国際大学
INTERNATIONAL UNIVERSITY OF JAPAN

Economics & Management Series *EMS-2024-06*

The Roles of Location and Education in Income Inequality: An Analytical Framework

Takahiro Akita
International University of Japan

July 2024

IUJ Research Institute
International University of Japan

The Roles of Location and Education in Income Inequality: An Analytical Framework

Takahiro Akita*
Professor Emeritus, International University of Japan, akita@iuj.ac.jp

Abstract

This study develops a two-stage hierarchical inequality decomposition method as an analytical framework for the examination of the roles of rural/urban location and education in income inequality. It compares this method with a non-hierarchical inequality decomposition method. In the two-stage hierarchical inequality decomposition method, a hierarchical structure of a country is considered, where individuals are classified first into the rural and urban sectors and then into several education groups. Using the Theil indices, overall income inequality is decomposed hierarchically into the following three components: the between-sector, within-sector between-group, and within-sector within-group inequality components. The between-sector component evaluates income disparity between the rural and urban sectors, while the within-sector between-group component evaluates income inequality among the education groups, but adjusted for rural-urban differences in the structure of educational attainment. The within-sector within-group component assesses inequality within the education groups. In the non-hierarchical inequality decomposition method, overall income inequality is decomposed simultaneously but non-hierarchically based on individual attributes. In the context of location and education, overall income inequality, as measured by the Theil indices, is decomposed non-hierarchically into the following four components: the between-sector, between-group, location-education interaction, and within-sector within-group inequality components. The location-education interaction component assesses the extent of rural-urban differences in the income disparity among the education groups. It can take a negative value. Based on nation-wide household surveys, this study also presents the results of hierarchical and non-hierarchical inequality decomposition analyses for Indonesia and the Philippines.

Key words: roles of location and education in income inequality, Theil indices, two-stage hierarchical inequality decomposition method, non-hierarchical inequality decomposition method, Indonesia, the Philippines

* The author is grateful to the Japan Society for the Promotion of Science for its financial support (Grant-in-Aid for Scientific Research: 23K01409). The author would like to thank Raquel Celeste for providing household survey data from the Philippines.

1. Introduction

Based on household or labor force surveys, numerous empirical studies have been conducted to explore the factors of income inequality in both developing and developed countries. These studies often decomposed overall income inequality by population subgroups, employing the generalized entropy class of inequality measures such as the Theil indices and the squared coefficient of variation.[1] They examined age, gender, ethnicity, occupation, rural/urban location, region, and education as potential factors contributing to income inequality; but, many of them identified income disparities between rural and urban locations and among education groups as major factors of income inequality (see, for example, Ikemoto, 1985; Glewwe, 1986; Ching, 1991; Mishra and Parikh, 1992; Tsakloglou, 1993; Estudillo, 1997; Akita, et al., 1999; Akita and Szeto, 2000; Liu, 2001; Mukhopadhaya, 2003; Rao, et al., 2003; Balisacan and Fuwa, 2004; Eastwood and Lipton, 2004; Shorrocks and Wan, 2005; Elbers, et al., 2008; Chongvilaivan and Kim, 2016).[2]

Against this background, this study develops an analytical framework to examine the roles of rural/urban location and education in income inequality. Since significant differences exist between rural and urban areas in the structure of educational attainment, this framework replies on a two-stage hierarchical inequality decomposition method developed by Akita and Miyata (2008). In the two-stage hierarchical inequality decomposition method, we consider a hierarchical structure of a country, where individuals are classified first into the rural and urban sectors and then into several education groups. Using the Theil indices, overall income inequality is decomposed hierarchically into three components: the between-sector, within-sector between-group, and within-sector within-group inequality components. The between-sector component evaluates income disparity between the rural and urban sectors, while the within-sector between-group component evaluates income inequality among the education

[1] They include Fishlow (1972), Fields (1979), Mookherjee and Shorrocks (1982), Cowell (1984), Ikemoto (1985), Glewwe (1986), Ching (1991), Mishra and Parikh (1992), Tsakloglou (1993), Cowell and Jenkins (1995), Estudillo (1997), Akita, et al. (1999), Akita and Szeto (2000), Dickey (2001), Liu (2001), Gray, et al. (2003), Mukhopadhaya (2003), Rao, et al. (2003), Balisacan and Fuwa (2004), Eastwood and Lipton (2004), Shorrocks and Wan (2005), Borooah, et al. (2006), Motonishi (2006), Elbers, et al. (2008), Chongvilaivan and Kim (2016).

[2] Some of these studies used household consumption expenditures instead of household incomes. According to the studies, the disparity between rural and urban locations accounted for 10-20% of overall income inequality, while the disparity among education groups explained 20-30% of overall income inequality.

groups, but adjusted for rural-urban differences in the structure of educational attainment. The within-sector within-group component assesses inequality within the education groups.

This paper is organized as follows. The next section introduces a two-stage hierarchical inequality decomposition method as an analytical framework for the examination of the roles of rural/urban location and education. It also compares this method with a non-hierarchical inequality decomposition method. Using hypothetical examples, section 3 performs hierarchical and non-hierarchical inequality decomposition analyses to examine the roles of rural/urban location and education in income inequality. Based on nation-wide household surveys in 2018, section 4 presents the results of hierarchical and non-hierarchical inequality decomposition analyses for Indonesia and the Philippines. The final section provides concluding remarks.

2. Hierarchical and Non-hierarchical Inequality Decomposition Methods

This section presents a two-stage hierarchical inequality decomposition method as an analytical framework for the examination of the roles of rural/urban location and education. It also compares this method with a non-hierarchical inequality decomposition method proposed by Tang and Petrie (2009).

2.1. Two-stage Hierarchical Decomposition of Income Inequality by Location and Education

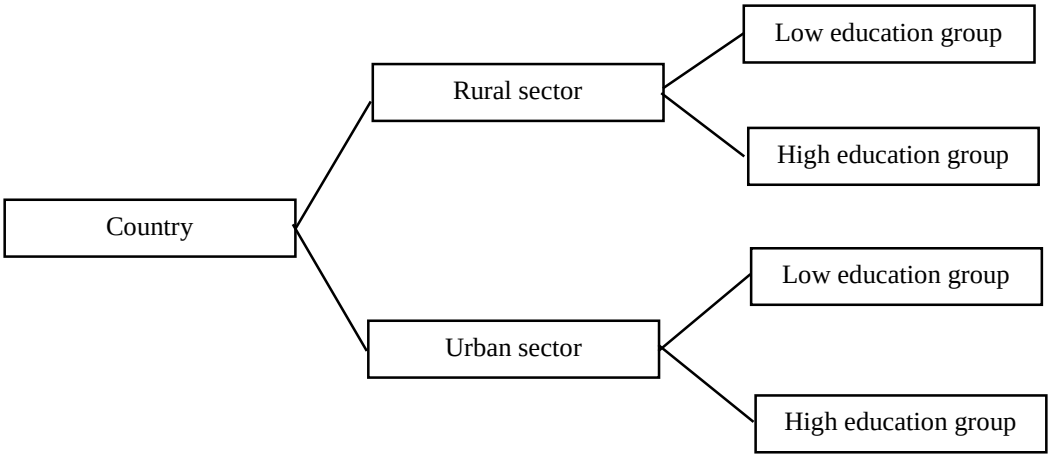
Consider a country consisting of the rural and urban sectors, each with two levels of education (low and high). Suppose that all individuals are first grouped into the urban and rural sectors, and then, individuals in each sector are classified into the low and high education groups (see Fig. 1). If we let y_{ijk}, n_{ij}, n, and μ be, respectively, income of individual k in education group j of sector i, total number of individuals in education group j of sector i, total number of all individuals, and mean income of all individuals ($i = 1, 2; j = 1,2$), then overall income inequality can be measured by the Theil L index as follows.[3]

[3] The Theil indices belong to the generalized entropy class of measures and satisfy three fundamental principles: income homogeneity (or mean independence), population homogeneity (or population independence) and the Pigou-Dalton principle of transfers (Anand, 1983; Fields, 2001; Akita and Kataoka, 2022). Income homogeneity implies that an inequality measure remains unchanged if everyone's income is changed by the same proportion,

$$L = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu}{y_{ijk}}\right) \quad (1)$$

where $\mu = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} y_{ijk}$ and $n = \sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}$.

