Economics & Management Series EMS-2024-01

Urbanization and Income Inequality

Takahiro Akita
International University of Japan

March 2024

IUJ Research Institute
International University of Japan

Urbanization and Income Inequality

Takahiro Akita*

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

Abstract

Using decomposable inequality measures, this study presents quantitatively the inverted U-shaped pattern of income inequality that emerges during a population shift from the low-income rural to the high-income urban sector (Kuznets process of urbanization). It investigates the effects of changes in the urban-rural income ratio and within-sector inequalities on the Kuznets process of urbanization. This study also examines urbanization and expenditure inequality in Indonesia using household-level data for 1996-2018. Our analysis reveals that if the urban-rural income ratio is relatively small while the urban-rural difference in income inequality is relatively large, then overall income inequality is likely to increase for a longer period of time as urbanization proceeds. Conversely, if the urban-rural income ratio is relatively large while the urban-rural difference in income inequality is relatively small, then overall income inequality is likely to peak at earlier stages of urbanization. Our analysis also reveals that the contribution of urban inequality to overall income inequality tends to increase as urbanization proceeds, though there may be some fluctuations due to changes in within- and between-sector inequalities. In Indonesia, the share of urban households has risen gradually from 36% to 55%. However, no systematic relationship is observed between the share of urban households and overall expenditure inequality, meaning that Indonesia's household-level data does not support the Kuznets inverted-U hypothesis. However, Indonesia's household-level data shows that urbanization has been associated with a rising contribution of urban inequality and a declining contribution of between-sector inequality.

Key words: urbanization, income inequality, Kuznets process of urbanization, decomposition of the Theil indices, household survey, Indonesia

* The author is grateful to the Japan Society for the Promotion of Science for its financial support (Grant-in-Aid for Scientific Research: 23K01409).

1. Introduction

In his seminal article on economic development and income inequality, Kuznets (1955) advanced the hypothesis that during the early stages of economic development there is an increase in inequality in the personal distribution of income (income inequality), which eventually peaks and then declines during the later stages of economic development. Kuznets argued that this phenomenon occurs due to a population shift from the low-income traditional (rural) sector to the high-income modern (urban) sector over the course of economic development, where the former sector typically has a smaller level of income inequality than the latter. When income inequality is plotted against the level of economic development, this overall process exhibits a bell-shaped or an inverted U-shaped pattern; thus, the hypothesis is referred to as the Kuznets inverted-U hypothesis (Tsakloglou, 1988; Ram, 1991; Hsing and Smyth, 1994; Thornton, 2001; Bahmani-Oskooee and Gelan, 2008; Angeles, 2010).

Using the variance of log income as a measure of inequality, Robinson (1976) introduced a simple model that can generate the inverted U-shaped pattern of inequality during a population shift from the low-income to the high-income sector. Anand and Kanbur (1993) termed this inverted U-shaped pattern the Kuznets process. They employed six commonly used decomposable inequality measures - the Theil *T* index, the Theil *L* index, the squared coefficient of variation, the Atkinson index, the Gini coefficient and the variance of log income - to derive the functional forms of the Kuznets process and obtained the condition under which income inequality attains the peak.[1] Using these functional forms, they also examined empirically the Kuznets hypothesis based on cross-section data for 60 countries.

The first objective of this study is to formally present the Kuznets process using the Theil indices. Because we consider a country divided into the rural and urban sectors and examine inequality changes during a population shift from the low-income rural sector to the high-income urban sector, we describe the Kuznets process of urbanization, where the rural sector typically has a lower within-sector income inequality than the urban sector.

Kuznets (1955) has triggered a large number of empirical studies on the relationship between economic development and income inequality. Due to the lack of long-term time

[1] The Theil indices, the Atkinson index and the Gini coefficient were introduced by Theil (1967), Atkinson (1970) and Gini (1921), respectively.

series data for individual countries, most of these studies have used cross-section or panel datasets to test the Kuznets inverted-U hypothesis.[2] The results of these studies are however mixed. Early studies using cross-country or pooled cross-country datasets, such as Paukert (1973), Ahluwalia (1976a, 1976b), Papanek and Kyn (1986), Campano and Salvatore (1988), Ram (1988), Tsakloglou (1988) and Jha (1996), found evidence of the Kuznets inverted-U relationship, though with a sample containing only developing countries or developed countries, some of these studies lost support for the hypothesis.[3] On the other hand, more recent studies using large comprehensive datasets allowing for panel data and/or country-by-country regressions, such as Hsing and Smyth (1994), Deininger and Squire (1998), Matyas, Konya and Macquarie (1998), Fraser (2006) and Angeles (2010), provided little support for the Kuznets inverted-U hypothesis.

There are two key assumptions that can generate the Kuznets process of urbanization. They are: (1) the ratio of mean income between the rural and urban sectors (urban-rural income ratio) remains unchanged and (2) inequalities within the rural and urban sectors remain unchanged. However, these assumptions are overly restrictive for the long-term dynamic process. During a population shift from the low-income rural sector to the high-income urban sector, both the urban-rural income ratio and within-sector inequalities are likely to change. As a result, most recent empirical studies using large comprehensive datasets have disproved the Kuznets inverted U hypothesis.

The second objective of this study is thus to examine how changes in the urban-rural income ratio and within-sector inequalities impact the Kuznets process of urbanization. Additionally, it tries to demonstrate that income inequality within the urban sector (urban inequality) is becoming increasingly significant in determining overall income inequality.

This paper is organized as follows. Before presenting the Kuznets process of urbanization, Section 2 discusses the characteristics of inequality measures including the Gini coefficient and the generalized entropy class of measures (*GE*), where the *GE* contains

[2] They include Paukert (1973), Ahluwalia (1976a and 1976b), Saith (1983), Papanek and Kyn (1986), Campano and Salvatore (1988), Tsakloglou (1988), Ram (1988 and 1991), Anand and Kanbur (1993), Hsing and Smyth (1994), Jha (1996), Deininger and Squire (1998), Matyas, Konya and Macquarie (1998), Fraser (2006) and Angeles (2010). Note that these studies used per capita GDP, the share of the population employed outside agriculture or the share of the population living in urban areas as a proxy for economic development to test the Kuznets hypothesis. These variables are highly positively correlated with each other.

[3] These studies used the income share of the bottom 20, 40 or 60% of population as a measure of inequality in addition to the income share of the top 20% of population or the Gini coefficient. In the case where the income share of the bottom 20, 40 or 60% is used, they examined the U hypothesis rather than the inverted-U hypothesis.

the Theil indices and the squared coefficient of variation as its members. Section 3 shows that the Theil indices are decomposable by population subgroups, that is, they can be decomposed into the within-group and between-group inequality components. Moreover, Section 3 performs an inequality decomposition analysis by location (rural and urban areas) using a simple example.

Section 4 delineates the Kuznets process of urbanization using the Theil indices. Section 5 investigates the effects of changes in the urban-rural income ratio and within-sector inequalities on the Kuznets process of urbanization. Section 6 examines the extent to which urban inequality contributes to overall income inequality. Section 7 considers Indonesia's urbanization as a case study and examines urbanization and expenditure inequality using household survey (*Susenas*) data from 1996 to 2018. Finally, Section 8 concludes this paper.

2. Characteristics of Inequality Measures

As a measure of income inequality, researchers have often used the Gini coefficient and the generalized entropy class of measures, because they fulfill 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, 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.

For a distribution of incomes in a country consisting of n individuals, $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, an inequality measure can be defined formally as a function of the form

$$I = I(y_1, y_2, \cdots, y_n) \geq 0.$$

This function assigns a non-negative number to a vector of incomes. Using this function, three principles are described as follows.

(P1) Income homogeneity: for any income distribution and any positive number α, we require

$$I(y_1, y_2, \cdots, y_n) = I(\alpha y_1, \alpha y_2, \cdots, \alpha y_n).$$

(P2) Population homogeneity: for any income distribution, we require

$I(y_1, y_2, \cdots, y_n) = I(y_1, y_2, \cdots, y_n;\ y_1, y_2, \cdots, y_n)$.

(P3) The Pigou-Dalton principle of transfers: for any income distribution and any transfer of income $\delta > 0$ from *k*th individual to *j*th individual where $y_j < y_k$, we require

$I(y_1, \cdots, y_j, \cdots, y_k, \cdots, y_n) > I(y_1, \cdots, y_j + \delta, \cdots, y_k - \delta, \cdots, y_n)$,

where $y_j + \delta \leq y_k - \delta$.

For an income distribution $\boldsymbol{y} = (y_1, y_2, \cdots, y_n) > 0$ with the mean income given by $\mu = \frac{1}{n}\sum_{i=1}^{n} y_i > 0$, the Gini coefficient can be defined by

$$G = \frac{2}{n\mu}\operatorname{cov}(\boldsymbol{y}, i(\boldsymbol{y})), \tag{1}$$

where $\operatorname{cov}(\boldsymbol{y}, i(\boldsymbol{y}))$ is the covariance between incomes $\boldsymbol{y}$ and the ranking of these incomes, $i(\boldsymbol{y})$. The Gini coefficient can also be defined by

$$G = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j|. \tag{2}$$

We can show that these two definitions are equivalent.[4]

On the other hand, the generalized entropy class of inequality measures (*GE*) is defined, for $\alpha \neq 0,1$, by

$$GE_\alpha = \frac{1}{\alpha(\alpha-1)}\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\mu}\right)^\alpha - 1\right], \tag{3}$$

where the parameter α is an indicator of inequality aversion, that is, a smaller α indicates more averse to inequality (Sen, 1997). When $\alpha = 0$ and 1, we have, respectively

$$GE_0 = L = \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{\mu}{y_i}\right), \text{ and} \tag{4}$$

$$GE_1 = T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu}\ln\left(\frac{y_i}{\mu}\right), \tag{5}$$

where $\ln(x)$ is the natural logarithm of x. GE_0 and GE_1 are usually called the Theil L and T indices, respectively.[5] We should note that GE_2 is one half of the squared coefficient of variation (CV^2), that is,

$$GE_2 = \frac{1}{2}CV^2 = \frac{1}{2}\left(\frac{\sigma}{\mu}\right)^2, \tag{6}$$

where $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2}$ is the standard deviation of incomes.

