Evaluating Distributional Differences in Income Inequality

Tim Futing Liao

Abstract
Income and wage inequality between gender and other social groups is commonly evaluated by the difference in average income or the decomposition of the total amount of inequality into between-group and within-group components using an inequality measure such as the Theil index. Either approach ignores the dispersional difference between social groups, often manifested in glass-ceiling and glass-floor effects. The author introduces a refined Theil decomposition that offers two ways of capturing such between-group distributional differences. An analysis of income data from 10 European countries illustrates the usefulness of the proposed evaluation and decomposition methods.

Keywords
inequality, Theil index, measurement, gender inequality, decomposition, glass ceiling, glass floor

This article concerns the measurement of inequality in general and the evaluation of gender wage inequality in particular. Research has shown that macro-level gender inequality can have effects on micro-level behavior and attitudes (Fuwa 2004). Macro-level gender inequality can manifest itself in several ways, one of which is a gender wage gap. Two of the inequality measures commonly used for evaluating income inequality are the Gini coefficient and the Theil index (see, e.g., Charles-Coll 2011), though the Theil index is preferred (Allison 1978). The Theil index of inequality has been widely applied in the social sciences; one reason for its popularity is its decomposability (Akita 2003).

Gender wage inequality can be simply viewed as a gap between the average men’s and women’s wages. Whereas gender pay gaps have gradually narrowed in many developed societies, such as the United Kingdom, it is at the top end where the gap is much wider, as recently observed in the news media: wages for those British women who are presumed to be at the peaks of their professional careers are much lower than those of their male counterparts (De Piero 2015). The phenomenon of a wider pay gap between the sexes at the top is better known as the “glass ceiling” effect and has been the focus of much recent empirical research (Albrecht, Björklund, and Vroman 2003; Arulampalam, Booth, and Bryan 2007; Christofides, Polycarpou, and Vrachimis 2013; Cotter et al 2003). A similar pattern can be found at the bottom end of the pay distribution, known as the “sticky floor” or “glass floor” effect, indicating one social group is less likely than another to be low-income earners (Christofides et al. 2013; McKnight 2015; Reeves and Howard 2013).

It follows that a conventional inequality measure, despite its decomposability, fails to capture such glass-ceiling and glass-floor effects. In other words, gender inequality in income cannot be understood as a simple difference in mean income between the sexes. Rather, to properly evaluate gender inequality when glass-ceiling and glass-floor effects are present, one must focus not just on differences in the mean but also differences in dispersion and distribution. Note that there are other approaches to studying different segments of distribution, especially the method of quantile regression pioneered by Koenker and Basset (1978). Quantile regression can be readily applied to analyzing glass-ceiling and glass-floor effects in income distributions (see, e.g., Christofides et al. 2013; Melly 2005). Such an approach can provide a detailed analysis of the mechanisms generating the inequalities reflected by glass-ceiling and glass-floor effects. However, an inequality index such as the Gini coefficient or the Theil index remains a powerful and useful method for measuring inequality. The Theil index, despite its popularity (due largely to its decomposability), is as yet incapable of recording dispersional differences in income distributions.

1University of Illinois at Urbana-Champaign, Urbana, IL, USA

Corresponding Author:
Tim Futing Liao, University of Illinois, Department of Sociology, 3120 Lincoln Hall, 702 S. Wright Street, Urbana, IL 61801, USA.
Email: tfliao@illinois.edu

Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (http://creativecommons.org/licenses/by-nc/3.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
The objective of this research, then, is to fill the gap and refine the Theil index so that it can reflect both the average and the distributional differences and variations. To illustrate the problem at hand in a plain manner, consider a simple hypothetical income data set consisting of two groups with 15 cases each in G1 and G2:

G1: [1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5, 6, 8, 10]
G2: [1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 6, 7, 8]

The means of the two groups in the data set are equal, with the mean of G1 = the mean of G2 = 3.933. Theil’s (first) measure of inequality gives a value of .134. Performing the standard decomposition of the Theil index, we obtain a between-group component of .000 and a within-group component of .134, indicating that there is no inequality between the two groups, with the inequality found entirely attributed to within-group variations. Figure 1 gives a visual comparison of the kernel density distributions of the two groups, showing that the first group has a heavier upper tail than the second. (Kernel density is a nonparametric method for estimating smooth probability density distributions of empirical data.)

In this simple example, the conventional Theil decomposition provides no information about the shapes of inequality beyond a simple comparison of the means. Examining the distributions in the groups, we see that G1 has a case with a value 10, above the maximum or the ceiling of the value 8 in G2. It becomes apparent that differences in dispersion, manifested here in a ceiling effect, should be taken into account when evaluating inequality.

To anticipate the conclusion, we apply one of the two approaches introduced later to the data. This method further decomposes the within-group component into the two sub-components of inequality contributed by shared dispersion and by floor and ceiling effects, or differences in dispersion. Applying this method, the within-group component from G1 and G2 is further decomposed into a shared-dispersion component of .055 and a different-dispersion component of .079. This particular approach focuses on shared versus different dispersions. We will introduce another, quantile-based approach that decomposes inequality in each of the quantiles separately. Income distributions are typically nonnormal, and a quantile analysis of income inequality has been the choice of many researchers. Naturally, evaluating income inequality can benefit from a quantile-based approach.

