Journal of Business and Economics, ISSN 2155-7950, USA
November 2018, Volume 9, No. 11, pp. 983-998
DOI: 10.15341/jbe(2155-7950)/11.09.2018/007

http://www.academicstar.us

Decomposition of Inequality in Vietnam 2012

Tadashi Kikuchi
(Faculty of Economics, Teikyo University, Tokyo, Japan)

Abstract: Income distribution in developing countries is changing dynamically due to changes in human activities on national development process. For political analysis, it is necessary to ascertain the main people's socioeconomic activities influential on the change of income distribution. This paper implements the decomposition of inequality in Vietnam and calculates inequality indices to study which factors caused income distribution changes of Vietnam in 2012. Under the market-oriented "Doi moi" economic reforms since 1986, author shows that Vietnamese household income distribution dynamically changed in itself through human activities changes. Author also mentions that it resulted in the wide-gap between the rich and the poor. For fixing and decreasing this unbalance on the income distribution, author insists that studying influence of multiple households' key characters, for instance, household's residential area, household-head's occupation and education level on income distribution helps Vietnam set back on the right path for its sustainable development with a long term perspective.

Key words: household data analysis; inequality; regional development
JEL codes: D

1. Introduction

Between the end of the Vietnamese war in 1975 and the mid-eighties, poverty in citizen society was so prevalent that practical methods to rectify the situation had limited effectiveness and interest in questions concerning inequality in Vietnam. Consequently, two main recognizable facts on inequality and economic development were documented in the Vietnamese history. First, when the Vietnamese war ended, the national economy had been sluggish for long under the system of central economic management. During a decade of this period when the economy was in a deadlock, the central governors and representatives from the national assembly of the Communist Party attempted to lead the national economy. However, their efforts were not successful to improve people's living standard in a nationwide at that time. Second, after all the political declaration of "Doi moi" reforms, in 1986, the Vietnamese economy started seeking new business opportunities in the domestic and overseas markets through the process of changing a central economic management structure to a market oriented one. On January 11, 2007, Vietnam officially joined the World Trade Organization (WTO) as its 150th member. These political and structural shift from a closed economy to a more open one resulted in providing business chances to all people and bridging economic inequality to some extent due to the nature of the market mechanism involved. Under the new perspectives for Vietnamese economic and social development, it is opportune to analyze

Tadashi Kikuchi, Associate Professor, Faculty of Economics, Teikyo University; research areas/interests: business capability, development economics, Asian economy. E-mail: kikuchi@main.teikyo-u.ac.jp.

the income distribution as well as to study the changes and impacts in inequality. Recently, in the background of Vietnam's expanding economy with high annual growth rates — an 8.3% GDP growth rate from 1994 to 1998, 6.9% from 1999 to 2002 and 6.7 % from 2000 to 2012, serious claims for the redistribution of wealth and solution to the problem of disparity are emerging among the public as pressing issues. In view of such social circumstances, there are several useful political studies on Vietnamese inequality, such as Dollar, Glewwe, and Litvack (1998), and Fritzen (2002). Pham and Reilly (2007) pointed out a sharp reduction in gender pay gap disparities for the wage employed. The author, however, believes that in order to meet the demand for a more accurate analysis, intertemporal changes in inequality as well as changes in the share of population have to be examined. It is widely accepted that inequality in Vietnam exists in geographically different regions and also between urban and rural areas. The author also agrees that it is important to examine inequality at the geographical level. In another section of the paper, the author outlines how "poverty" and "inequality" are rooted in some definite human factors that influence economic status. Therefore, in the case of establishing the existence of inequality in various regional areas, it is also necessary to ascertain the relationship of inequality with geographical factors and personal characteristics such as the household owner's age and educational level, which play a significant role in establishing economic status. This paper attempts to answer the following questions — "What are the sources of inequality in Vietnam?" and "How did inequality in Vietnam changed between 1993 and 2012?" — by using micro data of 4,612 and 7,542 households obtained from the Household Expenditure Surveys conducted in 1993, and 2012 (hereafter, 1993 HES and 2012 HES, respectively) under the cooperation of the World Bank, United Nations Development Program, and the General Statistic Office of Vietnam (GSO)[1]. The second section deals with methodological issues and presents the results of the measurement and decomposition of Vietnamese inequality in 1993 and 2012. The third section estimates ordered probit model to study the relation between income gap and principle characteristics among households. The last section summarizes the findings and presents the conclusions of the paper. The author concludes that multidimensional aspects concerning the extent to which the educational level of household owners is associated with the inequality between rural and urban areas. The author wishes that these outputs of the data analysis here would be useful for researchers as well as policy makers who deal with the formulation of a policy for poverty reduction and sustainable development of Vietnam.

2. Changes in Inequality between 1993 and 2012

2.1 Statistical Analysis

In this section, by using the decomposability of inequality, the author evidences the widening of inequality in the Vietnamese economy at the household level, two decades after 1993. Analytical methods and notations for decomposability that are utilized here were employed in Tsakloglou (1993); further, in line with him, several adjustments were made to the original data before proceeding with the estimation of inequality indices. First, the income data of 4,612 and 7,542 households from the 1993 and 2012 HES[2] was used. Second, In order to calculate

[1] The Vietnam Living Standards Survey 1993 and 2012, the 1989 Population Census surveyed by United Nations Population Fund (UNFPA) are used for their sampling base on one hand. An every ten-year later, UNFPA conducted survey in 1999, 2009 and the Vietnam Household Living Standard Survey 1993, 1998, 2002, 2012. This means that there exist population dynamics between dataset 1993 and 2012. The author analyzes this point in this paper. For more detail on the data information, see also Pham and Reilly (2007).

[2] While constructing data sets for analysis in this study, 187 (3.9%) and 1,857 (19.7%) households' data from 1993 and 2012 HES, respectively, has been excluded due to missing data. In Tsakloglou (1993), in order to analyze a representative household behavior, the equivalence scales for the cost of children were used. The author here thinks over much more the change in distribution of

the inequality, the author uses the two Theil indices (T and N) and the variance of the logarithms (L), which are defined as follows[3]:

Theil index (T)

$$T = \Sigma_j (n_j m_j/nm)\ T_j + \Sigma_j (n_j m_j/nm)\ \ln(m_j/m) \quad (1\text{-}1)$$

$$T_j = (1/n_j)\ \Sigma_i (y_i/m_j)\ \ln(y_j/m_j) \quad (1\text{-}2)$$

Theil index (N)

$$N = \Sigma_j (n_j/n) N_j + \Sigma_j (n_j/n) \ln(m/m_j) \quad (2\text{-}1)$$

$$N_j = (1/n_j) \Sigma_i \ln(m_j/y_i) \quad (2\text{-}2)$$

Variance of logarithm (L)

$$L = \Sigma_j (n_j/n)\ L_j + \Sigma_j (n_j/n)(\ \ln m_j^* - \ln m^*)^2 \quad (3\text{-}1)$$

$$L_j = (1/n_j)\ \Sigma_i (\ln m_j^* - \ln y_i)^2 \quad (3\text{-}2)$$

where y_i is the income of person i ($i = 1,\ldots, n$), n_j is the population size of group j, and m and m* are the arithmetic and geometric mean income of the population, respectively. These indices are known to satisfy the axioms of symmetry, mean independence, and population independence. In addition, T_j and N_j also satisfy the transfer axiom; however, L_j does not always satisfy it[4]. After calculating T_j, N_j, and L_j and summing up each group j, we can create three "aggregate" inequality indices, i.e., T, N, and L. This also implies that T, N, and L are additively decomposable inequality indices. In any case, aggregate inequality can be expressed as a weighted sum of the same index for the different groups ("within groups" component) if the income of every person in a group is equal to the mean income of that group ("between groups" component). The first term in each equation, (1-1), (2-1), and (3-1), is the "within groups" component of inequality and the second term is the "between groups" component. Since the index of T is only weakly additively decomposable, the group income shares $n_j m_j/nm$[5] are regarded as the weights in the "within groups" component of T. In contrast, the corresponding weights in N and L are the group population shares n_j/n, and hence, they are strictly additively decomposable. However, L is decomposable around the geometric mean.