We can show that the Gini coefficient and the generalized entropy class of measures (*GE*) satisfy aforementioned three principles. If an inequality measure satisfies income

[4] There are some other definitions for the Gini coefficient, one of which can be obtained based on the Lorenz curve.

[5] They are also called the mean logarithmic deviation and the Theil's entropy index (Theil, 1967), respectively.

homogeneity, then it is a *relative* inequality measure, because with this principle, we need only *relative* income shares to measure income inequality.[6] The Gini coefficient and the *GE* are *relative* inequality measures. Note that the variance of log income, another commonly used inequality measure, satisfies income homogeneity and population homogeneity.[7] Thus, it is a relative inequality measure. But it does not meet the Pigou-Dalton transfer principle. On the other hand, if an inequality measure does not satisfy income homogeneity, then it is an *absolute* inequality measure. Standard deviation and variance are *absolute* measures though they satisfy population homogeneity and the Pigou-Dalton principle of transfers.

3. Decomposition of the Theil Indices by Population Subgroups

In addition to satisfying three principles mentioned above, the generalized entropy class of measures (*GE*) is also decomposable by population subgroups; that is, overall income inequality as measured by the *GE* can be decomposed into the within-group and between-group inequality components (Bourguignon, 1979; Shorrocks, 1980; Anand, 1983).[8] To obtain inequality decomposition equations for $GE_0 = L$ and $GE_1 = T$, suppose that *n* individuals in a country are classified into *m* mutually exclusive and collectively exhaustive groups where group *i* has n_i individuals. Then, the distribution of incomes can be given by a vector of *m* income distributions as follows:

$$\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_m),$$

where $\boldsymbol{y}_i = \left(y_{i1}, y_{i2}, \cdots, y_{in_i}\right)$ is an income distribution of group *i* whose mean income is given by $\mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$. If overall mean income is given by $\mu = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}$, then we have $\mu = \sum_{i=1}^{m}\frac{n_i}{n}\mu_i$ where $n = \sum_{i=1}^{m} n_i$. The Theil L and T indices are now given, respectively, by

$$L = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \ln\left(\frac{\mu}{y_{ij}}\right) \text{ and} \tag{7}$$

$$T = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{y_{ij}}{\mu}\ln\left(\frac{y_{ij}}{\mu}\right). \tag{8}$$

---

[6] Consider a distribution of incomes in a country consisting of five individuals given by $\boldsymbol{y} = (4, 8, 10, 12, 16)$. Then total income is 50. If we set $\alpha = {}^{1}/_{50}$, then income homogeneity implies that $I(4, 8, 10, 12, 16) = I\left({}^{4}/_{50}, {}^{8}/_{50}, {}^{10}/_{50}, {}^{12}/_{50}, {}^{16}/_{50}\right) = I(0.08, 0.16, 0.20, 024, 0.32)$.

[7] The variance of log income is defined by $V = \frac{1}{n}\sum_{i=1}^{n}(\ln(y_i) - \ln(\mu))^2$.

[8] Like the generalized entropy class of measures, the variance of log income is also decomposable by population subgroup. On the other hand, the Gini coefficient cannot generally be decomposed into the within-group and between-group inequality components (Lambert and Aronson, 1993). There is an exception, however. If there are no overlaps in the distributions of incomes between groups, then the Gini coefficient can be decomposed into the within-group and between-group components. Otherwise, the residual component appears in the decomposition equation.

We can now decompose these inequality measures into the within-group and between-group inequality components as follows (detailed derivation of Eqs. (9) and (10) is presented in Appendix 1).

$$L = \sum_{i=1}^{m} \frac{n_i}{n} L_i + \sum_{i=1}^{m} \frac{n_i}{n} \ln\left(\frac{\mu}{\mu_i}\right) = L_W + L_B, \qquad (9)$$

where $L_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \ln\left(\frac{\mu_i}{y_{ij}}\right)$ is the within-group inequality of group i. In Eq. (9), $L_W = \sum_{i=1}^{m} \frac{n_i}{n} L_i$ is the within-group inequality component, which is the weighted average of within-group inequalities with the weights being population shares, while $L_B = \sum_{i=1}^{m} \frac{n_i}{n} \ln\left(\frac{\mu}{\mu_i}\right)$ is the between-group inequality component.

$$T = \sum_{i=1}^{m} \frac{n_i}{n} \frac{\mu_i}{\mu} T_i + \sum_{i=1}^{m} \frac{n_i}{n} \frac{\mu_i}{\mu} \ln\left(\frac{\mu_i}{\mu}\right) = T_W + T_B \qquad (10)$$

where $T_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{y_{ij}}{\mu_i} \ln\left(\frac{y_{ij}}{\mu_i}\right)$ is the within-group inequality of group i. In Eq. (10), $T_W = \sum_{i=1}^{m} \frac{n_i}{n} \frac{\mu_i}{\mu} T_i$ is the within-group inequality component, which is the weighted average of within-group inequalities with the weights being income shares, while $T_B = \sum_{i=1}^{m} \frac{n_i}{n} \frac{\mu_i}{\mu} \ln\left(\frac{\mu_i}{\mu}\right)$ is the between-group inequality component.

Let Y_i and Y be the total income of group *i* and the total income of the country as a whole where $Y_i = \sum_{j=1}^{n_i} y_{ij}$ and $Y = \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ij}$. Then, we have $n_i \mu_i = Y_i$ and $n\mu = Y$. Thus, we can rewrite Eqs. (9) and (10) as follows.

$$L = \sum_{i=1}^{m} \frac{n_i}{n} L_i + \sum_{i=1}^{m} \frac{n_i}{n} \ln\left(\frac{n_i/n}{Y_i/Y}\right) = L_W + L_B, \qquad (11)$$

$$T = \sum_{i=1}^{m} \frac{Y_i}{Y} T_i + \sum_{i=1}^{m} \frac{Y_i}{Y} \ln\left(\frac{Y_i/Y}{n_i/n}\right) = T_W + T_B. \qquad (12)$$

In other words,

$$L = \sum_{i=1}^{m} (\text{pop.share})_i \times L_i + \sum_{i=1}^{m} (\text{pop.share})_i \times \ln\left(\frac{(\text{pop.share})_i}{(\text{income share})_i}\right)$$

$$T = \sum_{i=1}^{m} (\text{income share})_i \times T_i + \sum_{i=1}^{m} (\text{income share})_i \times \ln\left(\frac{(\text{income share})_i}{(\text{pop.share})_i}\right)$$

where $(\text{pop.share})_i$ and $(\text{income share})_i$ are, respectively, the population and income shares of group *i*.

An analysis of income inequality often raises the question of how much of overall income inequality is due to income differences between population subgroups, such as age groups, education groups, rural and urban areas and regions. To address this question, the

GE is commonly used, because it can be decomposed into the within-group and between-group inequality components, as demonstrated earlier.

We illustrate how an inequality decomposition analysis is conducted by the Theil indices using a simple example. Suppose that a country has 10 individuals, of which 6 are in rural areas and 4 in urban areas (Table 1). Using all individuals, we can first obtain overall income inequality. By the Theil *L* index, it is 0.110 (see Eq. 4).

Table 1. Income Distribution for 10 Individuals

Location	Individual	Income	Population Share (%)	Income share (%)
Rural	R1	4	10.0	4.0
	R2	12	10.0	12.0
	R3	6	10.0	6.0
	R4	10	10.0	10.0
	R5	6	10.0	6.0
	R6	10	10.0	10.0
Urban	U1	10	10.0	10.0
	U2	6	10.0	6.0
	U3	16	10.0	16.0
	U4	20	10.0	20.0
	Total (Y)	100	100.0	100.0
	Mean income (μ)	10		
	Standard deviation (σ)	4.73		

(Note) Standard deviation is the population standard deviation, not the sample standard deviation.

Next, using Table 2, income inequalities within rural and urban areas (rural and urban inequalities) are, respectively, 0.069 and 0.099 by the Theil *L* index. The within-group inequality component is the weighted average of rural and urban inequalities with the weights being rural and urban population shares (see Eq. 11).

$$(\text{pop. share})_\text{R} \times (\text{rural inequality}) + (\text{pop. share})_\text{U} \times (\text{urban inequality})$$
$$= (0.6)(0.069) + (0.4)(0.099) = 0.081$$

where R and U denote rural and urban areas, respectively. On the other hand, the between-group inequality component is calculated using the following formula (see Eq. 11).

$$(\text{pop. share})_\text{R} \times \ln\left(\frac{(\text{pop.share})_\text{R}}{(\text{income share})_\text{R}}\right) + (\text{pop. share})_\text{U} \times \ln\left(\frac{(\text{pop.share})_\text{U}}{(\text{income share})_\text{U}}\right)$$
$$= (0.60)\ln\left(\frac{0.60}{0.48}\right) + (0.40)\ln\left(\frac{0.40}{0.52}\right) = 0.029.$$

The sum of the within-group and between-group inequality components (0.081 + 0.029) is, in fact, equal to overall income inequality obtained above (0.110).

Table 2. Income Distribution for 10 Individuals Classified into Rural and Urban Areas

Location	Individuals	Income	Population Share (%)	Income Share (%)
Rural	R1	4	16.7	8.3
	R2	12	16.7	25.0
	R3	6	16.7	12.5
	R4	10	16.7	20.8
	R5	6	16.7	12.5
	R6	10	16.7	20.8
	Sub-total (Y_R)	48	100.0	100.0
	Rural mean (μ_R)	8		
	Standard deviation (σ_R)	2.83		
Urban	U1	10	25.0	19.2
	U2	6	25.0	11.5
	U3	16	25.0	30.8
	U4	20	25.0	38.5
	Sub-total (Y_U)	52	100.0	100.0
	Urban mean (μ_U)	13		
	Standard deviation (σ_U)	5.39		

(Note) Standard deviation is the population standard deviation, not the sample standard deviation.