Broadly speaking, the research reported here follows up on studies of the evaluation of between-group and between-class inequality (Jasso and Kotz 2008; Liao 2006). It also follows up on a research tradition that deals with group heterogeneity within the world and within a country by applying methods of decomposition that break down individual inequality into between-group and within-group components to study inequality across nations or blocks of nations (Bourguignon and Morrison 2002; Firebaugh 1999, 2000; Goesling 2001; Schultz 1998) or across social groupings within nations such as gender, race/ethnicity, or social class (Jenkins 1995; Schultz 1998; Yitzhaki 2002). The research in this paper is perhaps more related to a report on the status of inequality in the United Kingdom, where differences are found not only between the mean levels of outcomes such as education, income, and wealth of social groups such as gender, ethnicity, and age but also between the ranges or variations in such outcomes of these groups (Hills et al. 2010). A lesson learned from their analysis is that studying only averages can be misleading without also analyzing these variational or dispersionsal differences, which are the focus of this article.

In the section below, we provide a brief review of the Theil index and its decomposition, followed by a section introducing the two refined approaches, one distinguishing between shared and different dispersions and the other using a quantile-based method for evaluating different dispersions in inequality. We then apply the two methods to income data on 10 European countries from the Luxembourg Income Study (LIS). The proposed methods and their application allow us to draw some conclusions in a final section.

**Theil’s Decomposition**

The Theil index of inequality, originally proposed by Theil (1967), is one of the two most widely applied inequality measures (the other being the Gini coefficient). Unlike the Gini coefficient, however, the total amount of inequality measured by the Theil index can be cleanly decomposed into the two additive parts of between-group and within-group inequality.
The total amount of inequality measured by Theil’s $T$ or first measure is

$$T = \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\bar{x}} \ln \frac{x_i}{\bar{x}}, \quad (1)$$

where $x_i$ is the income of individual $i$, $\bar{x}$ is the overall mean income, and $N$ is the sample size. Because our focus is on Theil’s $T$, or the first Theil measure, we subsequently drop the subscript $T$. Our presentation focuses on the first Theil measure, even though the R program for the proposed methods (Appendix A) implements as well the proposed methods for the second Theil measure, or Theil’s $L$. Hereafter, Theil index refers to Theil’s first measure. Using the same notation, the Theil index in equation 1 can be decomposed into a between-group component,

$$T_b = \sum_{k=1}^{K} y_k \ln \frac{\bar{x}_k}{\bar{x}}, \quad (2)$$

where $y_k$ is the $k$th group’s income share expressed as a proportion of the sample or population total income, and $\bar{x}_k$ is the mean income of group $k$, and a within-component,

$$T_w = \sum_{k=1}^{K} y_k \sum_{i=1}^{n_k} \frac{y_{ik}}{\bar{x}_k} \ln \frac{x_{ik}}{\bar{x}_k}, \quad (3)$$

where $y_{ik}$ is the income share of the $i$th individual within the $k$th group, and $x_{ik}$ is the $i$th individual’s income within group $k$. The decomposition is additive because equations 2 and 3 sum to equation 1. Theil’s decomposition has been popular in the economic and social sciences and has long been incorporated in popular statistical software packages such as Stata (Jenkins 1995, 1999; Jenkins and Van Kerm 2009). Theil’s decomposition finds many applications, such as Stata (Jenkins 1995, 1999; Jenkins and Van Kerm 2009). A similar formulation for the second Theil measure, or Theil’s $L$, is presented by

$$T_w = \sum_{k=1}^{K} y_{k,a} \sum_{i=1}^{n_k} \frac{y_{ik,a}}{\bar{x}_k} \ln \frac{x_{ik,a}}{\bar{x}_k} + \sum_{k=1}^{K} \sum_{i=1}^{n_k} y_{ik,b} \ln \frac{x_{ik,b}}{\bar{x}_k}, \quad (4)$$

where the $y$ terms with subscript $a$ are income shares based on the cases satisfying the shared-spread criterion of $[l, u]$, the $x$ terms with subscript $a$ are corresponding individual income values, and the $y$ and $x$ terms with subscript $b$ are income shares and income values, respectively, that do not satisfy the shared-dispersion criterion.

**Method 2**

Because income distributions are typically skewed, social scientists often study them with quantile analysis of various kinds, from quantile-quantile plots to quantile regression. Thus, it should be informative to analyze $G$ income distributions by quantile. For example, to understand the glass-ceiling and glass-floor effects, we may wish to study how men and women differ in the top decile or in the bottom decile of the income distribution. To accomplish this objective, we rely on the recursive nature of the Theil index by first decomposing the overall inequality into the two components of between-quantile and within-quantile inequality. Using the same notation, the Theil index in equation 1 is now decomposed into a between-quantile component and a within-quantile component as the first and the second items on the right-hand side of equation 5:

$$T = \sum_{q=1}^{Q} y_q \ln \frac{x_q}{\bar{x}} + \sum_{q=1}^{Q} y_q T_q, \quad (5)$$

where $y_q$ is the income share of the $q$th quantile group out of the total income of the $Q$ number of groups, $\bar{x}_q$ is the mean income of quantile $q$, and $T_q$ is the inequality within quantile group $q$. Following the notation of equations 2 and 3, the two components on the right-hand side of equation 5 are represented by $T_b$ and $T_w$, respectively. We then further decompose $T_q$ into between-group and within-group components for each quantile as the first and the second item on the right-hand side of equation 6:

**Capturing Glass-ceiling and Glass-floor Effects in Inequality**

To refine our evaluation of inequality by taking into account glass-ceiling and glass-floor effects, we consider two different approaches. The first focuses on a further decomposition of the within-group component of inequality (equation 3) into two subcomponents, one measuring shared dispersion by the groups and the other recording different dispersion between the groups. The second method takes a quantile-based approach.
In this figure, we use the same color (or hue) to indicate the same type of components. Method 1 of the refined Theil decomposition further divides the within-group component into two subcomponents: a different-dispersion component that captures the contribution of differences in dispersion (in blue) and a true within-group, shared-dispersion component (in green).