The decomposition is achieved by each of the four groups mentioned in this paper. These four groups are separated into two main categories on the basis of their characteristics: residential characteristic (region, locality) and household owner's characteristic (age, educational level) categories. The results of the measurement and decomposition of inequality are presented in Table 1.

The estimates of T, N, and L are reported for all the socioeconomic groups in both the survey years, 1993 and 2012, along with their population shares and annual arithmetic mean income. The percentages below the decomposed indices T_j, N_j, and L_j, represent the ratio of inequality of each group and the aggregate inequalities T, N, and L. The contributions of the "within groups" and "between groups" inequalities to aggregate inequality, which are calculated using the first and second terms in equations (1-1), (2-1), and (3-1), respectively, are also reported.

Vietnamese household and neither a household behavior nor the cost of children.

[3] With regard to this formulation, see Anand (1983) and Appendix 2.

[4] However, as Creedy (1977) points out, the probability of a "violating transfer" is very low for most empirical distributions.

[5] In the latter part of Table 2, we see that because of the dynamic change of "population share," the contribution of each index was over or under evaluated.

Table 1 Inequality Indices of Vietnam: 1993-2012

Characteristic of household member or household head		Population share nj/n		Group mean income mj a		Theil index T			Theil index N			Variance of Logarithms L		
		1993	2012	1993	2012	1993	2012	%	1993	2012	%	1993	2012	%
Region	Red river delta	0.244	0.157	2,549	98,573	0.193	0.303	55.6%	0.166	0.315	89.2%	0.283	0.690	144.2%
						15.7%	12.5%		14.2%	13.5%		12.5%	14.7%	
	Northeast	0.143	0.184	2,075	78,617	0.083	0.322	281.8%	0.082	0.309	269.2%	0.165	0.598	257.2%
						6.7%	13.3%		7.0%	13.2%		7.3%	12.8%	
	Northwest	0.027	0.053	2,163	54,210	0.069	0.248	264.3%	0.070	0.241	250.9%	0.141	0.460	232.1%
						5.6%	10.2%		6.0%	10.3%		6.2%	9.8%	
	Northcentral coast	0.134	0.105	1,978	66,230	0.094	0.241	158.1%	0.088	0.255	189.1%	0.167	0.546	221.6%
						7.6%	9.9%		7.5%	10.9%		7.4%	11.7%	
	Southcentral coast	0.090	0.091	2,887	87,953	0.215	0.300	40.8%	0.204	0.308	52.1%	0.413	0.650	59.0%
						17.5%	12.4%		17.4%	13.2%		18.2%	13.9%	
	Central highlands	0.020	0.047	2,952	66,339	0.162	0.275	75.2%	0.183	0.286	62.8%	0.405	0.581	49.9%
						13.2%	11.4%		15.6%	12.3%		17.9%	12.4%	
	Southeast	0.140	0.161	4,117	127,574	0.226	0.411	30.7%	0.212	0.322	28.7%	0.395	0.571	37.1%
						18.4%	17.0%		18.1%	13.8%		17.4%	12.2%	
	Mecong delta	0.201	0.202	3,205	88,428	0.186	0.321	74.0%	0.165	0.298	82.2%	0.295	0.585	99.9%
						15.2%	13.3%		14.1%	12.8%		13.1%	12.5%	
	Within Groups' component of inequlity					0.196	0.326	66.9%	0.157	0.300	91.3%	0.282	0.597	111.4%
						85.7%	93.1%		83.1%	92.5%		87.7%	93.7%	
	Between groups component of inequality					0.033	0.024	-25.4%	0.032	0.024	-23.5%	0.039	0.040	2.5%
						14.3%	6.9%		16.9%	7.5%		12.3%	6.3%	
	Within groups contribution to ineq.. Reduction							106.8%			105.5%			99.7%
	Between groups contribution to ineq.. Reduction							-6.8%			-5.5%			0.3%
Locality	Urban	0.196	0.242	4,616	132,423	0.187	0.364	56.2%	0.178	0.294	48.9%	0.337	0.525	49.8%
						56.4%	55.5%		58.1%	50.1%		59.0%	46.7%	
	Rural	0.804	0.758	2,337	76,167	0.145	0.292	94.6%	0.129	0.293	125.7%	0.235	0.599	156.3%
						43.6%	44.5%		41.9%	49.9%		41.0%	53.3%	
	Within Groups' component of inequlity					0.159	0.318	78.6%	0.138	0.293	104.6%	0.255	0.581	126.0%
						77.4%	90.6%		76.9%	90.6%		80.3%	91.2%	
	Between groups component of inequality					0.046	0.033	-34.4%	0.041	0.031	-32.0%	0.063	0.056	-13.4%
						22.6%	9.4%		23.1%	9.4%		19.7%	8.8%	
	Within groups contribution to ineq.. Reduction							114.1%			110.6%			102.7%
	Between groups contribution to ineq.. Reduction							-14.1%			-10.6%			-2.7%
Farm or nonfarm	Farm	0.700	0.437	2,223	67,397	0.126	0.315	131.7%	0.115	0.306	156.7%	0.217	0.599	174.2%
						38.0%	46.4%		37.7%	51.1%		38.1%	53.3%	
	nonfarm	0.300	0.242	4,093	107,114	0.205	0.364	52.2%	0.190	0.292	43.9%	0.352	0.525	47.0%
						62.0%	53.6%		62.3%	48.9%		61.9%	46.7%	
	Within Groups' component of inequlity					0.161	0.318	74.5%	0.138	0.293	102.8%	0.258	0.581	123.1%
						78.4%	90.6%		76.7%	90.6%		81.1%	91.2%	
	Between groups component of inequality					0.044	0.033	-28.5%	0.042	0.031	-30.1%	0.060	0.056	-8.6%
						21.6%	9.4%		23.3%	9.4%		18.9%	8.8%	
	Within groups contribution to ineq.. Reduction							110.7%			109.7%			101.7%
	Between groups contribution to ineq.. Reduction							-10.7%			-9.7%			-1.7%
Age of household head	Less than 25	0.036	0.014	2,306	52,059	0.212	0.369	74.1%	0.175	0.358	105.6%	0.300	0.660	122.9%
						14.8%	14.3%		13.8%	27.7%		13.3%	13.6%	
	25-34	0.258	0.148	2,870	74,400	0.201	0.334	65.4%	0.173	0.304	74.0%	0.296	0.553	83.3%
						14.0%	13.0%		13.6%	23.5%		13.1%	11.4%	
	35-44	0.258	0.292	3,106	90,502	0.212	0.286	3.3%	0.185	0.259	23.7%	0.323	0.483	44.1%
						14.7%	11.1%		14.6%	20.0%		14.3%	10.0%	
	45-54	0.167	0.294	2,748	99,951	0.195	0.300	33.5%	0.168	0.289	57.8%	0.293	0.578	89.8%
						13.6%	11.7%		13.3%	22.3%		13.0%	11.9%	
	55-64	0.159	0.168	2,456	102,333	0.176	0.467	170.4%	0.163	0.376	134.7%	0.306	0.688	129.3%
						12.2%	18.1%		12.9%	29.0%		13.6%	14.2%	
	65-74	0.094	0.060	2,503	65,633	0.199	0.369	86.8%	0.178	0.391	119.5%	0.325	0.836	154.7%
						13.9%	14.3%		14.0%	30.2%		14.4%	17.2%	
	More than 74	0.029	0.025	2,637	45,944	0.241	0.450	94.7%	0.226	0.500	131.4%	0.412	1.052	167.2%
						16.8%	17.5%		17.8%	38.6%		18.3%	21.7%	
	Within Groups' component of inequlity					0.224	0.338	50.7%	0.176	0.309	67.9%	0.310	0.593	88.0%
						98.2%	96.3%		97.7%	95.5%		97.6%	93.1%	
	Between groups component of inequality					0.004	0.013	214.8%	0.004	0.015	247.8%	0.008	0.044	614.3%
						1.8%	3.7%		2.3%	4.5%		2.4%	6.9%	
	Within groups contribution to ineq.. Reduction							93%			92.3%			88.1%
	Between groups contribution to ineq.. Reduction							7%			7.7%			11.9%
Educational level of household head	University graduate	—	0.582	—	104,633	—	0.300		—	0.297		—	0.388	
						—	21.5%		—	20.9%		—		
	Upper secondary edu. Completed	—	0.105	—	66,230	—	0.307		—	0.307		—	0.373	
						—	22.0%		—	21.7%		—		
	Lower secondary edu. Completed	—	0.053	—	61,793	—	0.346		—	0.335		—	0.276	
						—	24.8%		—	23.6%		—		
	Primary edu. Completed	—	0.184	—	78,617	—	0.208		—	0.246		—	0.304	
						—	14.9%		—	17.4%		—		
	Primary edu. No completed or	—	0.157	—	98,573	—	0.235		—	0.233		—	0.287	
						—	16.8%		—	16.4%		—		
	Within Groups' component of inequlity					—	0.326		—	0.297		—	0.580	
						—	93.0%		—	91.6%		—	91.1%	
	Between groups component of inequality					—	0.024		—	0.027		—	0.057	
						—	7.0%		—	8.4%		—	8.9%	
	Within groups contribution to ineq.. Reduction												85.8%	
	Between groups contribution to ineq.. Reduction												14.2%	

