Economics & Management Series EMS-2013-15

# The Roles of Location and Education in the Distribution of Economic Well-being in Indonesia: Hierarchical and Non-hierarchical Inequality Decomposition Analyses

Takahiro Akita
*International University of Japan*

Sachiko Miyata
*Rikkyo University*

September 2013

IUJ Research Institute
International University of Japan

**The Roles of Location and Education in the Distribution of Economic Well-being in Indonesia: Hierarchical and Non-hierarchical Inequality Decomposition Analyses**

Takahiro Akita*
Graduate School of International Relations
International University of Japan, Japan
777 Kokusai-cho, Minami Uonuma, Niigata 949-7277, Japan
e-mail: akita@iuj.ac.jp
Tel: 025-779-1414

and

Sachiko Miyata
College of Business
Rikkyo University
Tokyo, Japan

**Abstract**

This paper analyzes the roles of location (rural and urban sectors) and education in the distribution of economic well-being in Indonesia by employing the hierarchical and non-hierarchical decomposition methods of the Theil indices. This is done by using household expenditure data from the national socio-economic survey (*Susenas*) in 2008. It shows that there are large expenditure disparities across education levels but that these are more pronounced in the urban sector than the rural sector. When there are differences in educational structure between the rural and urban sectors, the hierarchical decomposition method appears to offer a better approach than the non-hierarchical method.

Keywords: Inequality; Hierarchical and non-hierarchical decompositions; Theil indices; Urban and rural locations; Education; Indonesia

JEL classification: O15, O18, R12

Running title: Hierarchical and non-hierarchical inequality decompositions by location and education

* Takahiro Akita is grateful to the Japan Society for the Promotion of Science for its financial support (Grant-in-Aid for Scientific Research No. 24530274).

## 1 Introduction

The Theil indices are often used to measure inequality in the distribution of economic well-being. They belong to the generalized entropy class of inequality measures, and satisfy several desirable properties, such as anonymity, income homogeneity, population homogeneity, and the Pigue-Dalton principle of transfers. They are also additively decomposable by population sub-group, i.e., overall inequality, as measured by the Theil indices, is expressed as the sum of the within-group and between-group inequality components (see, for example, Bourguignon (1979), Shorrocks (1980)). In a bivariate framework, Akita and Miyata (2008) proposed the hierarchical decomposition of the Theil indices, based on the two-stage nested decomposition method developed by Akita (2003) in the context of regional income inequality, while Tang and Petrie (2009) developed the non-hierarchical decomposition method. The objective of this paper is to employ the hierarchical and non-hierarchical decomposition methods of the Theil indices to analyze the roles of location (rural and urban sectors) and education in the distribution of economic well-being in Indonesia. This is done by using household expenditure data from the national socio-economic survey (*Susenas*) in 2008.

In the hierarchical decomposition method, all households are first classified into the rural and urban sectors, and then into educational groups for each sector based on household head's educational attainment. By doing this, one can analyze inequality due to differences in educational attainments after removing the effect of rural-urban differences in educational endowments on inequality. In the hierarchical decomposition method, however, the order of decomposition matters, i.e., all households can be first classified into educational groups, and then into the rural and urban sectors for each educational group. In order to rectify this problem, Tang and Petrie (2009) suggested the non-hierarchical decomposition method, in which overall inequality, as measured by the Theil indices, is decomposed, simultaneously, according to some categorical variables. In this framework, however, interaction terms among variables emerge.

In the present study, we conduct both hierarchical and non-hierarchical decomposition analyses by focusing on location and education as major factors of inequality and make a comparison between these two methods in a bivariate (location and education) framework. The paper shows that when there are differences in educational structure between the rural and urban sectors, the hierarchical decomposition method appears to offer a better approach than the non-hierarchical method.

This study differs from Tang and Petrie (2009) in that it applies the methods to household-level data rather than percentile or decile group data and considers location and education as major factors of inequality rather than gender and ethnicity. It should be noted that differences in location and education are found to account for a relatively large portion of income or expenditure inequality in decomposition analyses (see, for example, Glewwe (1986), Ikemoto (1991), Tsakloglou (1993), Estudillo (1997), Rao et al. (2003), Balisacan and Fuwa (2004), Shorrocks and Wan (2005), and Akita and Miyata (2008)).