Method 2 is quantile based and involves a two-step decomposition. It decomposes the total inequality into a between-quantile component (in gold) and a within-quantile component (in various shades of turquoise and olive colors) and then further into a set of within-quantile between-group components (in various turquoise shades) and a set of within-quantile within-group components (in various olive shades). The schematic example in Figure 2 shows four quantiles (or quartiles here), with each shade representing a specific quantile (or quartile).

To calculate twice the natural logarithms of the ratio between two $wb_q$ terms for an analysis in which quintiles are used, for example, we compute $2\log(wb_5/wb_1)$ and $2\log(wb_6/wb_5)$, respectively, for capturing glass-ceiling and glass-floor effects. When the analysis uses deciles as the quantile of choice, we compute $2\log(wb_{10}/wb_1)$ and $2\log(wb_{10}/mean[wb_2, wb_9])$, respectively, to evaluate glass-ceiling and glass-floor effects. Both methods 1 and 2 (including the alternative method 2b) are programmed as an R function for both Theil’s first measure and Theil’s second measure (Appendix A). The plan is to make this function also available as a Stata ado module in the future.

### Analyzing Gender Wage Inequality in 10 European Countries

#### Descriptive Analysis

For an empirical analysis, we examine 10 European countries—Austria, Belgium, France, Greece, Hungary, Ireland, Italy, Luxembourg, Russia, and Spain—by using data from the LIS. The LIS harmonizes national survey data on income and wealth from many European countries and some non-European countries. For this example, we selected all LIS countries in Europe with surveys conducted close to one another in time (in or within one year of 2000) that provided net wage data in actual local currency values (instead of in grouped categories). Net or post-tax income or wage data in local currency for sampled individuals were gathered on surveys from these countries, and natural income values from the past 12 months for men and women who declared working as their main activity at the time of the survey were used in the analysis. There is no need to convert the local currencies to a common currency, because the Theil index, as a member of the generalized entropy family, is scale independent.

To get a sense of men’s and women’s net income differences, we summarize the median, top 1 percentile, and maximum income values of these countries by sex in Table 1.
Even though the income data are in local currencies, the chief purpose of this table is to compare incomes between the sexes in the same countries. A few quick observations are in order. First, men’s net wages were universally higher than women’s at any point of the income distribution, be it the median, the top 1 percentile, or the maximum. Second, this gender wage gap widens when moving up a distribution; the higher up in the income distribution, the wider the gap. Third, the distance between the maximum and the top 1 percentile is much greater for men than for women. Both the second and the third observations suggest a glass-ceiling effect. To assess the difference in spread or dispersion of the income distributions, we can also examine coefficients of variation, which are reported in Appendix B, though for the sake of brevity, we do not focus on them in the main text.

To visually examine the actual gender-specific income distributions from these 10 countries, we present kernel density plots of these income distributions by gender in Figure 3. The kernel density plots by gender allow us to compare income distributions in each country in terms of the shape and the spread of the distributions.

With the exception of Russia, the peaks of men’s wage distributions are all noticeably higher than the peaks of women’s distributions. With the exception of Italy, the left tails of women’s distributions are all heavier than men’s. In all countries, men’s distributions tend to spread out more to the right.

**Figure 2.** A Schematic Comparison of the Three Methods of Theil Decomposition.

Note. \( wb_1 \) = within the first quantile/between group, \( wb_2 \) = within the second quantile/between group, and so on; \( ww_1 \) = within the first quantile/within group, \( ww_2 \) = within the second quantile/within group, and so on.
Table 1. Medians, Top 1 Percentiles, and Maximums of Net Wages in Thousands by Sex in 10 European Countries, circa 2000.

| Country   | Sex | Median | Top 1% | Maximum | n  |
|-----------|-----|--------|--------|---------|----|
| Austria   | Men | 225    | 729    | 2,291   | 1,629 |
|           | Women | 144   | 431    | 689     | 1,135 |
| Belgium   | Men | 751    | 2,329  | 4,800   | 1,406 |
|           | Women | 542   | 1,246  | 3,964   | 1,126 |
| France    | Men | 110    | 465    | 1,148   | 5,719 |
|           | Women | 84    | 277    | 460     | 4,666 |
| Greece    | Men | 1,695  | 11,200 | 30,000  | 2,598 |
|           | Women | 1,750 | 6,735  | 10,555  | 1,392 |
| Hungary   | Men | 564    | 2,491  | 5,694   | 719  |
|           | Women | 467   | 2,608  | 7,140   | 692  |
| Ireland   | Men | 12     | 42     | 112     | 2,058 |
|           | Women | 10    | 31     | 67      | 1,318 |
| Italy     | Men | 23     | 80     | 350     | 5,018 |
|           | Women | 20    | 50     | 150     | 3,004 |
| Luxembourg| Men | 1,090  | 3,845  | 7,357   | 1,621 |
|           | Women | 685   | 2,266  | 3,156   | 969  |
| Russia    | Men | 16     | 169    | 339     | 1,629 |
|           | Women | 10    | 90     | 181     | 1,657 |
| Spain     | Men | 1,920  | 7,700  | 34,416  | 3,233 |
|           | Women | 1,488 | 5,760  | 9,760   | 1,813 |

Note. All wages are reported in local currency.

(i.e., earning higher wages). All these observations suggest a male advantage in earning power in all 10 countries.

