group component as well, arriving at a hypothetical situation where all within-group inequality is computed with the members of the group under consideration. This operation makes little sense for computing a hypothetical between-group value because a hypothetical population cannot have every member belonging to the same group yet still has a between-group component. In the empirical application of the method to the CPS data, we will see how the procedure works.

A Reanalysis of the Five Artificial Data Examples

To gain a sense of how these three individual inequality measures perform for evaluating the five artificial data examples, I present their between-group and within-group components for each data set by group in Table 3. To have a concise presentation, repeated values are indicated by vertical ellipses.

As we recall, Data 1 and Data 2 are identical except for the structural difference between the two groups (no structural difference in Data 1 and a sizable structural difference in Data 2). In addition, from Data 2 to Data 5, structural inequality decreases slightly while inequality between cases within groups increases. Thus, for assessing individual inequality components, we must keep these patterns in mind. For Data 1, the iGini$_b$ and iGini$_w$ (or g_b and g_w) components for the same person are identical, due to the identical income distribution in the two groups (cf. Table 1). The iTheilT$_b$ and iTheilL$_b$ (t_b and l_b) components are zero for all cases in both groups, reflecting the nil difference between the two groups. In this data situation of absolute identical income distributional shapes between groups, iGini$_b$ has material values, due to its definition of pairwise comparison between a group’s member and the members of all the other group(s). The within components, however, are consistently distributed across the three measures. That is, a high value in iGini$_b$ corresponds to a high value in iTheilT$_b$ and a low value in iTheilL$_b$, while a low value in iGini$_b$ corresponds to a low value in iTheil$_b$ and a high value in iTheilL$_b$.

Judged by the size of a rectangle relative to the sizes the other rectangles in the same group and in the other group in Figure 1, cases 7–10 in each of the two groups in Data 1 should have the highest values for both iGini and iTheilT (or lowest for iTheilL) for the within components (or t_w and l_w). For the between components, iGini satisfies this condition, but the two iTheil components do not. The pattern of a larger relative rectangle size, a greater

Table 3. iGini and iTheil Components for Data 1–5.

(A) iGini and iTheil Components for Data 1

Case	Group	iGini	iTheilT	iTheilL
			Between components	
1	1	.0083	0	0
⋮	⋮	⋮	⋮	⋮
6	1	.0083	0	0
7	1	.0125	0	0
⋮	⋮	⋮	⋮	⋮
10	1	.0125	0	0
1	2	.0083	0	0
⋮	⋮	⋮	⋮	⋮
6	2	.0083	0	0
7	2	.0125	0	0
⋮	⋮	⋮	⋮	⋮
10	2	.0125	0	0
			Within components	
1	1	.0083	−.0183	.0549
⋮	⋮	⋮	⋮	⋮
6	1	.0083	−.0183	.0549
7	1	.0125	.0693	−.0347
⋮	⋮	⋮	⋮	⋮
10	1	.0125	.0693	−.0347
1	2	.0083	−.0183	.0549
⋮	⋮	⋮	⋮	⋮
6	2	.0083	−.0183	.0549
7	2	.0125	.0693	−.0347
⋮	⋮	⋮	⋮	⋮
10	2	.0125	.0693	−.0347

(B) iGini and iTheil Components for Data 2

			Between components	
1	1	.0167	−.0183	.0549
⋮	⋮	⋮	⋮	⋮
10	1	.0167	−.0183	.0549
1	2	0	.0085	−.0255
2	2	0	.0085	−.0255
3	2	.0208	.0511	−.0255
⋮	⋮	⋮	⋮	⋮
10	2	.0208	.0511	−.0255

(continued)

Table 3. (continued)

(B) iGini and iTheil Components for Data 2

			Within components	
1	1	0	0	0
⋮	⋮	⋮	⋮	⋮
10	1	0	0	0
1	2	.0167	−.0268	.0805
2	2	.0167	−.0268	.0805
3	2	.0042	.0182	−.0091
⋮	⋮	⋮	⋮	⋮
10	2	.0042	.0182	−.0091

(C) iGini and iTheil Components for Data 3

			Between components	
1	1	.0143	−.0182	.0424
⋮	⋮	⋮	⋮	⋮
10	1	.0143	−.0182	.0424
1	2	0	.0097	−.0226
⋮	⋮	⋮	⋮	⋮
4	2	0	.0097	−.0226
5	2	.0238	.0527	−.0226
⋮	⋮	⋮	⋮	⋮
10	2	.0238	.0527	−.0226
			Within components	
1	1	0	0	0
⋮	⋮	⋮	⋮	⋮
10	1	0	0	0
1	2	.0143	−.0278	.065
⋮	⋮	⋮	⋮	⋮
4	2	.0143	−.0278	.065
5	2	.0095	.0461	−.0198
⋮	⋮	⋮	⋮	⋮
10	2	.0095	.0461	−.0198

(D) iGini and iTheil Components for Data 4

			Between components	
1	1	.0125	−.0173	.0347
⋮	⋮	⋮	⋮	⋮
10	1	.0125	−.0173	.0347

(continued)

Table 3. (continued)

(D) iGini and iTheil Components for Data 4

1	2	0	.0101	−.0203
⋮	⋮	⋮	⋮	⋮
6	2	0	.0101	−.0203
7	2	.0312	.0608	−.0203
⋮	⋮	⋮	⋮	⋮
10	2	.0312	.0608	−.0203
			Within components	
1	1	0	0	0
⋮	⋮	⋮	⋮	⋮
10	1	0	0	0
1	2	.0125	−.0275	.0549
⋮	⋮	⋮	⋮	⋮
6	2	.0125	−.0275	.0549
7	2	.0188	.104	−.0347
⋮	⋮	⋮	⋮	⋮
10	2	.0188	.104	−.0347