We can perform a similar decomposition analysis by the Theil T index (see Eqs. 5 and 12). The decomposition results are summarized in Table 3. By the Theil L index, % contributions of rural and urban inequalities are calculated respectively as follows.

$$\%\text{ contribution of rural inequality} = \frac{(\text{pop.share})_{\mathrm{R}} \times (\text{rural inequality})}{\text{Overall inequality}} \times 100 = 37.8,$$

$$\%\text{ contribution of urban inequality} = \frac{(\text{pop.share})_{\mathrm{U}} \times (\text{urban inequality})}{\text{Overall inequality}} \times 100 = 36.0.$$

On the other hand, % contribution of the between-group component is calculated as follows.

$$\%\text{ contribution of B-group component} = \frac{\text{B}-\text{group inequality component}}{\text{Overall inequality}} \times 100 = 26.2.$$

By the Theil T index, % contributions of rural and urban inequalities are calculated respectively as follows.

$$\%\text{ contribution of rural inequality} = \frac{(\text{income share})_{\mathrm{R}} \times (\text{rural inequality})}{\text{Overall inequality}} \times 100 = 29.0,$$

$$\%\text{ contribution of urban inequality} = \frac{(\text{income share})_{\mathrm{U}} \times (\text{urban inequality})}{\text{Overall inequality}} \times 100 = 43.6.$$

On the other hand, % contribution of the between-group component is calculated as follows.

$$\% \text{ contribution of B-group component} = \frac{\text{B-group inequality component}}{\text{Overall inequality}} \times 100 = 27.4.$$

These decomposition results show that income inequality between rural and urban areas accounts for 26-27% of overall income inequality. In other words, if rural-urban income inequality is eliminated, then overall income inequality could be reduced by 26-27%.

Table 3. Decomposition of Income Inequality by the Theil *L* and *T* indices

	Theil *L*		Theil *T*		Pop.	Income
	Value	% cont.	Value	% cont.	Share (%)	share (%)
Rural inequality	0.0695	37.8	0.0647	29.0	60.0	48.0
Urban inequality	0.0993	36.0	0.0899	43.6	40.0	52.0
W-group component	0.0814	73.8	0.0778	72.6		
B-group component	0.0289	26.2	0.0293	27.4		
Overall inequality	0.1103	100.0	0.1071	100.0	100.0	100.0

(Note) % cont. is the % contribution of each component. W-group component is within-group inequality component, while B-group component is the between-group inequality component.

4. Delineating the Kuznets Process of Urbanization

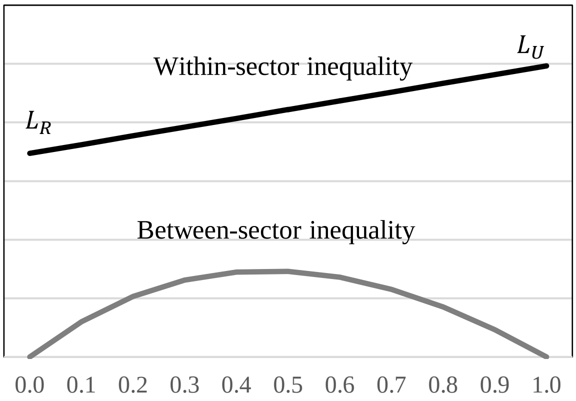
Using the Theil indices, we can delineate the Kuznets process of urbanization under the assumption that the urban-rural ratio of mean income (urban-rural income ratio) and inequalities within the rural and urban sectors remain unchanged throughout the urbanization process.

Consider a country divided into the rural and urban sectors. Suppose first that the rural and urban sectors have the same mean income, that is, there is no income disparity between these two sectors, thus eliminating the between-sector inequality. Then, overall income inequality is equal to the within-sector inequality component. Let p_U be the population share of the urban sector ($p_U = {}^{n_U}/_{n}$). By the Theil L index, the within-sector inequality component can be described as a function of p_U as follows (see Eq. 9).

$$L_w = (1 - p_U)L_R + p_U L_U \geq 0, \tag{13}$$

where L_U and L_R are, respectively, urban and rural inequalities. This is a linear function of p_U. If $p_U = 0$, that is, all individuals are in the rural sector, then $L_w = L_R$. On the other hand, if $p_U = 1$, that is, all individuals are in the urban sector, then $L_w = L_U$. If we assume that urban inequality is larger than rural inequality ($L_U > L_R$), then the within-sector inequality component increases linearly with p_U (see Fig. 1).

Figure 1. Within- and Between-sector Inequalities by the Theil *L* Index



(Note) The horizontal axis presents the population share of the urban sector (p_U), while the vertical axis shows income inequality by the Theil L index.

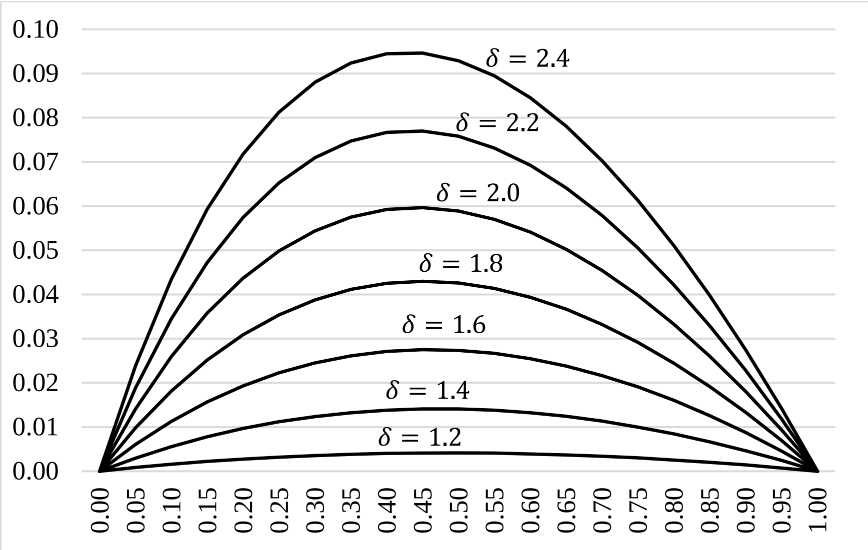
Suppose now that individuals in each sector (rural or urban) have their mean sector income, thus eliminating within-sector inequalities ($L_R = L_U = 0$). Then, overall income inequality is equal to the between-sector inequality component. Let δ be the urban-rural income ratio and assume that mean income is larger in the urban than in the rural sector, that is, $\delta = \frac{\mu_U}{\mu_R} > 1$. Since δ is assumed to be constant, by the Theil L index, the between-sector inequality component can be described as a function of p_U as follows (see Eq. 11).

$$L_B = \ln\big((1 - p_U) + \delta p_U\big) - p_U \ln(\delta) \geq 0 \; . \tag{14}$$

This is a non-linear function of p_U. If $p_U = 0$ or 1, then $L_B = 0$. Otherwise, $L_B > 0$. As shown in Fig. 1, the between-sector inequality component is a concave function.[9] The larger the urban-rural income ratio (δ) is, the more concave the function tends to be (see Fig. 2).

[9] Because $\delta > 1$, we have $\left.\frac{dL_B}{dp_U}\right|_{p_U=0} = (\delta - 1) - \ln(\delta) > 0$, while $\left.\frac{dL_B}{dp_U}\right|_{p_U=1} = \frac{(\delta-1)-\delta\ln(\delta)}{\delta} < 0$.

Figure 2. Between-sector Inequality by the Theil L Index for Different Values of the Urban-rural Income Ratio (δ)



(Note) The horizontal axis presents the population share of the urban sector (p_U), while the vertical axis shows the between-sector income inequality by the Theil L index.

By adding Eqs. (13) and (14), we now obtain the following equation for the Kuznets process of urbanization.

$$L = L_w + L_B = [(1 - p_U)L_R + p_U L_U] + \left[\ln\left((1 - p_U) + \delta p_U\right) - p_U \ln(\delta)\right]. \quad (15)$$

Appendix 2 provides the derivation of this equation. When all individuals are in the rural sector, overall inequality is equal to rural inequality. But it rises with a population shift from the rural to urban sector. If the following condition holds, then it attains the peak before all individuals are in the urban sector.

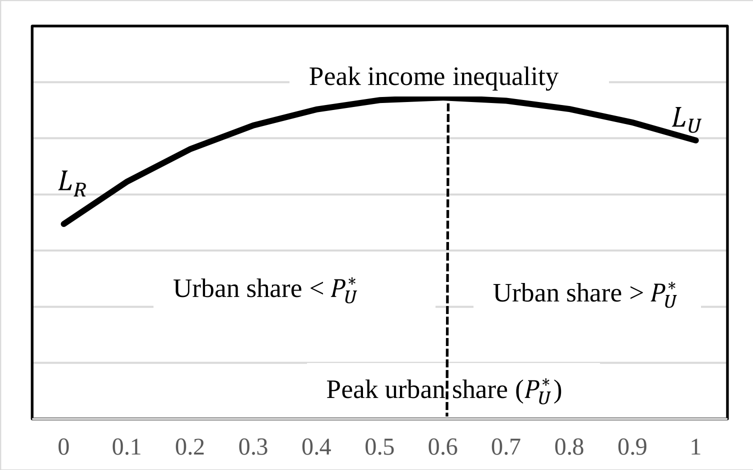
$$0 \leq L_U - L_R < \frac{\delta \ln(\delta) - (\delta - 1)}{\delta}. \quad (16)$$

After this turning point, overall inequality starts to decline. When all individuals are in the urban sector, it is equal to urban inequality. This overall process is depicted in Fig. 3. The peak inequality is attained at

$$p_U^* = \frac{(L_U - L_R) + (\delta - 1) - \ln(\delta)}{(\delta - 1)\left(\ln(\delta) - (L_U - L_R)\right)} \qquad (0 < p_U^* < 1). \quad (17)$$

If Eq. (16) does not hold, then the peak is attained when all individuals are in the urban sector.