Figure 1. Hierarchical Structure
Location (Rural & Urban Sectors) ⟹ Education (Low & High Education Groups)



Overall income inequality given by equation (1) can be decomposed hierarchically into three components as follows (see Fig. 2).

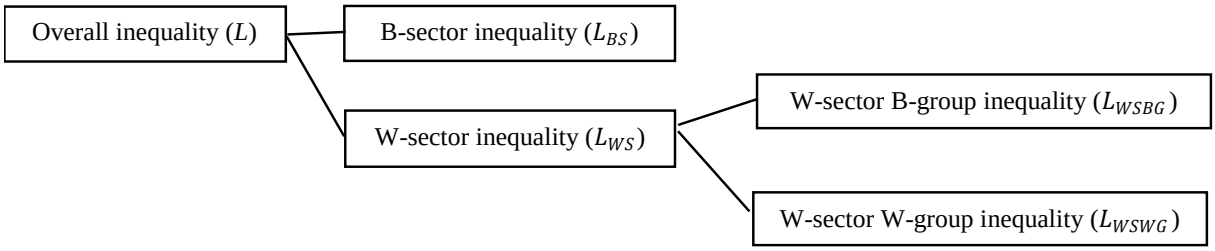
$$L = L_{BS} + L_{WSBG} + L_{WSWG}. \quad (2)$$

Eq. (2) presents a two-stage hierarchical inequality decomposition method (see Appendix 1 for the derivation of Eq. (2)). $L_{BS} = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)\ln\left(\frac{\mu}{\mu_{i\cdot}}\right)$ is the between-sector inequality component (inequality between the rural and urban sectors), where $n_{i\cdot} = \sum_{j=1}^{2} n_{ij}$ and $\mu_{i\cdot} = \frac{1}{n_{i\cdot}}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} y_{ijk}$ are total number of individuals in sector *i* and mean income of individuals in sector *i*, respectively. $L_{WSBG} = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)L_{BGi}$ is the within-sector between-group inequality

while population homogeneity denotes that an inequality measure remains unchanged if the number of individuals at each income level is changed by the same proportion. The Pigou-Dalton transfer principle implies that any income transfer from a richer to a poorer individual that does not reverse their relative ranks in income lowers the value of an inequality index. The Theil indices are also decomposable additively by population sub-groups, that is, can be expressed as the sum of the between- and within-group components (Bourguignon, 1979; Shorrocks, 1980; Anand, 1983).

component, where $L_{BGi} = \sum_{j=1}^{2} \left(\frac{n_{ij}}{n_{i\cdot}}\right) \ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)$ is inequality between the low and high education groups in sector *i* and $\mu_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$ is mean income of individuals in education group *j* of sector *i*. L_{WSBG} is the population-weighted average of L_{BGi}. Lastly, $L_{WSWG} = \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{n_{ij}}{n}\right) L_{ij}$ is the within-sector within-group inequality component, where $L_{ij} = \sum_{k=1}^{n_{ij}} \left(\frac{1}{n_{ij}}\right) \ln\left(\frac{\mu_{ij}}{y_{ijk}}\right)$ is inequality within education group *j* of sector *i*. L_{WSWG} is the population-weighted average of L_{ij}.

Figure 2. Two-stage Hierarchical Decomposition of Income Inequality (Theil *L*)
Location (Rural & Urban Sectors) ⟹ Education (Low & High Education Groups)



Overall income inequality can also be measured by the Theil *T* index as follows.

$$T = \frac{1}{n} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n_{ij}} \left(\frac{y_{ijk}}{\mu}\right) \ln\left(\frac{y_{ijk}}{\mu}\right). \quad (3)$$

Like the Theil *L* index, the Theil *T* index can be decomposed hierarchically into three components as follows (see Appendix 1 for the derivation of Eq. 4).

$$T = T_{BS} + T_{WSBG} + T_{WSWG}. \quad (4)$$

$T_{BS} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n} \frac{\mu_{i\cdot}}{\mu}\right) \ln\left(\frac{\mu_{i\cdot}}{\mu}\right)$ is the between-sector inequality component (inequality between the rural and urban sectors). $T_{WSBG} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n} \frac{\mu_{i\cdot}}{\mu}\right) T_{BGi}$ is the within-sector between-group inequality component, where $T_{BGi} = \sum_{j=1}^{2} \left(\frac{n_{ij}}{n_{i\cdot}} \frac{\mu_{ij}}{\mu_{i\cdot}}\right) \ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right)$ is inequality between the low and high education groups in sector *i*. T_{WSBG} is the income-weighted average of T_{BGi}. $T_{WSWG} =$

$\sum_{i=1}^{2}\sum_{j=1}^{2}\left(\frac{n_{ij}}{n}\frac{\mu_{ij}}{\mu}\right)T_{ij}$ is the within-sector within-group inequality component, where $T_{ij} = \sum_{k=1}^{n_{ij}}\left(\frac{1}{n_{ij}}\frac{y_{ijk}}{\mu_{ij}}\right)\ln\left(\frac{y_{ijk}}{\mu_{ij}}\right)$ is inequality within education group *j* of sector *i*. T_{WSWG} is the income-weighted average of T_{ij}.

In the two-stage hierarchical decomposition of income inequality, the order of decomposition can be reversed, that is, all individuals are first grouped into the low and high education groups, and then, individuals in each group are classified into the rural and urban sectors (see Fig. 3). In this two-stage hierarchical decomposition, overall income inequality, as measured by the Theil *L* index, can be decomposed hierarchically into the following three components (see Fig. 4).

$$L = L_{BG} + L_{WGBS} + L_{WGWS}. \quad (5)$$

L_{BG}, L_{WGBS}, and L_{WGWS} are the between-group, within-group between-sector inequality, and within-group within-sector inequality components, respectively. When measured by the Theil *T* index, we have

$$T = T_{BG} + T_{WGBS} + T_{WGWS}, \quad (6)$$

where T_{BG}, T_{WGBS}, and T_{WGWS} are the between-group, within-group between-sector, and within-group within-sector inequality components, respectively.

Figure 3. Hierarchical Structure
Education (Low & High Education Groups) ⟹ Location (Rural & Urban Sectors)

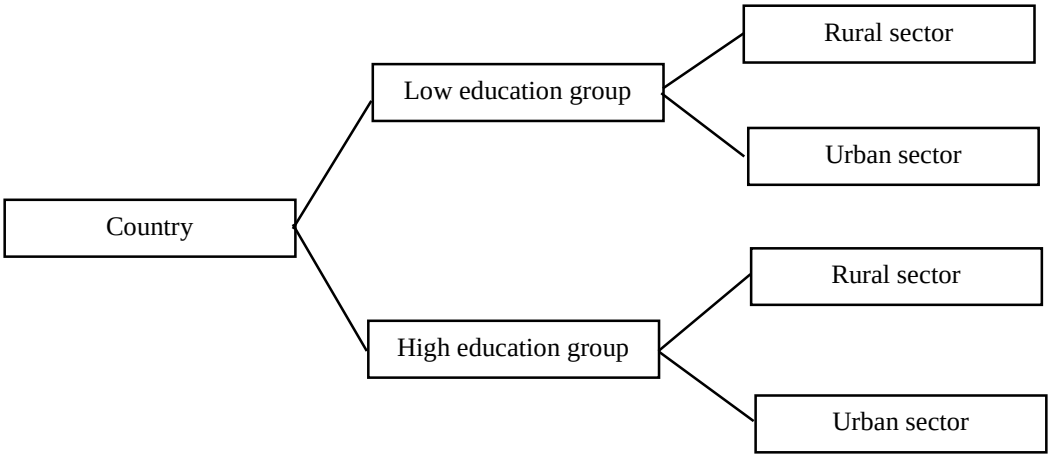
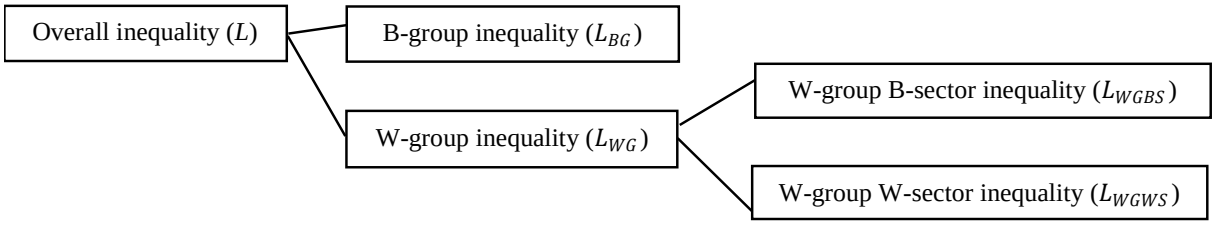


