WID.world WORKING PAPER N° 2018/13

Top Incomes in Korea: Update, 1933-2016

Nak-Nyeon Kim

November 2018

World Inequality Lab

Top Incomes in Korea: Update, 1933–2016[*]

Nak-Nyeon Kim[**1]

Abstract

This study updates data on Korea's top income shares (Kim and Kim 2014) in the World Inequality Database up to 2016. The national account statistics were revised in accordance with the System of National Accounts (SNA) in 2008, and income tax data have become more substantial, including information on tax exemptions. A new interpolation method for income tax data in tabular form called Generalized Pareto Curves was proposed. Top income shares are updated to reflect these changes in data and method. As a result of the update, the sharp rise in top income shares has been somewhat alleviated since the mid-1990s. The inequality of earned income has been showing constant improvement since 2010 because the income of workers in the bottom 50% has increased more rapidly than that of workers in the top 10%. On the other hand, the concentration of unearned income, consisting mainly of business income and financial income, has become weaker, but continues to grow. Accordingly, the concentration of total income for both tended to decrease or stagnate in the first half of the 2010s, but recently began to rise again, as the increase in the concentration of unearned income has been faster than the decline in the concentration of earned income.

Key Word: top income shares, income tax statistics, interpolation, inequality
JEL Classification: J3, N3

Ⅰ. Introduction

Top income shares indicate the percentages of total income taken by those in the top 0.1%, 1%, and 10% of the adult population in terms of income.[2] Estimates for 33 countries are provided by the World Inequality Database (WID) for global comparison. As for Korea, there have been estimates of top income shares limited to earned income (Kim 2012b) and top income shares in terms of total income that have added business or financial income to earned income (Kim 2012a), the results of which (Kim and Kim 2014) are provided in the WID's Korean data. However, the estimates cover only the period up to 2010 and 2012, respectively. This study provides the most recent top income shares by updating the data up to 2016.

Changes have occurred in the data in the meantime. Top income shares are obtained by dividing the incomes of the top income earners (numerator) by total income (denominator); the former can be obtained from income tax data and the latter by calculating incomes imputed to households in the national account. However, changes in the series have occurred due to the transition of the national income statistics from the 1993 SNA to the 2008 SNA, which must be applied to the most recent figures.

Income tax data, which previously included information only about taxpayers, have started

---

[*] This is an English version of the original work in Korean: "Han'guk ŭi sodŭk chipchungdo: Update, 1933-2016", *Han'gukkyŏngje p'orŏm*[*The Korean Economic Forum*], 11(1), Spring 2018, pp. 1-32.

[**1] Professor, Department of Economics, Dongguk University; nnkim@dongguk.edu

[2] Top income shares in this study are based on the pre-tax income of individuals. Income distribution indicators such as the Gini coefficient, calculated based on the Household Income and Expenditure Survey, are applied not to individuals but to households. In that case, the equivalized income calculated by dividing household income by the square root of household size is used, so that the welfare level can be compared among households with different household sizes. The income distribution indicator is obtained assuming that household members without income share this household income. Here, the indicator is obtained based on two types of income: market income and disposable income, which is a combination of market income and net public transfer income. Therefore, this difference must be considered when comparing the income distribution indicators of this study to the Household Income and Expenditure Survey.

including information about non-taxpayers and daily labor income earners since 2009. As a result, for earned income, income tax data show not only the incomes of the top income earners, but also the total income. While the numerator had been based on income tax data and the denominator on national accounts in calculating of top income shares, both the numerator and denominator have been based on income tax data since 2009, which increases the consistency of estimation. It is now possible to provide the total income distribution, including the middle and bottom in addition to the top income shares.

The estimation method has been improved as well. Income tax data are presented in tabular form, showing the number of persons and amounts of income in each income bracket instead of raw data. This shows, for example, to which income bracket someone in the top 1% belongs, but it is difficult to specify the exact position within the bracket. Interpolation methods such as Pareto interpolation and mean split histograms are applied to obtain the top income shares of the top x%. However, the WID has recently suggested an improved interpolation method called the "Generalized Pareto Curve" (GPC), which is also applied in this paper.

The structure of the rest of this paper is as follows. Section II provides the estimates of top income shares by applying the aforementioned data changes and methods in terms of earned income and shows how much they have changed. It also provides the total income distribution including not only the top but also the middle and bottom income earners after 2009. Section III presents the estimates of top income shares in terms of total income, including business and financial income in addition to earned income. Section IV summarizes the new findings of this study and outlines the remaining challenges that must be overcome.

II. Concentration of earned income

1. Revision of earned income statistics

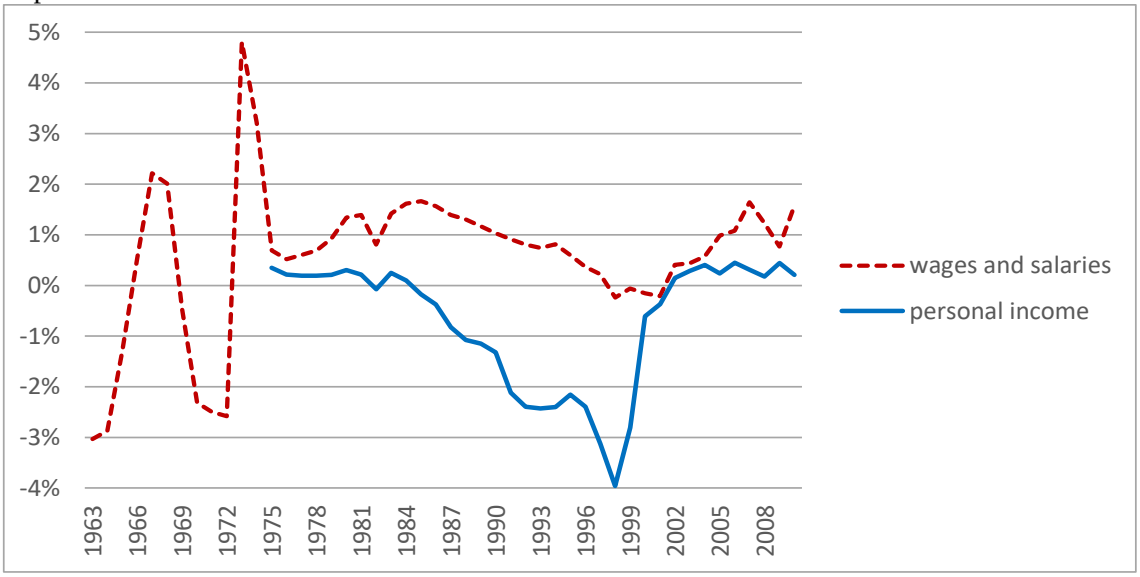
The concentration of earned income reflects the percentage of total earned income taken up by the top x% of all workers. Data on the total number of workers come from Statistics Korea′s *Economically Active Population Survey*, and data on total earned income from wages and salaries come from figures in the national accounts (Kim 2012b). Data on the income of top income earners can be obtained from income tax data, which is sometimes presented as the tax base, instead of income amounts, depending on the data. In this case, it is necessary to convert it into income amounts by adding the income deduction to this tax base. The changes in the income statistics used in the previous estimations are described below.

First, the wages and salaries of the national accounts are replaced by the new series as the results of the transition from the 1993 SNA to the 2008 SNA. Fig. 1 shows the resulting annual changes in the figures. A data discontinuity occurs around 1975. Wages and salaries after 1975 generally increased by 1% while they decreased by 0.1 to 0.2% from 1998 to 2001. It seems that, as total wages and salaries (denominator) increased, the top income shares decreased, whereas they increased slightly from 1998 to 2001. As a result, while the top income shares rose sharply after the currency crisis in the previous estimates, the revision slowed down the increase in the top income shares. No statistics on wages and salaries are available before 1974, which is why the extension is made according to employee compensation, including the employer's social contribution. This shows that considerable changes occurred when the statistics based on the 1953 SNA were transformed into those based on the 2008 SNA.

Second, after 2009, total earned income data can be obtained from income tax data in addition to the wages and salaries data of the national accounts. Earned income year-end settlement reports and daily labor income payment records are regarded as earned income in the *Statistical Yearbook of National Tax* of the National Tax Service. The former is subdivided into taxpayer and non-taxpayer types (i.e., those who have determined tax amounts and those who do not). However, there are more people with earned income reported by the taxation authorities than the number of

employed workers. Combining data on year-end settlement tax returns and daily labor income earners in the taxation data produces 25.46 million people in 2015, whereas there are only 19.23 million workers. This indicates that, among the recorded annual income earners, a considerable number are not included as employed persons in the *Economically Active Population Survey*. Employed persons here refer to those who worked for an hour or more to earn income during the survey period, whereas the taxation data include all the income data that were recorded at least once within the year, which results in such a difference.

Fig. 1 Rate of changes in wages and salaries as well as personal income due to the 2008 SNA implementation



Note: 1) 1963–1974 shows an extension of the wages and salaries after 1975 based on employee compensation. Employee compensation before 1974 is converted to the new series based on the 2008 SNA from the 1953 SNA.

2) Personal income is obtained by subtracting imputed rent, financial intermediation services indirectly measured (FISIM), and investment income disbursements from the sum of wages and salaries, operating surplus, and property income imputed to the household.