Analysis of Gender Wage Inequality via Method 1

Table 2 presents the Theil index decomposition, using both the conventional method and method 1 introduced earlier. The first three columns contain the standard Theil decomposition, and the last two columns contain a further decomposition of the conventional within-group component into the shared-dispersion subcomponent (the “Withinw” column) and the different-dispersion subcomponent (the “Withinb” column).

Judged by the total amount of inequality evaluated by the Theil index, Belgium has the lower amount of income inequality among the 10 countries around 2000, while Greece and Russia have the highest. Russia’s average gender wage gap appears to be the widest among the 10 countries (.027 of the between-group component), and Italy’s average gender wage gap is the lowest of the group (.002 of the between-group component). It is important to note that none of the between-group component values are above 10 percent of their respective total inequality.

An analysis of inequality such as this hides the glass-ceiling and glass-floor effects. To further investigate such effects, we applied method 1 by decomposing the within-group component into a shared-dispersion subcomponent and a different-dispersion subcomponent and report the results in the last columns of Table 2. The values in these two columns add up to the values in the within-group component of the middle column. This method enables us to see clearly the total effects of the glass ceiling and glass floor. Greece, for example, shows the second to lowest level of between-group inequality, even though it exhibits the highest level of total inequality. Relying on the conventional Theil decomposition, we would stop there. The new results show that about 17 percent of the within-group inequality or 16 percent of the total inequality is actually explained by differences in dispersion or distribution. The highest percentage explained out of the within-group inequality belongs to Luxembourg. However, we should not deduce that the glass-ceiling and glass-floor effects in that country were very high in 2000, because its total amount of inequality came out rather low (at .226). Because the Theil index is scale free, its decompositions are scale free too. The country with the lowest level of glass-ceiling and glass-floor effects is Belgium (a value of .005). After Greece (a value of .114), the country with the next highest level of glass-ceiling and glass-floor effects was Russia in 2000 (a value of .091).

Analysis of Gender Wage Inequality via Method 2

The results presented above reveal the combined effects of the glass ceiling and glass floor. To fine-tune our analysis, we present in Table 3 results from an application of method 2 (specifically method 2a), which enables the researcher to separately assess glass-ceiling and glass-floor effects. For the analysis presented in the table, we restricted the data by analyzing only those cases with wage values above zero. Zero income values would render the computation of the subcomponents of the lowest quantiles (in this case decile) of several countries undefined. Furthermore, glass-floor effects are typically defined for those with actual income (i.e., members of one social group earn a higher level of income and do not slip below members of another group for social structural reasons). In addition, the purposes of the two analyses are different, and we are not interested in a direct comparison of the two sets of results from the analyses.

In this table, we present the 90:50 and the 10:50 inequality ratios in their $2 \times \log$ inequality forms. Using the middle segment of the income distribution as a referent, the evidence of a glass-ceiling effect is very strong for 9 of the 10 countries ($2 \times \log[90:50] > 10$), and the evidence is still strong for Luxembourg ($2 \times \log[90:50] > 6$ and 10). The evidence for a glass-floor effect, however, is somewhat weaker for 9 of the 10 countries, except for Belgium, where the evidence for a glass-floor effect is even more substantial than that for a glass-ceiling effect. Of the 10 countries, the evidence for a glass-floor effect is very strong for Austria, Belgium, and Spain ($2 \times \log[10:50] > 10$), strong for France, Greece, Hungary, Ireland, and Luxembourg ($2 \times \log[10:50] > 6$ and 10), and positive for Russia ($2 \times \log[10:50] > 2$ and 4). There is no
evidence for a glass-floor effect for Italy \((2 \times \log_{10} [10:50 \text{ ratio}] < 2)\). Even though the size of a log-transformed ratio is comparable across countries, we caution here that the computation is relative to the amount of within-middle-quantile between-group inequality. For example, relative to the rather small amount of within-middle-quantile between-group inequality in Belgium, the amounts of glass-ceiling and of glass-floor inequality are comparatively rather high. Therefore, one may only interpret these findings within the context of each country. In comparison, the results produced by method 1 reported in Table 2 can be compared across countries in absolute terms.

**Summary Discussions**

What do the results suggest? A full explanation of the gender gap found is beyond the scope of this research, because there can be many social factors at work, including differences in the qualifications of men and women, differences in the treatment of equally qualified women, and the differential..
prices the labor market places on work skills in each country (Blau and Kahn 1992). In this article, we focus only on the institutional differences across countries. On a theoretical level, the amount of inequality embodied in gender-based differential distributions of income can be indicative of the type of the welfare state in a society, because the welfare state generates differentiation in de commodification and social stratification in a society, as discussed by the literature on welfare policy research initiated by Esping-Andersen (1990) and modified, refined, and reviewed by many other scholars (Arts and Gelissen 2002; Bonoli 1997; Ferrera 1996; Ferragina and Seeleib-Kaiser 2011; Leibfried 1992; Shalev 1996; Siaroff 1994). In the present example, we do not have any social democratic countries. Six of the 10 countries belong to the conservative democratic type: Austria, Belgium, France, Ireland, Italy, and Luxembourg (Italy can be grouped primarily with the conservative democratic type and secondarily with the Mediterranean type; see Ferragina and Seeleib-Kaiser 2011). Three countries stand out in the analyses reported in Tables 2 and 3: Greece, Russia, and Spain. Greece and Russia have the highest values of the component (Within\textsubscript{b}) contributed by different dispersion (Table 2), and Spain is found to have the highest combined 90:50 and 10:50 ratios (being the only country with the combined log-ratios greater than 30).