Figure 4. Two-stage Hierarchical Decomposition of Income Inequality (Theil *L*) Education (Low & High Education Groups) ⟹ Location (Rural & Urban Sectors)



2.2. Non-hierarchical Decomposition of Income Inequality by Location and Education

As an alternative multivariate decomposition method, Tang and Petrie (2009) suggested a non-hierarchical inequality decomposition method, where overall income inequality is decomposed simultaneously but non-hierarchically based on individual attributes such as age, gender, education, location, occupation and ethnicity. In the context of location and education, overall income inequality, as measured by the Theil *L* index, is decomposed non-hierarchically as follows (see Fig. 5).

$$L = L_{BS} + L_{BG} + L_{ISG} + L_{WSWG}, \quad (7)$$

where L_{ISG} is the location-education interaction component. Using (2) and (7), the interaction component is given by

$$L_{ISG} = L_{WSBG} - L_{BG}. \quad (8)$$

Since Eq. (8) is modified to

$$L_{ISG} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n}\right) L_{BGi} - L_{BG} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n}\right) (L_{BGi} - L_{BG}),$$

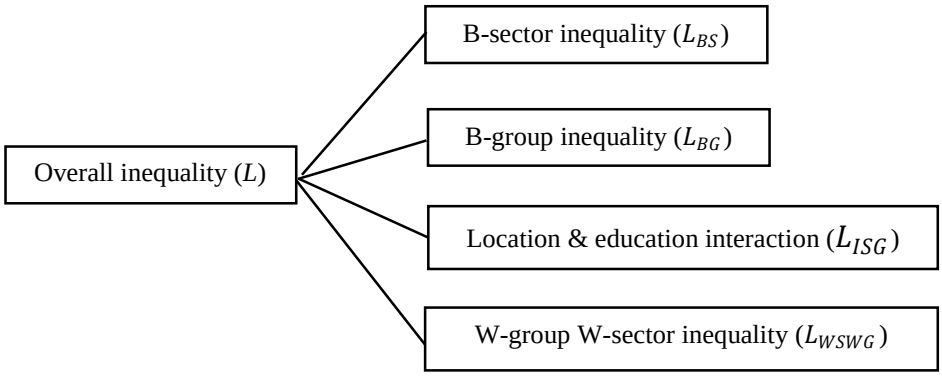
the location-education interaction component assesses the extent of rural-urban differences in the income disparity between the low and high education groups. It can take a negative value. When measured by the Theil *T* index, a non-hierarchical decomposition equation is given by

$$T = T_{BS} + T_{BG} + T_{ISG} + T_{WSWG}, \quad (9)$$

where T_{ISG} is the location-education interaction component. Using (4) and (9), the interaction component is given by

$$T_{ISG} = T_{WSBG} - T_{BG}. \quad (10)$$

Figure 5. Non-hierarchical Decomposition of Income Inequality by Location and Education (Theil *L*)



It is important to note that the within-sector within-group inequality component is equivalent to the within-group within-sector inequality component as follows.

$$L_{WSWG} = L_{WGWS} \text{ and } T_{WSWG} = T_{WGWS}. \quad (11)$$

Using this relationship and equating (2) and (5), we obtain

$$L_{WSBG} - L_{BG} = L_{WGBS} - L_{BS}. \quad (12)$$

Eq. (12) is the location-education (or education-location) interaction component given by Eq. (8). Similarly, equating (4) and (6), we have

$$T_{WSBG} - T_{BG} = T_{WGBS} - T_{BS}. \quad (13)$$

Eq. (13) is the location-education (or education-location) interaction component given by Eq. (10).

3. Hierarchical and Non-hierarchical Inequality Decomposition Analyses: Hypothetical Examples

Using hypothetical examples, this section performs hierarchical and non-hierarchical inequality decomposition analyses to examine the roles of rural/urban location and education in income inequality.

3.1. Two-stage Hierarchical Inequality Decomposition Analysis

To perform a two-stage hierarchical inequality decomposition analysis by location and education, let us consider a country comprising the rural and urban sectors, each with two levels

of education (low and high). Suppose that there are 20 individuals, of which 12 reside in the rural sector and 8 in the urban sector. Among the 12 rural individuals, 8 belong to the low education group, while 4 are in the high education group. Among the 8 urban individuals, 4 belong to the low education group, while 4 are in the low education group. Table 1 provides incomes for these 20 individuals.

Table 1. Distribution of Incomes for 20 Individuals (Location ⟹ Education)

Individual (ID)	Location	Education	Income	Population share	Income share
1	Rural	Low	4	0.05	0.04
2	Rural	Low	2	0.05	0.02
3	Rural	Low	3	0.05	0.03
4	Rural	Low	4	0.05	0.04
5	Rural	Low	4	0.05	0.04
6	Rural	Low	4	0.05	0.04
7	Rural	Low	2	0.05	0.02
8	Rural	Low	5	0.05	0.05
9	Rural	High	3	0.05	0.03
10	Rural	High	7	0.05	0.07
11	Rural	High	5	0.05	0.05
12	Rural	High	4	0.05	0.04
13	Urban	Low	4	0.05	0.04
14	Urban	Low	7	0.05	0.07
15	Urban	Low	4	0.05	0.04
16	Urban	Low	3	0.05	0.03
17	Urban	High	8	0.05	0.08
18	Urban	High	5	0.05	0.05
19	Urban	High	16	0.05	0.16
20	Urban	High	6	0.05	0.06
Total			100		
Mean			5		

Using incomes for all individuals, we obtain overall income inequality. By the Theil L index, overall income inequality is calculated as follows.

$$L = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu}{y_{ijk}}\right) = 0.122.$$

On the other hand, income inequality between the rural and urban sectors (L_{BS}) is calculated by

using the mean incomes for these sectors presented in Table 2.

$$L_{BS} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n}\right) \ln\left(\frac{\mu}{\mu_{i\cdot}}\right) = \left(\frac{12}{20}\right) \ln\left(\frac{5}{3.92}\right) + \left(\frac{8}{20}\right) \ln\left(\frac{5}{6.63}\right) = 0.034.$$

Next, to obtain the within-sector between-group inequality component (L_{WSBG}), we need to have the mean incomes for the low and high education groups in each of the rural and urban sectors. Using the mean incomes presented in Table 3, we obtain

$$L_{BG1} = \sum_{j=1}^{2} \left(\frac{n_{1j}}{n_{1\cdot}}\right) \ln\left(\frac{\mu_{1\cdot}}{\mu_{1j}}\right) = \left(\frac{8}{12}\right) \ln\left(\frac{3.92}{3.5}\right) + \left(\frac{4}{12}\right) \ln\left(\frac{3.92}{4.75}\right) = 0.011,$$

$$L_{BG2} = \sum_{j=1}^{2} \left(\frac{n_{2j}}{n_{2\cdot}}\right) \ln\left(\frac{\mu_{2\cdot}}{\mu_{2j}}\right) = \left(\frac{4}{8}\right) \ln\left(\frac{6.63}{4.5}\right) + \left(\frac{4}{8}\right) \ln\left(\frac{6.63}{8.75}\right) = 0.054.$$