Source: The author has calculated the figures above based on the data from the Vietnam HES in 1993 and 2012. The decomposition of groups based on the educational level is calculated from the HES in 1998 and 2012 because of the unavailability of sufficient data in 1993. Within and Between groups contribution to inequality reduction are calculated, respectively, A/C and B/C. A: Within groups component of inequality (2012) - Within groups component of inequality (1993). B: Between groups component of inequality (2012) - Between groups component of inequality (1993), C: A + B. The currency unit in table is 1,000 VND at 1998 price in 1993, and current price in 2012.

In the top panel of Table 1, the aggregate inequality is decomposed according to the eight regions in which household residences[6] are located. However, among the different categories, only the results of group mean regional income do not provide much useful information. For example, we can observe that in 1993, there was already a wide gap in the group mean income between the richest (Southeast) and poorest regions (Northwest); the figure in 1993 the richest region was as much as 2.0 times that in the poorest region. It In 2012, this difference was even more increased to 2.35 times the mean income in the poorest region. However, apart from these increased gap between regions, as can be seen later, it was the increase in the gap within regions that caused the widening in the inequality of the entire nation for the twenty years from 1993 to 2012, by 53% (T), 71% (N), and 98% (L)[7]. These results are intuitively understandable and consistent, this is the reason for the relatively much decreases in the between groups' contribution to inequality, for instance, index T (between region) is –6.8%, N (between region) was –5.52%, as seen Tn Table 1.

The first row of the region panel also provides evidence of wide disparities in each region. For instance, in 2012, Red river delta shows an increase of inequality by 55% and Northeast constitutes by 281%, respectively, of the Theil (T) index. It is these disparities occurred within each region that are influential on the total inequality indexes. With respect to the regional inequality, the inequality share of the "between regions" component accounted for only 6.9% (T), 7.5% (N) and 6.3% (L) of all inequality in 2012. In other words, more than 90% of regional disparities are explained by the "within regions" inequality in each of the three indices. Latter, the author proved that the change of the population share among the eight regions in the twenty years period due to the migration for higher wages helped to decrease the between regions inequality.

In the next panel of Table 1, inequality is decomposed according to locality, i.e., rural and urban. It is widely accepted that there are significant disparities between urban and rural areas; however, the ratio of the populations in rural and urban areas has changed in the twenty years period from 8.04:1.96 to 7.58:2.42. During the same period, the gap in the average group income between the rural and urban areas has decreased. The mean income in the urban areas is significantly higher than that in the rural areas. In addition, the inequality in urban areas is also relatively large, for instance, index $T_{j=\text{urban}}$ was 56.4% in 1993 and 55.5% in 2012. With regard to Vietnam study, the author suggests that significantly more attention should be paid to the facts that there exist more shrinking disparities between the different locality, urban-rural, for instance, index T (between locality) was 34.4% than are disparities between regions' index T (between regions) was 25.4%. These results warns us about the bipolarization that might be occurring urban and rural areas has a different influence on nation-wide inequality by region under a market oriented economy. It should be noted that the development of regional urban and rural areas does not by itself guarantee the resolution of the disparity problem under the market oriented economy[8] Because, more importantly, as will be observed later, there is evidence suggesting that urbanization is widening inequality through the labor inflow into urban areas.

In the fourth panel of Table 1, the households are grouped into seven categories, according to the age of the household owner: (1) less than 25 years old, (2) from 25 to 34 years old, (3) from 35 to 44 years old, (4) from 45 to 54 years old, (5) from 55 to 64 years old, (6) from 65 to 74 years old, and (7) more than 74 years old. Two

[6] The author separates Vietnam into eight regions based on the manner of classification employed by the Vietnamese government. See Appendix 1 for the provinces comprising each region.

[7] We can easily calculate the national inequality index by summing up T, N, and L indices of the subgroups. For instance, the aggregated inequality indices are 0.22 (T), 018 (N), and 0.32 (L) in 1993, and 0.35 (T), 032 (N), and 0.63 (L) in 2012.

[8] With regard to the bipolarization of inequality between the urban and rural areas, the author deems the building more primary and secondary school in rural areas in Vietnam to be effective. See also Figure 3 and Kikuchi (2007).

relationships become apparent from the estimates of this panel. First, the analyses here, as many studies on inequality in other countries, assure a nearly inverse U-shaped relationship between the age of the household owner and the total household income in Vietnam[9] in 2012. Second, both the share in population and the income of households headed by young people aged below 25 years are smaller than the corresponding values of other households. Such a result suggests that there is a higher probability of the incidence of poverty among households with young heads; further, it partly explains the fact that the inequality of households headed by young persons, i.e., less than 25 years, increased by 74.1% (T_j), 105.6% (N_j), and 122.9% (L_j), respectively, in the twenty years. In this case, a majority of the disparities are evident within groups, too.