(E) iGini and iTheil Components for Data 5

			Between components	
1	1	.0111	−.0163	.0294
⋮	⋮	⋮	⋮	⋮
10	1	.0111	−.0163	.0294
1	2	0	.0102	−.0184
⋮	⋮	⋮	⋮	⋮
8	2	0	.0102	−.0184
9	2	.0556	.0919	−.0184
10	2	.0556	.0919	−.0184
			Within components	
1	1	0	0	0
⋮	⋮	⋮	⋮	⋮
10	1	0	0	0
1	2	.0111	−.0265	.0478
⋮	⋮	⋮	⋮	⋮
8	2	.0111	−.0265	.0478
9	2	.0444	.3104	−.0621
10	2	.0444	.3104	−.0621

iGini or iTheilT component value, and a smaller iTheilL component value can be observed from Data 2 to Data 5. Take, for example, cases 3–10 in Group 2 in Data 2, cases 6–10 in Group 2 in Data 3, cases 7–10 in Group 2 in Data 4, and cases 9 and 10 in Group 2 in Data 5 all follow this pattern. The only exception to this rule is exhibited by the iTheilL between component, which would have a similar low value for the other cases in the same group that do not share the same relative rectangle size.

In sum, in spite of their general similar patterns of behavior, the between- and within-group components of the three inequality measures have their strengths and weaknesses. In exactly identical distributions between groups, an extremely rare case as in Data 1, iGini_b shows material values while the other two have zeros. This is because individual values in one group are compared with their individual counterparts in the other group in the iGini_b computation, while for iTheilT_b or iTheilL_b, they only get prorated an average difference between group, which is zero in this case. Furthermore, iGini_b does reflect relative sizes correctly for those cases with larger values in such situations. The iTheilL between-group component fails to distinguish those cases with different values, as noted earlier. Because an identical between-group distribution is rare for most practical data situations, both iGini and iTheilT should perform equally well. We next examine an application of these individual component measures of inequality to a real-world data set.

An Analysis of the March 2007 and 2017 CPS ASEC Data

For an empirical example, I analyze below the 2007 and 2017 March CPS ASEC data with sampling weights. Gender and race form the basic foundation of occupational segregation—and wage inequality—in the United States (del Río and Alonso-Villar 2015). To assess a possible rise in wage inequality after the 2008 recession, I selected those men and women in the labor force in the four racial groups of whites, blacks, Hispanics, and Asian Americans, with a sample of 88,991 for 2007 and a sample of 77,969 for 2017. Zero income cases are excluded because zero is undefined for the Theil L index even though they can be included not just for computing the Gini index but also for calculating the Theil T index, contrary to a common misbelief (for further details, see Liao 2016b). Furthermore, by focusing on those in the labor force only, the analysis is not subject to the problem of differential shares of working age populations that Firebaugh and Goesling (2004) discussed.

Note that the current empirical analysis differs from many earlier reports of income inequality in the United States in an important way. Earlier research typically reports household income inequality, while the unit of analysis here is the individual. In 2007 and in 2017, income inequality measured by the Gini coefficient is about .47 (or .465 and .472, more exactly), with a small 1.5 percent increase over the 10 years span. Theil's T and Theil's L also show an increase over the 10 years span, from .336 to .355 (5.6 percent) and from .454 to .461 (1.5 percent), respectively. To display how the three inequality measures perform after conducting a component analysis with sampling weight, I present iGini, iTheilT, and iTheilL density distributions in Figure 2. It appears that, judged by the first two measures, there has been an increase in the spread of iGini or iTheilT values over the 10 years span (2007: blue; 2017: red), thereby indicating an increase in income inequality. The increase in spread is not obvious, as judged by the third measure, though it can be deduced from the lower peak of the red curve (the longer red tail can be misleading because the red curve was plotted after the blue in order to have the correct *Y*-axis range for the blue curve).

Because of the long tails of the density distributions in Figure 2, I chose to use beanplots to present the individual component analysis. The beanplot combines the violin plot, a variation of the boxplot, with the scatterplot by showing outlying values as well as inlying density. I present these plots by contrasting the individual inequality components from the two different years, separately for the eight race and gender groups: Asian women, Asian men, black women, black men, Hispanic women, Hispanic men, white women, and white men. The eight groups are the foundation for the inequality decomposition. Figure 3 shows the iGini, iGini between, and iGini within values by year and by race and gender groups, with the darker shade indicating 2007 and lighter shade indicating 2017.

From the first panel of Figure 3, it appears that all groups have experienced an increase in inequality, except Hispanics for whom the averages represented by the middle bars do not differ much, though for Hispanic men, the spread is a bit wider in the later time. The second and the third panels show the between and the within components of the iGini measure. Comparing the beanplots in these panels, we can see that for whites, the increase in inequality over the 10 years span is primarily driven by between-group inequality, while for Hispanics, within-group inequality actually increased even though the overall iGini measure does not display a clear increase in the first panel. For Asian Americans, the increase in within-group inequality is greater than that in between-group inequality, and the same can be said about African Americans. Comparing across racial groups, whites have the highest