Figure 3. Kuznets Process of Urbanization by the Theil *L* Index



(Note) The horizontal axis presents the population share of the urban sector (p_U), while the vertical axis presents overall income inequality by the Theil *L* index.

By the Theil *T* index, we can obtain the following equation for the Kuznets process of urbanization (see Eq. 12).

$$T = \left[\left(\frac{1-p_U}{(1-p_U)+\delta p_U}\right)T_R + \left(\frac{\delta p_U}{(1-p_U)+\delta p_U}\right)T_U\right] + \left[\frac{\delta \ln(\delta) p_U}{(1-p_U)+\delta p_U} - \ln\left((1-p_U)+\delta p_U\right)\right] \tag{18}$$

Appendix 2 provides the derivation of this equation. If the following condition holds, we can draw an inverted U-shaped curve by the Theil *T* index.

$$0 \leq T_U - T_R < (\delta - 1) - \ln(\delta). \tag{19}$$

The peak inequality is attained at

$$p_U^* = \frac{\delta(T_U - T_R) + \delta \ln(\delta) - (\delta - 1)}{(\delta - 1)^2} \qquad (0 < p_U^* < 1). \tag{20}$$

We should note that the right hand side of Eq. (19) is an increasing function of the urban-rural income ratio ($\delta > 1$).

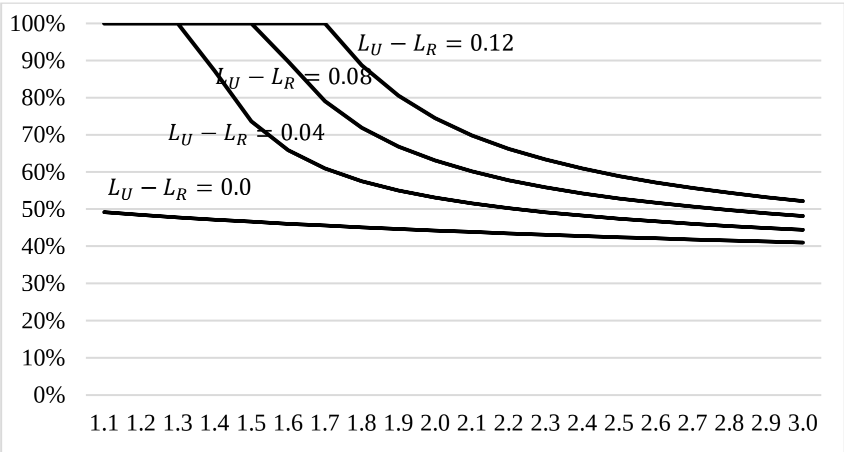
5. Effects of Changes in Urban-Rural Income Ratio and Within-sector Inequalities on the Kuznets Process of Urbanization

Using the Theil indices, the previous section described the Kuznets process of urbanization. As mentioned before, there are two key assumptions that produce this Kuznets process: urban-rural income ratio and inequalities within the rural and urban sectors remain unchanged throughout the urbanization process. However, these assumptions are overly restrictive. This section investigates numerically the effects of changes in the urban-rural income ratio and within-sector inequalities on the Kuznets process of urbanization.

According to Eqs. (16), (17), (19) and (20), there are two determinants of the urban population share that attains the peak inequality (peak urban share, p_U^*). They are the urban-rural income ratio (δ) and the urban-rural difference in income inequality ($L_U - L_R$ by the Theil L and $T_U - T_R$ by the Theil T). They determine whether the actual urban population share exceeds the peak urban share or not (see Fig. 3). If it does not exceed the peak urban share, then overall income inequality rises as urbanization proceeds. On the other hand, if it exceeds, then overall income inequality decreases as urbanization proceeds.

Previous empirical studies on income inequality using the Theil indices have suggested that the urban-rural income ratio ranges between 1.1 and 3.0, while the urban-rural difference in income inequality ranges between 0.02 and 0.12 (Ikemoto, 1985; Glewwe, 1986; Ikemoto and Limskul, 1987; Estudillo, 1997; Akita, Lukman and Yamada, 1999; Eastwood and Lipton, 2004; Liu, 2001; Akita and Miyata, 2018; Mahmud and Akita, 2018; Thein and Akita, 2019). Using these values, Fig. 4 presents the relationship between the urban-rural income ratio (δ) and the peak urban share (p_U^*) by the Theil L index for $L_U - L_R = 0.0$, 0.04, 0.08 and 0.12, where the urban-rural income ratio ranges between 1.1 and 3.0 (see Eqs. 16 and 17).

Figure 4. Relationship between U-R Income Ratio ($\boldsymbol{\delta}$) and Peak Urban Share ($\boldsymbol{p_U^*}$) by the Theil $\boldsymbol{L}$ Index for Different Values of U-R Inequality Difference ($\boldsymbol{L_U - L_R}$)



(Note) The horizontal axis presents the urban-rural income ratio, while the vertical axis presents the share of urban population that attains the peak inequality.

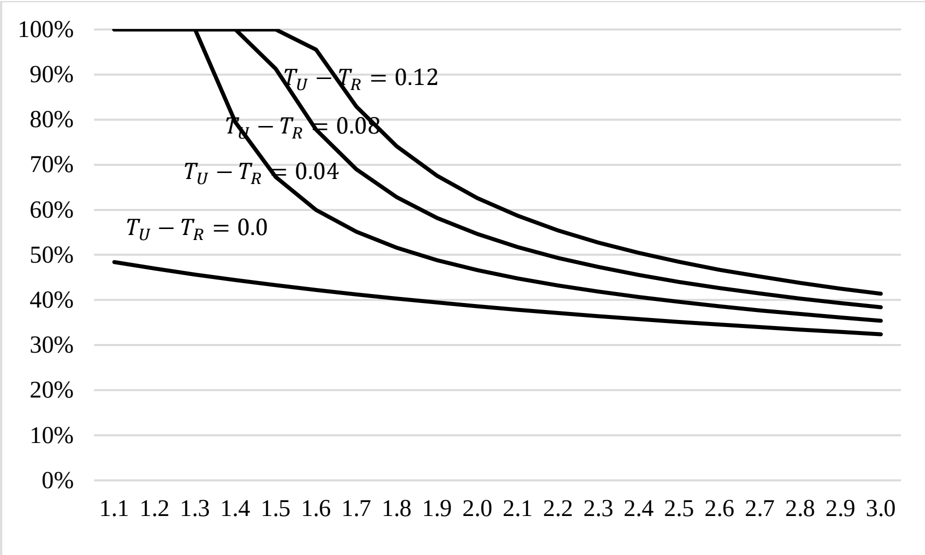
Major observations from Fig. 4 are as follows. First, when $L_U - L_R = 0$, the peak urban share (p_U^*) declines monotonically from 49% to 41% with the urban-rural income ratio (δ) rising from 1.1 to 3.0. Second, if $L_U - L_R > 0$, the peak urban share (p_U^*) is 100% unless the urban-rural income ratio (δ) exceeds the level where Eq. (16) holds. We should note here that the right hand side of Eq. (16) ($\ln(\delta) - (1 - 1/\delta)$) is an increasing function of the ratio (δ).[10] But if the ratio (δ) exceeds this level, the peak urban share (p_U^*) declines monotonically from 100% with the ratio (δ) increasing. Third, if the urban-rural income ratio (δ) remains constant, the peak urban share (p_U^*) rises as the urban-rural difference in income inequality ($L_U - L_R$) increases.

What do these observations imply? If the urban-rural income ratio (δ) is relatively small and the urban-rural inequality difference ($L_U - L_R$) is relatively large, overall income inequality is likely to increase for a longer period of time as urbanization proceeds. For example, if the urban-rural income ratio (δ) is kept smaller than 2 and the urban-rural inequality difference ($L_U - L_R$) is kept larger than 0.1, then overall income inequality, as measured by the Theil L index, rises until the urban population share reaches 70%. Conversely, if the urban-rural income ratio (δ) is kept greater than 2 and the urban-rural inequality difference ($L_U - L_R$) is kept smaller than 0.1, then overall income inequality is likely to decline before the urban population share reaches 70%.

As presented in Fig. 5, the Theil T index provides a similar result qualitatively. If the urban-rural income ratio (δ) is kept smaller than 2 and the urban-rural inequality difference ($T_U - T_R$) is kept larger than 0.1, then overall income inequality, as measured by the Theil T index, rises until the urban population share reaches 60%. Conversely, if the urban-rural income ratio (δ) is kept greater than 2 and the urban-rural inequality difference ($T_U - T_R$) is kept smaller than 0.1, then overall income inequality is likely to decline before the urban population share reaches 60%.

[10] $\frac{d}{d\delta}\left(\ln(\delta) - (1 - 1/\delta)\right) = \frac{\delta - 1}{\delta^2} > 0$ because $1.1 \leq \delta \leq 3.0$.

Figure 5. Relationship between U-R Income Ratio (δ) and Peak Urban Share (p_U^*) by the Theil T Index for Different Values of U-R Inequality Difference ($T_U - T_R$)



(Note) The horizontal axis presents the urban-rural income ratio, while the vertical axis presents the share of urban population that attains the peak inequality.

6. Contribution of Urban Inequality to Overall Income Inequality

It was shown in the previous section that the urban-rural income ratio and the urban-rural inequality difference determine the share of urban population that attains the peak income inequality. However, the previous section did not discuss overall income inequality and its components (within- and between-sector inequalities). Because urban inequality plays an important role in determining overall income inequality, this section examines the extent to which urban inequality contributes to overall income inequality.