Data Sources: Bank of Korea, ECOS; Bank of Korea (1982, pp. 170–173).

These income tax data show reports on wages and salaries paid by employers to workers and income tax after withholding to the taxation authorities. The earned income reported by the employer is approved as an expense and is thus reported to the National Tax Service, through which most earned income is detected by the taxation authorities. However, for employment in households, such as housekeepers and private tutors, the labor costs are not handled as an expense and are thus not reported in some cases.

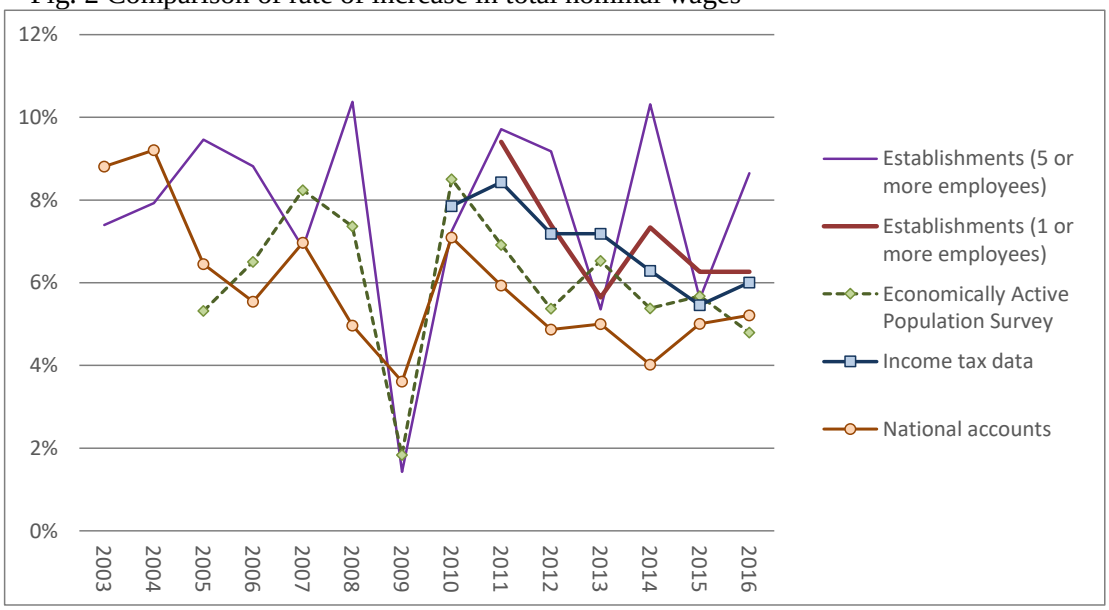
Then, which is greater? Earned income determined by income tax data, or wages and salaries in the national accounts? The income tax data were 5% lower than the national account data in 2009, but the gap narrowed after that and the total earned income from income tax data ended up 6% higher in 2016. The earned income trends in the two types of data have been very different, but which are more reliable? Fig. 2 shows various survey results presenting the rate of increase in total nominal wages besides income tax data and national accounts[3]. It shows that the rate of

[3] Since income tax data and the national accounts present the total amount of earned income, the rate of increase can be directly obtained from them. On the other hand, another survey determined the rate of increase in the

increase in the total amount of wages in the national accounts after 2010 remained at a considerably low level relative to that in the income tax data. Even when compared to the data surveyed by the Ministry of Employment and Labor and Statistics Korea, the rate of increase was lowest in the national accounts. Before 2009, it was higher than other surveys in some case, but, it was lower in other case, never slanting toward one side. However the underestimation bias in the national accounts has been cumulative since 2010.

As a result, wages and salaries in the national accounts recently turned out to be even lower than total earned income from the income tax data. The figures in the income tax data are not the estimates of a sample survey, but data on income that was actually reported and became the object of taxation (or tax exemption) by the National Tax Service in complete enumeration. Moreover, considering that there are omissions, the figures in the national accounts with the widest income survey scope may be expected to be greater than in the income tax data. However, the fact that it is 6% smaller indicates that the wages and salaries in the national accounts are underestimated. Fig. 2 shows that this bias is a result of an accumulated bias every year since 2010.

Fig. 2 Comparison of rate of increase in total nominal wages



Note: 1) Income tax data show the rate of increase in earned income (sum of year-end settlement earned income and daily labor income) obtained in the *Statistical Yearbook of National Tax*.

2) National accounts show the rate of increase in wages and salaries of the Bank of Korea.

3) Others show the rate of increase in total wages obtained from the wage growth rate and employment increase rate in each survey. Total wage increase rate = wage growth rate + employment increase rate + wage growth rate * employment increase rate.

Data Sources: National Tax Service, *Statistical Yearbook of National Tax*; Bank of Korea,

---

total amount of wages from the rates of increase in both wages and number of employees. For establishments with five or more persons, both wages and the number of workers were obtained from the *Survey Report on Labor Conditions by Employment Type*. Meanwhile, for establishments with one or more persons, wages were obtained from *the Survey Report on Labor Conditions by Employment Type*, while the number of workers was obtained from the *Report on Labor Force Survey at Establishments*. Statistics Korea's *Economically Active Population Survey* surveys wages in March and August every year (August wages are used here) in the additional survey by employment type, which is applied to all workers. The scope of the survey is in the following order: five or more persons < one or more persons < Economically Active Population Survey < income tax data < national account (the order of legends is shown in Fig. 2).

ECOS; Ministry of Employment and Labor, *Report on Labor Force Survey at Establishments*, *Survey Report on Labor Conditions by Employment Type*; Statistics Korea, *Economically Active Population Survey*.

Here, total income cannot be determined through the income tax data before 2008, which is why wages and salaries from the national accounts are used. However, the earned income surveyed in the income tax data will be used after 2009,[4] because consistency can be improved by using the same data for both the numerator and denominator of top income shares; moreover, the total earned income figures in the income tax data are more reliable than are the wages and salaries data in the national accounts. The decile distribution, including the middle and bottom income earners, can be obtained in addition to the top income shares at the top.

Third, the income of the top income earners can be obtained from the statistics by income bracket in the income tax data, but data have a limitation in the 1995–2004 period. In this period, the amounts of taxable wages and salaries can be found in the income tax data, but the distribution is presented only in the statistics arranged by tax base bracket. To determine the income of the top income earners, it is necessary to add the income deduction to the tax base and convert it to taxable wage and salary (and to total wage and salary).[5] Previous studies (Kim 2012a, 2012b) have applied the conversion rate by income bracket for 2005 (i.e., ratio of adjusted wage and salary/tax base and ratio of taxable wage and salary/adjusted wage and salary) to the 1995–2004 period, for which there are no data. This conversion rate changes along with the income deduction system or income distribution, which causes a gap between the income amounts restored in this method and the actual figures. Fortunately, the total amount of taxable wages and salaries can be found in the income tax data, and thus the conversion rate by income bracket can be adjusted so that the two figures are consistent, which can also reduce errors.

The conversion rate of total income can be obtained from the data, but additional information or assumptions are necessary for the conversion rate by income bracket. In this period, the income deduction has been gradually expanded, and the ratio of income deducted is higher in the lower income bracket. Therefore, as we go further back before 2005, the conversion rate for the year (= adjusted wage and salary/tax base) is lower than in 2005, and the effect is likely to be greater in the low income bracket. In previous methods, the conversion rate by income bracket was adjusted accordingly. However, since there was no clear evidence for the assumption about how different the conversion rate would be in each income bracket, this study made an improvement in a more rational way.[6]

---

[4] As mentioned, income tax data also include the income of those not considered employees. Thus, to obtain the top income shares of employed workers, it is necessary to exclude the income of the non-employed from the total earned income used in the denominator. Here, daily labor income earners with lower income are considered more likely to be non-employed, and thus total earned income was obtained by deducting their income. Due to the limited data available before 2008, the ratio of earned income imputed to them could not be determined, which is why the same ratio determined in 2009 (7.4%) is applied to all years before that.

[5] It is necessary to establish the scope of income in the *Statistical Yearbook of National Tax*. For earned income, the statistics provided are on the tax base, adjusted wage and salary (= tax base + income deduction), taxable wage and salary (= adjusted wage and salary + deduction for wage and salary income), and total wage and salary (= taxable wage and salary + nontaxable income).

[6] For example, the tax base of wage and salary in 2001 is 62 trillion KRW, adjusted wage and salary is 103 trillion KRW, and taxable wage and salary is 165 trillion KRW. First, the following shows how to convert the tax base into the amount of adjusted wage and salary. The amount of adjusted wage and salary obtained by applying the 2005 conversion rate (adjusted wage and salary/tax base) by income bracket to the 2001 tax base statistics by income bracket is 118 trillion KRW. This is overestimated by 15.3% (=118/103), which indicates that the actual conversion rate in 2001 was lower than that in 2005. However, we cannot determine the situation for each income bracket because this ratio concerns only the total conversion rate. In previous methods, the

Fourth, earned income tax data before 1985 are all established based on total wage and salary, unlike the aforementioned 1995–2004 data. However, Kim (2012b) misunderstood the fact that the statistics for 1963–1969 are based on the tax base and used the conversion rate (= total wage and salary/tax base) from 1970 that is close to this period in order to convert this to total wage and salary. However, this adjustment is not necessary because there was no income deduction system in the 1960s, and thus the amount of income was the same as the tax base. This study corrected this error.