Elbers et al. (2008) suggested a new measurement approach for the contribution of the between-group inequality component, as the between-group component hinges on the number of groups, the relative sizes of the groups, and differences in mean incomes or expenditures among the groups, and thus care should be taken to compare decomposition results based on different settings. In the new approach, observed between-group inequality is assessed against the maximum between-group inequality attainable given the number and relative sizes of the groups, rather than overall inequality that is used in the conventional approach. In the present study, we assess the between-group inequality not only against overall inequality but also the maximum between-group inequality in a bivariate hierarchical decomposition framework.

## 2 Method and the Data

### 2.1 Method

Suppose that all households are classified into the urban and rural sectors (sectors 1 and 2, respectively), and that households in each of these two sectors are classified into *m* mutually exclusive and collectively exhaustive education groups, for example, the primary, secondary and tertiary education groups. Let $y_{ijk}$ denote the per capita expenditure of household *k* in group *j* in sector *i*. Let *N*, *Y* and $N_{ij}$ be the number of all households, the total per capita expenditure of all households, and the number of households in group *j* in sector *i* respectively. Then using the two Theil indices *T* and *L*, overall inequality in per capita expenditure is given by:

$$T = \sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k=1}^{N_{ij}}\left(\frac{y_{ijk}}{Y}\right)\log\left(\frac{y_{ijk}/Y}{1/N}\right) \quad \text{and} \qquad (1)$$

$$L = \sum_{i=1}^{2}\sum_{j=1}^{m}\sum_{k=1}^{N_{ij}}\left(\frac{1}{N}\right)\log\left(\frac{1/N}{y_{ijk}/Y}\right) \tag{2}$$

Let $Y_{i\cdot}, N_{i\cdot}$, and $Y_{ij}$ denote the total per capita expenditure of households in sector *i*, the total number of households in sector *i*, and the total per capita expenditure of households in group *j* in sector *i*. Then, the Theil index *T* in equation (1) can be decomposed hierarchically into the between-sector inequality component ($T_{BS}$), the within-sector between-group inequality component ($T_{WSBG}$), and the within-sector within-group inequality component ($T_{WSWG}$) as follows:

$$T = T_{BS} + \sum_{i=1}^{2}\left(\frac{Y_{i\cdot}}{Y}\right)T_i \tag{3}$$

$$= T_{BS} + \sum_{i=1}^{2}\left(\frac{Y_{i\cdot}}{Y}\right)T_{Gi} + \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y}\right)T_{ij}$$

$$= T_{BS} + T_{WSBG} + T_{WSWG} \tag{4}$$

where $T_{BS} = \sum_{i=1}^{2}\left(\frac{Y_{i\cdot}}{Y}\right)\log\left(\frac{Y_{i\cdot}/Y}{N_{i\cdot}/N}\right)$, $T_i = T_{Gi} + \sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y_{i\cdot}}\right)T_{ij}$, $T_{Gi} = \sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y_{i\cdot}}\right)\log\left(\frac{Y_{ij}/Y_{i\cdot}}{N_{ij}/N_{i\cdot}}\right)$, and

$T_{ij} = \sum_{k=1}^{N_{ij}}\left(\frac{y_{ijk}}{Y_{ij}}\right)\log\left(\frac{y_{ijk}/Y_{ij}}{1/N_{ij}}\right)$ measure, respectively, inequality between the urban and rural sectors, inequality within sector *i*, inequality between groups in sector *i*, and inequality within group *j* in sector *i*. Equation (3) provides the ordinary one-stage decomposition equation, while equation (4) presents the two-stage hierarchical decomposition equation. It should be noted that the two-stage nested Theil decomposition method developed by Akita (2003) is similar to equation (4); but it uses district-level GDP data rather than household data and is based on a natural hierarchical structure, i.e., region-province-district, where each region consists of a different set of provinces and each province consists of a different set of districts. On the other hand, in equation (4), the urban and rural sectors have the same set of education groups.