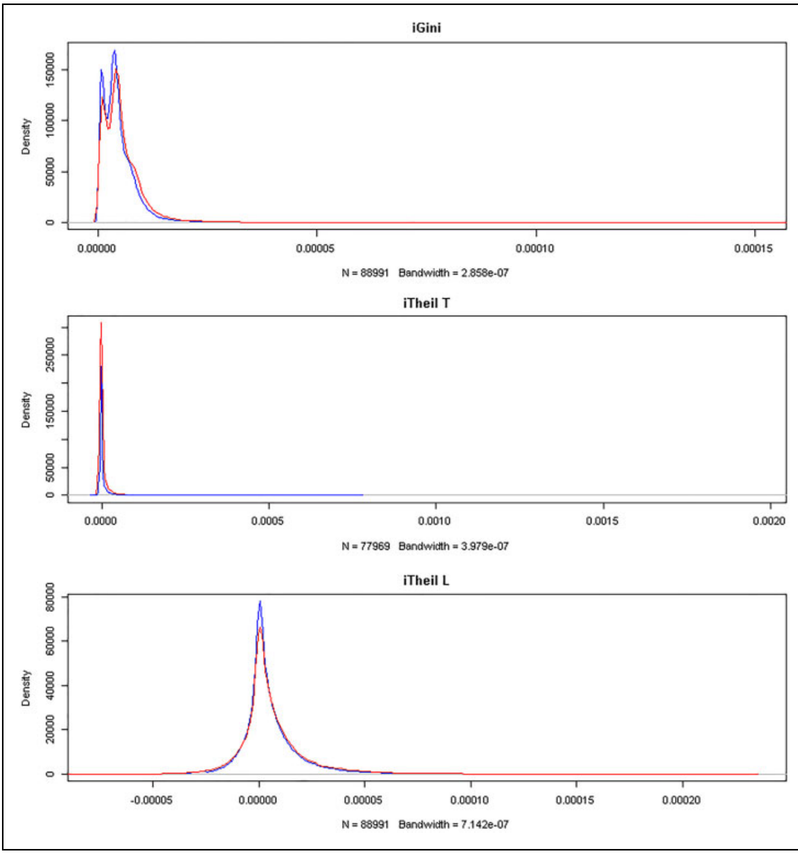

Figure 2. Density distributions of iGini, iTheil T, and iTheil L totals of income inequality in 2007 and 2017. Gini = .465 and .472; Theil's T = .417 and .440; Theil's L = .454 and .461 (N_{2007} = 98,816 and N_{2017} = 86,965).

within-group inequality of all while Asian American and Hispanic women have the lowest. We next examine the component analysis using the iTheilT and iTheilL measures in Figures 4 and 5.

Because the distributions of the iTheilT values, whether they represent the total or the between/the within component, are so concentrated, it is difficult to make the same observations as we did with the iGini components other than relying on the outlying cases. On the other hand, we probably should not draw any conclusions based on just a few extreme outliers. Although the

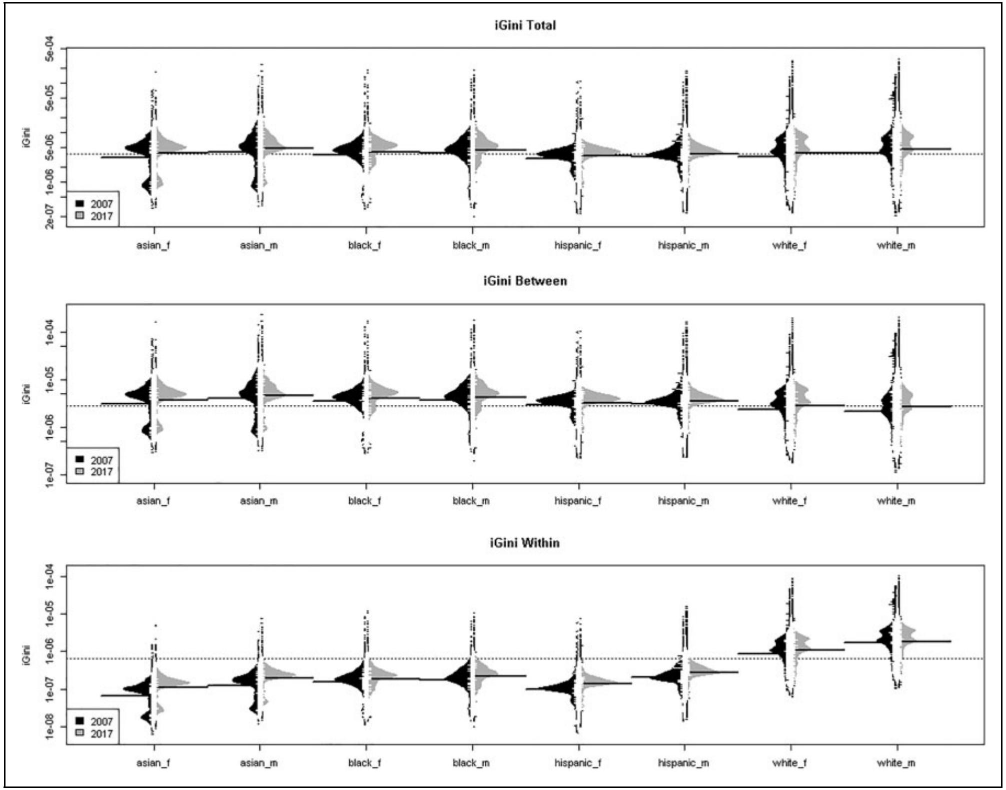

Figure 3. Beanplots of iGini total, iGini between, and iGini within by race and gender in 2007 and 2017. Gini = .465 and .472 (N_{2007} = 88,991 and N_{2017} = 77,969; subsample sizes: 2007: Asian f = 2,083, Asian m = 2,185, black f = 5,290, black m = 4,125, Hispanic f = 5,582, Hispanic m = 7,702, white f = 29,818, white m = 32,206; 2017: Asian f = 2,538, Asian m = 2,729, black f = 4,876, black m = 4,034, Hispanic f = 6,352, Hispanic m = 7,731, white f = 23,806, white m = 25,903).