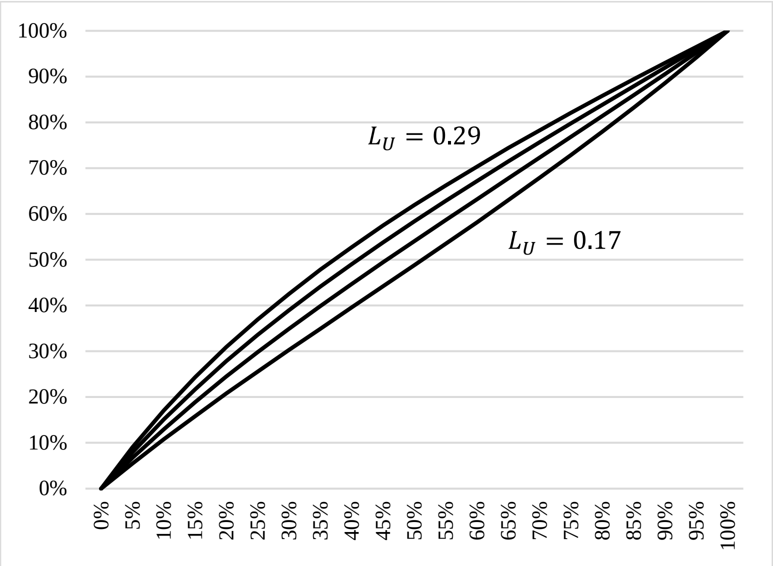
In rural areas, most people are engaged in agriculture and rural income inequality is relatively stable at a low level. We thus assume that rural inequality remains constant at 0.15 by the Theil L index over the Kuznets process of urbanization. We also assume that urban income inequality is larger than rural inequality. By the Theil L index, overall income inequality is given by the following equation (see Eq. 15).

$$L = [(1 - p_U)(0.15) + p_U L_U] + \left[\ln\left((1 - p_U) + \delta p_U\right) - p_U \ln(\delta)\right].$$

This is a function of urban inequality (L_U), urban-rural income ratio (δ) and urban population share (p_U).

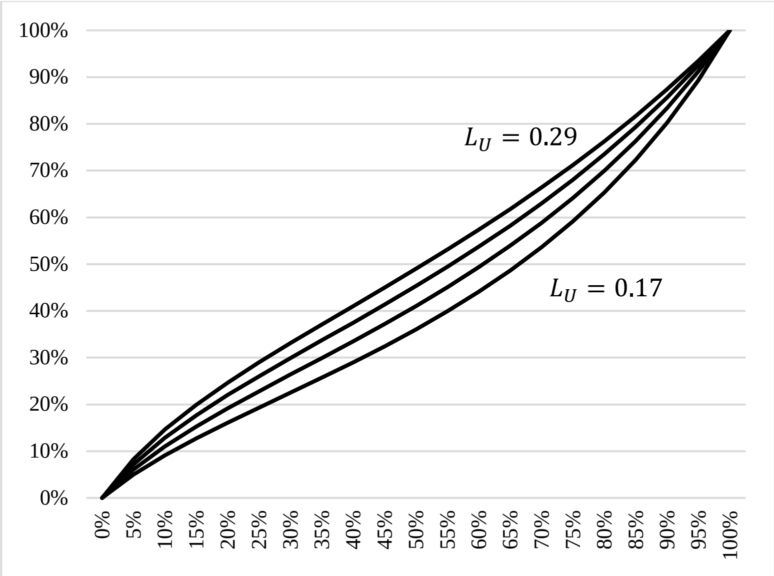
What is the contribution of urban inequality to overall income inequality $\left(\frac{p_U L_U}{L}\right)$? Figs. 6 and 7 present the % contribution of urban inequality for $L_U = 0.17$, 0.21, 0.25 and 0.29 when $\delta = 1.4$ and 2.2, respectively. It is observed that when the urban-rural income ratio (δ) and urban inequality (L_U) are kept constant, the contribution of urban inequality rises monotonically as urbanization proceeds. Second, for a given level of urban inequality (L_U), the smaller the urban-rural income ratio (δ) is, the faster the contribution reaches 50%. For example, when $L_U = 0.25$ and $\delta =$1.4, the contribution reaches 50% before $p_U = 45\%$. But, when $L_U = 0.25$ and $\delta =$2.2, the contribution reaches 50% after $p_U = 55\%$. Third, for a given level of the urban-rural income ratio (δ), the larger urban inequality (L_U) is, the faster the contribution reaches 50%. For example, when $L_U = 0.29$ and $\delta = 1.8$, the contribution reaches 50% before $p_U = 45\%$. When $L_U = 0.17$ and $\delta =$1.8, the contribution reaches 50% after $p_U = 55\%$. These observations suggest that the contribution of urban inequality exhibits an increasing trend over the course of urbanization, though there may be some fluctuations due to changes in within- and between-sector inequalities. This implies that urban inequality plays an increasingly important role in determining overall income inequality as urbanization proceeds.

Figure 6. Contribution of Urban Inequality to Overall Income Inequality by the Theil $\boldsymbol{L}$ index when $\boldsymbol{\delta = 1.4}$



(Note) The horizontal axis presents urban population share, while the vertical axis presents the contribution of urban inequality to overall income inequality.

Figure 7. Contribution of Urban Inequality to Overall Income Inequality by the Theil L index when $\delta = 2.2$



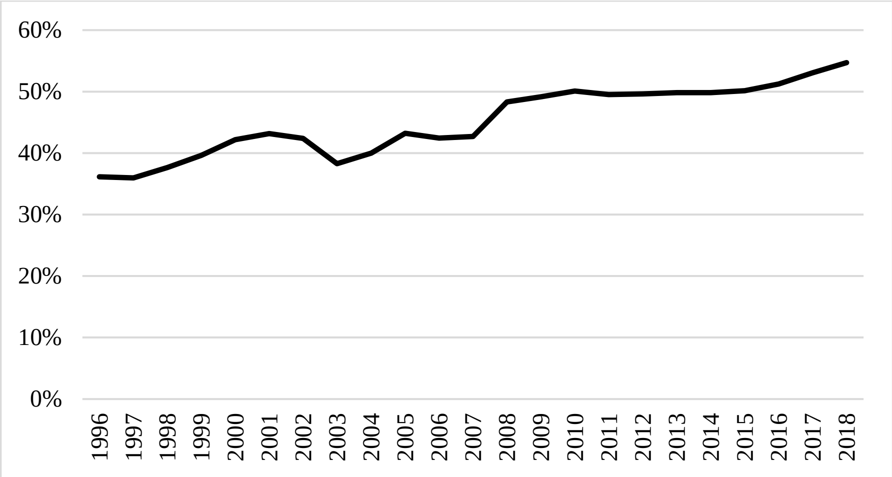
(Note) The horizontal axis presents urban population share, while the vertical axis presents the contribution of urban inequality to overall income inequality.

7. Urbanization and Expenditure Inequality in Indonesia

This section examines urbanization and inequality in the distribution of consumption expenditures (expenditure inequality) in Indonesia using household-level data from the National Socioeconomic Surveys (*Susenas*) for the period from 1996 to 2018.

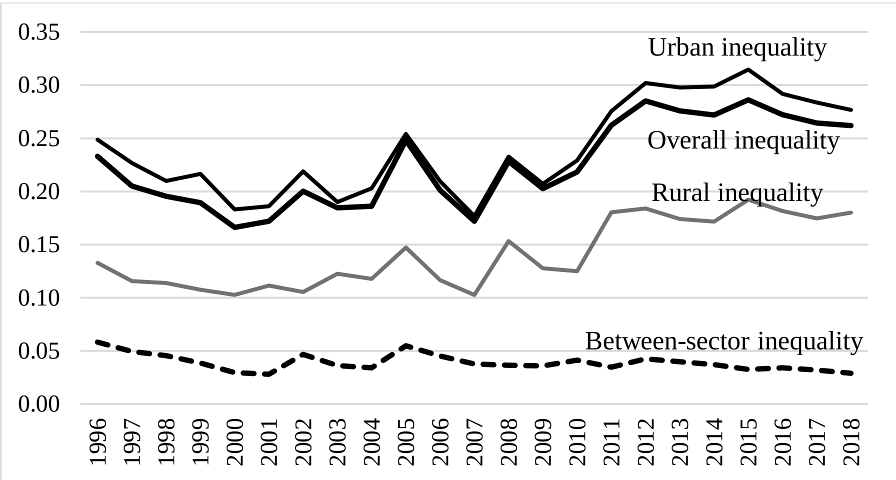
As shown in Fig. 8, urbanization has proceeded steadily over the period 1996-2018. Though there were some fluctuations, the share of urban households has risen from 36.2% to 54.7%. On the other hand, overall expenditure inequality exhibited a U-shaped pattern instead of an inverted U-shaped pattern (Fig. 9). Before 2000, it declined, partly due to the 1997/98 financial crisis. However, since 2000, it has been on the rise and peaked in 2005. Notably, between 2003 and 2005, the two wealthiest decile groups increased their expenditure shares while the shares of the other groups decreased. The Kuznets 20/20 ratio, which measures the ratio of the expenditure share of the richest 20% to that of the poorest 20%, increased significantly from 4.8 to 6.3. Yusuf, Sumner and Rum (2014) argued that one of the main reasons for the increase in expenditure inequality was the rise in domestic rice prices during 2003-2005, which had a more adverse impact on the poor than the rich.

Figure 8. Share of Urban Households



(Note) The vertical axis presents the share of urban households.

Figure 9. Expenditure Inequalities by the Theil *L* Index

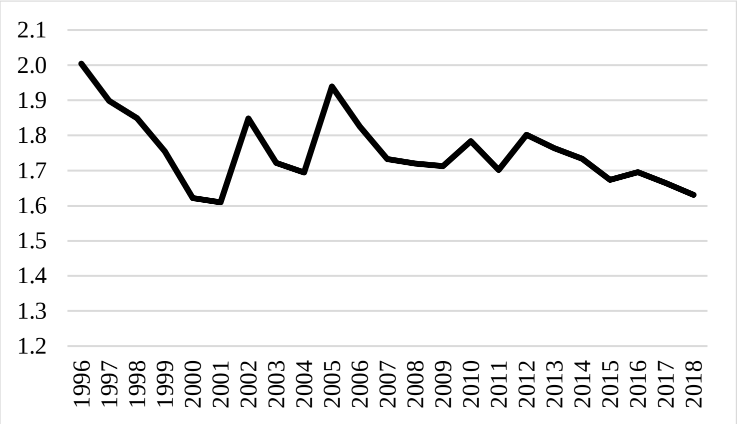


(Note) The vertical axis presents expenditure inequalities by the Theil *L* index.