Therefore, the within-sector between-group inequality component is calculated as follows.

$$L_{WSBG} = \sum_{i=1}^{2} \left(\frac{n_{i\cdot}}{n}\right) L_{BGi} = \left(\frac{12}{20}\right)(0.011) + \left(\frac{8}{20}\right)(0.054) = 0.028.$$

On the other hand, to obtain the within-sector within-group inequality component (L_{WSWG}), we need to calculate inequality within education group j of sector i: $L_{ij} = \sum_{k=1}^{n_{ij}} \left(\frac{1}{n_{ij}}\right) \ln\left(\frac{\mu_{ij}}{y_{ijk}}\right)$. Using individual incomes in Table 1 and mean incomes in Table 3, we obtain $L_{11} = 0.048$, $L_{12} = 0.048$, $L_{21} = 0.050$, and $L_{22} = 0.106$. Thus, we have

$$L_{WSWG} = \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{n_{ij}}{n}\right) L_{ij}$$
$$= \left(\frac{8}{20}\right)(0.048) + \left(\frac{4}{20}\right)(0.048) + \left(\frac{4}{20}\right)(0.050) + \left(\frac{4}{20}\right)(0.106) = 0.060.$$

Table 2. Mean Incomes for the Rural and Urban Sectors ($\mu_{i\cdot}$)

Location (i)	Mean income	Population share	Income share
Rural (1)	3.92	0.60	0.47
Urban (2)	6.63	0.40	0.53
Total	5.00	1.00	1.00

Table 3. Mean Incomes for the Low and High Education Groups in Each of the Rural and Urban Sectors (μ_{ij})

Location (i)	Education (j)	Mean income	Population share	Income share
Rural (1)	Low (1)	3.50	0.67	0.60
	High (2)	4.75	0.33	0.40
	Rural total	3.92	1.00	1.00
Urban (2)	Low (1)	4.50	0.50	0.34
	High (2)	8.75	0.50	0.66
	Urban total	6.63	1.00	1.00

In conclusion, we obtain the following two-stage hierarchical inequality decomposition equation by the Theil L index.

$L = L_{BS} + L_{WSBG} + L_{WSWG} = 0.034 + 0.028 + 0.060 = 0.122.$

Table 4 combines the decomposition of income inequality by location (rural and urban sectors) with the decomposition of income inequality by education (low and high education groups) in each sector, while Table 5 summarizes the result of the two-stage hierarchical inequality decomposition analysis (location $\Rightarrow$ education). Inequality between the rural and urban sectors contributes 27.8% to overall income inequality. Meanwhile, after adjusting for rural-urban differences in the structure of education, income inequality between the low and high education groups contributes 23.1%. The remaining 49.1% is attributed to the within-sector within-group inequality component.

Table 4. Decomposition of Income Inequality by Location (Rural & Urban Sectors) and Decomposition of Inequality by Education (Low & High Education Groups) in Each Sector

	Inequality (Theil *L*)	Contrib. (%)	Pop. share (%)		Inequality (Theil *L*)	Contrib. (%)	Pop. share (%)
Total (A) = (B) + (C)	0.122	100.0	100.0				
B-sector (B)	0.034	27.8					
W-sector (C) = (a) + (b)	0.088	72.2					
Rural sector (a) = (c) + (d)	0.059	28.9	60.0	Urban sector (b) = (e) + (f)	0.132	43.3	40.0
B-group (c)	0.011	5.3		B-group (e)	0.054	17.8	
W-group (d)	0.048	23.6		W-group (f)	0.078	25.5	
Low educ. group	(0.048)	15.7	40.0	Low educ. group	(0.050)	8.2	20.0
High educ. group	(0.048)	7.9	20.0	High educ. group	(0.106)	17.3	20.0

(Note) 'Contrib. (%)' is the % contribution of each component to overall income inequality, 'Pop. share' is the population share of each component to total population, 'B-sector' is the between-sector inequality component, 'W-sector' is the within-sector inequality component, 'B-group' is the between-group inequality component, and 'W-group' is the within-group inequality component.

Table 5. Two-stage Hierarchical Decomposition of Income Inequality
Location (Rural & Urban Sectors) ⟹ Education (Low & High Education Groups)

	Inequality (Theil *L*)	Contribution (%)
Total (L)	0.122	100.0
B-sector (L_{BS})	0.034	27.8
W-sector B-group (L_{WSBG})	0.028	23.1
W-sector W-group (L_{WSWG})	0.060	49.1
Low education group	(0.029)	23.9
High education group	(0.031)	25.2

(Note) 'B-sector' is the between-sector inequality component, 'W-sector B-group' is the within-sector between-group inequality component, and 'W-sector W-group' is the within-sector within-group inequality component.

Similarly, we can perform a two-stage hierarchical decomposition analysis by reversing the order of decomposition, that is, from education to location. There are 20 individuals, of which 12 belong to the low education group and 8 belong to the high education group (see Table 6). First, income inequality between the low and high education groups (L_{BG}) is calculated by using the mean incomes for these groups presented in Table 7.

$$L_{BG} = \sum_{j=1}^{2} \left(\frac{n_{\cdot j}}{n}\right) \ln\left(\frac{\mu}{\mu_{\cdot j}}\right) = \left(\frac{12}{20}\right) \ln\left(\frac{5}{3.83}\right) + \left(\frac{8}{20}\right) \ln\left(\frac{5}{6.75}\right) = 0.039.$$

Next, to obtain the within-group between-sector inequality component (L_{WGBS}), we need mean

incomes for the rural and urban sectors in each of the low and high education groups. Using the mean incomes presented in Table 8, we obtain

$$L_{BS1} = \sum_{i=1}^{2} \left(\frac{n_{i1}}{n_{\cdot 1}}\right) \ln\left(\frac{\mu_{\cdot 1}}{\mu_{i1}}\right) = \left(\frac{8}{12}\right) \ln\left(\frac{3.83}{3.5}\right) + \left(\frac{4}{12}\right) \ln\left(\frac{3.83}{4.5}\right) = 0.007,$$

$$L_{BS2} = \sum_{i=1}^{2} \left(\frac{n_{i2}}{n_{\cdot 2}}\right) \ln\left(\frac{\mu_{\cdot 2}}{\mu_{i2}}\right) = \left(\frac{4}{8}\right) \ln\left(\frac{6.75}{4.75}\right) + \left(\frac{4}{8}\right) \ln\left(\frac{6.75}{8.75}\right) = 0.046.$$

Therefore, the within-group between-sector inequality component is calculated as follows.

$$L_{WGBS} = \sum_{j=1}^{2} \left(\frac{n_{\cdot j}}{n}\right) L_{BSj} = \left(\frac{12}{20}\right)(0.007) + \left(\frac{8}{20}\right)(0.046) = 0.023.$$

Since the within-group within-sector inequality component is the same as the within-sector within-group inequality component (see Eq. 11), we obtain

$$L_{WGWS} = L_{WSWG} = 0.060.$$

Table 6. Distribution of Incomes for 20 Individuals (Education ⟹ Location)

Individual (ID)	Education	Location	Income	Population share	Income share
1	Low	Rural	4	0.05	0.04
2	Low	Rural	2	0.05	0.02
3	Low	Rural	3	0.05	0.03
4	Low	Rural	4	0.05	0.04
5	Low	Rural	4	0.05	0.04
6	Low	Rural	4	0.05	0.04
7	Low	Rural	2	0.05	0.02
8	Low	Rural	5	0.05	0.05
13	Low	Urban	4	0.05	0.04
14	Low	Urban	7	0.05	0.07
15	Low	Urban	4	0.05	0.04
16	Low	Urban	3	0.05	0.03
9	High	Rural	3	0.05	0.03
10	High	Rural	7	0.05	0.07
11	High	Rural	5	0.05	0.05
12	High	Rural	4	0.05	0.04
17	High	Urban	8	0.05	0.08
18	High	Urban	5	0.05	0.05
19	High	Urban	16	0.05	0.16
20	High	Urban	6	0.05	0.06
Total			100		
Mean			5		