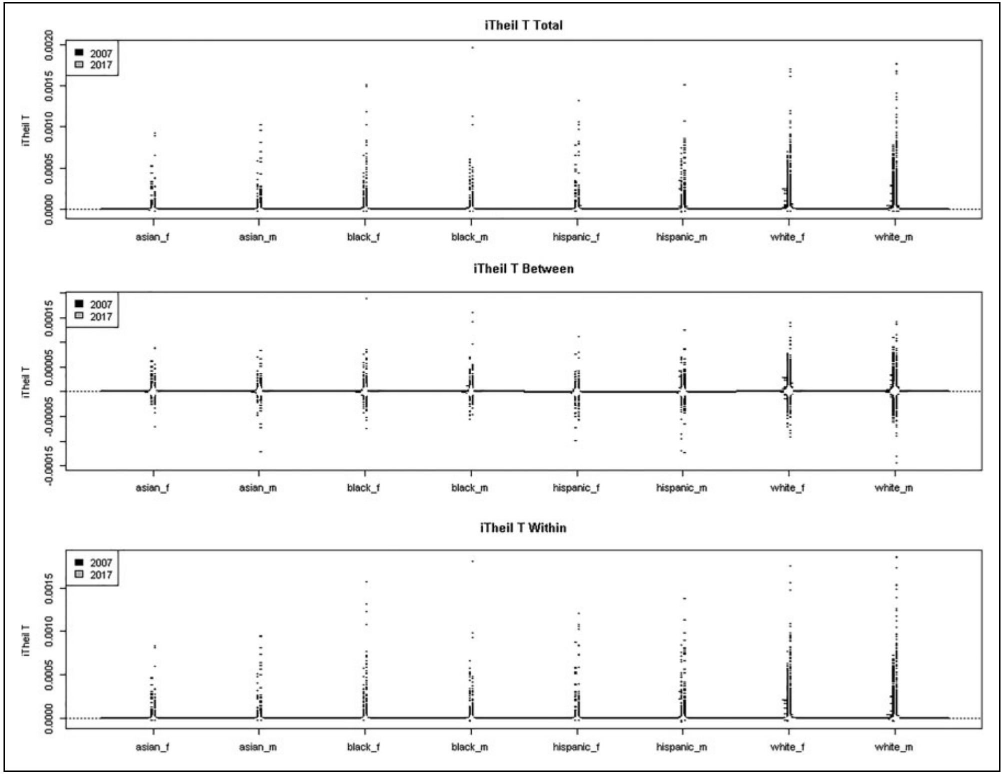

Figure 4. Beanplots of iTheil T total, iTheil T between, and iTheil T within by race and gender in 2007 and 2017. Theil's T = .418 and .440 ($N_{2007} = 88{,}991$ and $N_{2017} = 77{,}969$; subsample sizes: 2007: Asian f = 2,083, Asian m = 2,185, black f = 5,290, black m = 4,125, Hispanic f = 5,582, Hispanic m = 7,702, white f = 29,818, white m = 32,206; 2017: Asian f = 2,538, Asian m = 2,729, black f = 4,876, black m = 4,034, Hispanic f = 6,352, Hispanic m = 7,731, white f = 23,806, white m = 25,903).

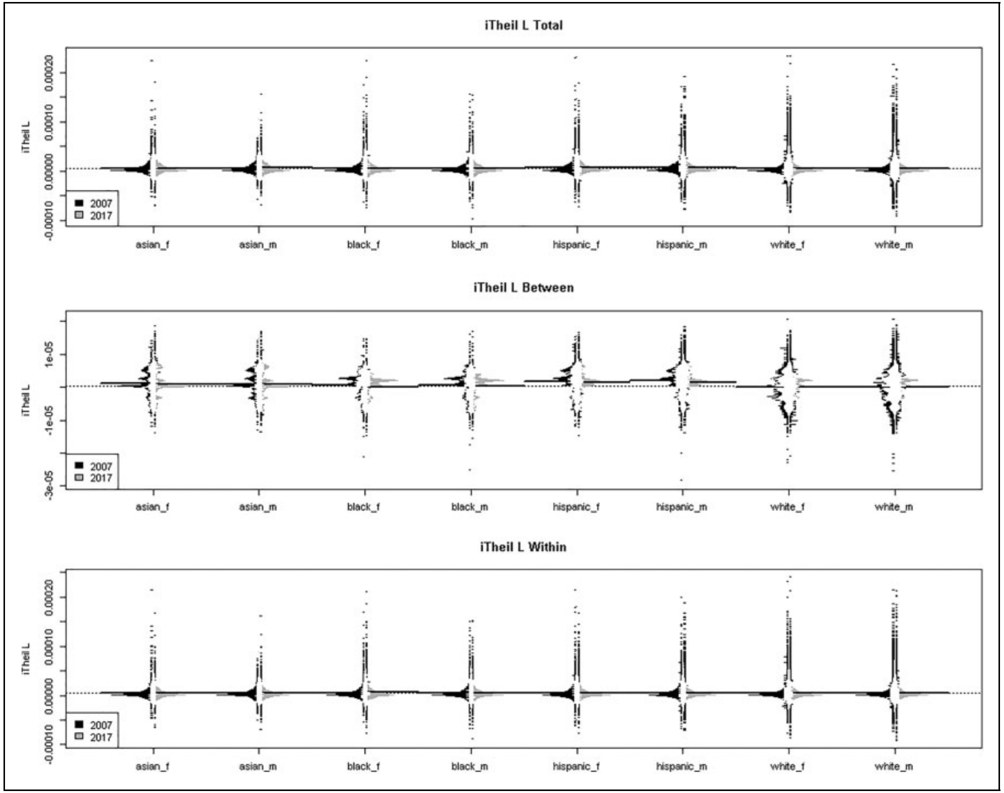

Figure 5. Beanplots of iTheil L total, iTheil L between, and iTheil L within by race and gender in 2007 and 2017. Theil's L = .454 and .461 (N_{2007} = 88,991 and N_{2017} = 77,969; subsample sizes: 2007: Asian f = 2,083, Asian m = 2,185, black f = 5,290, black m = 4,125, Hispanic f = 5,582, Hispanic m = 7,702, white f = 29,818, white m = 32,206; 2017: Asian f = 2,538, Asian m = 2,729, black fem = 4,876, black m = 4,034, Hispanic f = 6,352, Hispanic m = 7,731, white f = 23,806, white m = 25,903).

iTheilL components are a little less concentrated than the iTheilT component values, the same kind of observations we made about Figure 3 cannot be obtained unless we rely on the tails. For example, both Asian American men and women have much longer tails in 2017 than in 2007 (all four tails in the between components and three tails in the within components), and Hispanic American men and women exhibit five longer tails (of eight in 2017) than in 2007. Note, however, for iTheilL component values, the outlying values corresponding to the high values of iGini and iTheilT are located in the lower end instead of the higher end.

Finally, I provide an assessment of the race and gender contributions to income inequality in 2007 and 2017 by partitioning the overall Gini, Theil's T, and Theil's L values into group-specific contributions. As previously stated, such group-specific contributions are a function of group size. Therefore, I summed the iGini, iTheilT, and iTheilL components by groups and also used a version of equation (10) to compute the hypothetical Gini, Theil's T, and Theil's L values by race, gender, and year (Table 4).