Greece and Spain belong to a type of welfare systems termed the Latin rim (Leibfried 1992), late female mobilization (Siaroff 1994), or southern (Bonoli 1997; Ferrera 1996), which is also known as the Mediterranean type (Ferragina and Seeleib-Kaiser 2011). The criticism of the original Esping-Andersen (1990) typology of insufficient sensitivity to gender inequality and the proposal by Siaroff (1994) of a late female mobilization type of welfare state regime are particularly relevant here. Citizens from this Mediterranean or late female mobilization type were found to desire equality more than their counterparts in the conservative democratic countries, while citizens from social democratic regimes were found to desire equality less than the reference category (Arts and Gelissen 2001). The findings there suggest that people in more stratified societies tend to desire more what is wanted, and vice versa.

Russia is a more complicated case. Fenger (2007) expanded the research on welfare state typology to include Eastern European countries by establishing a former Soviet Union type that includes Belarus, Estonia, Latvia, Lithuania, Russia, and Ukraine and found in this type a level of inequality higher than that for the conservative democratic and the social democratic regimes (though not quite as high as that for the liberal regime). In addition, inequality, especially gender inequality, increased in transitional former socialist societies such as China and Russia following the onset of their respective market reform or transition. Therefore, it is not surprising that the total between-gender distributional difference is found to be higher in Russia, a postsocialist society in transition, than most other countries.

**Conclusion**

In this article, we identified a characteristic problem of conventional inequality measures, especially the Theil index and
its decomposition, when applied to analyzing group-based inequality. Such measures and their decomposition hide differences in dispersion and distribution, in particular the well-known glass-ceiling and glass-floor effects in gender wage inequality. To deal with the problem, we introduced two new approaches. Method 1 further decomposes the conventional within-group Theil inequality component into the two subcomponents representing shared dispersion and different dispersion. This method does not distinguish between glass-ceiling and glass-floor effects.

To separate the two effects, we then introduced a quantile-based Theil decomposition that assesses glass-ceiling and glass-floor effects separately via two times the log-transformed 90:50 and 10:50 ratios. Applying the two methods in an analysis of the income data from 10 European countries, we were able to gain some new knowledge about gender-based income inequality that was not possible using the conventional measures. We were able to compute the relative amount of inequality (in terms of the Theil measure) out of the total amount of inequality via method 1, and the subcomponent of between-group dispersionsal difference inequality is comparable across countries. Method 2 provides additional insight into the relative size of glass-ceiling and glass-floor effects within the context of a specific country. This relative understanding is useful in its own right: if a country’s gender-based inequality is small in the middle range of the distribution but relatively large at the bottom and at the top, then the government of the country should continue to work toward removing the barriers of such glass ceilings and glass floors.

Similar to the application of conventional inequality measures, sample sizes including those based on national surveys must be sufficiently large so that the results will not be greatly affected by a few influential cases (as shown in the simple illustrative example in the introduction). Also, it is desirable to use the two methods together for studying the shape of inequality, gender based or otherwise. Similar analysis can be applied to studying race-based inequality, country-of-origin-based inequality, and inequality based on another social-structural factor. Similar to a multidimensional decomposition of a conventional Theil index, we may perform the refined analysis here by taking into account of both gender-based and race-based inequality and by applying the methods introduced using a cross-classified grouping variable. An analysis of both gender-based and race-based inequality may provide a positive answer to the twin questions raised by Browne and Misra (2003) of whether the two systems of gender and racial wage inequality intersect and, if so, whether the processes creating racial wage inequality exacerbate gender inequality.

From a social policy point of view, to achieve equality, a significant level of between-group inequality (in group means) calls for Hills et al.’s (2004) between-group redistribution, and a significant level of between-group inequality in dispersion requires Hills et al.’s (within-group) vertical redistribution differentially for the groups. To solve the problem at its root, fundamental reforms are necessary to implement a fair and equitable policy of equal participation and equal pay for all jobs and occupations.

### Appendix A: R Program for the Proposed Methods

```r
### Version 1 of TheilNew, an R program for decomposing the Classical Theil measure
### by separating the within-component into two parts, either value or quantile based.
### The code implements the methods proposed in Tim F. Liao’s published paper in
### Socius (2016). For input: x=input data of all groups, g=grouping variable;
### alpha=0 (Theil’s 2nd measure) or 1 (Theil’s first measure, the default);
### method=0 (common spread), 1 (equal quantiles, the default), or 2 (chosen percentile);
### q=quantile number (for method=1) or percentile (for method=2). Note that Method 0 here
### is Method 1 in the published paper, Method 1 is Method 2a, and Method 2 is Method 2b.
### Author: Tim Futing Liao, University of Illinois, version 1.0, December 2015
TheilNew <- function(x,g,alpha=1,method=1,q=5)
{
    G <- length(unique(g)) # number of groups
    if (G<=1)
        stop("data must have 2 or more groups to use this function")
    N <- length(x) # total number of cases
    sx <- sum(as.numeric(x)) # total sum
    mx <- mean(as.numeric(x)) # grand mean
    if (method==0) { # find common spread for method 0
        min <- min(x[g==1])
        max <- max(x[g==1])
```

The rest of the code continues with similar calculations and definitions for the other methods.
for (k in 2:G) {
    min <- max(min,min(x[g==k]))
    max <- min(max,max(x[g==k]))
}

mxa <- mean(x[x>=min & x<=max])
mxb <- mean(x[x<min | x>max])
sxa <- sum(as.numeric(x[x>=min & x<=max]))
sxb <- sum(as.numeric(x[x<min | x>max]))
}

# assigning cuts and quantiles to data
# Q quantiles for method 1 percentiles for method 2
else if (method==1) {
    cuts=seq(0,1,1/q)
    Q <- q
}
else if (method==2) {
    cuts=c(0,q,1)
    Q <- 2
}

if (method==1 | method==2) {
    qnt <- .bincode(x,breaks=quantile(x,probs=cuts),include.lowest=TRUE)