Table 7. Mean Incomes for the Low and High Education Groups ($\mu_{\cdot j}$)

Education (j)	Mean income	Population share	Income share
Low	3.83	0.60	0.46
High	6.75	0.40	0.54
Total	5.00	1.00	1.00

Table 8. Mean Incomes for the Rural and Urban Sectors in Each of the Low and High Education Groups (μ_{ij})

Education (j)	Location (i)	Mean income	Population share	Income share
Low (1)	Rural (1)	3.50	0.67	0.61
	Urban (2)	4.50	0.33	0.39
	Low total	3.83	1.00	1.00
High (2)	Rural (1)	4.75	0.50	0.35
	Urban (2)	8.75	0.50	0.65
	High total	6.75	1.00	1.00

In conclusion, we obtain the following two-stage hierarchical inequality decomposition equation by the Theil L index.

$$L = L_{BG} + L_{WGBS} + L_{WGWS} = 0.039 + 0.023 + 0.060 = 0.122.$$

Table 9 combines the decomposition of income inequality by education (low and high education groups) with the decomposition of income inequality by location (rural and urban sectors) in each education group, while Table 10 summarizes the result of the two-stage hierarchical inequality decomposition analysis (education $\Rightarrow$ location). Inequality between the low and high education groups contributes 32.3% to overall income inequality. Meanwhile, after adjusting for differences between the low and high education groups in the structure of rural and urban locations, income inequality between the rural and urban sectors contributes 18.6%. The remaining 49.1% is attributed to the within-group within-sector inequality component.

Table 9. Decomposition of Income Inequality by Education (Low & High Education Groups) and Decomposition of Inequality by Location (Rural & Urban Sectors) in Each Group

	Inequality (Theil *L*)	Contrib. (%)	Pop. share (%)		Inequality (Theil *L*)	Contrib. (%)	Pop. share (%)
Total (A) = (B) + (C)	0.122	100.0	100.0				
B-group (B)	0.039	32.3					
W-group (C) = (a) + (b)	0.083	67.7					
Low (a) = (c) + (d)	0.056	27.4	60.0	High (b) = (e) + (f)	0.123	40.3	40.0
B-sector (c)	0.007	3.5		B-sector (e)	0.046	15.1	
W-sector (d)	0.049	23.9		W-sector (f)	0.077	25.2	
Rural sector	(0.048)	15.7	40.0	Rural sector	(0.048)	7.9	20.0
Urban sector	(0.050)	8.2	20.0	Urban sector	(0.106)	17.3	20.0

(Note) 'Contrib. (%)' is the % contribution of each component to overall income inequality, 'Pop. share' is the population share of each component to total population, 'B-group' is the between-group inequality component, 'W-group' is the within-group inequality component, 'B-sector' is the between-sector inequality component, and 'W-sector' is the within-sector inequality component.

Table 10. Two-stage Hierarchical Decomposition of Income Inequality
Education (Low & High Education Groups) ⟹ Location (Rural & Urban Sectors)

	Inequality (Theil *L*)	Contribution (%)
Total (L)	0.122	100.0
B-group (L_{BG})	0.039	32.3
W-group B-sector (L_{WGBS})	0.023	18.6
W-group W-sector (L_{WGWS})	0.060	49.1
Rural sector	(0.029)	23.6
Urban sector	(0.031)	25.5

(Note) 'B-group' is the between-group inequality component, 'W-group B-sector' is the within-group between-sector inequality component, and 'W-group W-sector' is the within-group within-sector inequality component.

We can perform a similar two-stage hierarchical inequality decomposition analysis by location and education using the Theil *T* index. The result is presented in Appendix 2.

3.2. Non-hierarchical Inequality Decomposition Analysis

To conduct a non-hierarchical inequality decomposition analysis using individual incomes given in Table 1 (or Table 6), we need to calculate the location-education interaction

component (L_{ISG}) defined by Eq. (8). Using the values for L_{WSBG} and L_{BG} presented in Tables 5 and 10, respectively, we obtain

$$L_{ISG} = L_{WSBG} - L_{BG} = 0.028 - 0.039 = -0.011.$$

The following presents the result of the non-hierarchical decomposition analysis by location and education.

$$L = L_{BS} + L_{BG} + L_{ISG} + L_{WSWG} = 0.034 + 0.039 - 0.011 + 0.060 = 0.122.$$

Table 11 compares the result of the non-hierarchical inequality decomposition analysis with that of the two-stage hierarchical inequality decomposition analysis (location ⇒ education) presented in Table 5. The contribution of the location-education interaction component to overall income inequality is negative at -9.2%, indicating that there is a rural-urban difference in the income disparity between the low and high education groups.

Table 11. Hierarchical and Non-hierarchical Inequality Decomposition Analyses

	Hierarchical decomposition		Non-hierarchical decomposition	
	Inequality (Theil *L*)	Contribution (%)	Inequality (Theil *L*)	Contribution (%)
Total (L)	0.122	100.0	0.122	100.0
B-sector (L_{BS})	0.034	27.8	0.034	27.8
B-group (L_{BG})			0.039	32.3
W-sector B-group (L_{WSBG})	0.028	23.1		
Interaction component (L_{ISG})			-0.011	-9.2
W-sector W-group (L_{WSWG})	0.060	49.1	0.060	49.1
Low education group	(0.029)	23.9	(0.029)	23.9
High education group	(0.031)	25.2	(0.031)	25.2

We can perform a non-hierarchical inequality decomposition analysis by location and education using the Theil *T* index. The result is presented in Appendix 2.

4. Hierarchical and Non-hierarchical Inequality Decomposition Analyses: Empirical Evidence from Indonesia and the Philippines

Based on nation-wide household surveys in 2018, this section presents the results of hierarchical and non-hierarchical inequality decomposition analyses for Indonesia and the Philippines. Tables 12 and 13 combine the decomposition of expenditure inequality by location with the decomposition of expenditure inequality by education in each sector in Indonesia and

the Philippines, respectively. On the other hand, Tables 14 and 15 compare the result of the two-stage hierarchical inequality decomposition analysis (location ⇒ education) with that of the non-hierarchical inequality decomposition analysis in Indonesia and the Philippines,

In 2018, Indonesia and the Philippines have almost the same level of overall expenditure inequality at 0.26 by the Theil L index. But, its determinants are different between these two countries. While the rural-urban disparity (L_{BS}) constitutes around 11% of overall expenditure inequality in both countries, the contribution of the disparity between the four education groups is much smaller in Indonesia than in the Philippines. According to the non-hierarchical inequality decomposition analysis (Tables 14 and 15), the contribution of the location-education interaction component to overall expenditure inequality is negative in both countries, indicating that there is a rural-urban difference in the expenditure disparity between the four education groups. After adjusting for this rural-urban difference, the expenditure disparity between the four education groups (L_{WSBG}) contributes 13.0% in Indonesia while 20.6% in the Philippines. As a consequence, Indonesia's within-sector within-group inequality component (L_{WSWG}) has a larger contribution to overall inequality than the Philippines' (76.0% vs. 68.2%).

Table 12. Decomposition of Income Inequality by Location and Decomposition of Inequality by Education in Each Sector in Indonesia

	Inequality (Theil L)	Contrib. (%)	Pop. share (%)		Inequality (Theil L)	Contrib. (%)	Pop. share (%)
Total (A) = (B) +(C)	0.262	100.0	100.0				
B-sector (B)	0.029	11.0					
W-sector (C) = (a) + (b)	0.233	89.0					
Rural sector (a) = (c) + (d)	0.180	31.1	45.3	Urban sector (b) = (e) + (f)	0.277	57.8	54.7
B-group (c)	0.016	2.7		B-group (e)	0.049	10.3	
W-group (d)	0.165	28.5		W-group (f)	0.227	47.5	
No or incomplete primary	0.163	8.6	13.9	No	0.247	8.4	8.9
Primary	0.154	9.6	16.4	Primary	0.205	9.6	12.2
Secondary	0.173	8.9	13.5	Secondary	0.229	23.9	27.3
Tertiary	0.218	1.3	1.6	Tertiary	0.235	5.6	6.2

(Source) National Socioeconomic Survey in 2018.