Similarly, the Theil index *L* in equation (2) can be decomposed into:

$$L = L_{BS} + \sum_{i=1}^{2}\left(\frac{N_{i\cdot}}{N}\right)L_i \tag{5}$$

$$= L_{BS} + \sum_{i=1}^{2}\left(\frac{N_{i\cdot}}{N}\right)L_{Gi} + \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{N_{ij}}{N}\right)L_{ij}$$

$$= L_{BS} + L_{WSBG} + L_{WSWG} \tag{6}$$

In the two-stage hierarchical inequality decomposition, the order of decomposition can be reversed, i.e., overall inequality can be decomposed into the between-group inequality component ($T_{BG}$), the within-group between-sector inequality component ($T_{WGBS}$), and the within-group within-sector inequality component ($T_{WGWS}$) as follows:

$$T = T_{BG} + T_{WGBS} + T_{WGWS} \tag{7}$$

Similarly, we have

$$L = L_{BG} + L_{WGBS} + L_{WGWS} \tag{8}$$

It should be noted that $T_{WSWG}$ in equation (4) is the same as $T_{WGWS}$ in equation (7), since $T_{WSWG} = \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y}\right)T_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{2}\left(\frac{Y_{ij}}{Y}\right)T_{ij} = T_{WGWS}$.

Decomposition equations (4), (6), (7) and (8) present the hierarchical decomposition of the Theil indices. But, as mentioned above, the order of decomposition matters in this decomposition. In order to rectify this limitation, Tang and Petrie (2009) proposed the non-hierarchical decomposition method, in which the Theil indices are decomposed, simultaneously, according to some categorical variables, such as location, education, gender, ethnicity and age. Using two variates, for example, location (rural and urban sectors) and education, the non-hierarchical decomposition is expressed as:

$$T = T_{BS} + T_{BG} + T_{ISG} + T_{WSWG} \tag{9}$$

$$L = L_{BS} + L_{BG} + L_{ISG} + L_{WSWG} \tag{10}$$

where $T_{ISG}$ and $L_{ISG}$ are the sector-group interaction terms in the Theil indices *T* and *L*, respectively, which are expressed, respectively, as:

$$T_{ISG} = \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{Y_{ij}}{Y}\right)\left(\ln\left(\frac{Y_{ij}/Y}{N_{ij}/N}\right) - \ln\left(\frac{Y_{i\cdot}/Y}{N_{i\cdot}/N}\right) - \ln\left(\frac{Y_{\cdot j}/Y}{N_{\cdot j}/N}\right)\right)$$

$$L_{ISG} = \sum_{i=1}^{2}\sum_{j=1}^{m}\left(\frac{N_{ij}}{N}\right)\left(\ln\left(\frac{N_{ij}/N}{Y_{ij}/Y}\right) - \ln\left(\frac{N_{i\cdot}/N}{Y_{i\cdot}/Y}\right) - \ln\left(\frac{N_{\cdot j}/N}{Y_{\cdot j}/Y}\right)\right)$$

Since $T_{WSBG} = T_{BG} + T_{ISG}$ and $L_{WSBG} = L_{BG} + L_{ISG}$ from equations (4), (6), (9) and (10), the interaction term is given by $T_{ISG} = T_{WSBG} - T_{BG}$ and $L_{ISG} = L_{WSBG} - L_{BG}$, which can be positive or negative.

As suggested by Elbers et al. (2008), we assess observed between-group inequality not only against overall inequality but also the maximum between-group inequality. In the hierarchical decomposition framework of two sectors (rural and urban sectors) and three educational groups (primary, secondary and tertiary groups), there are six mutually exclusive and collectively exhaustive groups: the rural-primary, rural-secondary, rural-tertiary, urban-primary, urban-secondary, and urban-tertiary groups. Given the distribution of households in per capita expenditure, these six groups are first arranged in an ascending order of mean per capita expenditure. All households are then reclassified in an ascending order of per capita expenditure into these groups, from the smallest to the largest mean per capita expenditure group. The maximum between-group inequality can be obtained based on these non-overlapping and rank-preserving groups.

The contribution of observed between-group inequality to the maximum between-group inequality, as measured by the Theil indices *T* and *L*, is given, respectively, by

$$\overline{CT_{B(SG)}} = (T_{BS} + T_{WSBG}) / T_{B(SG)}^{max} \quad \text{and} \quad \overline{CL_{B(SG)}} = (L_{BS} + L_{WSBG}) / L_{B(SG)}^{max} \qquad (11)$$

In the conventional approach, the contribution of observed between-group inequality is assessed against overall inequality as follows:

$$CT_{B(SG)} = (T_{BS} + T_{WSBG}) / T \quad \text{and} \quad CL_{B(SG)} = (L_{BS} + L_{WSBG}) / L \,. \qquad (12)$$