Table 4 has six columns, with the first three containing statistics for 2007 and the next three for 2017. The column with the heading Part contains group-specific partitions of, or contributions to, the overall inequality index value, and the column with the heading Total reports the hypothetical inequality values using equation (10). Group sizes are also reported for the two years.

It is obvious that the group-specific partitioned values of income inequality measures in the first and the fourth columns are a function of group size, with the larger groups having the higher values. For a given year, these group-specific contributions sum up to the overall inequality value of the year given in the row below the name of an index for each of the three indices. The hypothetical values computed with equation (10) show a different picture. Before the 2008 Great Recession, Asian men contributed the most to the overall income inequality in terms of average contributions, followed by white men, according to the Gini decomposition, with Hispanics contributing the least. According to Theil's T, white females contributed the most on average, followed by white men, with the two Hispanic groups again contributing the least. According to Theil's L, however, the two Hispanic groups contributed the most on average, while the two white groups contributed the least. By comparing the total hypothetical values with the observed inequality in the headings (i.e., .465, .336, and .454) for 2007, we can gain a sense of how the level of inequality would look if the population is entirely composed of a particular race/gender group.

Table 4. Gini, and Theil's T, and Theil's L Indices by Race, Gender, and Year.

	2007			2017		
Group	Part	Total	*N*	Part	Total	*N*
			Gini			
		.465			.472	
Asian female	.010	.415	2,083	.014	.419	2,538
Asian male	.014	.587	2,185	.020	.564	2,729
Black female	.024	.408	5,290	.026	.413	4,876
Black male	.022	.466	4,125	.025	.480	4,034
Hispanic female	.020	.323	5,582	.026	.321	6,352
Hispanic male	.033	.376	7,702	.038	.382	7,731
White female	.138	.412	29,818	.136	.444	23,806
White male	.204	.564	32,206	.188	.567	25,903
			Theil's T			
		.336			.336	
Asian female	.008	.332	2,083	.012	.379	2,538
Asian male	.008	.331	2,185	.013	.382	2,729
Black female	.018	.297	5,290	.034	.539	4,876
Black male	.014	.313	4,125	.021	.399	4,034
Hispanic female	.010	.160	5,582	.024	.296	6,352
Hispanic male	.019	.219	7,702	.030	.300	7,731
White female	.167	.499	29,818	.147	.482	23,806
White male	.173	.479	32,206	.159	.478	25,903
			Theil's L			
		.454			.461	
Asian female	.011	.474	2,083	.015	.465	2,538
Asian male	.011	.454	2,185	.017	.493	2,729
Black female	.027	.458	5,290	.030	.480	4,876
Black male	.022	.466	4,125	.024	.466	4,034
Hispanic female	.040	.642	5,582	.044	.535	6,352
Hispanic male	.055	.631	7,702	.053	.530	7,731
White female	.140	.417	29,818	.130	.426	23,806
White male	.149	.411	32,206	.149	.447	25,903

The 2017 values reported in the Total column can be used in two ways. They can be used to compare across groups, as we did earlier for 2007, or they can be compared with the 2007 figures. We will focus on the latter interpretation here. Using the hypothetical overall inequality computation of equation (10), we see that income inequality increased for all race/gender groups over the 10 years span, except for Hispanic women according to at least two of the three inequality measures. The inequality measure used may yield different results. Here, the knowledge of Theil's L was more sensitive to the lower end of the distribution, while Theil's T, the upper, should be taken into consideration. It can be useful to see whether such an increase or decrease is due to changes of within-group inequality, which I report in Table 5.

According to either Theil's T or Theil's L, African American women experienced the greatest average increase in within-group inequality over the 10 years span. The white groups, in comparison, increased less in within-group inequality. In terms of year-specific evaluations, the two white groups show mostly higher inequality values than the observed within-group totals using the average interpretation according to the Gini and Theil's T measures (except for white men according to the Gini). I should emphasize that the hypothetical overall (or overall within group) inequality relies on an average interpretation. If one is indeed interested in how much a race/gender group contributes to inequality overall as observed, then the partitioned values in the first and fourth columns can and should be used instead. It is just that when using these partitioned values, researchers should bear in mind that we are not speaking of greater inequality but greater overall shares of inequality due to more people contributing to inequality.

We must also emphasize the difference between the meaning of the iGini between-group component and that of the typical between-group component for either the conventional Gini decomposition with a residual component (not treated here) or the conventional generalized entropy decomposition such as those of the Theil indices. Those other decompositions give an individually proportionated between-group average difference, while for the iGini, between-group component is based on the pairwise contrasts of *all members* from different groups. Therefore, the between-group component here will have a higher value. Returning to the overall assessment of income inequality in 2007 versus 2017, we find that, despite an almost identical overall Gini value, income inequality increased both between race–gender groups and within them according to the iGini components as shown in Figure 3 and Table 4. Such a finding reveals two important features of inequality in America: that the very essence of income inequality in this country is driven by the intersection of gender and race accounts for the bulk of inequality in this land and that the 2008 recession deepened income inequality in this country across all race and gender groups, albeit to a different degree.

Table 5. Gini, and Theil's T, and Theil's L Within-Group Components by Race, Gender, and Year.