However, the third and fourth factors above are related to converting the tax base to the amount of total wage and salary, and the gap between the two is not very big because the ratio of income deduction decreases in the higher income bracket. For example, when estimating the income share of the top 10%, errors that may occur in the assumption of the conversion rate may have some effect, but this effect is likely to become more insignificant in higher income groups such as the top 1% and even the top 0.1%.

2. Application of the GPC interpolation method

Income tax data show how the number of people and their income is distributed in each income bracket. Table 1 presents the case in 2015.[7] For example, it shows that those in the top 1% are included in the 100–200 million income bracket and that they are within the 3.77%–14.9% income share,[8] but the exact position cannot be specified. Either a Pareto interpolation or mean split histogram is used to estimate this. The former method may come up with various estimates depending on how it is applied, and the GPC has recently been proposed as an alternative. This section briefly compares these methods and outlines how they differ.

A Pareto distribution is given in Formula (1):

$$1\text{-}F(y)=(z/y)^{a} \tag{1}$$

conversion rate was adjusted based on the assumption that overestimation would be greater in the lower income bracket. However, this assumption was unsatisfactory due to its arbitrariness, which is why this study took a different approach. The difference between the amount of adjusted wage and salary and tax base obtained by applying the 2005 conversion rate to the 2001 tax base was 56 (=118-62) trillion KRW, which was overestimated compared to the actual value of 41 (=103-62) trillion KRW. Here, 0.723 (= 41/56), the ratio of both, is equally multiplied by the difference between the estimated amount of adjusted wage and salary and the tax base of each income bracket. Unlike before, the overestimated ratio was obtained with the difference, which was applied equally to all income brackets, making the assumption simpler and easier to understand. Comparing the 2001 conversion rate by income bracket with that in 2005, the conversion rate declined more sharply in the lower brackets, as assumed by the previous methods, but the decrement has increased. The figures for the two income brackets above 80 million KRW and below 10 million KRW show that, while the conversion rates of adjusted wage and the salary/tax base in 2005 were 1.119 and 2.943 respectively, those in 2001 using the previous methods decreased to 1.079 and 2.481, and those in the results of this study decreased to 1.088 and 2.405 respectively. The above shows the conversion of the tax base to the amount of adjusted wage and salary, but the same approach is taken to convert the amount of the adjusted wage and salary to the amount of taxable wage and salary (and to total wage and salary).

[7] As mentioned, the sum of the subjects of earned income year-end settlement (taxpayers and non-taxpayers) and daily labor income earners in the *Statistical Yearbook of National Tax* exceeds the total number of workers in the Economically Active Population Survey. This study considers that daily labor income earners with low income may not have been included as employed people. Thus, the total number of workers and earned income are obtained by excluding them and the income imputed to them. Moreover, the standard was changed to total salaries by adding the nontaxable income excluded from taxable income. For example, this is why extra numbers are added such as 100.681 million KRW for 100 million KRW in the income interval of Table 1.

[8] The cumulative distribution (p^N) of the number of people in the interval of 100–200 million KRW in Table 1 is 0.9686-0.9967, which is the top 3.14%-0.33% when accumulated from the highest ($1\text{-}p^N$). The cumulative distribution (p^Y) of income that corresponds to this is 0.8510-0.9623, which is 14.9%-3.77% when accumulated from the highest ($1\text{-}p^Y$).

Here, y is income, F(y) is a cumulative distribution function, z is the lower limit of income where the Pareto distribution is applicable, and a is the Pareto coefficient. 1-F(y) is the ratio of inclusion in the top x% when accumulated from the highest ($1-p^N$ when marked as shown in Table 1). For example, the formula is 1-F(y) = 0.01 for obtaining the top 1% income (y).

There are various ways to obtain the Pareto coefficient. First, one of the important features of Pareto distribution is the fact that the average income of those with y or more income is b times y. Here, b is an inverted Pareto coefficient, and b = a/(a-1)[9]. Piketty and Saez (2003) obtained b in each income bracket using this method, through which it is also possible to obtain the threshold income to be included in the top x% and their income share. This method is hereafter referred to as "Pareto 1." Table 1 presents b as b1 using the 2015 earned income data. When marked as shown in Table 1, $b1_i = (\sum_i^k Y_i/\sum_i^k N_i)/y_i$. Here, $_i$ is the i^{th} income bracket from the lowest income bracket, and the highest income bracket becomes the k^{th} bracket.

Second, a log-linear interpolation is used by Feenberg and Poterba (1992) (hereafter referred to as "Pareto 2"). Regarding the two income brackets close to Formula (1) above, $1-F(y_i) = (z/y_i)^a$ and $1-F(y_{i+1}) = (z/y_{i+1})^a$, from which $a = \log[(1-F(y_i))/(1-F(y_{i+1}))]/\log[y_{i+1}/y_i]$. When marked as shown in Table 1, $a = \log((1-p^N_i)/(1-p^N_{i+1})/\log(y_{i+1}/y_i)$. Here, b obtained from a is presented as b2 in Table 1.

Table 1 Earned income distribution by income bracket and inverted Pareto coefficient (2015)

lower limit	upper limit	no. of people	income	no. of people	income	inverted Pareto coefficient		
(thousand KRW)		(thousand	(1 billion	cumulative distribution				
y		people) N	KRW) Y	p^N	p^Y	b1	b2	b3
	10,158	3,760	19,490	-	-			6.690
10,158	15,152	2,662	33,759	0.1955	0.0320	3.752		3.202
15,152	20,183	2,298	40,454	0.3340	0.0874	2.864		2.617
20,183	30,221	3,337	82,732	0.4535	0.1538	2.430		2.183
30,221	40,256	2,196	76,511	0.6270	0.2897	1.996		1.878
40,256	45,297	791	33,738	0.7412	0.4153	1.777	3.147	1.738
45,297	50,343	655	31,246	0.7823	0.4707	1.700	2.639	1.668
50,343	60,396	1,023	56,457	0.8163	0.5220	1.637	2.139	1.586
60,396	80,541	1,328	92,265	0.8695	0.6147	1.549	1.618	1.508
80,541	100,681	578	51,648	0.9386	0.7662	1.497	1.497	1.493
100,681	200,695	540	67,783	0.9686	0.8510	1.495	1.440	1.644
200,695	300,462	38	9,057	0.9967	0.9623	1.817	1.787	1.829
300,462	500,280	17	6,429	0.9987	0.9771	1.839	1.805	1.847
500,280	1,000,184	6	4,059	0.9996	0.9877	1.865	1.929	1.848
1,000,184		2	3,447	0.9999	0.9943	1.805		
total		19,230	609,075	1.0000	1.0000			

Note: 1) The cumulative distribution of the number of people and income (p^N, p^Y) is obtained by dividing the cumulative number of people (or income) up to the previous income bracket by the total number of people (or income).

2) b1, b2, and b3 indicate inverted Pareto coefficients obtained with the three Pareto interpolation methods explained in the text.

Data Source: National Tax Service, *Statistical Yearbook of National Tax*, 2016.

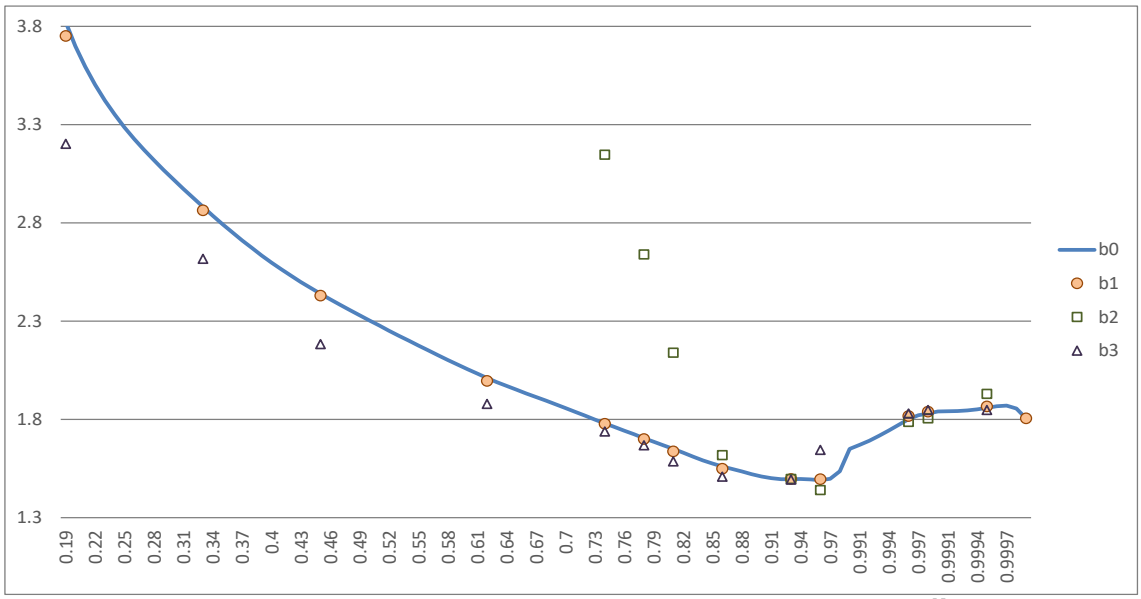
[9] The formula is not provided due to space limitations. For details, see Kim (2012a, pp. 9–11).