Between 2005 and 2007, overall expenditure inequality sharply decreased and reached its lowest point in 2007. However, from 2007 to 2012, it increased again, and the Kuznets 20/20 ratio rose substantially from 4.9 to 6.9, similar to the period of 2003-2005. According to Yusuf, Sumner and Rum (2014), the increasing fuel subsidies had a dis-equalizing effect on expenditures since their impact on incomes was regressive. Additionally, changes in formal labor market regulations, such as rising minimum wages, strengthening labor unions, and increasing retirement benefits, were likely to have benefited the rich more than the poor, leading to an increase in inequality. Since 2012, overall inequality has remained relatively stable but at a high level.

Due largely to urbanization, the levels and trends of overall expenditure inequality resemble very closely those of urban inequality, particularly between 2003 and 2012. Like other Asian countries, rural inequality was much smaller than urban inequality. But its rising and declining trends were similar to those of urban inequality. The urban-rural difference in expenditure inequality ($L_U - L_R$) ranged between 0.07 and 0.13. On the other hand, inequality between the rural and urban sectors (between-sector inequality) was relatively small and exhibited a slight declining trend. The urban-rural expenditure ratio (δ) has declined from 2.0 to 1.6 though there were fluctuations (Fig. 10). As suggested by Knight and Sabot (1983), urbanization appears to have narrowed the urban-rural income gap in Indonesia.[11]

Figure 10. Urban-Rural Expenditure Ratio (δ)



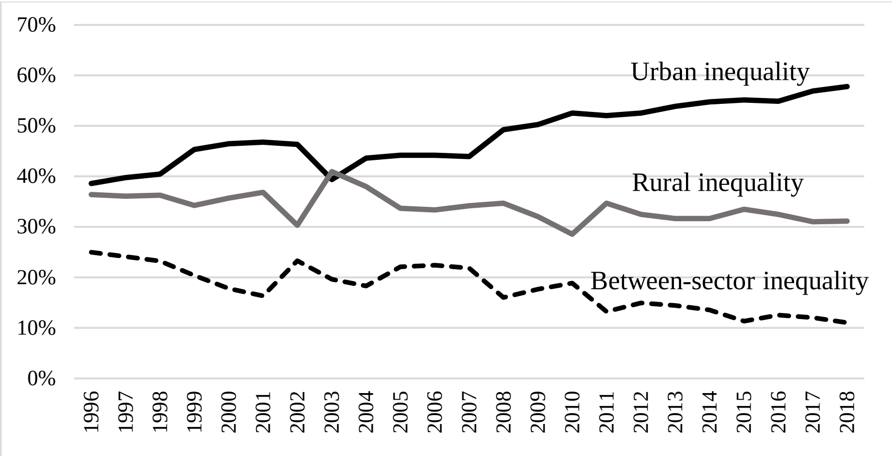
(Note) The vertical axis presents the urban-rural expenditure ratio (δ).

Using Eq. (9), overall expenditure inequality is decomposed into the within- and between-sector components. The result is presented in Fig. 11. Despite some fluctuations, the contribution of urban inequality exhibited an upward trend, increasing substantially from 39% to 58%, while that of rural inequality has declined from 36% to 31%. On the other

[11] Knight and Sabot (1983) introduced two forces of inequality changes: the composition and compression effects. In the urban-rural dual framework, the composition effect refers to the effect of the expansion of the urban sector on wage inequality holding the structure of urban and rural wages constant, while the compression effect refers to the effect of the compression of that wage structure on wage inequality holding the composition of the urban and rural sectors constant. However, they did not consider the effect of changes in urban and rural inequalities on inequality. In our analysis, the composition effect denotes the effect of the change in the proportion of the urban sector (p) on expenditure inequality, while the compression effect refers to the effect of the change in the urban-rural expenditure ratio (δ) on expenditure inequality.

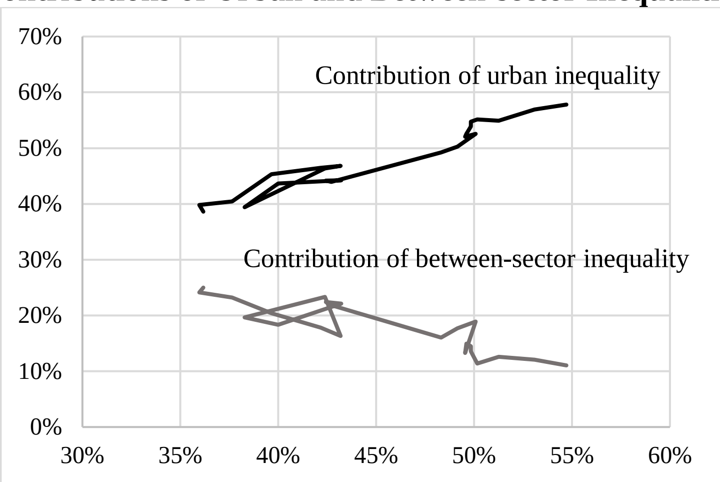
hand, the contribution of between-sector inequality has decreased notably from 25% to 11%. Fig. 12 presents the relationship between the share of urban households and the contributions of urban and between-sector inequalities to overall expenditure inequality. The results suggest that urbanization has been associated with the rising contribution of urban inequality and the declining contribution of between-sector inequality.

Figure 11. Decomposition of Overall Expenditure Inequality by the Theil *L* Index (Contribution of Each Component)



(Note) The vertical axis presents the contribution to overall expenditure inequality.

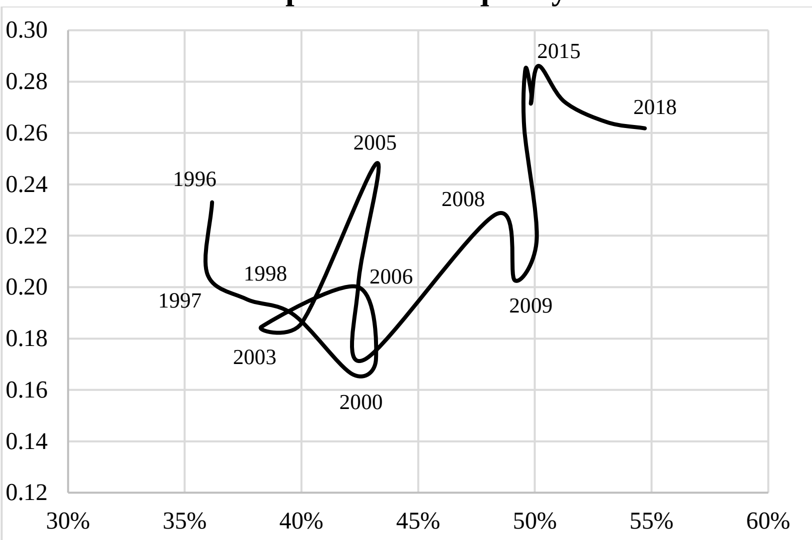
Figure 12. Relationship between the Share of Urban Households and the Contributions of Urban and Between-sector Inequalities



(Note) The horizontal axis presents the share of urban households, while the vertical axis presents the contributions of urban and between-sector inequalities to overall expenditure inequality.

Section 3 used the Theil indices to describe quantitatively the Kuznets process of urbanization, which refers to the inverted U-shaped relationship between income inequality and the share of urban population. This raises the question of how Indonesia has experienced changes in expenditure inequality as urbanization proceeds. Fig. 13 shows the relationship between the share of urban households and overall expenditure inequality for the period 1996-2018. The share of urban households has risen from 36.2% to 54.7%. But, no systematic relationship is observed between the share of urban households and overall expenditure inequality. In other words, Indonesia's household-level data does not support the Kuznets inverted-U hypothesis. It is worth noting that the period of 1996-2018 may be too short to draw definitive conclusions about the existence of an inverted U-shaped relationship between them in Indonesia.

Figure 13. Relationship between the Share of Urban Households and Overall Expenditure Inequality

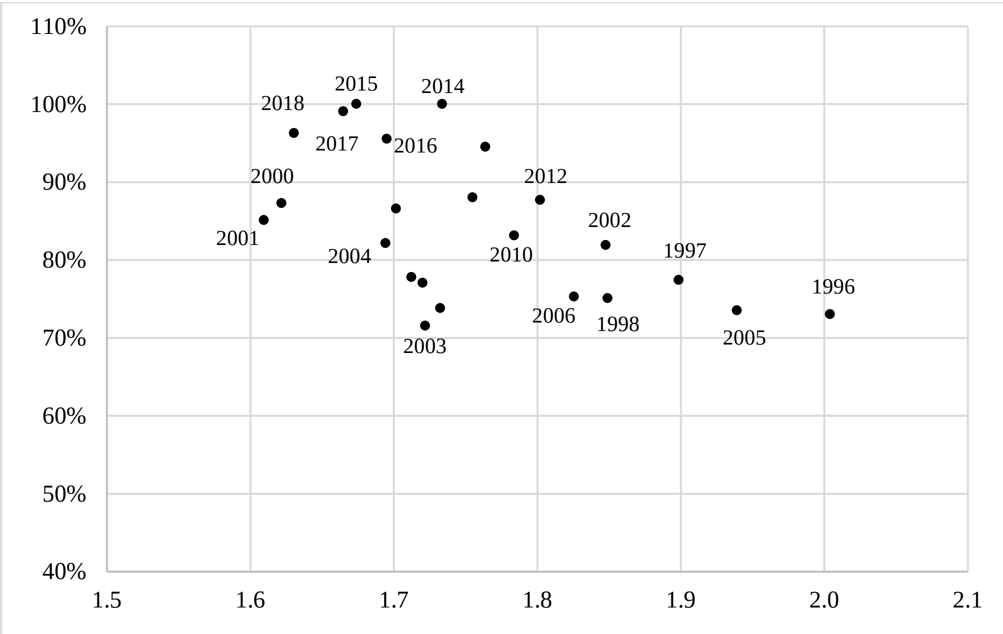


(Note) The horizontal axis presents the share of urban households, while the vertical axis presents overall expenditure inequality.