The last panel of Table 1 provides the result of the measurement and decomposition of inequality between 1993 and 2012, according to the educational levels of household owners grouped in five categories: (1) incomplete primary education, (2) complete primary education, (3) complete lower secondary education, (4) complete upper secondary education, and (5) university/graduate school level. Since the data available on the educational levels of household owners in 1993 was insufficient, the analysis used 7,542 household data of the year 2012. Considering that the household head provides the main source of income in a household, it is reasonable to assume that improving his/her educational level is closely associated with the quality and level of living standards of the rest of the household members. However, it should be reminded that the educational level and age of the household head are intrinsic characteristics and it is hard to improve a whole household's living standard immediately. On the other hand, household locations based on the eight regions or urban or rural areas are changeable anytime depending on the household head's decision where to live with an expectation of high wage. Therefore, as planning any long term policy that targets the educational levels of household owners for reducing inequality, it should be attach great importance to this point. In Table 1, the results of the author's analysis indicate a direct link between education level and inequality. First, there is a strong positive relationship between the educational level of the household owner and household income. Second, there are substantial differences in the income levels among groups. For instance, in 2012 the mean household income for households whose owners are university graduates is more than 1.69 times that of households whose owners have completed lower secondary education. Third, although there exist wide differences in the mean income among groups, the "between groups" component of inequality in 2012 is relatively small, for instance, 7.0% (T), 8.4% (N), and 8.9% (L). Before considering the elements that cause relatively large and complex "within groups" inequality, the author demonstrates the results of intertemporal changes of inequality of Vietnam between 1993 and 2012, as below.

2.2 Intertemporal Change Analysis

As already mentioned, each index — T, N, or L — between 1993 and 2012 contains a certain parameter that is influenced by the change in the population share. Therefore, for a more accurate estimation, the change in the population share between the sampled 1993 and 2012 HES data must be considered thoroughly. As observed above, the eight regional groups show an increase in inequality. However, according to our results, as we see later, relatively small 1.5% (T) of the change on the regional inequality was accounted for by the change in the population share in the market oriented economy. On the other hand, with regard to the farmer-non-farmer groups, this change increased to close to 66% (T). As already pointed out, these differences and the diversity in residential characteristics emphasize the need to take necessary steps and formulate a government policy instead of counting

[9] See Cowell (1984) for the inverse U-shaped relationship between the age of household head and the total household income or expenditure.

on the role of a certain providence to reduce inequality nationwide. In the following analysis, new notations are assumed, i.e., $v_j = n_j/n$, $k_j = m_j/m$, and $k_j^* = m_j^*/m^*$, in order to derive (1-1), (2-1), and (3-1).[10]

Theil index (T)

$$\Delta T = \Sigma_j v_j k_j \Delta T_j + \Sigma_j k_j (T_j + \ln k_j)\Delta v_j - \Sigma_j v_j (T_j + \ln k_j + 1)(\Sigma_j k_j \Delta v_j) + \Sigma_j v_j k_j (T_j + \ln k_j + 1)(\Delta \ln m_j - \Sigma_j v_j k_j \Delta \ln m_j) \quad (1\text{-}3)$$

Theil index (N)

$$\Delta N = \Sigma_j v_j \Delta N_j + \Sigma_j N_j \Delta v_j + \Sigma_j (k_j - \ln k_j)\Delta v_j + \Sigma_j v_j (k_j - 1)\Delta \ln m_j \quad (2\text{-}3)$$

Variance of logarithm (L)

$$\Delta L = \Sigma_j v_j \Delta L_j + \Sigma_j L_j \Delta v_j + \Sigma_j (\ln k_j^*)^2 \Delta v_j + \Sigma_j 2 v_j \ln k_j^* \Delta \ln m_j^* \quad (3\text{-}3)$$

where Δ represents the change in the variable form period t (1993) to period $t + 1$ (2012). Equations (2-3) and (3-3) decompose the change in inequality into four terms that can be interpreted as the effect of intertemporal changes in the "within groups" inequality ($\Sigma_j v_j \Delta N_j$, $\Sigma_j v_j \Delta L_j$), the effect of changes in population shares on the "within groups" component of inequality ($\Sigma_j N_j \Delta v_j$, $\Sigma_j L_j \Delta v_j$), the effect of changes in population shares on the relative mean income of the population groups ($\Sigma_j (k_j - \ln k_j) \Delta v_j$, $\Sigma_j (\ln k_j^*)^2 \Delta v_j$), and the effect of changes in the relative mean income of the population groups ($\Sigma_j v_j (k_j - 1) \Delta \ln m_j$, $\Sigma_j 2 v_j \ln k_j^* \Delta \ln m_j^*$). Evidently, the overall effect of demographic changes is given by the sum of the second and the third terms. The decomposition of ΔT can not be interpreted in a similar way because T is weakly additively decomposable.

Table 2 Decomposition of the Changes in Aggregate Inequality: 1993-2012

Characteristic of household member or household head	Contribution to changes in inequality due to changes in					
	Index of Inequality	Within groups inequality (a)	(b)	Population shares	(c)	Mean group income (d)
Region	Theil's T	133.1		1.6		-24.2%
		1.2		1.5%		-21.9%
	Theil's N	141.2	-2.1		2.9	
		116%				
	Variance of logarithms L	325.3	-0.4		2.7	-20.0
		106%		0.7%		-6.5%
Locality	Theil's T	138.3		10.4		14.2
		84.9%		6.4%		8.7%
	Theil's N	149.8	2.8		6.0	11.0
		88%		5.2%		6.5%
	Variance of logarithms L	327.1	4.9		10.1	15.1
		92%		4.2%		4.2%
Farm or nonfarm	Theil's T	154.7		327.5		8.1
		31.6%		66.8%		1.7%
	Theil's N	157.1	-43.4		-333.4	6.5
		-74%		177%		-3.1%
	Variance of logarithms L	316.1	-78.7		-15.5	4.1
		140%		-42%		1.8%
Age of household head	Theil's T	116.8		-53.6		-0.4
		186.1%		-85.4%		-0.7%
	Theil's N	133.3	12.1		57.4	-0.6
		66%		34%		-0.3%
	Variance of logarithms L	290.7	19.1		-0.2	-2.2
		95%		6.17%		-0.7%

Source: The author has calculated the figures above based on the Vietnam HES in 1993, 1998, and 2012. The percentages show the weight of total changes. Note that the values in the table are multiplied by 1,000 for see clearly.

[10] See Appendix 2 for more detail and the expanded forms of equations (1-3), (2-3), and (3-3).