    # compute Theil between quantiles
    thq <- 0
    for (i in 1:Q) {
        xq <- x[qnt==i]
        mq <- mean(xq)
        if (alpha==1) thq <- thq + {sum(xq)/sx}*log(mq/mx)
        else if (alpha==0) thq <- thq + {length(xq)/N}*log(mx/mq)
    }
}

if (alpha==1) { # module for Theil’s first measure
    Thb1=Thw1=Thw1a=Thw1b=Thb1a=Thb1b <- 0 # variable initiation
    Thl <- sum(na.omit(as.vector({1/N}*{x/mx}*log({x/mx})))) # Theil total
    for (k in 1:G) {
        xk <- as.numeric(x[g==k])
        nk <- length(xk)
        sk <- sum(as.numeric(xk))
        mk <- mean(xk)
        Sk <- sk/sx
        Thb1 <- sum(Thb1,na.omit(Sk*log({mk/mx}))) # Theil between
        Thk1 <- sum(na.omit({xk/sk}*log({xk/mk}))) # Theil within step 1
        Thw1 <- sum(Thw1,{Sk*Thk1}) # Theil within step 2
        if (method==0) { # compute within-decomposition for method 0
            xka <- (xk[xk>=min & xk<=max])
            xkb <- (xk[xk<min | xk>max])
            ska <- sum(as.numeric(xka))
            skb <- sum(as.numeric(xkb))
            mka <- mean(xka)
            mkb <- mean(xkb)
            Ska <- ska/sxa
            Skb <- skb/sxb
            Thk1a <- sum(na.omit({xka/sk}*log({xka/mk})))
            Thk1b <- sum(na.omit({xkb/sk}*log({xkb/mk})))
            Thw1a <- sum(Thw1a,{Sk*Thk1a})
        }
    }
}
if (method==1 | method==2) { # compute within-decomposition for methods 1 & 2
  Qt.t=Qt.b=Qt.wt=Qt.w<-rep(0,Q) # quantile-based Theil component initiation
  for(i in 1:Q) {
    xq <- x[qnt==i]
    mq <- mean(xq)
    nq <- length(xq)
    sq <- sum(na.omit(xq))
    Qt.t[i] <- 
      sum(Qt.t[i],na.omit(as.vector((sq/sx)*(1/nq)*(xq/mq)*log((xq/mq)))))
    for (k in 1:G) {
      xqk <- as.numeric(x[qnt==i & g==k])
      nqk <- length(xqk)
      sqk <- sum(as.numeric(xqk))
      mqk <- mean(xqk)
      Sqk <- sqk/sum(xq)
      Qt.b[i] <- sum(Qt.b[i],na.omit((sq/sx)*Sqk*log((mqk/mq)))) # Theil between
    }
  }
}
else if (alpha==0) { # module for Theil’s second measure
  Thb1=Thw1=Thk1a=Thk1b=Thw1a=Thw1b <- 0 # variable initiation
  Th1 <- sum(na.omit(as.vector((1/N)*log((mx/x))))) # Theil total
  for (k in 1:G) {
    xk <- as.numeric(x[g==k])
    nk <- length(xk)
    mk <- mean(xk)
    Sk <- sum(xk)/sx
    Nk <- nk/N
    Thb1 <- sum(Thb1,na.omit(Nk*log((mx/mk)))) # Theil between
    Thk1 <- sum(na.omit((1/nk)*log((mk/xk)))) # Theil within step 1
    Thw1 <- sum(Thw1,{Nk*Thk1}) # Theil within step 2
    if (method==0) { # compute within-decomposition for method 0
      xka <- (xk[xk>=min & xk<=max])
      xkb <- (xk[xk<min | xk>max])
      Thk1a <- sum(na.omit((1/nk)*log((mk/xka))))
      Thk1b <- sum(na.omit((1/nk)*log((mk/xkb))))
      Thw1a <- sum(Thw1a,{Nk*Thk1a})
      Thw1b <- sum(Thw1b,{Nk*Thk1b})
    }
    if (method==1 | method==2) { # compute within-decomposition for methods 1 & 2
      Qt.t=Qt.b=Qt.wt=Qt.w<-rep(0,Q) # quantile-based Theil component initiation
      for(i in 1:Q) {
        xq <- x[qnt==i]
        mq <- mean(xq)
        nq <- length(xq)


\[ Nq \leftarrow \frac{nq}{N} \]
\[ Qt.t[i] \leftarrow \sum\{Qt.t[i], na.omit(as.vector(Nq*\{1/nq}*\log(\{mq/xq\}))\)} \]

```r
for (k in 1:G) {
  xqk <- as.numeric(x[qnt==i & g==k])
  nqk <- length(xqk)
  mqk <- mean(xqk)
  Nqk <- nqk/nq
  Qt.b[i] <- sum(Qt.b[i], na.omit(Nq*Nqk*log(\{mq/mqk\}))) # Theil between
for quantile i
Qt.wt[i] <- sum(na.omit(\{1/nqk\}*\log(\{mqk/xqk\}))) # Theil within for quantile i step 1
Qt.w[i] <- sum(Qt.w[i], \{Nq*Nqk*Qt.wt[i]\}) # Theil within for quantile i step 2
}
}
```