To analyze the role of educational endowments in the rural-urban disparity, we conduct a Blinder-Oaxaca decomposition analysis, which was developed by Blinder (1973) and Oaxaca (1973). Let $y_U$ and $y_R$ be the natural log of per capita expenditure of urban and rural households, respectively. Given the linear regression model,

$$y_k = \boldsymbol{X}_k{}'\boldsymbol{\beta}_k + e_k \qquad E(e_k) = 0 \qquad k = U, R$$

where $\boldsymbol{X}_k$ is a vector of explanatory variables, $\boldsymbol{\beta}_k$ is a vector of coefficients associated with $\boldsymbol{X}_k$, and $e_k$ is the error term, we let $\hat{\boldsymbol{\beta}}_k$ be a vector of the least-squares estimates for $\boldsymbol{\beta}_k$ ($k = U, R$), obtained separately from the urban and rural samples and $\overline{\boldsymbol{X}}_k$ be the estimate for $E(\boldsymbol{X}_k)$. Then, based on the twofold decomposition suggested by Newmark (1988), the estimated urban-rural difference in mean per capita expenditure is given by:

$$\hat{D} = \bar{y}_U - \bar{y}_R = (\overline{\boldsymbol{X}}_U - \overline{\boldsymbol{X}}_R)'\hat{\boldsymbol{\beta}}* + \left(\overline{\boldsymbol{X}}_U'(\hat{\boldsymbol{\beta}}_U - \hat{\boldsymbol{\beta}}*) + \overline{\boldsymbol{X}}_R'(\hat{\boldsymbol{\beta}}* - \hat{\boldsymbol{\beta}}_R)\right) \qquad (13)$$

where $\hat{\boldsymbol{\beta}}*$ is a vector of the least-squares estimates of the coefficients, which are obtained from the pooled sample of urban and rural households. The first term in equation (13) is the part that is explained by urban-rural differences in the explanatory variables (endowments effect), while the second term is the unexplained part.

## 2.2 The Data

To perform hierarchical and non-hierarchical decomposition analyses, this study uses monthly household expenditure data from the expenditure module of the 2008 National Socio-Economic Survey (*Susenas*) in Indonesia, which was conducted by the Central Bureau of Statistics. Table 1 presents the number of households by location and education in the 2008 *Susenas* sample.[1] The sample size is 282,387 households, of which 181,147 and 101,240 households are in the rural and urban sectors, respectively. This sample size is much larger than the sample sizes of previous surveys (e.g., 60,591, 64,406, and 62,551 households in the 1999, 2002, and 2005 surveys, respectively), since the government has tried to obtain more reliable estimates of the population values at the provincial and district levels.

Table 1 also provides the estimated number of households by location and education in 2008, which is obtained by using sample weights. The table indicates that rural households are over represented in the sample with 64.1% of households from the rural sector compared to an estimated 51.4% in Indonesia. According to the estimate, the total number of households is 57.6 million in Indonesia, of which 29.6 and 28.0 million are, respectively, in the rural and urban sectors. It should be noted that almost three quarters of rural

[1] The primary education group includes households whose heads have either no education, incomplete primary education or primary education. The secondary group consists of households whose heads completed junior high school or senior high school, whereas the tertiary group includes households whose heads completed two-year junior college, three-year junior college, four-year university/college, or graduate school (master’s or doctoral program).

households are in the primary education group in 2008, i.e., have either no education, incomplete primary education or primary education, while 57% of urban households have attained at least secondary education.

## 3 Empirical Results[2]

Table 2 shows the result of the one-stage inequality decomposition of the Theil *T* index by location (see equations (3) and (5)).[3] The urban sector not only has a larger mean per capita expenditure but also a larger within-sector inequality than the rural sector. The ratio between the urban and rural sectors in mean per capita expenditure is 1.7. According to the conventional approach, where the between-sector inequality is assessed against overall inequality (equation (12)), observed between-sector inequality accounts for 13% by the Theil T. However, if we employ an alternative approach where the maximum attainable between-sector inequality is used as the denominator rather than overall inequality (equation (11)), the contribution jumps to 27% by the Theil *T*, signifying the prominence of rural-urban differences in expenditure inequality.

Table 2 indicates that a large inequality exists within the urban sector. To examine the determinants of urban inequality, urban inequality is decomposed further with respect to education. The decomposition result is also shown in Table 2, together with the result for the rural sector. About 23% of urban inequality, as measured by the Theil *T*, is attributable to the disparity in mean per capita expenditure between the three education groups: the primary, secondary and tertiary groups. According to the alternative approach, however, the contribution increases to 30%, indicating the importance of educational differences in urban inequality. The ratio in mean per capita expenditure between the tertiary and primary educational groups is very high at 2.8. It should be noted that both within-group inequality and mean per capita expenditure increase as we move from the primary to the tertiary group. In contrast, educational differences do not seem to play an important role in rural inequality, as the contribution to rural inequality amounts to around 9%. Though the contribution rises to 12% by the alternative approach, educational differences are not so prominent in the rural sector.