The results of the decomposition of the changes in aggregate inequality, according to (1-3), (2-3), and (3-3) are shown in Table 2 (Mookhjee & Shorrocks, 1982). For the sake of clarity, we will mention the original values that were multiplied by 1,000. The second line of each row reports the percentage contribution of each component to the observed changes in aggregate inequality. If a particular component contributes to an increase in inequality, its sign will be positive; if not, the sign of its percentage contribution to inequality reduction is negative. We can make three findings from the results of Table 2. First, there are crucial differences among groups, characteristics of household member, in 1993 and 2012, before and after excluding the population changes. For instance, in the second raw of the locality groups in rural-urban level, if there were no changes in the population share during twenty years, ceteris paribus, the inequality among rural-urban groups would have decreased much more, i.e., by 6.4% (T), 5.2% (N), and 4.2% (L)[11]. In other words, it shows that the inflow of labor into the urban areas widened the disparity between the urban and rural areas. Second, as we can see in Table 1, the share of the population on non-farm increased substantially from 0.30 in 1993 to 0.56 in 2012 by 87%. It proves that more Vietnamese people stopped farmer and moved to other occupations in these twenty years. Based on these back ground of job changes on market economy, Table 2 reveals that the change in the population share of households contributed to a larger extent to increasing the inequality — 66% (T), 177% (N)[12] — than that occurred by the change in the population share of households in locality, urban-rural group. This appears natural because of the much closer and stronger relation between human economic activities and job selection, than resident place decision. It should be considered, from the view point of preventing from widening national inequality, that the smooth job-changing programs from agriculture to industrialization on a long term basis and the fact that the Vietnamese labor market is unstable[13]. Although the influence of the change in the population share of the household owner's age on an inequality is unclear, these findings add new dimensions that facilitate a better understanding of the dynamic structural changes in inequality that have occurred from 1993 to 2012. In this regard, we can neither overestimate the effect of market mechanisms nor ignore the impact of migration on increasing inequality from local into urban areas. On the other hand, we need to broadly study multiple aspects of household characteristics. Any developing strategies in absence of these multiple observations can adversely and unexpectedly affect the Vietnamese economy[14] in a manner that would prevent the decrease in inequality in society and hinder the sustainable development.

3. Where the Rich and Poor Household Exists

In this section, to analyze the influence of residential and household owner's characteristics on household income, the author uses here ordered probit model estimation. The 7,542 household data obtained from the 2012 HES are used.

At the onset, let y^* denote the latent variable indicating the critical borders of the household income distribution in 2012, which is clearly not visible here. Further, let y denote this distribution's ordered group from

[11] These changes are easily calculated by firstly summing up the contributions of changes in the population share, Table 2, and secondly, dividing it 1,000.

[12] Contribution of variance of logarithm (L) is negative and not clear in interpretation. This is because, this index is not exactly separated into three part. See also equation (3-3).

[13] For instance, according to the data of *Key Indicators* of the Asian development Bank, the unemployment rate of Vietnam fluctuated with values of 4.5% in 1998, 2.3% in 2002, and 2.1% in 2015.

[14] Kikuchi (2007) used the Vietnam Living Standards Survey 1998 for quantitative analysis and showed that the existing primary school structure in rural areas is effective in reducing inequality, when measured by the Gini coefficient.

the poorest to the richest, which assumes the following values: 0 for the first group (the poorest 10% people in rural area); 1 for the second (the richest 10% people in rural area); 3 for the fourth (the poorest 10% people in urban area); 4 for the fifth (the richest 10% people in urban area); and 2 for the other group who does not belong to the four groups above mentioned.

We will assume that the latent regression, ordered probit model, has the following form below.

$$y_i^* = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \ldots\ldots + X_{i17}\beta_{17} + \varepsilon_i.$$

where an independent variable X_{i1} is the logarithm of the household *i's* income[15]. And X_{i2},.., X_{i17} are dummy variables that take one if household *i*'s regional residence, urban or rural area, and household characteristics on age and educational level is truly correspondent to the dummies categories. Finally, the term ε_i is assumed to indicate a normal distribution. In short, we can rewrite the latent regression as $y_i^* = X_{ij}\beta_j + \varepsilon_i$ (i = 1,.. 7,542, j = 0, ..,17). Note that we denote constant as $X_{i0}\beta_0 = \beta_0$ in the equation. We can also define α_1, α_2, and α_3 as the three cutoff points. For the estimation, we have used the method of maximum likelihood estimation[16].

In accordance with the general notation, our model can be expressed as follows (McCullagh, 1980; Hosmer & Lemeshow, 1989). Φ denotes the distribution function:

$$\begin{cases} P(y_i = 0|X_{ij}) = \Phi(-X_{ij}\beta_j) \\ P(y_i = 1|X_{ij}) = \Phi(\alpha_1 - X_{ij}\beta_j) - \Phi(-X_{ij}\beta_j) \\ P(y_i = 2|X_{ij}) = \Phi(\alpha_2 - X_{ij}\beta_j) - \Phi(\alpha_1 - X_{ij}\beta_j) \\ P(y_i = 3|X_{ij}) = \Phi(\alpha_3 - X_{ij}\beta_j) - \Phi(\alpha_2 - X_{ij}\beta_j) \\ P(y_i = 4|X_{ij}) = 1 - \Phi(\alpha_3 - X_{ij}\beta_j) \quad (i = 0, 1, 2, 3, 4) \end{cases}$$

From the results analyzed in the paper, we already know that household's regional residence, urban or rural area as well as household characteristics of age and education level has each influence on the Vietnamese national inequality. The result of regression shows that most these dependent variables are statistically significant and also describes well observed ratio of grouped data (y = 0, 1, 2, 3, 4) from the fitted valuable. The author shows the results in Table 3 for simplicity.

Among 7,542 household data in 2012 HES, there are 182 observations in the richest 10% and the poorest 10% people in urban area[17]. There are, in the same way, 572 observations in the richest10% and the poorest10% people in rural area. According to the results of ordered probit analysis, the exactly estimated household number of the poorest10% people in urban and rural area is, respectively, 106 observations and 319. This goodness of fit in Table 4 tells that we can choose and target, especially, the poor who need donor's support among the huge population[18].

[15] It is well known that income distribution is well described by log normal distribution. The author mentioned it on the 1998 HES in Kikuchi (2007).

[16] The program used TSP and the program can be obtained from the author, if necessary.

[17] As we see in Table 1, the ratio of the populations in rural and urban areas in 2012 is 7.58: 2.42. The household data 7,542 is composed of households in rural, 5,718 observation, and in urban 1,824, respectively.

[18] Base on the result here in Table 4, the poorest 10% in rural could be selected appropriately with 55.8 % (= B/A) and 58 % in urban.

Table 3 Output of Estimation: Ordered Probit Model

Independent variables	Estimated parameters	Independent variables	Estimated parameters
Income of household (*i*)	0.32310*** (0.019154)	Intercept	–2.73274*** (0.23456)
Regional dummy (True =1, False = 0)		Age group dummy (True =1, False = 0)	
Northeast	0.11559*** (0.05168)	25 ≦ age < 34	0.60064*** (0.12647)
Northwest	0.06054 (0.07721)	35 ≦ age < 44	0.57248*** (0.12423)
North central coast	0.0660 (0.06001)	45 ≦ age < 54	0.47612*** (0.12429)
South central coast	0.20254*** (0.06295)	55 ≦ age < 64	0.41024*** (0.12665)
Central highlands	0.11956 (0.07932)	65 ≦ age < 74	0.37774** (0.13507)
Southeast	0.16108** (0.054008)	75 ≦ age	0.04906 (0.15164)
Mekong Delta	0.05348 (0.05226)		
Education level (True =1, False = 0)		Threshold parameters	
Primary education completed	017202 (0.14038)	α_1	0.45673*** (0.01826)
Lower secondary education completed	0.01736 (0.09109)	α_2	3.20521*** (0.03346)
Upper secondary education Completed	0.01247 (0.06376)	α_3	3.55295*** (0.04004)
University graduate	0.10642* (0.06358)	Number of observations = 7,542 Pseudo R-squared = 0.09048 Log likelihood = – 5,377	

Source: The Household Expenditure Surveys 2012. The author estimated these results using the TSP software. The pseudo R-squared introduced in Cragg and Uhler (1970) has been calculated by the authors. ***, ** and * imply that the parameter is significant at the 1%, 5% and 10% critical levels, respectively. The standard error is mentioned within parentheses. Note that the independent variable of income of household is taken logarithm. Threshold parameter α_i imply: y = 0, if y* < 0; y = 1, if 0 < y* ≦ α_1; y = 2, if α1 < y* ≦ α_2; y = 3, if α_2 < y* ≦ α_3; y = 4, if α_3 < y*. See also Table 4.