# collect results for output
if (method==0)
results <-
list(Theil=Th1,Theil.Between=Thb1,Theil.Within=Thw1,Theil.Within.W=Thw1a,Theil.Within.B=Thw1b)
else if (method==1) {
if (length(Qt.b)%%2==1) m <- Qt.b[ceiling(length(Qt.b)/2)]
else if (length(Qt.b)%%2==0) m <-
mean(Qt.b[length(Qt.b)/2], Qt.b[\{length(Qt.b)/2\}+1])
results <-
list(Theil=Th1,Theil.Between=Thb1,Theil.Within=Thw1,Theil.Between.Quantiles=thq,
Theil.Quantile.Total=Qt.t, Theil.Quantile.Between=Qt.b, Theil.Quantile.Within=Qt.w,
Theil.WithinQuantile.total=sum(Qt.t), Theil.WithinQuantile.Between=sum(Qt.b),
Theil.WithinQuantile.Witin=sum(Qt.w), Theil.Between.TopMiddle.Ratio=2*log(tail(Qt.b,n=1)/m),
Theil.Between.BottomMiddle.Ratio=2*log(head(Qt.b,n=1)/m))
}
else if (method==2)
results <-
list(Theil=Th1,Theil.Between=Thb1,Theil.Within=Thw1,Theil.Between.Quantiles=thq,
Theil.Quantile.Total=Qt.t, Theil.Quantile.Between=Qt.b, Theil.Quantile.Within=Qt.w,
Theil.WithinQuantile.total=sum(Qt.t), Theil.WithinQuantile.Between=sum(Qt.b),
Theil.WithinQuantile.Witin=sum(Qt.w))
return(results)
}

### Appendix B: Coefficients of Variation in Net Wages for 10 European Countries, circa 2000

| Country  | CoV   | n     | CoV   | n     |
|----------|-------|-------|-------|-------|
| Austria  | .732  | 1,629 | .724  | 1,135 |
|          | .574  | 1,410 | .516  | 943   |

(continued)
Acknowledgments

I wish to thank Thierry Kruten and Teresa Munzi at the LIS for their assistance with data processing; the anonymous reviewers and the editors of Socius for their comments; and Glenn Firebaugh, Leo Goodman, Shin-Kap Han, Stephen Jenkins, and Shlomo Yizhaki for their feedback on earlier versions of a related paper. The current version is entirely rewritten, with the core measures redeveloped. Therefore, any errors remain my own.

Funding

This research was supported in part by a University of Illinois sabbatical leave and a visiting scholarship at the LIS.

References

Akita, Takahiro. 2003. “Decomposing Regional Income Inequality in China and Indonesia Using Two-stage Nested Theil Decomposition Method.” Annals of Regional Science 37(1):55–77.

Albrecht, James, Anders Björklund, and Susan Vroman. 2003. “Is There a Glass Ceiling in Sweden?” Journal of Labor Economics 21(1):145–77.

Allison, Paul D. 1978. “Measures of Inequality.” American Sociological Review 43(6):865–80.

Arts, Wil, and John Gelissen. 2001. “Welfare States, Solidarity, and Justice Principles: Does the Type Really Matter?” Acta Sociologica 44(4):283–99.

Arts, Wil, and John Gelissen. 2002. “Three Worlds of Welfare Capitalism or More? A State-of-the-art Report.” Journal of European Social Policy 12(2):137–58.

Arulampalam, Wiji, Alison L. Booth, and Mark L. Bryan. 2007. “Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution.” Industrial and Labor Relations Review 60(2):163–86.

Blau, Francine D., and Lawrence M. Kahn. 1992. “The Gender Earnings Gap: Learning from International Comparisons.” American Economic Review 82(2):533–38.

Bonoli, Giuliano. 1997. “Classifying Welfare States: A Two-dimension Approach.” Journal of Social Policy 26(3):351–72.

Bourguignon, François, and Christian Morrison. 2002. “Inequality among World Citizens: 1820–1992.” American Economic Review 92(4):727–44.

Browne, Irene, and Joya Misra. 2003. “The Intersection of Gender and Race in the Labor Market.” Annual Review of Sociology 29:487–513.

Charles-Coll, Jorge A. 2011. “Understanding Income Inequality: Concept, Causes and Measurement.” International Journal of Economics and Management Sciences 1(1):17–28.

Christofides, Louis N., Alexandros Polycarpou, and Konstantinos Vrachimis. 2013. “Gender Wage Gaps, ‘Sticky Floors’ and ‘Glass Ceilings’ in Europe.” Labour Economics 21:86–102.

Cutler, David A., Joan M. Hermsen, Seth Ovadia, and Reeve Vanneman. 2001. “The Glass Ceiling Effect.” Social Forces 80(2):655–82.

De Piero, Gloria. 2015. “The Gender Pay Gap Is a Scandal. David Cameron Must Act on His Pledge to End It.” The Guardian, August 26, 2015. Retrieved September 15, 2015 (http://www.theguardian.com/commentisfree/2015/aug/26/gender-pay-gap-david-cameron).

Esping-Andersen, Gausta. 1990. The Three Worlds of Welfare Capitalism. Cambridge, UK: Polity.