Table 13. Decomposition of Income Inequality by Location and Decomposition of Inequality by Education in Each Sector: the Philippines

	Inequality (Theil *L*)	Contrib. (%)	Pop. share (%)		Ineq1uality (Theil *L*)	Contrib. (%)	Pop. share (%)
Total (A) = (B) +(C)	0.265	100.0	100.0				
B-sector (B)	0.030	11.2					
W-sector (C) = (a) + (b)	0.235	88.8					
Rural sector (a) = (c) + (d)	0.223	40.3	47.8	Urban sector (b) = (e) + (f)	0.246	48.5	52.2
B-group (c)	0.051	9.1		B-group (e)	0.058	11.5	
W-group (d)	0.173	31.2		W-group (f)	0.188	37.0	
No or incomplete primary	0.150	7.1	12.6	No	0.186	4.6	6.5
Primary	0.155	6.0	10.3	Primary	0.172	4.4	6.8
Secondary	0.167	10.8	17.1	Secondary	0.161	13.8	22.6
Tertiary	0.247	7.3	7.8	Tertiary	0.232	14.2	16.3

(Source) Family Income and Expenditure Survey in 2018.

Table 14. Hierarchical and Non-hierarchical Inequality Decomposition Analyses: Indonesia

	Hierarchical decomposition		Non-hierarchical decomposition	
	Inequality (Theil *L*)	Contrib. (%)	Inequality (Theil *L*)	Contrib. (%)
Total (L)	0.262	100.0	0.262	100.0
B-sector (L_{BS})	0.029	11.0	0.029	11.0
B-group (L_{BG})			0.051	19.4
W-sector B-group (L_{WSBG})	0.034	13.0		
Interaction component (L_{ISG})			-0.017	-6.4
W-sector W-group (L_{WSWG})	0.199	76.0	0.199	76.0

(Source) National Socioeconomic Survey in 2018.

Table 15. Hierarchical and Non-hierarchical Inequality Decomposition Analyses: the Philippines

	Hierarchical decomposition		Non-hierarchical decomposition	
	Inequality (Theil L)	Contrib. (%)	Inequality (Theil L)	Contrib. (%)
Total (L)	0.265	100.0	0.265	100.0
B-sector (L_{BS})	0.030	11.2	0.030	11.2
B-group (L_{BG})			0.071	26.9
W-sector B-group (L_{WSBG})	0.055	20.6		
Interaction component (L_{ISG})			-0.017	-6.3
W-sector W-group (L_{WSWG})	0.181	68.2	0.181	68.2

(Source) Family Income and Expenditure Survey in 2018.

5. Concluding Remarks

This study developed a two-stage hierarchical inequality decomposition method as an analytical framework for the examination of the roles of rural/urban location and education in income inequality. It compared this method with a non-hierarchical inequality decomposition method.

In the two-stage hierarchical inequality decomposition method, a hierarchical structure of a country is considered, where individuals are classified first into the rural and urban sectors and then into several education groups. Using the Theil indices, overall income inequality is decomposed hierarchically into the following three components: the between-sector, within-sector between-group, and within-sector within-group inequality components. The between-sector component evaluates income disparity between the rural and urban sectors, while the within-sector between-group component evaluates income inequality among the education groups, but adjusted for rural-urban differences in the structure of educational attainment. The within-sector within-group component assesses inequality within the education groups.

On the other hand, in the non-hierarchical inequality decomposition method, overall income inequality is decomposed simultaneously but non-hierarchically based on individual attributes. In the context of location and education, overall income inequality, as measured by the Theil indices, is decomposed non-hierarchically into the four components: the between-sector, between-group, location-education interaction, and within-sector within-group

inequality components. The location-education interaction component assesses the extent of rural-urban differences in the income disparity among the education groups. It can take a negative value.

Based on nation-wide household surveys in 2018, this study also presented the results of hierarchical and non-hierarchical inequality decomposition analyses for Indonesia and the Philippines. Indonesia and the Philippines have almost the same level of overall expenditure inequality at 0.26 by the Theil *L* index. But, its determinants are different between these two countries. While the rural-urban disparity constitutes 11% of overall expenditure inequality in both countries, the contribution of the disparity between the four education groups is much smaller in Indonesia than in the Philippines. According to the non-hierarchical inequality decomposition analysis, the contribution of the location-education interaction component is negative in both countries, indicating that there is a rural-urban difference in the expenditure disparity between the four education groups. After adjusting for this rural-urban difference, the expenditure disparity between the four education groups contributes 13% in Indonesia while 21% in the Philippines. As a consequence, Indonesia's within-sector within-group inequality component has a larger contribution to overall inequality than the Philippines' (76% vs. 68%).

Appendix 1: Two-stage Hierarchical Inequality Decomposition Method

Consider a country consisting of the rural and urban sectors, each with *m* levels of education.

1. Theil *L* Index

When measured by the Theil *L* index, overall income inequality is given by

$$L = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu}{y_{ijk}}\right).$$

This can be modified to

$$L = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\frac{1}{n_{i\cdot}}\right)\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left[\ln\left(\frac{\mu}{\mu_{i\cdot}}\right) + \ln\left(\frac{\mu_{i\cdot}}{y_{ijk}}\right)\right]$$

$$= \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)\left[\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu}{\mu_{i\cdot}}\right) + \frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu_{i\cdot}}{y_{ijk}}\right)\right]$$

$$= \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)\ln\left(\frac{\mu}{\mu_{i\cdot}}\right) + \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)\left[\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} \ln\left(\frac{\mu_{i\cdot}}{y_{ijk}}\right)\right]$$

since $\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\ln\left(\frac{\mu}{\mu_{i\cdot}}\right)=\ln\left(\frac{\mu}{\mu_{i\cdot}}\right)$.

Let $L_{BS}=\sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)\ln\left(\frac{\mu}{\mu_{i\cdot}}\right)$ and $L_{WSi}=\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\ln\left(\frac{\mu_{i\cdot}}{y_{ijk}}\right)$ be, respectively, income inequality between the rural and urban sectors and income inequality within sector *i*. Then, we have

$$L=L_{BS}+\sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\right)L_{WSi}\ .$$

Next, L_{WSi} can be modified to

$$L_{WSi}=\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\frac{1}{n_{ij}}\right)\sum_{k=1}^{n_{ij}}\left[\ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)+\ln\left(\frac{\mu_{ij}}{y_{ijk}}\right)\right]$$

$$=\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)+\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\right)\left[\left(\frac{1}{n_{ij}}\right)\sum_{k=1}^{n_{ij}}\ln\left(\frac{\mu_{ij}}{y_{ijk}}\right)\right]$$

since $\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)=\ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)$.

Let $L_{BGi}=\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu_{ij}}\right)$ and $L_{ij}=\sum_{k=1}^{n_{ij}}\left(\frac{1}{n_{ij}}\right)\ln\left(\frac{\mu_{ij}}{y_{ijk}}\right)$ be, respectively, inequality between *m* education groups in sector *i* and inequality within education group *j* of sector *i*. Then, we obtain

$$L_{WSi}=L_{BGi}+\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\right)L_{ij}.$$

Substituting this into *L*, we finally obtain

$$L=L_{BS}+\sum_{i=1}^{2}\frac{n_{i\cdot}}{n}\left(L_{BGi}+\sum_{j=1}^{m}\left(\frac{n_{ij}}{n_{i\cdot}}\right)L_{ij}\right)$$

$$=L_{BS}+\sum_{i=1}^{2}\frac{n_{i\cdot}}{n}L_{BGi}+\sum_{i=1}^{2}\sum_{j=1}^{m}\frac{n_{ij}}{n}L_{ij}$$

$$=L_{BS}+L_{WSBG}+L_{WSWG}.$$

2. Theil T Index

When measured by the Theil *L* index, overall income inequality is given by

$$T=\frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu}\right)\ln\left(\frac{y_{ijk}}{\mu}\right).$$

This can be modified to

$$T=\sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{n_{i\cdot}\mu_{i\cdot}}\right)\left[\ln\left(\frac{\mu_{i\cdot}}{\mu}\right)+\ln\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\right]$$

$$=\sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\left[\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{n_{i\cdot}\mu_{i\cdot}}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu}\right)+\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\ln\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\right]$$

$$= \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu}\right) + \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\left[\frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\ln\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\right]$$

since $\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{n_{i\cdot}\mu_{i\cdot}}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu}\right) = \ln\left(\frac{\mu_{i\cdot}}{\mu}\right)$.