Table 4 Goodness of Fit: Ratio of the Group

	Observed fraction (A)	Estimated fraction (B)	B/A
The poorest 10% of rural areas: y = 0	0.0758	0.0423	0.5580
The richest 10% of rural areas: y = 1	0.0758	0.0472	0.6227
Other groups: y = 2	0.8000	0.8800	1.1000
The poorest 10% of urban areas: y = 3	0.0242	0.0141	0.5806
The richest 10% of urban areas: y = 4	0.0242	0.0164	0.6793

Note: The share of population in urban and rural area is 24.2% and 75.8% in 2012 HES.

Source: The Household Expenditure Surveys 2012. See also Table 3.

3. Conclusions

This paper examined the inequality in the distribution of income in Vietnam for twenty years 1993 and 2012. The decomposition analysis yielded several findings. First, the aggregate inequality increased between 1993 and 2012, and the inequality within groups dominated the aggregate inequality more than that between groups. Second, diversities and specific characteristics exist within each group. One of the findings establishes that urban-rural disparities would have decreased further if there was no change in the population shares between the urban and rural areas in 1993-2012. This result has also consistency with the recent trend that a migration into large cities is widening the rich and poor gap in the urban area (Ravallion, 2002; Ravallion et al., 2007).

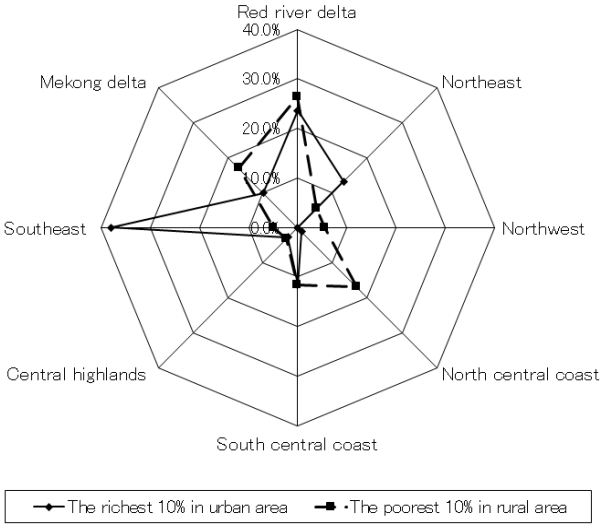
These findings may appear complex when viewed in the light of formulating a policy on inequality reduction in Vietnam in the future. It is imperative to address this issue as after thirty years of “Doi moi” reforms aimed toward a market oriented economy, bipolarization of the rich and the poor between the urban and rural areas is

currently becoming a concerning issue with lots of discussion. Considering these various aspects, the author suggests formulating a policy; further, targeting groups should be assigned high priority on the policy agenda for inequality reduction. For this purpose, it is useful to pay attention to and study households' characteristics, such as the ages and educational levels of the household heads who could be promising players of economic activities and also making a breakthrough towards sustainable market economy[19].

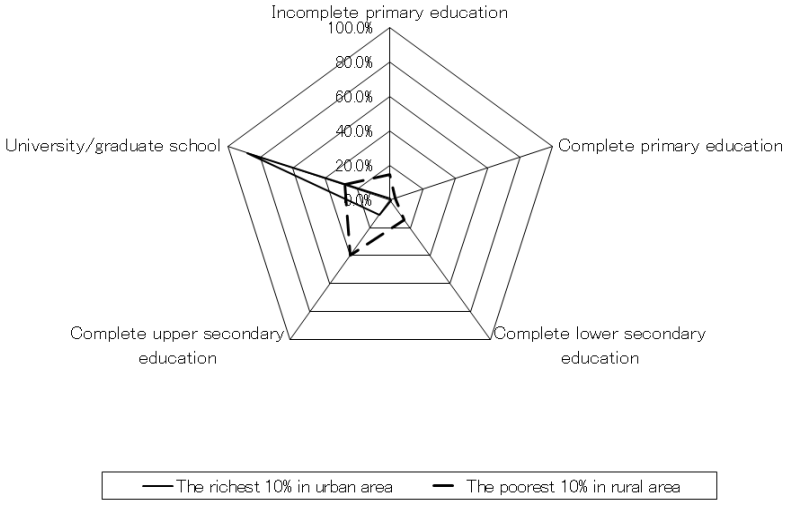
The author also shows in the second section that these four characteristics, residential characteristic (region, locality) and household owner's characteristic (age, educational level) can well describe who are poor and rich in a whole distribution.

Here, the author briefly presents the contrast in characteristics between the poorest 10% group in rural areas and the richest 10% group in urban areas below.

By regional residence



By household owner's educational level



[19] World economy, according with "Global Economic Prospects: Broad Based Upturn, but for How Long?, by the World Bank research, in 2018 is running favorably for many developing countries and will earn 3.1% economic growth. Also for the pioneering empirical analysis on Vietnamese family size, see Knodel, Friedman, Truong, and Bui (2000).

Figure 1 Multiple Characteristics of the Richest and the Poorest 10% of Urban-Rural Areas: 2012
Source: The Household Expenditure Surveys 2012. See also Tables 1 and 4.

These figures projects rich households in the urban areas as those with household heads who are around 50 (between 45 and 54) years of age, and who have completed at least more than lower secondary school, and live in the Southeast region. On the other hand, the image of poor households in the rural areas comprises household heads who are less than 45 years of age, have not completed upper secondary school at most, and live in local regions such like the Red river delta, Mekong delta or North central coast region. These findings are also consistent with author's stance in this paper: that a mere regional development program is not enough[20] for sustainable economic development. The process to sustainable economic development should be formed in a well-balanced distribution of living standard, it is necessary to seek out relevant characteristics of households to reduce inequality in the light of the dynamic change of population and income distributions of developing countries. According to the author, the application of the decomposition methods here in the analysis of Vietnam is not only useful in determining the causes of inequality and compare from multiple perspectives, but also helpful when we target specific groups for inequality reduction effectively. At last but not least, author would like to express, without considering this issue dynamically from multiple viewpoints, any political program that is advocated may turn out to be unrealistic and fruitless.

Acknowledgements

The author appreciates the thoughtful advice and useful comments to the former naive version paper provided by Professor Emeritus Kazuhuro Takanashi (Keio University), Professor Emeritus Toshio Ogata (Chuo University), Professor Tran Dinh Lam (Hochiminh University of Socio Sciences and Humanities) as well as with the audience at the International Conference on Business and Economics, held in Hochiminh City, Vietnam on November, 2017. The author would also like to express gratitude to valuable knowledge and excellent skill of

[20] For a more accurate estimation, a dynamic economic model of Vietnam and the specification of household economic activities have to be analyzed. The author only notes them here as issues deserve of studying in near future.

econometrics provided by Professor Emeritus Chiohiko Minotani (Keio University).

References

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

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

Creedy J. (1977). "The principle of transfers and the variance of the logarithms", *Oxford Bulletin of Economics and Statistics*, Vol. 39, pp. 152-158.