In Section 4, we analyzed the relationship between the urban-rural income ratio (δ) and the urban population share that attains the peak inequality (p_U^*) for different values of the urban-rural difference in income inequality ($L_U - L_R$). Our analysis revealed that the peak urban share (p_U^*) is smaller than 100% if the urban-rural income ratio (δ) exceeds the level where Eq. (16) holds. What is the relationship between the urban-rural expenditure ratio (δ) and the share of urban households that attains the peak expenditure inequality (p_U^*) in Indonesia? Fig. 14 shows the relationship over the period 1996-2018. In Indonesia, the urban-rural difference in expenditure inequality ($L_U - L_R$) ranges between 0.07 and 0.13

(see Fig. 9), while the urban-rural expenditure ratio (δ) ranges between 1.6 and 2.0. In 2014 and 2015, $L_U - L_R < \frac{\delta \ln(\delta)-(\delta-1)}{\delta}$ does not hold (see Eq. 16); thus, the peak urban share is 100%. Otherwise, the peak urban share is smaller than 100%. It seems that the peak urban share has declined as the urban-rural expenditure ratio (δ) increases. This is an illustration of what we observed quantitatively in Section 4.

Figure 14. Relationship between the Urban-rural Expenditure Ratio (δ) and the Urban Share of Households that attains the Peak Inequality (p_U^*)



(Note) The horizontal axis presents the urban-rural expenditure ratio, while the vertical axis presents the share of urban households that attains the peak expenditure inequality.

8. Concluding Remarks

Using decomposable inequality measures, this study presented quantitatively the inverted U-shaped pattern of income inequality that emerges during a population shift from the low-income rural to the high-income urban sector (the Kuznets process of urbanization). However, this is under the assumption that the ratio of mean income between the rural and urban sectors and inequalities within these sectors remain unchanged. Because these assumptions are highly restrictive in the long-term dynamic process, the study also investigated the effects of changes in the urban-rural income ratio and within-sector inequalities on the Kuznets process of urbanization.

Our analysis revealed that if the urban-rural income ratio is relatively small while the urban-rural difference in income inequality is relatively large, then overall income inequality is likely to increase for a longer period of time as urbanization proceeds. Conversely, if the urban-rural income ratio is relatively large while the urban-rural difference in income inequality is relatively small, then overall income inequality is likely to peak at earlier stages of urbanization. Furthermore, our analysis revealed that the contribution of urban inequality to overall income inequality tends to increase as urbanization proceeds, though there may be some fluctuations due to changes in within- and between-sector inequalities. This means that urban inequality plays an increasingly important role in determining overall income inequality as urbanization proceeds.

Finally, this study examined urbanization and expenditure inequality in Indonesia using household-level data from the National Socioeconomic Surveys (*Susenas*) for the period 1996-2018. The share of urban households has risen gradually from 36.2% to 54.7%, with some fluctuations. However, no systematic relationship is observed between the share of urban households and overall expenditure inequality, meaning that Indonesia's household-level data does not support the Kuznets inverted-U hypothesis. The period 1996-2018 may be too short to judge whether there is an inverted U-shaped relationship between them. However, Indonesia's household-level data showed that urbanization has been associated with a rising contribution of urban inequality and a declining contribution of between-sector inequality. This implies that urban inequality has played an increasingly important role in determining overall expenditure inequality in Indonesia.

Appendix 1: Decomposition of Inequality Measures by Population Subgroups

1. Theil *L* Index

$$L = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\ln\left(\frac{\mu}{y_{ij}}\right) = \sum_{i=1}^{m}\frac{n_i}{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\ln\left(\frac{\mu_i}{y_{ij}}\frac{\mu}{\mu_i}\right)$$

$$= \sum_{i=1}^{m}\frac{n_i}{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\ln\left(\frac{\mu_i}{y_{ij}}\right) + \ln\left(\frac{\mu}{\mu_i}\right)\right]$$

$$= \sum_{i=1}^{m}\frac{n_i}{n}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\ln\left(\frac{\mu_i}{y_{ij}}\right)\right] + \sum_{i=1}^{m}\frac{n_i}{n}\ln\left(\frac{\mu}{\mu_i}\right)\left[\frac{1}{n_i}\sum_{j=1}^{n_i}(1)\right]$$

$$= \sum_{i=1}^{m}\frac{n_i}{n}L_i + \sum_{i=1}^{m}\frac{n_i}{n}\ln\left(\frac{\mu}{\mu_i}\right) = L_W + L_B \ \text{ since } \frac{1}{n_i}\sum_{j=1}^{n_i}(1) = 1.$$

2. Theil *T* Index

$$T = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{y_{ij}}{\mu}\ln\left(\frac{y_{ij}}{\mu}\right) = \sum_{i=1}^{m}\frac{n_i}{n}\frac{\mu_i}{\mu}\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{y_{ij}}{\mu_i}\ln\left(\frac{y_{ij}}{\mu_i}\frac{\mu_i}{\mu}\right)$$

$$= \sum_{i=1}^{m} \frac{n_i}{n}\frac{\mu_i}{\mu}\frac{1}{n_i}\sum_{j=1}^{n_i} \frac{y_{ij}}{\mu_i}\left[\ln\left(\frac{y_{ij}}{\mu_i}\right) + \ln\left(\frac{\mu_i}{\mu}\right)\right]$$

$$= \sum_{i=1}^{m} \frac{n_i}{n}\frac{\mu_i}{\mu}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} \frac{y_{ij}}{\mu_i}\ln\left(\frac{y_{ij}}{\mu_i}\right)\right] + \sum_{i=1}^{m} \frac{n_i}{n}\frac{\mu_i}{\mu}\ln\left(\frac{\mu_i}{\mu}\right)\left(\frac{1}{n_i}\sum_{j=1}^{n_i} \frac{y_{ij}}{\mu_i}\right)$$

$$= \sum_{i=1}^{m} \frac{n_i}{n}\frac{\mu_i}{\mu}T_i + \sum_{i=1}^{m} \frac{n_i}{n}\frac{\mu_i}{\mu}\ln\left(\frac{\mu_i}{\mu}\right) = T_W + T_B \text{ since } \frac{1}{n_i}\sum_{j=1}^{n_i} \frac{y_{ij}}{\mu_i} = \frac{\mu_i}{\mu_i} = 1.$$

3. Generalized Entropy Class of Measures

The generalized entropy class of inequality measures is defined by:

$$GE_\alpha(\boldsymbol{y}) = \frac{1}{\alpha(\alpha-1)}\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\left(\frac{y_{ij}}{\mu}\right)^\alpha - 1\right],$$

where $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, $n = \sum_{i=1}^{m} n_i$ and $\mu = \sum_{i=1}^{m} \frac{n_i}{n}\mu_i$. This equation can be decomposed into the within- and between-group components as follows.

$$GE_\alpha(\boldsymbol{y}) = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i}{\mu}\right)^\alpha GE_\alpha(\boldsymbol{y}_i) + \frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{m} \frac{n_i}{n}\left[\left(\frac{\mu_i}{\mu}\right)^\alpha - 1\right] = GE_W + GE_B,$$

where $GE_\alpha(\boldsymbol{y}_i) = \frac{1}{\alpha(\alpha-1)}\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\left(\frac{y_{ij}}{\mu_i}\right)^\alpha - 1\right]$ is the within-group inequality of group *i* and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \cdots, y_{in_i})$. In this equation, $GE_W = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i}{\mu}\right)^\alpha GE_\alpha(\boldsymbol{y}_i)$ is the within-group inequality component, which is the weighted sum of within-group inequalities, while $GE_B = \frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{m} \frac{n_i}{n}\left[\left(\frac{\mu_i}{\mu}\right)^\alpha - 1\right]$ is the between-group inequality component.

When $\alpha = 2$, we have

$$GE_2(\boldsymbol{y}) = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i}{\mu}\right)^2 GE_2(\boldsymbol{y}_i) + \frac{1}{2}\sum_{i=1}^{m} \frac{n_i}{n}\left[\left(\frac{\mu_i}{\mu}\right)^2 - 1\right].$$

Since $GE_2(\boldsymbol{y}) = \frac{1}{2}CV(\boldsymbol{y})^2$ and $\mathrm{GE}_2(\boldsymbol{y}_i) = \frac{1}{2}CV(\boldsymbol{y}_i)^2$, this decomposition equation can be rewritten as

$$CV(\boldsymbol{y})^2 = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i}{\mu}\right)^2 CV(\boldsymbol{y}_i)^2 + \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i-\mu}{\mu}\right)^2 = CV_W^2 + CV_B^2\ ,$$

where $CV(\boldsymbol{y}_i)^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(\frac{y_{ij}-\mu_i}{\mu_i}\right)^2$ is the within-group inequality of group *i*. This equation is the inequality decomposition equation for the squared coefficient of variation. $CV_W^2 = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i}{\mu}\right)^2 CV(\boldsymbol{y}_i)^2$ is the within-group inequality component and $CV_B^2 = \sum_{i=1}^{m} \frac{n_i}{n}\left(\frac{\mu_i-\mu}{\mu}\right)^2$ is the between-group inequality component.