Let $T_{BS} = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu}\right)$ and $T_{WSi} = \frac{1}{n_{i\cdot}}\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)\ln\left(\frac{y_{ijk}}{\mu_{i\cdot}}\right)$ be, respectively, income inequality between the rural and urban sectors and income inequality within sector *i*. Then, we have

$$T = T_{BS} + \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)T_{WSi}\ .$$

Next, T_{WSi} can be modified to

$$T_{WSi} = \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{n_{ij}\mu_{ij}}\right)\left[\ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right) + \ln\left(\frac{y_{ijk}}{\mu_{ij}}\right)\right]$$

$$= \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)\ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right) + \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)\left[\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu_{ij}}\right)\ln\left(\frac{y_{ijk}}{\mu_{ij}}\right)\right]$$

since $\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{n_{ij}\mu_{ij}}\right)\ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right) = \ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right)$.

Let $T_{BGi} = \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)\ln\left(\frac{\mu_{ij}}{\mu_{i\cdot}}\right)$ and $T_{ij} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu_{ij}}\right)\ln\left(\frac{y_{ijk}}{\mu_{ij}}\right)$ be, respectively, inequality between *m* education groups in sector *i* and inequality within education group *j* of sector *i*. Then, we obtain

$$T_{WSi} = T_{BGi} + \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)T_{ij}.$$

Substituting this into *T*, we finally obtain

$$T = T_{BS} + \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)\left(T_{BGi} + \sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n_{i\cdot}\mu_{i\cdot}}\right)T_{ij}\right)$$

$$= L_{BS} + \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)T_{BGi} + \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{n_{ij}\mu_{ij}}{n\mu}\right)T_{ij}$$

$$= T_{BS} + T_{WSBG} + T_{WSWG}.$$

We should note that if we let Y, $Y_{i\cdot}$, and Y_{ij} be total income, total income of sector *i*, total income of education group *j* in sector *i*, then we have the following relationships.

$Y = n\mu$, $Y_{i\cdot} = n_{i\cdot}\mu_{i\cdot}$, and $Y_{ij} = n_{ij}\mu_{ij}$.

Therefore, $\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu} = \frac{Y_{i\cdot}}{Y}$ is income share of sector *i*, while $\frac{n_{ij}\mu_{ij}}{n\mu} = \frac{Y_{ij}}{Y}$ is income share of education group j in sector i. $T_{WSBG} = \sum_{i=1}^{2}\left(\frac{Y_{i\cdot}}{Y}\right)T_{BGi}$ is the income-weighted average of T_{BGi}, while $T_{WSWG} = \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y}\right)T_{ij}$ is the income-weighted average of T_{ij}.

Appendix 2: Hierarchical and Non-hierarchical Decomposition Analyses by Location and Education using the Theil *T* Index

By the Theil *T* index, overall income inequality is calculated as follows.

$$T = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}}\left(\frac{y_{ijk}}{\mu}\right)\ln\left(\frac{y_{ijk}}{\mu}\right) = 0.137.$$

On the other hand, income inequality between the rural and urban sectors is calculated as follows.

$$T_{BS} = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}}{n}\frac{\mu_{i\cdot}}{\mu}\right)\ln\left(\frac{\mu_{i\cdot}}{\mu}\right) = \left(\frac{47}{100}\right)\ln\left(\frac{3.92}{5}\right) + \left(\frac{53}{100}\right)\ln\left(\frac{6.63}{5}\right) = 0.034$$

Next, to obtain the within-sector between-group inequality component (T_{WSBG}), we need to have the mean incomes for the low and high education groups in each of the rural and urban sectors. Using the mean incomes presented in Table 3, we obtain

$$T_{BG1} = \sum_{j=1}^{2}\left(\frac{n_{1j}}{n_{1\cdot}}\frac{\mu_{1j}}{\mu_{1\cdot}}\right)\ln\left(\frac{\mu_{1j}}{\mu_{1\cdot}}\right) = \left(\frac{28}{47}\right)\ln\left(\frac{3.5}{3.92}\right) + \left(\frac{19}{47}\right)\ln\left(\frac{4.75}{3.92}\right) = 0.011$$

$$T_{BG2} = \sum_{j=1}^{2}\left(\frac{n_{2j}}{n_{2\cdot}}\frac{\mu_{2j}}{\mu_{2\cdot}}\right)\ln\left(\frac{\mu_{2j}}{\mu_{2\cdot}}\right) = \left(\frac{18}{53}\right)\ln\left(\frac{4.5}{6.63}\right) + \left(\frac{35}{53}\right)\ln\left(\frac{8.75}{6.63}\right) = 0.052.$$

Therefore, the within-sector between-group inequality component is calculated as follows.

$$T_{WSBG} = \sum_{i=1}^{2}\left(\frac{n_{i\cdot}\mu_{i\cdot}}{n\mu}\right)T_{BGi} = \left(\frac{47}{100}\right)(0.011) + \left(\frac{53}{100}\right)(0.052) = 0.033.$$

On the other hand, to obtain the within-sector within-group inequality component (T_{WSWG}), we need to calculate inequality within education group *j* of sector *i*: $T_{ij} = \sum_{k=1}^{n_{ij}}\left(\frac{1}{n_{ij}}\frac{y_{ijk}}{\mu_{ij}}\right)\ln\left(\frac{y_{ijk}}{\mu_{ij}}\right)$.

Using individual incomes in Table 1 and mean incomes in Table 3, we obtain $T_{11} = 0.044$, $T_{12} = 0.048$, $T_{21} = 0.052$, and $T_{22} = 0.111$. Thus, we have

$$T_{WSWG} = \sum_{i=1}^{2}\sum_{j=1}^{2}\left(\frac{n_{ij}\mu_{ij}}{n\mu}\right)T_{ij}$$

$$= \left(\frac{28}{100}\right)(0.044) + \left(\frac{19}{100}\right)(0.048) + \left(\frac{18}{100}\right)(0.052) + \left(\frac{35}{100}\right)(0.111) = 0.070.$$

In sum, we obtain the following two-stage hierarchical inequality decomposition equation by the Theil *T* index.

$$T = T_{BS} + T_{WSBG} + T_{WSWG} = 0.034 + 0.033 + 0.070 = 0.137.$$

Table 16 presents the decomposition of income inequality by location (rural and urban sectors) and the decomposition of income inequality by education (low and high education groups) in each of the rural and urban sectors, while Table 17 summarizes the result of the two-stage hierarchical inequality decomposition analysis (location ⇒ education). Inequality between the rural and urban sectors contributes 25.2% to overall income inequality. Meanwhile, after adjusting for rural-urban differences in the structure of education, income inequality between the low and high education groups contributes 24.1%. The remaining 50.7% is attributed to the within-sector within-group inequality component.