Third, by differentiating Formula (1) and obtaining the density function, it is possible to obtain the ratio of the number of people with higher income than y as well as the ratio of the income imputed to those people. This leads to the formula that can obtain b from the number of people and income data in adjacent brackets.[10] When marked as shown in Table 1, $\log((1-p^{N}_{i})/(1-p^{N}_{i+1}))/\log((1-p^{Y}_{i})/(1-p^{Y}_{i+1}))=a/(a-1)=b$. This method is referred to as "Pareto 3," and b here is presented as b3.

There are differences between the b obtained using the aforementioned method, due to the difference in the information used. Unlike for the lower limit of the income bracket, upper limit information was not used in Pareto 1. Only data on income brackets and number of people were used in Pareto 2; income information was not used. In Pareto 3, only data on the number of people and income for adjacent brackets were used. In Pareto 1, b can be obtained even in the bracket without the upper limit (income bracket of 1 billion KRW or above in Table 1); in Pareto 2 and 3, this is applied only to income brackets with both the lower and upper limit.

According to Table 1, b is changing according to the income bracket. The coefficients (a or b) are originally fixed in a strict Pareto distribution. If b is 2, the average income of earners of 100 million or more is 200 million, and the same applies to other brackets (e.g., the average income of earners of 50 million or 1 billion KRW or more is twice that amount, thus 100 million or 2 billion KRW). As shown in Table 1, however, this is not so in reality; rather, b changes according to the income bracket. Fig. 3 shows this in a graph, where X-axis shows p^{N}, the cumulative distribution of the number of people, instead of income.

Fig. 3 Comparison of inverted Pareto coefficients (b) estimated according to interpolation methods



Note: 1) X-axis indicates cumulative distribution of the number of people (p^{N} in Table 1). It is divided into the unit of 0.01 to 0.99, and then subdivided into 0.001 to 0.99–0.999, and then 0.0001 to 0.999–1.

2) b1, b2, and b3 indicate b in Table 1, and b0 is obtained using the GPC method.

3) This deals with 2015 earned income.

Data Source: National Tax Service, *Statistical Yearbook of National Tax*, 2016.

[10] The formula is not provided due to space limitations. For details, see Kim (2012a, pp. 9–11, 34–35).

As the figure shows, b declines in the higher income brackets, but the cumulative distribution shows a U-shaped curve as it goes back up around 0.95 (or top 5%). When b is on the rise, this indicates that inequality is increasing in the higher income brackets. Empirically, some differences are observed in other years and countries, but they generally show U-shaped curves similar to those in Fig. 3.

To capture the characteristics of such income (or wealth) distributions, Blanchet, Fournier, and Piketty (2017) presented a more flexible Generalized Pareto Curve instead of a Pareto distribution where b is fixed. There, it is defined as the curve of an inverted Pareto coefficient, $b(p) = E[X>Q(p)]/Q(p)$. Here, p is the cumulative distribution of the number of people (p^N in Table 1) within the range of $0 \leq p < 1$, and Q(p) is income (y) corresponding to p. b(p) shows how many times greater than y(p) is the average income for those with income higher than y(p), corresponding to the cumulative distribution p of the number of people. In this sense, the method is the same as that by which b1 was obtained above. However, b(p) is defined as a continuous function, where all income tax data are reflected as a constraint condition in estimations, in contrast to the aforementioned b1–b3.

Fig. 3 presents the b(p) estimated based on the 2015 earned income as b0[11]. Unlike b1–b3, in which b changes in tiers according to the income bracket, b0 changes continuously. There was a huge gap between the b according to the estimation method in the middle and bottom income brackets, but b is generally getting closer in the top 10%. However, in the income bracket that includes the top 1% (between 0.9686 and 0.9967 in terms of p^N as shown in Table 1), the gap between the b increased. Thus, there may be a gap in the top income shares of the top 1% depending on which b is used in the estimation.

The mean split histogram approach is not based on certain distributions, like the Pareto distribution. The idea of this method is as follows. The number of people tends to decrease when the income is higher in the top brackets; in which case, the average is located on the left side of the center in the relevant bracket. In other words, when the lower and upper limits of the income bracket is y_i and y_{i+1} and the average income in the bracket is y_{mi}, then $y_{mi}<(y_i+y_{i+1})/2$. When this bracket is divided into two intervals on the left and right of y_{mi} into $[y_i, y_{mi}]$ and $[y_{mi}, y_{i+1}]$, more people are distributed in the left interval than in the right. Moreover, it is assumed that the speed at which the number of people decreases when the income increases tends to be constant in each interval. With just this simple assumption, it is possible to obtain results close to the actual distribution of the number of people within the income bracket.[12]

Which of the various methods above will produce results closest to the actual status? Blanchet, Fournier, and Piketty (2017, pp. 18–23) compare four methods (the aforementioned Pareto 1, 2, GPC, and mean split histogram) in the US (1962–2014) and France (1944–2012), where raw data on income tax can be obtained. First, they acquired income data in which the p obtained from the

---

[11] WID (http://wid.world/gpinter) is convenient because, if you enter the data required for applying the GPC, WID automatically calculates and provides the result.

[12] The following is the coordinate in which the x-axis represents income and the y-axis represents $1-p^N$, which is the cumulative distribution of the highest income earners. Suppose that the three coordinates, which consist of the lower limit of the income interval (y_i), average of interval (y_{mi}), upper limit (y_{i+1}), and the corresponding $1-p^N_i$, $1-p^N_{mi}$, and $1-p^N_{i+1}$, are $L(y_i, 1-p^N_i)$, $M(y_{mi}, 1-p^N_{mi})$, and $H(y_{i+1}, 1-p^N_{i+1})$. Here, $1-p^N$ decreases as y increases, and thus the slope of the line connecting these coordinates is (-). The mean split histogram assumes; thus, there are more people distributed below the average within the interval (i.e., the slope of LM is steeper than MH), and the decreasing speed of the people within each interval is constant (i.e., LM and MH are straight lines). However, there is no guarantee that the actual distribution of people starts to change at the average point, or that the decrease of $1-p^N$ along with the increase in income is shown along the straight line. For an explanation using a graph, see Atkinson (2005, p. 333) and Kim (2014, pp. 63–73).

raw data are 10%, 50%, 90%, and 99%, applied each interpolation method above to estimate the threshold income and income share to be included in the top 5%, 25%, and 70%, and compared this to the actual values in the raw data. The results showed that Pareto 1 and 2 had a relatively large gap, followed by the mean split histogram, while the GPC result was closest to reality. Pareto 1 was relatively close to reality in the estimation of the top income share, but showed a huge gap in estimating threshold income. The mean split histogram obtained results similar to reality in the middle and bottom income share. Meanwhile, GPC showed results superior to other methods, as the error did not exceed 1% in most cases for either threshold income or income share. It is remarkable that results close to reality could be obtained with such strictly limited income data of four points. Of course, the gap in each method is reduced by using more specific and subdivided information. Considering the above, this study adopts the GPC method.

Since raw income tax data are not available in Korea, the relative merits of each method cannot be verified, as mentioned above. Instead, this study finds the gap between the income share and threshold income by quantile obtained using each method for the 2015 earned income. Table 2 shows the cases with at least a 1% difference from the GPC method in light shades (dark gray if the gap is at least 3%). The income share results are generally close to reality, except for Pareto 2. The mean split histogram showed results closest to GPC, except for the top 1% income share. Pareto 1 and 3 also were not very different in obtaining the top income shares, except for the top 1%. However, when estimating threshold income, Pareto 1 and 3 showed a difference from the GPC greater than that in the estimation of income share. Here, the mean split histogram was closest to GPC in intervals of the 10th quantile or lower, but the gap increased in the top 1% or higher, which increased up to 7%. As shown in Fig. 3, the gap turned out to be greater around the top 1% because the income bracket in the income tax data is becoming wider even though the trend of b is changing in that bracket;[13] the available information is not sufficiently subdivided to use it to track the trend of b.