	2007			2017		
Group	Part	Total	*N*	Part	Total	*N*
			Gini			
		.129			.117	
Asian female	.000	.009	2,083	.000	.012	2,538
Asian male	.000	.016	2,185	.001	.021	2,729
Black female	.001	.019	5,290	.001	.020	4,876
Black male	.001	.021	4,125	.001	.025	4,034
Hispanic female	.001	.011	5,582	.001	.014	6,352
Hispanic male	.002	.024	7,702	.003	.029	7,731
White female	.040	.118	29,818	.038	.125	23,806
White male	.084	.232	32,206	.071	.214	25,903
			Theil's T			
		.381			.409	
Asian female	.008	.335	2,083	.012	.365	2,538
Asian male	.008	.318	2,185	.013	.383	2,729
Black female	.017	.283	5,290	.031	.488	4,876
Black male	.015	.320	4,125	.019	.362	4,034
Hispanic female	.015	.245	5,582	.029	.354	6,352
Hispanic male	.026	.305	7,702	.035	.352	7,731
White female	.142	.425	29,818	.128	.420	23,806
White male	.150	.414	32,206	.143	.430	25,903
			Theil's L			
		.418			.430	
Asian female	.009	.384	2,083	.013	.388	2,538
Asian male	.01	.389	2,185	.015	.419	2,729
Black female	.024	.409	5,290	.030	.474	4,876
Black male	.019	.407	4,125	.023	.443	4,034
Hispanic female	.030	.481	5,582	.034	.414	6,352
Hispanic male	.040	.463	7,702	.040	.408	7,731
White female	.139	.416	29,818	.130	.427	23,806
White male	.147	.405	32,206	.146	.438	25,903

To compare the three individual-level inequality measures between the two years further, I estimated a series of regression models to illustrate their differences. To continue the assessment of the race and gender groups'

Table 6. Descriptive Statistics of the Current Population Survey Data for 2007 and 2017.

	2007		2017	
	Mean	*SD*	Mean	*SD*
Income ($)	42,335.371	51,106.58	53,880.367	71,648.123
iGini total	5.225	6.883	6.055	8.765
iGini between	3.776	4.888	4.551	6.487
iGini within	1.450	2.412	1.504	2.751
iTheilT total	4.693	33.721	5.645	47.89
iTheilT between	0.407	5.444	0.393	5.828
iTheilT within	4.286	31.698	5.252	46.320
iTheilL total	5.105	14.420	5.916	16.331
iTheilL between	0.410	3.527	0.402	3.671
iTheilL within	4.695	14.154	5.514	16.02
	%	*N*	%	*N*
Less than high school	12.173	10,833	8.751	6,823
High school	28.988	25,797	25.918	20,208
Some college	28.298	25,183	28.542	22,254
College graduation	20.068	17,859	22.994	17,928
Graduate education	10.472	9,319	13.795	10,756

Note. $N = 88{,}991$ (2007) and $N = 77{,}969$ (2017). All iGini, iTheilT, and iTheilL as well as their components are multiplied by a constant of 1,000,000 to make them displayable with the other statistics.

contributions to inequality, the explanatory variable of interest in the regressions is education, which has five categories, with less than high school education as the reference category. Education is important to examine because of the sociological question of degree to which income inequality can be explained by education (Breen and Chung 2015). In the regressions to follow, the outcome variables represent individual contributions to income inequality and such contributions are specific to between or within race–gender group considerations. Prior to present the regression results, I report some relevant descriptive statistics in Table 6.

All the statistics reported in the table are for the variables used in the regressions, but income is listed as a point of reference because it is the base for computing the iGini, iTheilT, and iTheilL measures. The means and standard deviations (*SD*s) are provided for the nine outcome variables plus income while subsample percentages and numbers are given for the educational categories (all statistics unweighted, with the sample limited to those in the labor force with nonzero income).

Even though the average income increased over the 10 years span, the amount of inequality increased even more, as evidenced by the greater *SD* values for all the iGini, iTheilT, and iTheilL totals and their component measures (as well as the *SD* for income) in the later time. Where does this increase in inequality come from? It appears that the increase in the *SD* comes from both the between and the within gender–race groups, as shown by the *SD*s for these two components compared over time. We next examine the relation between education and inequality.

Altogether 18 linear regressions were fitted with sampling weights, with 9 to the 2007 CPS ASEC data and 9 to the 2017 CPS ASEC data. For each year, three of the nine regressions have iGini total, iTheilT total, and iTheilL total, as the dependent variable, respectively. The next three have their between components as the outcome variable, respectively, while the final three have their within components as the outcome variable, respectively. All the dependent variables were multiplied by a constant of 1,000,000 to obtain estimates in an easily displayable range. Coefficient estimates with their 95 percent conference intervals from the regressions for 2007 and 2017 are presented in Tables 7 and 8, respectively.

We can make several general observations about the properties of these individual inequality measures and their educational effects. First and foremost, education explains a bit over 5 percent income inequality in 2007 but only 3 percent in 2017, judged by the two R^2 statistics from the two regressions of iGini total. This shows that the importance of education declined from 2007 to 2017. Although the significant R^2 statistics are too small from the iTheilT and iTheilL regressions, we can easily see that the number of statistically significant (at the .05 percent) estimates for the educational categories reduces from all five to none for the iTheilT total regression and from five to two for the iTheilL regression. Second, income inequality between race–gender-specific groups as well as within them that can be explained by education also declined over the 10 years span, judged by the regression models in the second and third columns in the two tables. This observation can be made with the R^2 statistics when they manifest their nonzero values; it can still be made with the coefficient estimates when they do not. Last but not least, judged by the R^2 statistics from the regressions of the iGini components, educational inequality in income between the eight race–gender groups is just slightly greater than that within the groups. However, judged by either the iTheilT or the iTheilL component models, educational inequality in income is much more consequential between the

Table 7. Linear Regression of iGini, iTheilT, and iTheilL Totals and Their Between- and Within-gender and Race Components on Education, 2007 (Weights Applied).