Table 3 presents the result of the hierarchical decomposition by location and education.

---

[2] Except the Blinder-Oaxaca decomposition, all the results are weighted results based on sample weights.

[3] The paper presented only the decomposition results of the Theil *T*, since the implications of the Theil *L* results are similar to those of the Theil *T* results. The Theil *L* results are available from the authors upon request.

After removing the effect of rural-urban differences in educational endowments, the disparity between the primary, secondary and tertiary educational groups accounts for 16% of overall inequality by the Theil T. But this is the weighted average of the between-group inequalities in the rural and urban sectors (see equations (4) and (6)). As mentioned above, while educational differences do not play an important role in rural inequality, they contribute prominently to urban inequality. It should be noted that 29% (= 13% + 16%) of overall inequality is accounted for by the disparity between the following six groups: the rural-primary, rural-secondary, rural-tertiary, urban-primary, urban-secondary, and urban-tertiary groups. Conversely, 71% of overall inequality is due to within-sector within-group inequalities. However, if we assess the disparity between these six groups according to the alternative approach, the contribution increases to about 35% (see equation (11)).

As mentioned in the method section, the order of decomposition can be reversed, i.e., overall inequality can be first decomposed by education rather than location. In this hierarchical decomposition, overall inequality is decomposed into the between-group inequality component, the within-group between-sector inequality component, and the within-group within-sector inequality component (see equations (7) and (8)). The decomposition result is given in Table A1 in the Appendix. If we remove the effect of locational differences between the three educational groups, the rural-urban disparity accounts for merely 5% of overall expenditure inequality by the Theil *T*.

Table 3 also shows the result of the non-hierarchical decomposition by location and education. Inequality due to differences in mean per capita expenditure between the three educational groups is 0.063 by the Theil *T*, accounting for 24% of overall inequality (one-stage decomposition analysis). However, there is a large negative interaction effect between location and education: -7.7% of overall inequality by the Theil *T*, meaning that much of the disparity between the rural and urban sectors is due to differences in educational endowments between them. This is indicated by Table 4, which presents the estimated number of households and mean per capita expenditure by education (the number of years of education) in the rural and urban sectors. In the primary education group, more households are in the rural sector, while in the secondary and tertiary education groups, more households are in the urban sector. Furthermore, mean per capita expenditure increases gradually with the number of years of education in both the rural and urban sectors.

To further analyze the role of educational endowments in the rural-urban disparity, we conduct a Blinder-Oaxaca decomposition analysis, where we consider, as the explanatory variables, *edyear* (years of education of household head), *age* (age of household head), *age2* (square of age of household head), *size* (household size), *gender* (gender of household head: female = 0; male = 1), and *job* (job of household head: agriculture = 0; non-agriculture = 1).

Table 5 presents the result of the Blinder-Oaxaca decomposition. The mean of natural log of per capita expenditure is 13.13 for urban households and 12.67 for rural households, resulting in an urban-rural expenditure difference of 0.46.[4] This is decomposed into two parts. The first part, i.e., the explained part (endowments effect), shows the increase in mean per capita expenditure if rural households had the same endowments as urban households, given that rural and urban households have the same coefficients of the explanatory variables, which are estimated by the pooled sample of urban and rural households. The increase of 0.224 means that differences in endowments (education, age, household size, gender, and job type) as a whole account for 49% of the urban-rural expenditure difference. However, much of the increase is due to differences in educational endowments, accounting for 33%. This confirms the assertion described above. It should be noted that differences in job types (agriculture versus non-agriculture) also contribute to the urban and rural difference, though to a much lesser extent than education.

## 4 Concluding Remarks

To analyze the roles of location and education in the distribution of economic well-being in Indonesia, this paper employed two inequality decomposition methods in a bivariate framework: the hierarchical and non-hierarchical decompositions of the Theil indices by location and education. The analysis was conducted by using household expenditure data from the 2008 national socio-economic survey (*Susenas*).