Table 16. Decomposition of Income Inequality by Location (Rural & Urban Sectors) and Decomposition of Inequality by Education (Low & High Education Groups) in Each Sector: Theil *T* Index

	Inequality (Theil *T*)	Contrib. (%)	Income share (%)		Inequality (Theil *T*)	Contrib. (%)	Income share (%)
Total (A) = (B) + (C)	0.137	100.0					
B-sector (B)	0.034	25.2					
W-sector (C) = (a) + (b)	0.102	74.8					
Rural sector (a) = (c) + (d)	0.056	19.3	47.0	Urban sector (b) = (e) + (f)	0.143	55.5	53.0
B-group (c)	0.011	3.8		B-group (e)	0.052	20.3	
W-group (d)	0.045	15.5		W-group (f)	0.091	35.2	
Low educ. group	(0.044)	8.9	28.0	Low educ. Group	(0.052)	6.8	18.0
High educ. Group	(0.048)	6.6	19.0	High educ. Group	(0.111)	28.4	35.0

(Note) 'Contrib. (%)' is the % contribution of each component to overall income inequality, 'Income share' is the income share of each component to total income, 'B-sector' is the between-sector inequality component, 'W-sector' is the within-sector inequality component, 'B-group' is the between-group inequality component, and 'W-group' is the within-group inequality component.

Table 17. Two-stage Hierarchical Decomposition of Income Inequality by Theil *T* Index Location (Rural & Urban Sectors) ⇒ Education (Low & High Education Groups)

	Inequality (Theil *T*)	Contribution (%)
Total (T)	0.137	100.0
B-sector (T_{BS})	0.034	25.2
W-sector B-group (T_{WSBG})	0.033	24.1
W-sector W-group (T_{WSWG})	0.070	50.7
Low education group	(0.022)	15.7
High education group	(0.048)	35.0

(Note) 'B-sector' is the between-sector inequality component, 'W-sector B-group' is the within-sector between-group inequality component, and 'W-sector W-group' is the within-sector within-group inequality component.

To perform a non-hierarchical inequality decomposition analysis by location and education, we need to calculate income inequality between the low and high education groups (T_{BG}). Since we have

$$T_{BG} = \sum_{j=1}^{2} \left(\frac{n_{.j}\mu_{.j}}{n\mu}\right) \ln\left(\frac{\mu_{.j}}{\mu}\right) = \left(\frac{46}{100}\right) \ln\left(\frac{3.83}{5}\right) + \left(\frac{54}{100}\right) \ln\left(\frac{6.75}{5}\right) = 0.040,$$

The location and education interaction component is

$$T_{ISG} = T_{WSBG} - T_{BG} = 0.033 - 0.040 = -0.007.$$

Table 18 compares the result of the non-hierarchical inequality decomposition analysis with that of the two-stage hierarchical inequality decomposition analysis (location ⇒ education) presented in Table 17.

Table 18. Hierarchical and Non-hierarchical Inequality Decomposition Analyses by Location and Education: Theil *T* Index

	Hierarchical decomposition		Non-hierarchical decomposition	
	Inequality (Theil *T*)	Contribution (%)	Inequality (Theil *T*)	Contribution (%)
Total (*T*)	0.137	100.0	0.137	100.0
B-sector (L_{BS})	0.034	25.2	0.034	25.2
B-group (L_{BG})			0.040	29.2
W-sector B-group (T_{WSBG})	0.033	24.1		
Interaction component (T_{ISG})			-0.007	-5.1
W-sector W-group (T_{WSWG})	0.070	50.7	0.070	50.7
Low education group	(0.022)	15.7	(0.022)	15.7
High education group	(0.048)	35.0	(0.048)	35.0

References

Akita, T., & Kataoka, M. (2022). *Regional Inequality and Development: Measurement and Applications in Indonesia*, Springer.

Akita, T., Lukman, R. A., & Yamada, Y. (1999). Inequality in the distribution of household expenditures in Indonesia: A Theil decomposition analysis. *The Developing Economies*, 37(2), 197-221.

Akita, T., & Miyata, S. (2008). Urbanization, educational expansion, and expenditure inequality in Indonesia in 1996, 1999, and 2002. *Journal of the Asia and Pacific Economy*, 13(2), 146-165.

Akita, T., & Szeto, J. J. K. (2000). Inpres Desa Tertinggal (IDT) program and Indonesian regional inequality. *Asian Economic Journal*, 14(2), 167-186.

Anand, S. (1983). *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.

Balisacan, A. M., & Fuwa, N. (2004). Changes in spatial income inequality in the Philippines: An exploratory analysis. WIDER/UNU, Research Paper No. 2004/34,

Borooah, V. K., Gustafsson, B., & Shi, L. (2006). China and India: Income inequality and poverty north and south of the Himalayas. *Journal of Asian Economics*, 17, 797–817.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47(4), 901-20.

Ching, P. (1991). Size distribution of income in the Philippines. In T. Mizoguchi (Ed.) *Making Economies More Efficient and More Equitable: Factors Determining Income Distribution*, Tokyo: Kinokuniya Company, 157-178.

Chongvilaivan, A., & Kim, J. (2016). Individual income inequality and its drivers in Indonesia: A Theil decomposition reassessment. *Social Indicators Research*, 126, 79–98.

Cowell, F. A. (1984). The structure of American income inequality. *Review of Income and Wealth*. 30(3), 351-375.

Cowell, F. A., & Jenkins, S. P. (1995). How much inequality can we explain? A methodology and an application to the USA. *Economic Journal*, 105, 421–430.

Dickey, H. (2001). Regional earnings inequality in Great Britain: A decompositiaon Analysis. *Regional Studies*, 35(7), 605–612.

Eastwood, R., & Lipton, M. (2004). Rural and urban income inequality and poverty: Does convergence between sectors offset divergence within them? in Giovanni Andrea Cornia (ed.) *Inequality, Growth, and Poverty in an Era of Liberalization and Globalization*, Oxford: Oxford University Press, 112-141.

Elbers, C., Lanjouw, P., & Mistiaen, J. A., & Özler, B. (2008). Reinterpreting between-group inequality. *Journal of Economic Inequality*, 6, 231–245.

Estudillo, J. P. (1997). Income inequality in the Philippines, 1961–91. *The Developing Economies*. 35(1), 68–95.

Fields, G. S. (1979). Decomposing LDC inequality. *Oxford Economic Papers*, 31(3), 437-459.

Fields, G. S. (2001). *Distribution and Development*. Cambridge, MA, MIT Press

Fishlow, A. (1972). Brazilian sixe distribution of income. *The American Economic Review*, 62(1/2), 391-402

Glewwe, P. (1986). The distribution of income in Sri Lanka in 1969–70 and 1980–81: A decomposition analysis. *Journal of Development Economics*. 24(2), 255–74.

Gray, D., Mills, J. A., & Zandvakili, S. (2003). Statistical analysis of inequality with decompositions: The Canadian experience. *Empirical Economics*, 28, 291–302.

Ikemoto, Y. (1985). Income distribution in Malaysia: 1957–80. *The Developing Economies*, 23(4), 347–67.

Liu, A. Y. C. (2001). Markets, inequality and poverty in Vietnam. *Asian Economic Journal*, 15(2), 217-235.

Mishra, P., & Parikh, A. (1992). Household consumer expenditure inequalities in India: A decomposition analysis. *Review of Income and Wealth*, 38(2), 225-236.

Mukhopadhaya, P. (2003). Trends in Total and Subgroup Income Inequality in the Singaporean Workforce. *Asian Economic Journal*, 17(3), 243-264.

Mookherjee, D. & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal*, 92, 886-902.

Motonishi, T. (2006). Why has income inequality in Thailand increased? An analysis using surveys from 1975 to 1998. *Japan and the World Economy*, 18, 464–487.

Rao, V.V. B., Banerjee, D.S., & Mukhopadhaya, P. (2003). Earnings inequality in Singapore. *Journal of the Asian Pacific Economy*, 8(2), 210-228.

Shorrocks, A. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3), 613-25.

Shorrocks, A. and Wan, G. (2005). Spatial decomposition of inequality. *Journal of Economic Geography*, 5(1), 59-81.

Tang, K.K., Petrie, D. (2009). Non-hierarchical bivariate decomposition of Theil indexes. *Economics Bulletin*, 29(2), 928-927.

Tsakloglou, P. (1993). Aspects of inequality in Greece: Measurement, decomposition and intertemporal change, 1974, 1982. *Journal of Development Economics*, 40(1), 53–74.