Deaton A. (1980). "Measurement of welfare: Theory and practical guidelines", World Bank Living Standards Measurement Study Working Paper No. 7, World Bank.

Deaton A. (1997). *The Analysis of Household Surveys*, The Johns Hopkins University Press.

Dikhanov Y. and Ward M. (2001). "Evolution of the global distribution of income in 1970-1999", in: The Fourth Meeting of the Expert Group on Poverty Statistics, Rio de Janeiro, Brazil.

Dollar D., Glewwe P. and Litvack J. (1998). *Household Welfare and Vietnam's Transition*. World Bank.

Fritzen S. (2002). "Growth, inequality and the future of poverty reduction in Vietnam", *Journal of Asian Economics*, Vol. 13, pp. 635-657.

Gibson J., Le T. and Kim B. (2017). "Prices, Engel curves, and time-space deflation: impacts on poverty and inequality in Vietnam", *The World Bank Economic Review*, Vol. 31, No. 2, pp. 504-530.

Knodel J., Friedman J., Truong S. A. and Cuong, T. B. (2000). "Intergenerational exchanges in Vietnam: Family size, sex composition, and the location of children", *Population Studies*, Vol. 54, pp. 89-104.

Kikuchi T. (2007). "An analysis of the impacts of development on Gini inequality using grouped and individual observations: Examples from the 1998 Vietnamese household expenditure data", *Journal of Asian Economics*, Vol. 18, pp. 537-552.

Sen A. K. (1976). "Real national income", *Review of Economic Studies*, Vol. 43, pp. 19-39.

Mookherjee D. and Shorrocks A. F. (1982). "A decomposition of the trend in U.K. income inequality", *Economic Journal*, Vol. 92, pp. 886-902.

Pham T. H. and Barry Reilly (2007). "The gender pay gap in Vietnam, 1993–2002: A quantile regression approach", *Journal of Asian Economics*, Vol. 18, pp. 775-808.

Ravallion M. (2002). "On the urbanization of poverty", *Journal of Development Economics*, Vol. 82, No. 1, pp. 1-42.

Ravallion M., S. Chen and P. Sangraula (2007). "New evidence on the urbanization of global poverty", in: *The PEGNet's Recent Conference*, Berlin, Germany.

Tsakloglou P. (1993). "Aspects of inequality in Greece measurement, decomposition and intertemporal change: 1974, 1982", *Journal of Development Economics*, Vol. 40, pp. 53-74.

Appendix 1

List of Provinces in Region Groups

1	Red River Delta	Ha Noi	Vinh Phuc	Bak Ninh	Ha Tay	Hai Duong	Hai Phong	Hung Yen
		Thai Binh	Ha Nam	Nam Dinh	Ninh Binh			
2	Northeast	Ha Giang	Cao Bang	Bac Can	Tuyen Quang	Lao Cai	Yen Bai	Thai Nguyen
		Lang Son	Quang Ninh	Bac Giang	Phu Tho			
3	Northwest	Lai Chau	Son La	Hoa Binh				
4	North Central Coast	Thanh Hoa	Nghe An	Ha Tinh	Quang Binh	Quang Tri	Thua Thien Hue	
5	South Central Coast	Da Nang	Quang Nam	Quang Ngai	Binh Dinh	Phu Yen	Khanh Hoa	
6	Central Highlands	Kontum	Gia Lai	Dak Lak	Lam Dong			
7	Southeast	T.P. Ho Chi Minh	Ninh Thuan	Binh Thuan	Binh Phuoc	Tay Ninh	Binh Duong	Dong Nai
		Ba Ria Vung Tau						
8	Mekong Delta	Long An	Tien Giang	Ben Tre	Tra Vinh	Vinh Long	Dong Thap	An Giang
		Kien Giang	Can Tho	Soc Trang	Bac Lieu	Ca Mau		

Source: Statistical Yearbook 2004, General Statistic Office of Vietnam, Hanoi.

Appendix 2

Some supplementary information and expansion of equations (1-3), (2-3), and (3-3) are provided below. The author also referred to Anand (1983) and Tsakloglou (1993). Here, *i* denotes a household belonging to a group j. Then, the Theil index (T) of equation (1-2) is written as

$$T = \Sigma_j\Sigma_i (y_{ji}/Y) \ln \frac{y_{ji}/Y}{n_{ji}/n}$$

$$Y = \Sigma_j\Sigma_i\, y_{ji} = \Sigma_j\, Y_j$$

$$n = \Sigma_j\Sigma_i\, n_{ji} = \Sigma_j\, n_j$$

The logarithm of the RHS is written as

$$T = \Sigma_j\, Y_j/Y \Sigma_i\, y_{ji}/Y_j \left[\ln \frac{y_{ji}/Y_j}{n_{ji}/n_j} + \ln \frac{Y_j/n_j}{Y/n} \right] \quad (\Sigma_i\, y_{ji}/Y_j = 1 \text{ for each j})$$

$$= \Sigma_j\, Y_j/Y \left[\sum_i \frac{y_{ji}}{Y_j} \ln \frac{y_{ji}/Y_j}{n_{ji}/n_j} \right] + \Sigma_j\, Y_j/Y \ln \frac{y_j/Y}{n_j/n}$$

$$= \Sigma_j\, Y_j/Y[T_j] + \Sigma_j\, Y_j/Y \ln \frac{y_j/Y}{n_j/n}$$

$$= T_W + T_B$$

This corresponds to equation (1－1). Now, replacing $v_j = n_j/n$ and $k_j = m_j/m$ and taking the first difference of the equation above, we obtain

$$\Delta T = \Sigma_j\, v_j k_j \Delta T_j + \Sigma_j\, k_j (T_j + \ln k_j)\, \Delta v_j + \Sigma_j v_j (T_j + \ln k_j + 1)\, \Delta k_j$$

$$(\Delta\, n_j m_j/nm = v_j k_j,\ y_j/Y = n_j m_j/nm)$$

Expanding the last term on the RHS as

$$\Sigma_j v_j (T_j + \ln k_j + 1)\Delta k_j = \Sigma_j v_j (T_j + \ln k_j + 1) k_j \Delta \ln (m_j/m)$$

$$= \Sigma_j q_j \Delta \ln m_j - \Sigma_j q_j \Delta \ln (\Sigma_j\, m_j v_j) \qquad (\Delta \Sigma_j m_j v_j = m)$$

rewrite, here, $q_j = v_j k_j (T_j + \ln k_j + 1)$

$$= \Sigma_j q_j \Delta \ln m_j - \Sigma_j q_j\, (\Sigma_j (m_j/m) \Delta v_j + \Sigma_j (v_j/m)\, \Delta m_j)$$

$$(\Delta m = \Sigma_j\, m_j v_j)$$

$= \Sigma_j q_j \Delta \ln m_j - \Sigma_j q_j(\Sigma_j(k_j \Delta v_j) - \Sigma_j q_j(\Sigma_j v_j k_j \Delta \ln m_j)$

$(\Delta k_j = m_j / m)$

Now, equation (1-3) is derived.