Fenger, H. J. M. 2007. “Welfare Regimes in Central and Eastern Europe: Incorporating Post-communist Countries in a Welfare

Appendix B (continued)

|       | CoV  | n    | CoV  | n    |
|-------|------|------|------|------|
| Belgium | .568 | 1,406 | .538 | 1,126 |
|        | .468 | 1,296 | .442 | 1,044 |
| France  | .793 | 5,719 | .704 | 4,698 |
|         | .685 | 5,162 | .624 | 4,366 |
| Greece  | 1.230 | 2,598 | 1.001 | 1,392 |
|         | .654 | 1,477 | .486 | 859 |
| Hungary | .773 | 719  | .973 | 692  |
|         | .754 | 706  | .966 | 687  |
| Ireland | .842 | 2,058 | .631 | 1,318 |
|         | .549 | 1,567 | .549 | 1,227 |
| Italy   | .863 | 5,018 | .684 | 3,004 |
|         | .537 | 3,706 | .475 | 2,508 |
| Luxembourg | .634 | 1,621 | .661 | 969  |
|         | .578 | 1,543 | .617 | 931  |
| Russia  | 1.371 | 1,629 | 1.222 | 1,657 |
|         | 1.159 | 1,326 | 1.056 | 1,406 |
| Spain   | .890 | 3,233 | .793 | 1,813 |
|         | .652 | 2,571 | .613 | 1,530 |

Note. Coefficients of variation are obtained for each country for those without income (first row) and for those with nonzero income (second row).
Regime Typology.” *Contemporary Issues and Ideas in Social Sciences* 3(2):1–30.

Ferrera, Maurizio. 1996. “The ‘Southern’ Model of Welfare in Social Europe.” *Journal of European Social Policy* 6(1):17–37.

Ferragina, Emanuele and Martin Seeleib-Kaiser. 2011. “Welfare Regime Debate: Past, Present, Futures?” *Policy & Politics* 39(4):583–611.

Firebaugh, Glenn. 1999. “Empirics of World Income Inequality.” *American Journal of Sociology* 104(6):1597–1630.

Firebaugh, Glenn. 2000. “The Trend in Between-nation Inequality.” *Annual Review of Sociology* 26:323–39.

Fuwa, Makiko. 2004. “Macro-level Gender Inequality and the Division of Household Labor in 22 Countries.” *American Sociological Review* 69(6):751–67.

Goesling, Brian. 2001. “Changing Income Inequalities within and between Nations: New Evidence.” *American Sociological Review* 66(5):745–61.

Hills, John. 2004. *Inequality and the State*. Oxford, UK: Oxford University Press.

Hills, John, Mike Brewer, Stephen Jenkins, Ruth Lister, Ruth Lipton, Stephen Machin, Colin Mills, Tariq Modood, Teresa Rees, and Sheila Riddell. 2010. “An Anatomy of Economic Inequality in the UK: Report of the National Equality Panel.” London: Centre for Analysis of Social Exclusion, London School of Economics and Political Science.

Jasso, Guillermina, and Samuel Kotz. 2008. “Two Types of Inequality: Inequality between Persons and Inequality between Subgroups.” *Sociological Methods & Research* 37(1):31–74.

Jenkins, Stephen P. 1995. “Accounting for Inequality Trends: Decomposition Analysis for the UK.” *Economica* 62(1):29–64.

Jenkins, Stephen P. 1999. “INEQDECO: Stata Module to Calculate Inequality Indices with Decomposition by Subgroup.” Revised May 26, 2008. Statistical Software Components S366002. Boston: Boston College Department of Economics.

Jenkins, Stephen P., and Philippe Van Kerm. 2009. “DSGINIDECO: Stata Module to Compute Decomposition of Inequality Change into Pro-poor Growth and Mobility Components.” Statistical Software Components S457009. Boston: Boston College Department of Economics.

Kass, Robert E., and Adrian E. Raftery. 1995. “Bayes Factor.” *Journal of the American Statistical Association* 90(430):773–95.

Koenker, Roger, and Gilbert Bassett, Jr. 1978. “Regression Quantiles.” *Econometrica* 46:33–50.

Leibfried, Stephan. 1992. “Towards a European Welfare State? On Integrating Poverty Regimes into the European Community.” Pp. 245–80 in *Social Policy in a Changing Europe*, edited by Zsuzsa Ferge and Jon-Eivind Kolberg. Frankfurt-am-Main, Germany: Campus Verlag.

Leigh, Andrew. 2007. “How Closely Do Top Income Shares Track Other Measures of Inequality?” *Economic Journal* 117(524):F619–33.

Liao, Tim Futing. 2006. “Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-based Clustering.” *Sociological Methodology* 36(1):201–24.

McKnight, Abigail. 2015. “Downward Mobility, Opportunity Hoarding and the ‘Glass Floor.’” Research Report. London: Centre for Analysis of Social Exclusion, London School of Economics.

Melly, Blaise. 2005. “Decomposition of Differences in Distribution Using Quantile Regression.” *Labour Economics* 12:577–590.

Reeves, Richard V., and Kimberly Howard. 2013. “The Glass Floor: Education, Downward Mobility, and Opportunity Hoarding.” Research Paper. Washington, DC: Center on Children and Families, Brookings Institution.

Schultz, T. Paul. 1998. “Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why.” *Journal of Population Economics* 11:307–344.

Shalev, Michael. 1996. *The Privatization of Social Policy? Occupational Welfare and the Welfare State in America, Scandinavia and Japan*. London: Macmillan.

Shiroff, Alan. 1994. “Work, Welfare and Gender Equality: A New Typology.” Pp. 82–100 in *Gendering Welfare States*, edited by Diana Sainsbury. London: Sage Ltd.

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

Yitzhaki, Shlomo. 2002. “Do We Need a Separate Poverty Measurement?” *European Journal of Political Economy* 18: 61–85.

**Author Biography**

Tim Futing Liao is professor of sociology and professor of statistics at the University of Illinois at Urbana-Champaign. He served as head of the Sociology Department from 2004 to 2009. He was editor of SAGE’s Quantitative Applications in the Social Sciences series from 2004 to 2009 and editor of *Sociological Methodology* from 2009 to 2015. His current research focuses on structures and shapes of inequality, historical demography, and life-course sequence analysis.