Appendix 2: Kuznets Process of Urbanization

By the Theil L index, the within-group inequality component is given by

$$L_W = p_R L_R + p_U L_U = (1 - p_U)L_R + p_U L_U.$$

On the other hand, because $\mu = \mu_R p_R + \mu_U p_U$ and $\delta = \frac{\mu_U}{\mu_R}$, the between-group inequality component is given by:

$$L_B = p_R \ln\left(\frac{\mu}{\mu_R}\right) + p_U \ln\left(\frac{\mu}{\mu_U}\right) = p_R \ln\left(\frac{\mu_R p_R + \mu_U p_U}{\mu_R}\right) + p_U \ln\left(\frac{\mu_R p_R + \mu_U p_U}{\mu_U}\right)$$

$$= (1 - p_U)\ln\big((1 - p_U) + \delta p_U\big) + p_U \ln\left(\frac{1 - p_U}{\delta} + p_U\right)$$

$$= (1 - p_U)\ln\big((1 - p_U) + \delta p_U\big) + p_U \ln\big((1 - p_U) + \delta p_U\big) - p_U \ln(\delta)$$

$$= \ln\big((1 - p_U) + \delta p_U\big) - p_U \ln(\delta).$$

By adding these two equations, we obtain the equation for the Kuznets process of urbanization by the Theil *L* index.

$$L = L_w + L_B = [(1 - p_U)L_R + p_U L_U] + \big[\ln\big((1 - p_U) + \delta p_U\big) - p_U \ln(\delta)\big]$$

By differentiating this equation with respect to p_U, we have

$$\frac{\partial L}{\partial p_U} = (L_U - L_R) + \frac{(\delta - 1) - \big((1 - p_U) + \delta p_U\big)\ln(\delta)}{(1 - p_U) + \delta p_U}.$$

Because $L_U > L_R$ and $\delta > 1$ by assumption, we have $\left.\frac{\partial L}{\partial p_U}\right|_{p_U=0} = (L_U - L_R) + (\delta - 1) - \ln(\delta) > 0$. But, $\left.\frac{\partial L}{\partial p_U}\right|_{p_U=1} = (L_U - L_R) + \frac{(\delta - 1) - \delta \ln(\delta)}{\delta}$ can be positive or negative because $\frac{(\delta - 1) - \delta \ln(\delta)}{\delta} < 0$ for $\delta > 1$. If $0 \le L_U - L_R < \frac{\delta \ln(\delta) - (\delta - 1)}{\delta}$, then *L* attains the peak before all individuals are in the urban sector. The peak value is obtained by setting this equation to 0 and solving for p_U. Let p_U^* denote the urban population share that achieves the peak Theil *L* value. Then, we have

$$p_U^* = \frac{(L_U - L_R) + (\delta - 1) - \ln(\delta)}{(\delta - 1)\big(\ln(\delta) - (L_U - L_R)\big)}.$$

By substituting the value of p_U^* into the equation for *L* given above, we obtain the peak Theil *L* value.

By the Theil *T* index, the within-group inequality component is given by:

$$T_W = \frac{n_R}{n}\frac{\mu_R}{\mu}T_R + \frac{n_U}{n}\frac{\mu_U}{\mu}T_U = p_R \frac{\mu_R}{\mu_R p_R + \mu_U p_U} T_R + p_U \frac{\mu_U}{\mu_R p_R + \mu_U p_U} T_U$$

$$= \left[\frac{1 - p_U}{(1 - p_U) + \delta p_U}\right] T_R + \left[\frac{\delta p_U}{(1 - p_U) + \delta p_U}\right] T_U.$$

On the other hand, the between-group inequality component is given by:

$$T_B = \frac{n_R}{n}\frac{\mu_R}{\mu}\ln\left(\frac{\mu_R}{\mu}\right) + \frac{n_U}{n}\frac{\mu_U}{\mu}\ln\left(\frac{\mu_U}{\mu}\right)$$

$$= p_R \frac{\mu_R}{\mu_R p_R + \mu_U p_U}\ln\left(\frac{\mu_R}{\mu_R p_R + \mu_U p_U}\right) + p_U \frac{\mu_U}{\mu_R p_R + \mu_U p_U}\ln\left(\frac{\mu_U}{\mu_R p_R + \mu_U p_U}\right)$$

$$= \frac{1 - p_U}{(1 - p_U) + \delta p_U}\ln\left(\frac{1}{(1 - p_U) + \delta p_U}\right) + \frac{\delta p_U}{(1 - p_U) + \delta p_U}\ln\left(\frac{\delta}{(1 - p_U) + \delta p_U}\right)$$

$$= \frac{\delta \ln(\delta) p_U}{(1-p_U)+\delta p_U} - \ln\big((1-p_U)+\delta p_U\big).$$

By adding these two equations, we obtain the following equation for the Kuznets process of urbanization by the Theil *T* index.

$$T = \left[\left(\frac{1-p_U}{(1-p_U)+\delta p_U}\right) T_R + \left(\frac{\delta p_U}{(1-p_U)+\delta p_U}\right) T_U\right] + \left[\frac{\delta \ln(\delta) p_U}{(1-p_U)+\delta p_U} - \ln\big((1-p_U)+\delta p_U\big)\right].$$

By differentiating this equation with respect to p_U, we have

$$\frac{\partial T}{\partial p_U} = \frac{\delta(T_U - T_R) + \delta \ln(\delta) - (\delta-1)\big((1-p_U)+\delta p_U\big)}{\big((1-p_U)+\delta p_U\big)^2}.$$

Because $T_U > T_R$ and $\delta > 1$ by assumption, we have $\left.\frac{\partial T}{\partial p_U}\right|_{p_U=0} = \delta(T_U - T_R) + \delta\ln(\delta) - (\delta-1) > 0$. But, $\left.\frac{\partial T}{\partial p_U}\right|_{p_U=1} = \frac{(T_U-T_R)+\ln(\delta)-(\delta-1)}{\delta}$ can be positive or negative because $\ln(\delta) - (\delta-1) < 0$ for $\delta > 1$. If $0 \le T_U - T_R < (\delta-1) - \ln(\delta)$, then *T* attains the peak before all individuals are in the urban sector. The peak value is obtained by setting this equation to 0 and solving for p_U. Let p_U^* denote the urban population share that achieves the peak Theil *T* value. Then, we have

$$p_U^* = \frac{\delta(T_U-T_R)+\delta\ln(\delta)-(\delta-1)}{(\delta-1)^2}.$$

By substituting the value of p_U^* into the equation for *T* given above, we obtain the peak Theil *T* value.

References

Ahluwalia, M. S. (1976a). Income distribution and development: Some stylized facts. The *American Economic Review*. 66(2), 128-135.

Ahluwalia, M. S. (1976b). Inequality, poverty and development. Journal of development economics. *Journal of Development Economics*, 3, 307-342.

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. (2018). Spatial inequalities in Indonesia. 1996-2010: A hierarchical decomposition analysis. *Social Indicators Research*, 138(3), 829-852.

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

Anand, S., & Kanbur, S. M. R. (1993). The Kuznets process and the inequality-development relationship. *Journal of Development Economics*, 40(1), 25-52.

Angeles, L. (2010). An alternative test of Kuznets' hypothesis. *The Journal of Economic Inequality*. 8, 463–473.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2, 244-263.

Bahmani-Oskooee, M., & Gelan, A. (2008). Kuznets inverted-U hypothesis revisited: a time-series approach using US data. *Applied Economics Letters*, 15, 677–681.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47(4), 901–920.

Campano, F., and Salvatore, D. (1988). Economic development, income inequality and Kuznets' U-shaped hypothesis. *Journal of Policy Modeling*, 10(2), 265-280.

Deininger, K., & Squire. L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics*, 57, 259–287.

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

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

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

Fraser, G. (2006). Inequality and development across and within countries. *World Development*, 34(9), 1459–1481.

Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, 31(121), 124-126.

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), 255–274.

Hsing, Y., & Smyth, J. S. (1994). Kuznets's inverted-U hypothesis revisited. *Applied Economics Letters*, 1(7), 111-113.

Ikemoto, Y. (1985). Income distribution in Malaysia: 1957-80. *The Developing Economies*, 23(4), 347-367.

Ikemoto, Y., & Limskul, K. (1987). Income inequality and regional disparity in Thailand 1962-81. *The Developing Economies*, 25(3), 249-269.

Jha, S. K. (1996). The Kuznets curve: A reassessment. *World Development*, 24(4), 773-780

Knight, J. B., & Sabot, R. H. (1983). Educational expansion and the Kuznets effect. *American Economic Review*, 73(5), 1132-1136.

Kuznets, S. S. (1955). Economic growth and income inequality. *American Economic Review*, 45(1), 1–28.

Lambert, P. J., & Aronson, J. R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, 103(420), 1221-1227.

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

Mahmud, S. S., & Akita, T. (2018). Urban and rural dimensions of the role of education in income inequality in Bangladesh. *Review of Urban and Regional Development Studies*, 30(3), 169-183.

Matyas, L., Konya, L, & Macquarie, L. (1998). The Kuznets U-curve hypothesis: some panel data evidence. *Applied Economics Letters*, 5(11), 693-697.

Papanek, G. F., & Kyn, O. (1986). The effect on income distribution of development, the growth rate and economic strategy. *Journal of Development Economics*, 23(1), 55-65.

Paukert, F. (1973). Income distribution at different levels of development: A survey of evidence, *International Labour Review*, 108, 97-125.

Ram, R. (1988). Economic development and income inequality: Further evidence on the U-curve hypothesis. *World Development*, 16(11), 1371-1376.

Ram, R. (1991). Kuznets's inverted-U hypothesis: Evidence from a highly developed country. *Southern Economic Journal*, 57(4), 1112–23.

Robinson, S. (1976). A note on the U hypothesis relating income inequality and economic development. *The American Economic Review*, 66(3), 437-440.

Saith, A. (1983). Development and distribution: A critique of the cross-country U-hypothesis. *Journal of Development Economics*, 13, 367-382.

Sen, A. (1997). *On Economic Inequality*, New York: Oxford University Press.

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

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North Holland.

Thein, A. T. Z., & Akita, T. (2019). Education and expenditure inequality in Myanmar: An analysis with the 2006 and 2012 Household Income and Expenditure Survey in an urban and rural setting. *Regional Science Policy & Practice*, 11, 55–70.

Thornton, J. (2001). The Kuznets inverted-U hypothesis: panel data evidence from 96 countries. *Applied Economics Letters*, 8(1), 15-16.

Tsakloglou, P. (1988). Development and inequality revisited. *Applied Economics*, 20(4), 509-531.

Yusuf, A A., Sumner, A., & Rum, I. A. (2014). Twenty years of expenditure inequality in Indonesia, 1993-2013. *Bulletin of Indonesian Economic Studies*, 50(2), 243-254.