Table 2 Comparison of income share and threshold income by quantile estimated by interpolation method (2015 earned income in Korea)

	Income share by quantile (%)					Threshold income (thousand KRW)				
	GPC	Pareto 1	Pareto 2	Pareto 3	mean split histogram	GPC	Pareto 1	Pareto 2	Pareto 3	mean split histogram
1st quantile	0.83	3.34		1.56	3.34					
2nd quantile	2.52			1.81		5,355			5,178	
3rd quantile	3.84	3.80		3.95	3.84	10,343	10,199		11,459	10,322
4th quantile	4.99	4.87		5.00	5.00	13,901	14,669		13,096	13,954
5th quantile	6.33	6.42		6.45	6.34	17,843	16,217		17,690	17,818
6th quantile	7.96	8.01		7.51	7.95	22,414	21,268		23,574	22,471
7th quantile	10.10	10.03		10.07	10.10	28,329	29,186		31,082	28,384
8th quantile	13.09	13.18		13.35	13.10	36,049	37,738		38,673	36,012
9th quantile	17.96	17.91		18.01	17.96	47,797	46,908		47,770	47,757
10th quantile	32.37	32.45	34.16	32.30	32.37	67,920	66,367	66,860	67,819	67,692
Top 5%	20.38	20.38	20.38	20.37	20.37	86,409	86,237	86,236	86,437	86,446
Top 1%	7.16	6.94	6.49	7.44	7.27	137,390	147,008	142,747	143,272	128,310
Top 0.5%	4.77	4.76	4.70	4.81	4.82	170,603	165,992	166,636	166,439	182,154
Top 0.1%	1.97	1.97	1.93	1.98	1.97	339,759	340,046	339,117	338,684	341,236
Top 0.05%	1.36	1.36	1.40	1.36	1.36	462,403	460,512	459,057	465,283	469,020
Top 0.01%	0.57	0.57			0.57	996,710	996,880			997,757

Note: 1) Pareto 1, 2, and 3 indicate Pareto interpolation methods corresponding to b1, b2 and b3 as shown in Table 1.

2) The light shades indicate 1 to 3% gap compared to GPC, and dark grey shade indicates

[13] As shown in Table 1, the income brackets up to 100 million KRW were either at 10 million or 20 million, whereas the income bracket after that increased up to 100 million KRW. The top 1% belongs here.

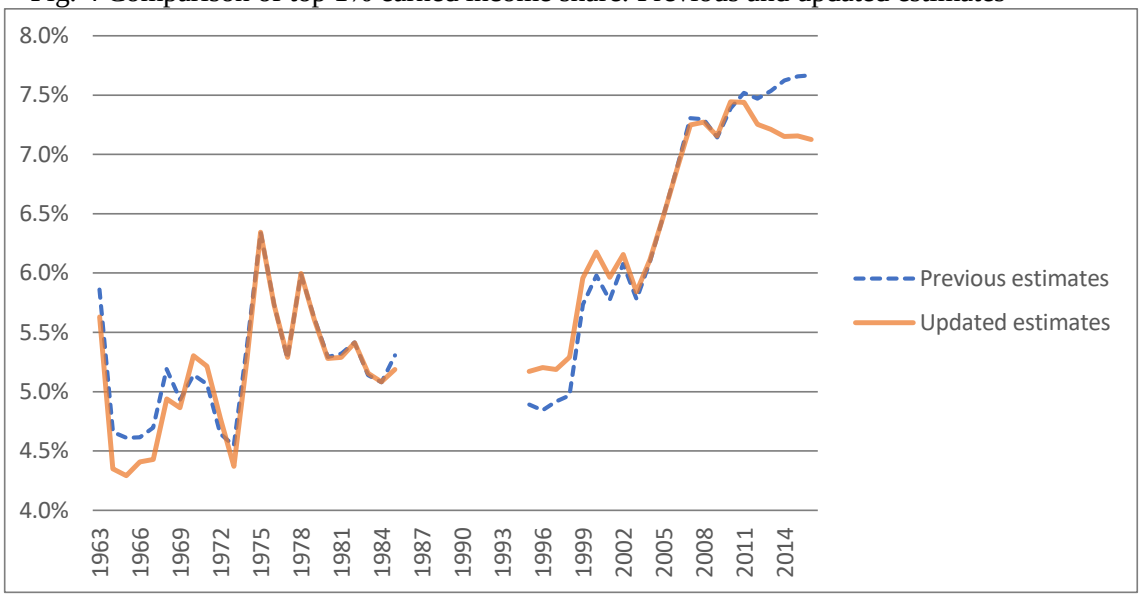
at least a 3% gap.

Data Source: National Tax Service, *Statistical Yearbook of National Tax*, 2016.

3. Estimation results

Appendix Table 1 shows the top x% workers' threshold income, average income, and income shares obtained as explained above by revising the earned income statistics and applying the interpolation method of GPC. Fig. 4 provides the income share of the top 1% and compares how this result has changed compared to previous estimates (Kim and Kim 2014). According to the results, there were relatively large differences from previous estimates in the 1960s and early 1970s, from 1995 to 2004, and in the 2010s. First, the gap caused by changing the interpolation method from Pareto 1 to GPC turned out to be less than 1% in most time periods. However, as shown in the case of 2015 in Table 2, the gap widened in some years (1963, 2011–2016) up to around 3%.

Fig. 4 Comparison of top 1% earned income share: Previous and updated estimates



Note: Previous estimates cover up to 2010, but the same method was applied to extend the data to 2016.

Data Source: Kim and Kim (2014); Appendix Table 1.

The gap mostly originates from the data revision. The gap in the 1960s and early 1970s results from the first and fourth factors mentioned in Section II-1 above—that is, applying the series of wages and salaries before 1974 based on the 2008 SNA (as shown in Fig. 1) and correcting the error made in the treatment of income tax statistics in the 1960s. The gap for 1995–2004 comes from revising the conversion rate to convert income tax data based on the tax base to income amount, mentioned as the third factor, aside from the transition to the 2008 SNA mentioned in Section II-1. As a result, the top income shares updated during this period increased slightly, which slowed down the rise of the top income shares accordingly.

One thing to note is that, while the top income shares of the previous estimates continue to rise since 2010, the updated estimates are falling because, as mentioned as the second factor in Section II-1, previous estimates continued to use wages and salaries in the national accounts for total earned income as the denominator in calculating the concentration, whereas the updated estimates

use income tax data. The updated estimates more clearly reflect reality, given that wages and salaries in the national accounts have been cumulatively underestimated in the 2010s; also, consistency can be maintained, as both the denominator and numerator are based on the same income tax data.

After 2009, the total income distribution can be determined without being limited to top income earners by using income tax data. Table 3 shows the ratio of earned income by decile and Gini coefficients based on the total number of workers, which are summarized in Fig. 5. The income share of the bottom 50% hit bottom in 2010, at 16.1%, and constantly increased afterwards up to 19.0% in 2016. The income share of the top 10% fell from 33.9% to 32.0% in the same year, while the middle 40% declined from 50% to 49%. As a result, the Gini coefficients have constantly decreased since 2010. Fig. 4 shows that the top income share of the top 1% has fallen since 2010, but the key factor was the relatively quick increase of income in the bottom 50%.[14]

Table 3 Earned income distribution by quantile (unit: %)

	2009	2010	2011	2012	2013	2014	2015	2016
1st quantile	0.592	0.520	0.679	0.774	0.740	0.801	0.828	0.824
2nd quantile	1.967	1.731	2.124	2.364	2.392	2.527	2.516	2.608
3rd quantile	3.382	3.323	3.447	3.681	3.785	3.902	3.842	3.971
4th quantile	4.585	4.546	4.592	4.773	4.870	5.034	4.992	5.138
5th quantile	5.967	5.952	5.980	6.116	6.225	6.350	6.333	6.464
6th quantile	7.814	7.756	7.746	7.820	7.905	7.968	7.964	8.054
7th quantile	10.178	10.101	10.054	10.074	10.135	10.136	10.102	10.122
8th quantile	13.534	13.454	13.289	13.210	13.186	13.108	13.094	13.022
9th quantile	18.784	18.735	18.459	18.227	18.103	17.895	17.958	17.787
10th quantile	33.197	33.882	33.630	32.961	32.658	32.279	32.372	32.012
Total average (thousand KRW)	24,501	25,489	27,065	28,569	29,733	30,770	31,673	33,078
Gini coefficient	0.491	0.499	0.489	0.476	0.472	0.464	0.465	0.459

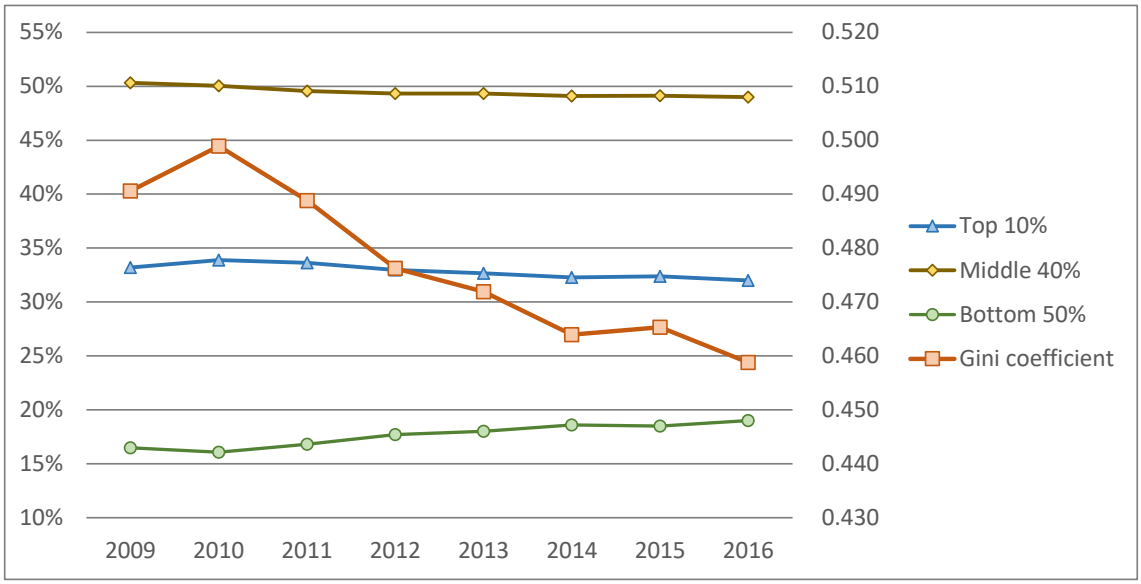
Note: 1) This is the income share of each quantile when dividing the total number of workers into 10 quantiles (or deciles) by income.