While the non-hierarchical decomposition method indicates, based on an interaction term, the significance of differences in educational endowments in the rural-urban disparity, it is not able to show explicitly differences in educational structure between the rural and urban sectors. In contrast, the hierarchical decomposition method is able to show the differences in educational structure, by performing a one-stage decomposition analysis by education for each sector.

---

[4] Regression results for the urban, rural and pooled models are given in Table A2 in the Appendix.

According to the 2008 *Susenas* expenditure data, rural-urban disparity explains 13% of overall inequality in per capita expenditure by the Theil T, meaning that 87% of overall inequality is attributable to inequalities within the urban and rural sectors. Further decomposition by education for each sector reveals, however, that while the expenditure disparity between the three educational groups (primary, secondary and tertiary groups) accounts for 23% of urban inequality, it explains only 9% of rural inequality. The average contribution of these between-group inequalities amounts to 16% of overall inequality. When there are differences in educational structure between the rural and urban sectors, the hierarchical decomposition method appears to offer a better approach than the non-hierarchical method.

It should be noted that by the Theil *T*, 29% (13% + 16%) of overall inequality is accounted for by the disparity between the following six groups: the rural-primary, rural-secondary, rural-tertiary, urban-primary, urban-secondary, and urban-tertiary groups. Conversely, 71% of overall inequality is due to within-sector within-group inequalities. However, if we assess the disparity between these six groups according to an alternative approach, proposed by Elbers et al. (2008), the contribution increases to 35%, signifying the important role of location and education in the distribution of economic well-being in Indonesia. In order to mitigate inequality among households in Indonesia, it is essential to narrow the gap in education between the rural and urban sectors by raising rural sector's general educational levels; but at the same time, it is necessary to reduce the disparity between educational groups in the urban sector and inequality among urban households with higher education.

This study has several limitations. First, it employed consumption expenditure rather than income as a measure of welfare. In the *Susenas* surveys, it is widely believed that nonfood expenditures are understated progressively by higher-income households, especially in the urban sector; thus, expenditure inequalities are underestimated if they are measured based on the *Susenas* data. However, welfare levels at any point in time are likely to be better indicated by current expenditure than by current income and consumption expenditure is more reliable than income as an indicator of a household's permanent income, particularly in less developed countries, because it does not vary as much as income in the short term. Secondly, our study did not adjust expenditure data for the cost-of-living differentials between the urban and rural sectors; thus, urban-rural disparity might be exaggerated.

## References


Akita, T.: Decomposing regional income inequality in China and Indonesia using two-stage nested Theil decomposition method. The Annals of Regional Science 37(1), 55-77 (2003)

Akita, T., Miyata, S.: Urbanization, educational expansion, and expenditure inequality in Indonesia in 1996, 1999, and 2002. Journal of the Asia Pacific Economy 13(2), 147-67 (2008)

Balisacan, A. M., Fuwa, N.: Changes in spatial income inequality in the Philippines. UNU/WIDER Research Paper No. 2004/X, Helsinki (2004)

Blinder, A.S.: Wage discrimination: Reduced form and structural estimates. Journal of Human Resources 8(4), 436–55 (1973)

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

Elbers, C., Lanjouw, P., Mistiaen, J.A., Ozler, B.: Reinterpreting between-group inequality. Journal of Economic Inequality 6(3), 231-45 (2008)

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

Glewwe, P.: The distribution of income in Sri Lanka in 1969–70 and 1980–81: a decomposition analysis. Journal of Development Economics 24(2), 255–274 (1986)

Ikemoto, Y.: Income Distribution in Thailand: Its Changes, Causes, and Structure. Tokyo: University of Tokyo Press (1991)

Neumark, D.: Employers' discriminatory behavior and the estimation of wage discrimination. Journal of Human Resources 23(3), 279–95 (1988)

Oaxaca, R.: Male‑female wage differentials in urban labor markets. International Economic Review 14(3), 693–709 (1973)

Rao, V.V.B., Banerjee, D.S., Mukhopadhaya, P.: Earnings inequality in Singapore. Journal of the Asia Pacific Economy 8(2), 210–228 (2003)