	iGini Total (1)		iGini Between (2)		iGini Within (3)	
Predictor	*b*	*b* 95 Percent CI	*b*	*b* 95 Percent CI	*b*	*b* 95 Percent CI
(Intercept)	5.63**	[5.48, 5.79]	4.38**	[4.27, 4.49]	1.26**	[1.20, 1.32]
High school	0.49**	[0.30, 0.67]	−0.03	[−0.16, 0.10]	0.52**	[0.45, 0.58]
Some college	0.74**	[0.56, 0.93]	0.15*	[0.02, 0.28]	0.59**	[0.53, 0.66]
College graduation	3.03**	[2.83, 3.22]	1.68**	[1.54, 1.81]	1.35**	[1.28, 1.42]
Graduate education	6.30**	[6.08, 6.53]	4.02**	[3.87, 4.18]	2.28**	[2.20, 2.36]
	R^2 = .053**		R^2 = .049**		R^2 = .044**	
	iTheilT Total (1)		iTheilT Between (2)		iTheilT Within (3)	
(Intercept)	3.71**	[3.06, 4.37]	−0.05	[−0.15, 0.06]	3.76**	[3.14, 4.37]
High school	1.38**	[0.61, 2.15]	0.70**	[0.57, 0.82]	0.68	[−0.04, 1.41]
Some college	0.93*	[0.16, 1.71]	0.41**	[0.28, 0.54]	0.52	[−0.20, 1.25]
College graduation	1.22**	[0.41, 2.04]	0.69**	[0.55, 0.82]	0.54	[−0.23, 1.30]
Graduate education	0.96*	[0.01, 1.90]	0.76**	[0.61, 0.92]	0.19	[−0.69, 1.08]
	R^2 = .000**		R^2 = .002**		R^2 = .000	
	iTheilL Total (1)		iTheilT Between (2)		iTheilL Within (3)	
(Intercept)	5.78**	[5.50, 6.06]	0.83**	[0.75, 0.90]	4.95**	[4.68, 5.22]
High school	−0.88**	[−1.20, −0.55]	−0.63**	[−0.71, −0.55]	−0.25	[−0.57, 0.07]
Some college	−0.68**	[−1.00, −0.35]	−0.34**	[−0.42, −0.25]	−0.34*	[−0.66, −0.02]
College graduation	−0.61**	[−0.95, −0.26]	−0.52**	[−0.61, −0.43]	−0.08	[−0.42, 0.25]
Graduate education	−1.02**	[−1.42, −0.62]	−0.69**	[−0.79, −0.59]	−0.33	[−0.72, 0.06]
	R^2 = .000**		R^2 = .003**		R^2 = .000	

Note. N = 88,991, Gini = .465, TheilT = .418, TheilL = .454. Each of the models 1, 2, and 3 estimates educational effects on the total iGini, iTheilT, and iTheilL and their between and within gender–race components (which are all multiplied by a constant to render the estimates, respectively; the educational category of less than high school is the reference).
*p < .05. **p < .01.

Table 8. Linear Regression of iGini, iTheilT, and iTheilL Totals and Their Between- and Within-gender and Race Components on Education, 2017 (Weights Applied).

	iGini Total (1)		iGini Between (2)		iGini Within (3)	
Predictor	*b*	*b* 95% CI	*b*	*b* 95% CI	*b*	*b* 95% CI
(Intercept)	6.24**	[5.98, 6.50]	5.04**	[4.84, 5.23]	1.20**	[1.12, 1.29]
High school	0.73**	[0.43, 1.03]	0.11	[−0.10, 0.33]	0.62**	[0.52, 0.71]
Some college	1.03**	[0.73, 1.32]	0.33**	[0.12, 0.55]	0.69**	[0.60, 0.79]
College graduation	3.21**	[2.91, 3.51]	1.81**	[1.59, 2.03]	1.40**	[1.30, 1.50]
Graduate education	5.93**	[5.61, 6.26]	3.94**	[3.70, 4.18]	2.00**	[1.89, 2.10]
	R^2 = .031**		R^2 = .030**		R^2 = .027**	
	iTheilT Total (1)		iTheilT Between (2)		iTheilT Within (3)	
(Intercept)	5.98**	[4.71, 7.25]	0.35**	[0.20, 0.51]	5.63**	[4.41, 6.85]
High school	0.86	[−0.59, 2.31]	0.59**	[0.41, 0.77]	0.27	[−1.12, 1.65]
Some college	−0.23	[−1.66, 1.20]	0.33**	[0.15, 0.50]	−0.56	[−1.93, 0.81]
College graduation	0.47	[−1.00, 1.94]	0.52**	[0.34, 0.70]	−0.05	[−1.45, 1.35]
Graduate education	0.30	[−1.29, 1.89]	0.58**	[0.39, 0.78]	−0.29	[−1.81, 1.23]
	R^2 = .000		R^2 = .001**		R^2 = .000	
	iTheilL Total (1)		iTheilT Between (2)		iTheilL Within (3)	
(Intercept)	6.15**	[5.74, 6.56]	0.63**	[0.53, 0.73]	5.52**	[5.12, 5.93]
High school	−0.69**	[−1.16, −0.22]	−0.73**	[−0.84, −0.62]	0.04	[−0.43, 0.50]
Some college	−0.41	[−0.88, 0.05]	−0.43**	[−0.54, −0.32]	0.02	[−0.44, 0.47]
College graduation	−0.68**	[−1.16, −0.21]	−0.64**	[−0.76, −0.53]	−0.04	[−0.51, 0.43]
Graduate education	−0.47	[−0.99, 0.04]	−0.70**	[−0.82, −0.57]	0.22	[−0.29, 0.73]
	R^2 = .000		R^2 = .003**		R^2 = .000	

Note. $N = 77,969$, Gini = .472, TheilT = .440, TheilL = .461. Each of the models 1, 2, and 3 estimates educational effects on the total iGini, iTheilT, and iTheilL and their between and within gender–race components (which are all multiplied by a constant to render the estimates, respectively; the educational category of less than high school is the reference).
*$p < .05$. **$p < .01$.

race–gender groups than this inequality within them. The difference is due to the different definition of the iGini and iTheil between components. We can also interpret individual estimates from these models. For example, close to 1.4 per million of the overall Theil T index of income inequality in 2007 can be explained by the difference between an individual with a high school education and another without it (Table 7, panel 2, column 1). The effect sounds small because the sample is large, with 88,991 individuals in the analysis, and because the total Theil's T is only .465 (resulting in a minute per person contribution to the total). When this total TheilT is decomposed between the eight race–gender groups, seven per million of the total TheilT can be explained by the difference between the two individuals with or without a high school education in the same year (Table 7, panel 2, column 2). This effect, however, declines to just under six per million in 2017 (Table 8, panel 2, column 2).