2) Here, the total number of workers is the one determined by the *Economically Active Population Survey* (Statistics Korea).

Data Sources: National Tax Service, *Statistical Yearbook of National Tax*, 2017; Statistics Korea, KOSIS.

Fig. 5 Earned income share by income group and Gini coefficients

[14] Regarding the earned income distribution, the survey of establishments in the Ministry of Employment and Labor's *Survey Report on Labor Conditions by Employment Type* is commonly used. Jeong (2017, pp. 64–66) used the microdata in this survey to estimate the inequality index of earned income for 1980–2015 for establishments with 10 or more employees. The results were contrary to those in this study (see Table 3 and Fig. 5), as the inequality index continued to rise after 2010. This matter can be fully reviewed only in further research. This study merely points out the possibility that this survey report may not properly reflect the real status of the earned income distribution. The working paper version (Kim 2018, pp. 14–16) of the study compares the distribution of the number of workers in this survey report with the relevant distribution of the income tax data, which is not included in this study due to limited space. The comparison shows that, while the survey report overestimated the number of workers in the middle income bracket (20−30 million KRW), the number of workers tended to decrease rapidly in the lower and higher income brackets. For example, according to income tax data, while there were 50,908 earners of earned income included in the 200 million KRW or higher annual income bracket, there were less than half that number in establishments with one or more employees, and only 15,873 in establishments with five or more employees.



Note: 1) The middle 40% is the group excluding the bottom 50% and top 10%.

2) Income share is shown in the scale of the left axis, and the Gini coefficient is shown in the scale of the right axis.

Data Source: Table 3.

Ⅲ. Concentration of total income

How about the concentration of total income including financial and business income in addition to earned income? In this case, the income of those in the top x%, which is the numerator of top income shares, is also obtained from the income tax data. The total amount of various forms in income earned by an individual can be found in the global income tax data; in many cases, however, the returns of global income tax are not reported. This is because, for those with earned income, taxation is completed with the year-end settlement without a report of the return of global income tax if other sources of income do not exceed a certain amount. Therefore, it is necessary to combine global income tax data and earned income tax year-end settlement data to determine the top income earners.[15]

Income imputed to households used to be extracted from the national accounts to determine the total income used as the denominator of top income shares. However, as revealed by the aforementioned earned income, it is desirable to determine the denominator based on income tax data if possible. Personal income in the national accounts can be categorized into earned income (wages and salaries in the income accounts by institutional sector; the same applies to the rest),

[15] For a more specific method, see Kim (2012a, pp. 78–93). When combining the global income tax data and earned income year-end settlement data, it is necessary to exclude those with earned income that appear redundantly in both sets of data. In this case, the situation is similar between Korean and Japanese data. Thus, Kim (2012a, p. 87) adopted the method applied to Japan by Moriguchi and Saez (2008) to estimate the Korean data. The National Tax Service disclosed combined income excluding the redundancies in the two data for certain time periods (2007–2012) upon the request of members of the National Assembly. The top income shares obtained using this data showed little difference from the data obtained by Kim′s (2012a) method (for example, the income share of the top 1% in 2012 was the same, at 11.7%, and that of the top 10% was 43.0% and 42.6%). The previous method is used here, as Kim′s (2012a) method does not distort the results, and it is necessary to retain consistency with the time periods for which data are not disclosed.

financial income (interest, dividends), and business income (operating surplus, rent, withdrawals from income of quasi-corporations). For earned income, it is first necessary to correct the aforementioned bias toward underestimation. This study revises the data so that the earned income after 2010 has increased not according to the rate of increase in the national accounts data as seen in Fig. 2 but according to the rate of increase in the income tax data, which is higher than that.

Since the interest and dividends included in financial income are all subject to withholding tax, the total amount can be obtained from income tax data. The interest income in the national accounts includes financial intermediation services indirectly measured (FISIM), which must be estimated and excluded, but the interest obtained here is almost the same as the interest in the income tax data. There is almost no difference in dividend income between the two types of data as well, which indicates that the financial income in the national accounts is based on the income tax data. Therefore, the national accounts data can be replaced by the income tax data for financial income.

However, while business income is generally included in the global income tax report, there are other types of business income that are subject to withholding tax or year-end settlement, and business income is also included in other income. Income tax data for each of these items are available, but some of them are redundant as they are included in the global income tax report. These redundancies must be eliminated, which is not easy due to data limitations. Small establishments also have undisclosed income. In addition, it seems that business income taxes are dodged in many cases, which might not have been captured by the income tax data. Therefore, it is difficult to determine the total amount of business income with only the income tax data. Thus, this study depends on the national account statistics. In the national accounts, operating surplus, rent, and withdrawals from the income of quasi-corporations are the business income items of a household. However, imputed rent for owner-occupied housing included in operating surplus is not a common income, and must therefore be estimated and deducted from the data.[16]

In using the national account statistics, the results were all updated up to the 2008 SNA implementation, and the aforementioned GPC was applied to the interpolation of income brackets[17]. The series before liberation (1933–1940) was extended to 1942,[18] and the GPC method is also applied here. Appendix Table 2 shows the threshold income, average income and income share of the top x% obtained by using the sum of earned income, financial income, and business income as the total income (denominator) and the population aged 20 or above as the total number of people[19]. Fig. 6 compares the top 1% income shares in previous estimates (Kim and Kim 2014) with the updated results. While previous estimates end at 2012, the same method was applied in extending the top income shares up to 2016.

---

[16] The operating surplus from residential services can be obtained from the Input–Output Tables of the Bank of Korea, which can be classified into actual and imputed rent using the owner-occupied and rented housing ratio of the Housing Census.

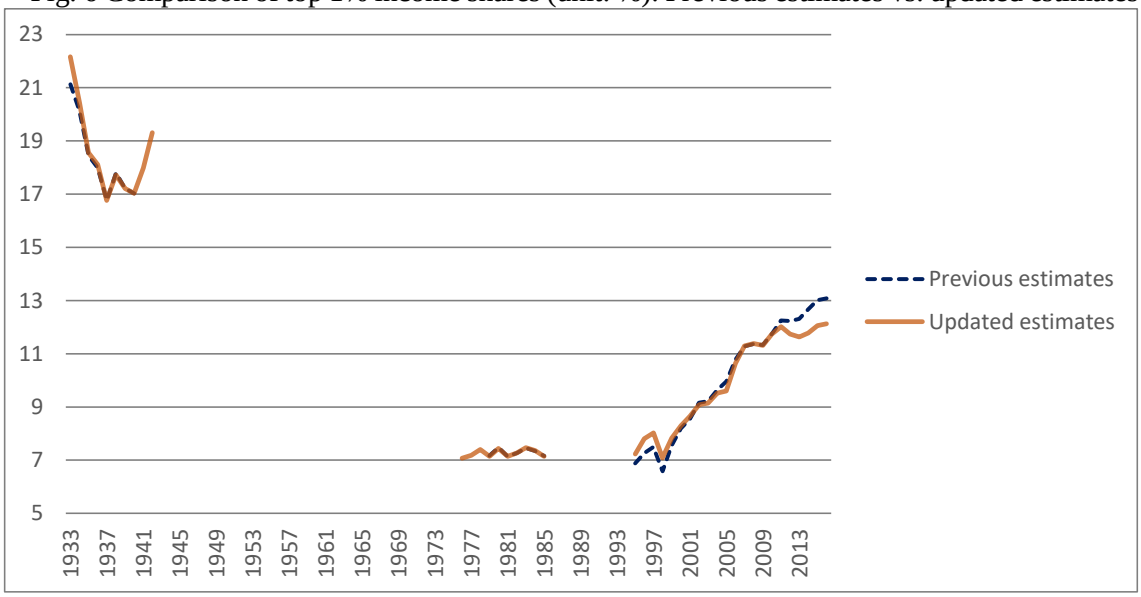
[17] To interpolate with GPC, it should be taken into consideration that there are many non-taxpayers whose income is larger than that of taxpayers, due to income deduction system. By assuming that there is no such case in the brackets where the lower threshold income of the brackets is greater than the twice of average income of adult population, the brackets with income less than the twice of average income are integrated into one bracket which includes the income of both taxpayers and non-taxpayers. In this case, the top 5% and 10% lies in the integrated bracket during 1976-79 and 1980-85, respectively. During 1933-36 when there were very small number of taxpayers, even top 1% lies in the integrated bracket. As a result, the threshold (or average) incomes of these top x% become larger than the previous estimates which did not make adjustments for non-taxpayers with relatively high income.