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

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

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

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

**Table 1**
**Sample Size, Estimated Number of Households and Mean Monthly Expenditure in 2008**

| Location | Education | Sample Size | | Estimated No of Households | | Mean PCE | Exp. Share |
|---|---|---|---|---|---|---|---|
| | | Value | Share (in %) | Value (in 1,000) | Share (in %) | (in 1,000) | (in %) |
| Rural | Primary | 125,280 | 44.4 | 21,839 | 37.9 | 317 | 25.7 |
| | Secondary | 50,655 | 17.9 | 7,056 | 12.3 | 419 | 11.0 |
| | Tertiary | 5,212 | 1.8 | 676 | 1.2 | 651 | 1.7 |
| | Sub-total | 181,147 | 64.1 | 29,571 | 51.4 | 349 | 38.4 |
| Urban | Primary | 39,455 | 14.0 | 12,057 | 21.0 | 406 | 18.2 |
| | Secondary | 50,475 | 17.9 | 13,056 | 22.7 | 641 | 31.2 |
| | Tertiary | 11,310 | 4.0 | 2,864 | 5.0 | 1,142 | 12.2 |
| | Sub-total | 101,240 | 35.9 | 27,977 | 48.6 | 591 | 61.6 |
| Total | | 282,387 | 100.0 | 57,548 | 100.0 | 467 | 100.0 |

(Note) Mean PCE is mean monthly per capita expenditure estimated using sample weights (in 1,000 Rupiah).

**Table 2**
**Decompositions of Theil T by Location and by Education in Rural and Urban Sectors**

| | Decomposition by Location | | | Decomp. by Education in Rural & Urban Sectors | | | |
|---|---|---|---|---|---|---|---|
| | | | | Rural Sector | | Urban Sector | |
| | Value | % Cont. | | Value | % Cont. | Value | % Cont. |
| Rural | 0.170 | 24.7 | | | | | |
| | | | Primary | 0.145 | 57.3 | | |
| | | | Secondary | 0.170 | 28.8 | | |
| | | | Tertiary | 0.206 | 5.2 | | |
| Urban | 0.267 | 62.4 | | | | | |
| | | | Primary | | | 0.173 | 19.3 |
| | | | Secondary | | | 0.200 | 37.8 |
| | | | Tertiary | | | 0.274 | 20.4 |
| WS | 0.229 | 87.1 | WS | 0.155 | 91.3 | 0.207 | 77.5 |
| BS | 0.034 | 12.9 | BS | 0.015 | 8.6 | 0.060 | 22.5 |
| Total | 0.263 | 100.0 | Total | 0.170 | 100.0 | 0.267 | 100.0 |
| BS | 0.034 | 27.3 | BS | 0.015 | 12.1 | 0.060 | 29.7 |
| Max BS | 0.124 | 100.0 | Max BS | 0.121 | 100.0 | 0.202 | 100.0 |

**Table 3**
**Hierarchical vs. Non-hierarchical Decompositions of Theil T Index**
**Location ⇒ Education**

| | Hierarchical Decomposition | | Non-hierarchical Decomposition | |
|---|---|---|---|---|
| | Value | % Cont. | Value | % Cont. |
| BS (Location) | 0.034 | 12.9 | 0.034 | 12.9 |
| BG (Education) | | | 0.063 | 23.9 |
| WSBG | 0.043 | 16.2 | | |
| ISB | | | -0.020 | -7.7 |
| WSWG | 0.186 | 70.9 | 0.186 | 70.9 |
| Total | 0.263 | 100.0 | 0.263 | 100.0 |
| B(SG) | 0.077 | 34.1 | | |
| Max B(SG) | 0.225 | 100.0 | | |

**Table 4**
**Estimated Number of Households and Mean Per Capita Expenditure**
**by Location (Rural and Urban Sectors) and Education (No. of Years of Education)**

in 1,000

| | | Rural | | Urban | |
|---|---|---|---|---|---|
| Education | Years of Education | No Hhd | Mean PCE | No Hhd | Mean PCE |
| Primary | 0 | 3,733 | 287 | 1,554 | 344 |
| | 3 | 8,182 | 320 | 3,440 | 390 |
| | 6 | 9,923 | 324 | 7,064 | 428 |
| Secondary | 9 | 3,949 | 380 | 4,642 | 531 |
| | 12 | 3,107 | 469 | 8,414 | 701 |
| Tertiary | 13 | 177 | 581 | 225 | 798 |
| | 15 | 108 | 661 | 607 | 1,042 |
| | 16 | 376 | 662 | 1,832 | 1,155 |
| | 18 | 15 | 1,135 | 200 | 1,718 |
| | Total | 29,571 | 349 | 27,977 | 591 |

(Note) 1. The number of years of education is determined according to the following: no schooling (0 year); incomplete primary school (3 years); general and Islamic primary schools (6 years); general and Islamic junior high schools (9 years); general, Islamic and vocational senior high schools (12 years); diploma I and II (13 years); diploma III (15 years); diploma IV (Bachelor’s degree) (16 years); and master’s or doctor’s degree (18 years).