Conclusion

In this article, I presented an individual-based decomposition of three common inequality measures, the Gini, Theil's T, and Theil's L measures. I used five artificial data examples and the 2007 and 2017 CPS ASEC data to illustrate the performance of these iGini, iTheilT, and iTheilL measures and their between-group and within-group components.

There are at least three usages of these iGini, iTheilT, and iTheilL measures for sociological research. First, as illustrated with the artificial data examples as well as the CPS survey data, these individual inequality measures can display shapes of inequality across the sample space, including the structural between-group as well as within-group features of such shapes, in graphic or tabular forms. Second, these individual measures can be aggregated by groups to evaluate the overall group-specific contributions to overall inequality (or to overall within-group inequality), as given by equation (10). Here, a group's contributions are based on an assumption of an entire population composed of the same group members, the only way group-specific contributions can be evaluated fairly unless one is simply interested in the total share of inequality contributions by a particular group, and in that case, the group-specific contributions can be appropriate. Finally, because each case represents its contribution to overall inequality (or to its component between-group or within-group inequality), such iGini, iTheilT, or iTheilL measures may enter into a further regression type of substantive analysis as an independent or the dependent variable. I illustrated this third usage by including these individual-level inequality measures as a dependent variable in a series of simple regression models explaining the relation between education and income inequality in the United States between 2007 and 2017. This application suggests a new, direct way for analyzing inequality: When income is the outcome variable in a regression model, one estimates the effect of an explanatory factor on income; when iGini, iTheilT, or iTheilL is the outcome variable in a model, one estimates the effect or contribution of an explanatory factor on the total amount of inequality as represented by Gini, Theil's T, or Theil's L.

When to use which individual-based inequality measure? As is well-known, Theil's T is more sensitive to the upper tail, while Theil's L is more sensitive to the lower end of the distribution. The same property applies to their individual components. It is also important to understand the definitional difference in the between-group component of these measures. For the two Theil measures, individual-based between-group measures are defined as individually proportionated overall between-group inequality, while for the Gini index, its individual between-group component is based on all possible pairwise comparisons between a specific member in a social group and all members of the other groups. In addition, each of the iGini total computation is based on *n* calculations, as compared with a single calculation for the two generalized entropy measures considered in this article. For these reasons, the iGini measure has a greater variation and (and a smaller standard error) than the iTheilT and iTheilL measures, which may show even less variation without using sampling weights (which were included in the empirical application in this article) and may differentiate individuals better. Therefore, the iGini components are useful in further substantive research in their own right as shown in the series of 18 regression models, particularly in analyses that typically would otherwise include only the conventional Gini coefficient measured at an aggregate level, a common practice in the social sciences that ignores individual heterogeneity in contributions to inequality.

Author's Note

This article also benefited from the constructive comments of the reviewers. An R package, iIneq (version 1.0.1), available at CRAN, computes the iGini, iTheilT, and iTheilL measures, and a Stata module, igini1, installable in a Stata session, computes the iGini components.

Acknowledgments

The author would like to acknowledge a fellowship from the Center for Advanced Study in the Behavioral Sciences (Stanford University) and a leave from the University of Illinois in 2017 to 2018 that facilitated the research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tim Futing Liao https://orcid.org/0000-0002-1296-7660

References

Akita, Takahiro. 2003. “Decomposing Regional Income Inequality in China and Indonesia Using Two-stage Nested Theil Decomposition Method.” *Annuals of Regional Science* 37:55-77.

Breen, Richard and Inkwan Chung. 2015. “Income Inequality and Education.” *Sociological Science* 2:454-77.

Ceriani, Lidia and Paolo Verme. 2015. “Individual Diversity and the Gini Decomposition.” *Social Indicators Research* 121:637-46.

Dagum, Camilo. 1997. “A New Approach to the Decomposition of the Gini Income Inequality Ratio.” *Empirical Economics* 22:515-31.

Dagum, Camilo. 1998. “Fondements de bien-être social et décomposition des mesures d’inégalité dans la répartition du revenu.” *Economie Appliquée* 51: 151-202.

del Río, Cora and Olga Alonso-Villar. 2015. “The Evolution of Occupational Segregation in the United States, 1940-2010: Gains and Losses of Gender and Race/ethnicity Groups.” *Demography* 52:967-88.

Firebaugh, Glenn and Brian Goesling. 2004. “Accounting for the Recent Decline in Global Income Inequality.” *American Journal of Sociology* 110:283-312.

Kenworthy, Lane. 2007. “Inequality and Sociology.” *American Behavioral Scientist* 50:584-602.

Liao, Tim Futing. 2016a. “Evaluating Distributional Differences in Income Inequality.” *Socius: Sociological Research for a Dynamic World* 2:1-14.

Liao, Tim Futing. 2016b. “Evaluating Distributional Differences in Income Inequality with Sampling Weights” *Socius: Sociological Research for a Dynamic World* 2:1-6.

Modalsli, Jørgen. 2017. “Decomposing Global Inequality.” *Review of Income and Wealth* 63:445-63.

Mussard, Stéphane, Françoise Seyte, and Michel Terraza. 2003. "Decomposition of Gini and the Generalized Entropy Inequality Measures." *Economics Bulletin* 4: 1-5.

Osberg, Lars. 1981. *Economic Inequality in Canada*. Toronto, Canada: Butterworths.

Osberg, Lars. 2017. "On the Limitations of Some Current Usages of the Gini Index." *Review of Income and Wealth* 63:574-84.

Theil, Henri. 1967. *Economics and Information Theory*. Amsterdam, the Netherlands: North-Holland.

Tilly, Charles. 2003. "Inequality, Democratization, and De-democratization." *Sociological Theory* 21:37-43.

Yitzhaki, Shlomo. 1994. "Economic Distance and Overlapping of Distributions." *Journal of Econometrics* 61:147-59.

Author Biography

Tim Futing Liao is Professor of Sociology (and of Statistics) at the University of Illinois at Urbana-Champaign. He is former Editor of *Sociological Methodology*, a current Deputy Editor of *Demography*, and an executive committee member of the Sequence Analysis Association. His current research interests include inequality measurement and analysis as well as life course trajectory survey and analysis.