$\Delta T = \Sigma_j v_j k_j \Delta T_j + \Sigma_j k_j (T_j + \ln k_j)\Delta v_j + \Sigma_j q_j \Delta \ln m_j - \Sigma_j q_j(\Sigma_j(k_j \Delta v_j) - \Sigma_j q_j(\Sigma_j v_j k_j \Delta \ln m_j)$

$= \Sigma_j v_j k_j \Delta T_j + \Sigma_j k_j (T_j + \ln k_j)\Delta v_j - \Sigma_j v_j k_j (T_j + \ln k_j + 1) + \Sigma_j v_j k_j (T_j + \ln k_j + 1)(\Delta \ln m_j - \Sigma_j v_j k_j \Delta \ln m_j)$

The Theil index (N), in the same way, is written as

$$N = \Sigma_j \Sigma_i (n_{ji}/n) \ln \frac{n_{ji}/n}{y_{ji}/Y}$$

$$= \Sigma_j n_j/n \Sigma_i n_{ji}/n_j \left[\ln \frac{n_{ji}/n_j}{y_{ji}/Y_j} + \ln \frac{n_j/n}{Y_j/Y} \right] \quad (\Delta \Sigma_i n_{ji}/n_j = 1 \text{ for each } j)$$

$$= \Sigma_j n_j/n \left[\sum_i \frac{n_{ji}}{n_j} \ln \frac{n_{ji}/n_j}{y_{ji}/Y_j} \right] + \Sigma_j n_j/n \ln \frac{n_j/n}{Y_j/Y}$$

$$= \Sigma_j n_j/n \, [N_j] + \Sigma_j n_j/n \ln \frac{n_j/n}{Y_j/Y}$$

$= N_W + N_B$

This corresponds to equation (2-1). Now, replacing $v_j = n_j/n$ and $k_j = m_j/m$ and taking the first difference of the equation above, we get

$\Delta N = \Sigma_j v_j \Delta N_j + \Sigma_j N_j \Delta v_j - \Sigma_j \ln k_j \Delta v_j - \Sigma_j v_j \Delta \ln k_j$

$(\Delta y_j/Y = n_j m_j/nm)$

Expanding the last term on the RHS as

$-\Sigma_j v_j \Delta \ln k_j = \Sigma_j v_j \Delta \ln(m/m_j)$

$= \Sigma_j v_j \Delta \ln(\Sigma_j m_j v_j) - \Sigma_j v_j \Delta \ln m_j \quad (\Delta \Sigma_j m_j v_j = m)$

$= \sum_j v_j \left[\sum_j \left(m_j / \sum_j m_j v_j \right) \Delta v_j + \sum_j \left(v_j / \sum_j m_j v_j \right) \Delta m_j \right]$

$- \Sigma_j v_j \Delta \ln m_j$

$= \Sigma_j v_j \Sigma_j (m_j/m) \Delta v_j + \Sigma_j \Sigma_j (v_j/m) m_j \Delta \ln m_j - \Sigma_j v_j \Delta \ln m_j$

$(\Delta m = \Sigma_j m_j v_j)$

$= \Sigma_j v_j \Sigma_j k_j \Delta v_j + \Sigma_j \Sigma_j v_j k_j \Delta \ln m_j - \Sigma_j v_j \Delta \ln m_j$

$= \Sigma_j k_j \Delta v_j + \Sigma_j k_j \Delta \ln m_j - \Sigma_j v_j \Delta \ln m_j \quad (\Delta \Sigma_j v_j = 1)$

Now, equation (2-3) is derived.

$\Delta N = \Sigma_j v_j \Delta N_j + \Sigma_j N_j \Delta v_j - \Sigma_j \ln k_j \Delta v_j + \Sigma_j k_j \Delta v_j + \Sigma_j k_j \Delta \ln m_j - \Sigma_j v_j \Delta \ln m_j$

$= \Sigma_j v_j \Delta N_j + \Sigma_j N_j \Delta v_j + \Sigma_j (k_j - \ln k_j) \Delta v_j + \Sigma_j v_j (k_j - 1) \Delta \ln m_j$

Finally, we can also expand the variance of logarithm (L) in the same manner. First, the definition on the distribution of a household i's income (expenditure) among a group j is given as $x_{ji} = \ln y_j$ (same for all i). We assume that every member of group j has the same distribution, as stated above. Then, L is calculated as a total variance of x_{ji} as

$V = (1/n)\Sigma_j \Sigma_i n_{ji} (\ln(m^*) - \ln(y_{ji}))^2$

$= (1/n)\Sigma_j \Sigma_i n_{ji} (x_{ji} - x..))^2$

$x.. = (1/n)\Sigma_j \Sigma_i n_{ji} x_{ji}$

$x_j. = \Sigma_i n_{ji} x_{ji} / \Sigma_i n_{ji}$

$= (1/n)\Sigma_j \Sigma_i n_{ji} [(x_{ji} - x_j.) + (x_i - x..)]^2$

$= (1/n)\Sigma_j \Sigma_i [n_{ji}(x_{ji} - x_j)^2 + n_{ji}(x_j - x..)^2 + 2n_{ji}(x_{ji} - x_j.)(x_j - x..)]^2$

$= \Sigma_j (n_j/n)[\Sigma_i n_{ji}/n_j (x_{ji} - x_j.)^2] + \Sigma_j (n_j/n)(x_j - x..)^2$

$(\Delta x_j. = \Sigma_i n_{ji} x_{ji} / \Sigma_i n_{ji}, \text{ then } \Sigma_i n_{ji}(x_{ji} - x_j) = 0)$

$= \Sigma_j(n_j/n)[V_j]+\Sigma_j(n_j/n)(x_j - x..)^2$
$= V_W + V_B$

Note the next relation on $x_j.$ and $x..$

$x_j. = \Sigma_i n_{ji}\, x_{ji}/\Sigma_i n_{ji} = \Sigma_i n_{ji} lny_j/\Sigma_i n_{ji} = lnm_j^*$
$x.. = (1/n)\Sigma_j\Sigma_i n_{ji} x_{ji} = (1/n)\Sigma_j\Sigma_i n_{ji} lny_j = lnm^*$

$V = \Sigma_j(n_j/n)L_j + \Sigma_j(n_j/n)(lnm_j^* - lnm^*)^2$
This corresponds to equation (3-1). Now, replacing $\nu_j = n_j/n$ and $k_{j*} = m_j^*/m^*$ and taking the first difference of the equation above, we obtain
$\Delta V = \Sigma_j \nu_j \Delta L_j + \Sigma_j L_j \Delta \nu_j + \Sigma_j (lnk_j^*)^2 \Delta \nu_j + \Sigma_j 2\nu_j lnk_j^* \Delta lnk_j^*$

Note the next equation. Expanding the last term on the RHS, we get
$\Sigma_j \nu_j lnk_j^* = \Sigma_j \nu_j (lnm_j^* - lnm^*) = \Sigma_j \nu_j lnm_j^* - lnm^* \Sigma_j \nu_j = lnm_j^* - lnm^* = 0$
$\Sigma_j 2\nu_j lnk_j^* \Delta lnk_j^* = \Sigma_j 2\nu_j lnk_j^* \Delta(lnm_j^* - lnm^*)$
$= \Sigma_j 2\nu_j lnk_j^* \Delta lnm_j^* - \Sigma_j 2\nu_j lnk_j^* lnm^*$
$= \Sigma_j 2\nu_j lnk_j^* \Delta lnm_j^*$
Now, equation (2-3) is expressed as
$\Delta L = \Sigma_j \nu_j \Delta L_j + \Sigma_j L_j \Delta \nu_j + \Sigma_j (lnk_j^*)^2 \Delta \nu_j + \Sigma_j 2\nu_j lnk_j^* \Delta lnm_j^*$