2. Mean PCE is mean monthly per capita expenditure estimated using sample weights (in 1,000 Rupiah).

**Table 5**
**Blinder-Oaxaca Decomposition of Rural-Urban Difference in Mean Per Capita Expenditure: Twofold Decomposition**

| | Coefficient | t-statistic | Contribution |
|---|---|---|---|
| Differential | | | |
| Prediction for Urban | 13.128 | 6,645.5 | |
| Prediction for Rural | 12.668 | 9,863.1 | |
| Difference | 0.460 | 195.3 | 100.0% |
| Explained | | | |
| edyear | 0.154 | 129.5 | 33.4% |
| age | -0.014 | -13.8 | -3.0% |
| age2 | 0.011 | 13.4 | 2.5% |
| size | 0.002 | 3.1 | 0.5% |
| gender | 0.000 | -3.7 | 0.0% |
| job | 0.071 | 62.0 | 15.4% |
| Explained Total | 0.224 | 138.0 | 48.7% |
| Unexplained | | | |
| edyear | 0.226 | 49.5 | 49.0% |
| age | -0.622 | -14.8 | -135.2% |
| age2 | 0.367 | 17.2 | 79.7% |
| size | 0.030 | 5.4 | 6.5% |
| gender | -0.124 | -22.1 | -26.9% |
| job | 0.065 | 14.5 | 14.2% |
| constant | 0.294 | 13.2 | 64.0% |
| Unexplained Total | 0.236 | 99.1 | 51.3% |
| R-squared: 0.287 for urban model; 0.189 for rural model; 0.319 for pooled model | | | |
| Number of Observations = 282,387 | | | |

## Appendix

**Table A1**
**Hierarchical vs. Non-hierarchical Decompositions of Theil T Index**
**Education ⇒ Location**

| | Hierarchical Decomposition | | Non-hierarchical Decomposition | |
|---|---|---|---|---|
| | Value | % Cont. | Value | % Cont. |
| BG (Education) | 0.063 | 23.9 | 0.063 | 23.9 |
| BS (Location) | | | 0.034 | 12.9 |
| WGBS | 0.014 | 5.2 | | |
| ISB | | | -0.020 | -7.7 |
| WGWS | 0.186 | 70.9 | 0.186 | 70.9 |
| Total | 0.263 | 100.0 | 0.263 | 100.0 |
| B(GS) | 0.077 | 34.1 | | |
| Max B(GS) | 0.225 | 100.0 | | |

**Table A2**
**Regression Results for the Blinder-Oaxaca Decomposition**

| | Coefficient | Std. Error | t-statistic |
|---|---|---|---|
| Model for Urban Sector | | | |
| edyear | 0.069 | 0.000 | 155 |
| age | 0.008 | 0.001 | 12 |
| age2 | 0.000 | 0.000 | -5 |
| size | -0.095 | 0.001 | -95 |
| gender | -0.070 | 0.005 | -14 |
| job | 0.215 | 0.005 | 41 |
| constant | 12.460 | 0.018 | 705 |
| R-squared = 0.2865 | | | |
| Number of Observations = 101,240 | | | |
| Model for Rural Sector | | | |
| edyear | 0.040 | 0.000 | 116 |
| age | 0.022 | 0.001 | 44 |
| age2 | 0.000 | 0.000 | -40 |
| size | -0.103 | 0.001 | -146 |
| gender | 0.074 | 0.004 | 19 |
| job | 0.138 | 0.002 | 56 |
| constant | 12.165 | 0.013 | 957 |
| R-squared = 0.1892 | | | |
| Number of Observations = 181,147 | | | |
| Pooled Model | | | |
| edyear | 0.052 | 0.000 | 191 |
| age | 0.017 | 0.000 | 42 |
| age2 | 0.000 | 0.000 | -34 |
| size | -0.100 | 0.001 | -172 |
| gender | 0.012 | 0.003 | 4 |
| job | 0.143 | 0.002 | 63 |
| urb (urban = 0; rural = 1) | -0.236 | 0.002 | -101 |
| constant | 12.691 | 0.011 | 1,109 |
| R-squared = 0.3189 | | | |
| Number of Observations = 282,387 | | | |