Evaluating distributional differences in income inequality
(Liao (Socius 2016))

Naoki TANI, Yuki UEMURA

Institute of Economic Research, Kyoto University

April 12, 2024

1. Introduction
2. Theory

Aim of the paper

- This paper develops a refined Theil decomposition index that offers two ways of capturing between-group distributional differences: the first method further decomposes the within-group component of inequality into shared dispersion by groups and different dispersion between groups: the second method takes a quantile based approach.
- Applying their method to income data from 10 European countries illustrates the usefulness of their indexes

Example: problem of conventional Theil index

- $G1$: $[1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5, 6, 8, 10]$
- $G2$: $[1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 6, 7, 8]$
- The means of the two groups are equal: 3.933.
- Theil's measure of inequality: 0.134.
- Theil's between-group component is 0.000 and its within-group component is 0.134.

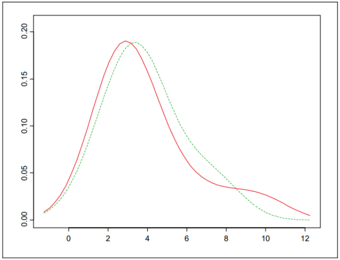

Figure 1. Kernel Density Distributions of the Two Groups in the Hypothetical Data Set.

Problem of conventional Theil index

- The conventional Theil index fails to capture
  - glass-ceiling effects: a phenomenon of a wider income gap between sexes at the top income group.
  - glass-floor effects: a phenomenon that one social group is less likely than another to be low-income earners.

Conventional Theil index

- The total amount of inequality measured by Theil index is

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

where x_i is the income of individual i, $\bar{x}$ is the overall mean income, and N is the sample size.

Conventional Theil index

- The between-group component of Theil index is

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

where y_k is the kth 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,

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

where y_{ik} is the income share of the ith individual within the kth group, and x_{ik} is the ith individual's income within group k.

Method 1 of the refined Theil index

- By using the observed shared range of the distributions of the G groups $[l, u]$, where l represents the largest (or the maximum) of the G number of minimum values of the G distributions and u represents the smallest (or the minimum) of the G number of maximum values of the same distributions,
- The within-component of conventional Theil index (Equation 3) can be separated into the two components:

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

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 of the refined Theil index

- In order to study how each group (eg. men and women) differs in the top decile or in the bottom decile of the income distribution,
- It first decomposes the overall inequality into the two components of between-quantile and within-quantile inequality: the Theil index in equation 1 is 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:

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

where y_q is the income share of the qth 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.

Method 2 of the refined Theil index

- It further decomposes 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:

$$T_q = \sum_{k=1}^{K} y_{kq} \ln \frac{\bar{x}_{kq}}{\bar{x}_q} + \sum_{k=1}^{K} y_{kq} \sum_{i=1}^{n_{kq}} y_{ikq} \ln \frac{x_{ikq}}{\bar{x}_{kq}}, \qquad (6)$$

where y_{kq} is the income share of the kth group within the qth quantile, $\bar{x}_{kq}$ is the mean income of the kth group within the qth quantile, y_{ikq} is the income share of the ith individual within the kth group and the qth quantile, and x_{ikq} is the income value of the ith individual within group k and quantile q.

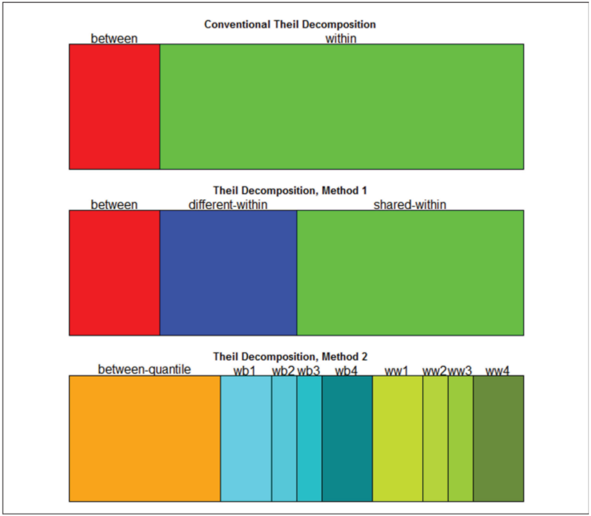

Figure 2. A Schematic Comparison of the Three Methods of Theil Decomposition.
Note. wb1 = within the first quantile (quartile here)/between group, *wb2* = within the second quantile/between group, and so on; *ww1* = within the first quantile/within group, *ww2* = within the second quantile/within group, and so on.

Result of method 1

Table 2. Theil Index Decomposition of Income Inequalities in 10 European Countries, circa 2000.

Country	Total	Between	Within	Within_w	Within_b
Austria	.308	.021	.287	.226	.061
				78.7%	21.3%
Belgium	.187	.017	.170	.165	.005
				97.1%	2.9%
France	.294	.013	.281	.227	.054
				80.9%	19.1%
Greece	.696	.003	.693	.579	.114
				83.5%	16.5%
Hungary	.255	.004	.251	.215	.036
				85.7%	14.3%
Ireland	.342	.003	.339	.324	.016
				95.4%	4.6%
Italy	.379	.002	.376	.368	.009
				97.6%	2.4%
Luxembourg	.226	.021	.205	.148	.057
				72.2%	27.8%
Russia	.633	.027	.606	.515	.091
				85.0%	15.0%
Spain	.377	.005	.372	.327	.046
				87.8%	12.2%

Note. The within-component decomposition reported in the first row for each country is performed using the shared-spread criterion and in the second row, the percentages of the within-within and the within-between components of the overall within-group component.

Result of method 2

Table 3. Between-gender 90:50 and 10:50 Inequality 2 × Log-ratios on the Basis of Theil Index Decomposition of Net Wage in 10 European Countries, circa 2000.

Country	90:50 Ratio	50:10 Ratio
Austria	15.523	12.772
Belgium	11.567	16.779
France	19.234	9.354
Greece	10.836	7.640
Hungary	17.863	8.802
Ireland	15.373	8.935
Italy	13.173	.076
Luxembourg	9.932	7.519
Russia	13.970	4.643
Spain	18.440	14.228

Note. Because zero income may render the bottom-quantile computation meaningless, the analysis is based on nonzero income data only. When the total number of quantiles is an even number, the middle quantile is taken to be the average of the two middle quantiles, here the average of the fifth and sixth deciles. The 2 × log-ratios can be interpreted as BIC differences.

Analysis

- In the result of method 1, none of the between-group component values are above 10 percent of their respective total inequality.
- But, the result of method 1 shows that about $10 - 20$ percent of the within-group inequality is actually explained by differences in dispersion or distribution.
- In the result of method 2, the evidence of a glass-ceiling effect is very strong for 9 of the 10 countries.