[18] Income tax data for 1941–1942 are available, but there are no total personal income statistics. Here, personal income is obtained by extending it to 1942 using per capita GDP of constant prices in 2010 (Kim, Park, Park, and Cha 2018, pp. 734–735) and the consumer price index (Park and Kim 2011).

[19] While the amount is deflated by the CPI for the earned income, the amount is converted into 2015 constant prices by the GDP deflator, following the WID. As a result, it is different from the figures in Appendix Table 2 of Korean version of this paper which used CPI to deflate income.

The gap between the updated and previous estimates before 2009 was affected by the transition of the national account statistics to the 2008 SNA, in addition to the fact that the demographics were revised after the 2000s. Fig. 1 shows how personal income has changed due to the 2008 SNA implementation. Personal income decreased up to 4% in the 1990s,[20] which resulted in an increase of top income shares during this period. The gap that appeared after the 2010s is due to the revision of the aforementioned underestimation of earned income in the national accounts. After the update, the increase in top income shares slowed after the mid-1990s, as did earned income.

Fig. 6 Comparison of top 1% income shares (unit: %): Previous estimates vs. updated estimates



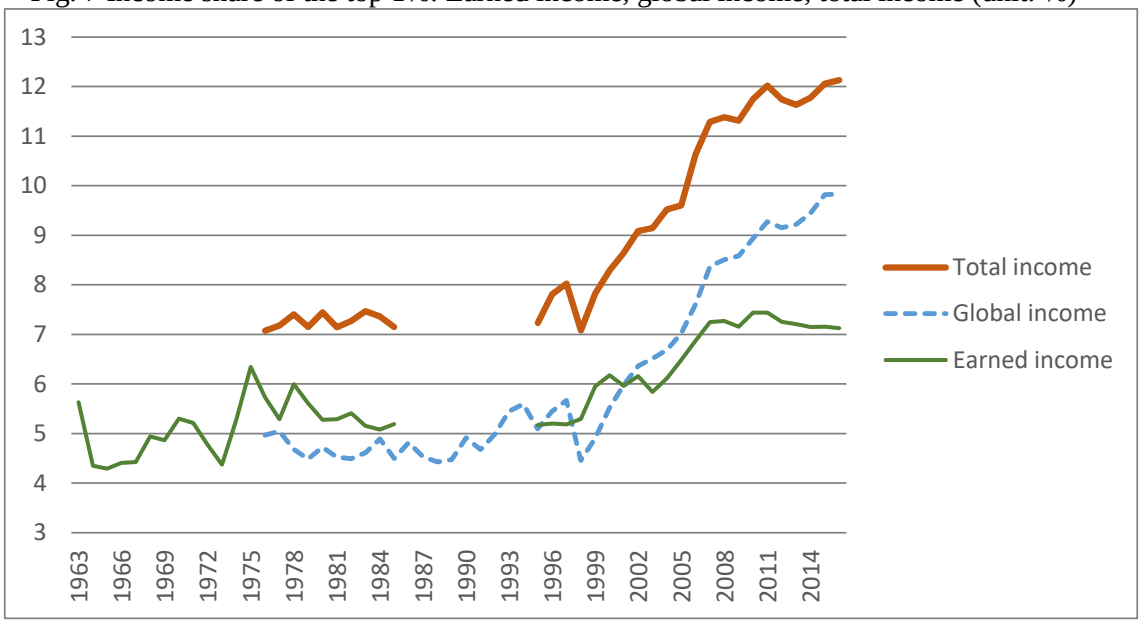
Data Source: Kim and Kim (2014); Appendix Table 2.

However, the concentration of total income would have been affected by unearned income such as business or financial income aside from earned income. Can the two be separated? Fig. 7 provides the top income shares obtained using the global income tax data along with the top 1% concentration of earned income and total income (see Figs. 4 and 6). The global income tax data include some on earned income (28% of total reported income was earned income in 2015), but it includes most business income and at least a certain amount of financial income (e.g., 20 million KRW or above), thereby reflecting the concentration of unearned income to a certain extent.

Here, the income share of the top 1% in global income is obtained from those who reported global income tax returns and are aged 20 or above, like for the concentration of total income. Accordingly, the concentration of global income shown in Fig. 7 is lower than that of total income, and the gap between the two indicates that there are quite a few high income earners among the top 1% earners who did not report global income tax returns. The graph is not provided here, but, if we change Fig. 7 to the top 0.1% income shares, the gap between total income and global income is reduced significantly, drawing the two closer. This is because most of the adult population within the top 0.1% income is subject to global income tax return reporting.

[20] The interest income of households increased significantly due to the high interest rates in the 1990s, which also includes income from investments managed by insurance companies. The 2008 SNA surveys this item as "investment income disbursements." This will be later imputed to individuals but is excluded from personal income in the relevant year. As a result of this adjustment, personal income during this period decreased as shown in Fig. 1.

Fig. 7 Income share of the top 1%: Earned income, global income, total income (unit: %)



Note: 1) Earned income is the share of income for the top 1% in earned income among total workers out of total earned income.

2) Global income is the share of income for the top 1% of the population aged 20 or above by income among those who report global income tax return.

3) Total income is the share of income for the top 1% of the population aged 20 or above by income combining the two above out of total income.

Data Source: Appendix Tables 1 and 2; the concentration of global income was calculated using global income tax data and the GPC interpolation method.

The global income shown in Fig. 7 represents the concentration of unearned income. The concentration remained stagnant during the mid-1970s and throughout the 1980s. In contrast, the concentration of earned income was declining in the same period. The concentration of unearned income rose slowly after the 1990s and temporarily decreased due to the currency crisis; since then, it has been constantly increasing at a high pace. The intensification of income inequality was accelerated up until the 2000s because the concentration of both earned and unearned income was increasing. On the other hand, the concentration of earned income began to decrease after the 2010s, whereas that of unearned income continued to rise. The concentration of total income was stagnant when the two were in a counterbalance, but the increase in the unearned income concentration has recently become faster than the decline in the earned income concentration, thereby showing signs of an increase in the concentration of total income.

Ⅳ. Conclusion

This section summarizes the findings of this study and briefly discusses the remaining challenges to be overcome in order to investigate and reveal the reality of income inequality.

Since Kim and Kim (2014), many changes have occurred in the data used to estimate top income shares in Korea. The national account statistics that was previously used to obtain total personal income changed to the 2008 SNA, and income tax data have become more substantial

since 2009. A new method, the GPC, was proposed for the interpolation of income brackets. This study updated the estimates of top income shares in light of the changes in data and method and extended the series up to 2016.

The updated top income shares show several differences from those in previous estimates. First, the top income shares became somewhat higher in the late 1990s for both earned and total income. This is mainly due to the revision of the income data, and caused the sharp increase in income inequality after the 1990s to be somewhat mitigated.

However, the trend in income inequality after the 2010s has changed. First, the concentration of earned income has been declining constantly since 2010 because income increase among the bottom 50% of workers was faster than that among the top income earners. On the other hand, the concentration of unearned income, consisting of business and financial income, continues to rise, although more slowly than before. Accordingly, the concentration of total income combining the two fell or was stagnant in the early 2010s, but turned to rise again. This is because, though the inequality of earned income is being reduced, the inequality of unearned income is growing so much that it is more than enough to cancel out the reduction.

The estimation result of this study, showing that the trend in income inequality has changed since the 2010s, raises the question of which data are more reliable. If the wages and salaries from national accounts are used as total income for the denominator in calculating the concentration of earned income, the top income shares will continue to rise and produce results contrary to this study. This is because the bias in the underestimation of the wages and salaries in national accounts has accumulated since the 2010s; as a result, this income is now 6% lower than the earned income calculated via the complete enumeration of the National Tax Service, which is unacceptable. It is necessary to increase the consistency of the top income share estimations by using the same income tax data for both the numerator and denominator. Furthermore, it is possible to expand the scope of income shares to the entire distribution, including the middle and bottom income earners in addition to the top income earners. This study made such an attempt in terms of earned income, but there are many difficulties in applying this to unearned income such as business income due to data limitations. The remaining challenge is to increase estimation consistency based on the income tax data for total income as well.

This study also raises doubts about how well wage surveys on establishments such as the Ministry of Employment and Labor's *Survey Report on Labor Conditions by Employment Type* reveal the distribution of wage income. It can be expected that lower income workers are likely to be excluded from the survey if the scope is limited to establishments that are greater than a certain size, but this is not the only problem. Sampling issues have been found when comparing the distribution by earned income bracket between this study and the survey report, such as the rapid decline in the reporting rate of the survey in the higher income brackets and the relative overestimation of middle income earners.

These issues were pointed out in household surveys such as the *Household Income and Expenditure Survey* of Statistics Korea as well (Kim and Kim 2013). Statistics Korea is attempting to supplement personal income data with administrative data including the taxation data of the National Tax Service in reorganizing the *Survey of Household Finances and Living Conditions*. The household surveys, establishment surveys, and income tax data all show an aspect of income distribution, but the results are inconsistent in many cases. It is important to track and understand the differences and then exploit the advantages and disadvantages of each type of data to reveal the reality of income distribution.

References

Atkinson, Anthony B. (2005), "Top Incomes in the UK over the 20th Century," *Journal of the Royal Statistical Society*, Vol. 168 No. 2, pp. 325–343.

Atkinson A., T. Piketty and E. Saez (2011), "Top Incomes in the Long Run of History", *Journal of Economic Literature*, Vol. 49 No. 1, pp. 3–71.

Bank of Korea, ECOS (http://ecos.bok.or.kr/).

------- (1982), *Han'guk ŭi kungmin sodŭk* [*National Income in Korea*].

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2017), "Generalized Pareto Curves: Theory and Applications," WID.world Working Paper 2017/3.

Feenberg, Daniel and James Poterba (1992), "Income Inequality and the Incomes of Very High Income Taxpayers: Evidence from Tax Returns", NBER Working Paper No. 4229.

Kim, Nak Nyeon (2012a), "Han'gugŭi sodŭk chipchungdo ch'uiwa kukche pigyo, 1976–2010: sodŭkse charyoe ŭihan chŏpkŭn" ["Income Concentration in Korea, 1976–2010: Evidence from Income Tax Statistics"], *Kyŏngje punsŏk* [*Economic Analysis*], Vol. 18 No. 3, pp. 75–114.

----- (2012b), "Han'gugŭi sodŭkpulp'yŏngdŭng, 1963-2010: kŭllosodŭgŭl chungsimŭro" ["Earned Income Inequality in Korea, 1963-2010"], *Kyŏngje palchŏn yŏn'gu* [*Journal of Korean Economic Development*], Vol. 18 No. 2, pp. 125–158.

----- (2014), "2013yŏn sodŭkseje kaep'yŏn'gwa kyech'ŭngbyŏl sodŭkse pudamnyul" ["Income Tax Reform in 2013 and Its Effect on Tax Burden by Income Group"], *Chaejŏngak yŏn'gu* [*Korean Journal of Public Finance*], Vol. 7 No. 2, pp. 59–93.

----- (2018), "Han'gugŭi sodŭk chipchungdo" ["Top Income Shares in Korea: Update, 1933-2016"], Naksungdae Institute of Economic Research Working Paper Series WP2018-01.

Kim, Nak-Nyeon and Jongil Kim (2013), "Han'guk sodŭk punbae chip'yoŭi chaegŏmt'o" ["Reexamining the Indices of Income Distribution in Korea"], Han'guk kyŏngjeŭi punsŏk [*Journal of Korean Economic Analysis*], Vol. 19 No. 2, pp. 1–64.

Kim, Nak Nyeon and Jongil Kim (2014), "Top Incomes in Korea, 1933–2010: Evidence from Income Tax Statistics", WID.world Working Paper 2014/2.

Kim, Nak-Nyeon, Park, Ki-joo, Park, Yi-taek and Myeong-soo Cha (2018), Han'guk ŭi changgi t'onggye [*Historical Statistics in Korea*], Haenam.

Jeong, Hyeok(2017), "Han'guk ŭi kungmin pulp'yŏngdŭng kujo ŭi ihae wa p'oyongjŏk sŏngjang chŏlyak" ["Empirical Understanding of Korea's Income Inequality Structure and Inclusive Growth Strategies"], Hwang, Su-kyeong et al., *Sodŭk punbae wa kyŏngje sŏngjang* [*Income Distribution and Economic Growth]*, National Research Council for Economics, Humanities and Social Sciences.

Ministry of Employment and Labor, *Survey Report on Labor Conditions by Employment Type* (Microdata), various years.

--------, *Report on the Occupational Labor Force Survey at Establishments*, various years.

Moriguchi, C. and E. Saez (2008), "The Evolution of Income Concentration in Japan: Evidence from Income Tax Statistics, 1886–2005", *The Review of Economics and Statistics*, Vol. 90 No. 4, pp. 713–734.

National Tax Service, *Statistical Yearbook of National Tax*, various years.

Park, Ki-joo and Nak-Nyeon Kim (2011), "Haebang chŏn (1907–1939) sobija mulga chisu ŭi ch'ugye" ["An Estimate of Consumer Price Index of Colonial Korea: 1907-1939"], *Kyŏngje punsŏk* [*Economic Analysis*], Vol. 17 No. 1, pp. 131–168.

Piketty, Thomas, and Emmanuel Saez (2003), "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, Vol. 118 No. 1, pp. 1–39.

Statistics Korea, *Economically Active Population Survey*, various years.

-----, KOSIS (http://kostat.go.kr/portal/index/statistics.action)

The World Inequality Database, (http://wid.world).

Appendix Table 1 Top income shares, threshold income, and average income: Earned income

	Top income shares (%)					Threshold income (thousand KRW)					Average income (thousand KRW)				
	P99.9–100	P99.5–100	P99–100	P95–100	P90–100	P99.9	P99.5	P99	P95	P90	P99.9–100	P99.5–100	P99–100	P95–100	P90–100
1963	1.221	3.670	5.630	15.685	26.361	27,105	12,004	10,313	6,082	5,916	34,313	20,628	15,824	8,816	7,409
1964	0.814	2.682	4.349	13.905	25.472	17,191	11,281	8,431	6,739	6,702	23,646	15,585	12,636	8,080	7,400
1965	0.843	2.656	4.292	13.205	23.671	18,339	10,998	8,683	6,182	6,134	24,740	15,592	12,600	7,753	6,949
1966	0.963	2.796	4.407	14.055	25.753	21,483	12,251	9,638	7,815	7,766	32,065	18,624	14,677	9,362	8,577
1967	0.960	2.791	4.428	13.817	24.609	23,361	14,092	10,904	8,505	7,831	36,313	21,124	16,757	10,458	9,313
1968	1.124	3.084	4.939	14.337	24.432	25,156	16,156	12,987	8,076	7,819	44,243	24,274	19,440	11,286	9,617
1969	1.007	3.006	4.865	15.020	24.401	27,557	18,005	15,384	8,615	8,005	44,369	26,488	21,437	13,236	10,751
1970	1.048	3.204	5.304	16.574	27.216	31,368	21,019	17,368	9,913	9,308	46,830	28,640	23,700	14,813	12,162
1971	1.124	3.228	5.215	16.036	25.619	34,297	20,161	16,546	10,232	8,304	52,081	29,920	24,167	14,863	11,873
1972	0.978	2.941	4.775	15.164	24.969	31,808	20,326	16,884	10,402	9,476	48,267	29,044	23,575	14,975	12,328
1973	0.918	2.728	4.371	13.602	23.991	36,998	21,823	18,280	12,621	12,483	55,430	32,929	26,381	16,418	14,480
1974	1.242	3.451	5.282	15.021	25.777	58,584	25,622	19,633	13,309	13,136	76,310	42,411	32,462	18,463	15,841
1975	1.256	4.191	6.343	16.080	25.364	54,779	34,474	22,110	11,895	11,369	77,740	51,886	39,271	19,909	15,702
1976	1.037	3.579	5.732	15.136	23.978	48,829	37,583	24,001	12,939	11,907	71,934	49,661	39,767	21,004	16,636
1977	1.060	3.230	5.289	15.350	24.358	54,992	34,711	28,532	14,401	13,493	80,437	49,007	40,116	23,287	18,477
1978	1.241	3.767	5.998	17.681	27.586	73,770	44,412	34,525	20,041	15,712	107,367	65,179	51,882	30,591	23,863
1979	1.201	3.564	5.608	16.867	27.406	77,153	46,116	35,584	21,364	19,299	113,592	67,439	53,063	31,917	25,930
1980	1.004	3.234	5.279	16.118	25.725	65,315	42,526	34,585	20,360	15,819	91,762	59,129	48,256	29,465	23,514
1981	1.104	3.341	5.290	15.957	25.728	72,376	39,634	32,600	19,778	17,027	100,675	60,917	48,224	29,094	23,455
1982	1.028	3.375	5.413	16.533	27.037	72,834	44,300	35,284	21,892	19,216	98,063	64,382	51,629	31,538	25,788
1983	1.049	3.206	5.156	15.771	25.375	73,210	47,609	38,176	23,692	19,120	112,462	68,715	55,257	33,806	27,197
1984	0.964	3.158	5.083	15.290	24.491	74,178	51,792	38,201	23,881	19,190	109,194	71,549	57,581	34,638	27,742
1985	1.083	3.230	5.188	15.729	25.093	82,582	51,739	40,459	25,140	19,275	125,141	74,672	59,970	36,360	29,004
1995	1.273	3.231	5.172	16.578	28.032	145,187	95,663	80,698	57,254	46,148	287,018	145,713	116,607	74,754	63,200
1996	1.022	3.128	5.204	16.404	27.789	160,190	111,307	90,188	58,056	52,076	245,271	150,145	124,898	78,737	66,693
1997	0.922	3.043	5.187	16.676	28.423	157,526	113,122	92,613	59,114	53,490	219,566	144,936	123,527	79,421	67,684
1998	1.163	3.225	5.293	16.239	27.498	155,260	105,501	85,618	54,236	49,689	266,690	147,859	121,343	74,460	63,043
1999	1.337	3.699	5.957	17.015	28.760	198,705	116,520	93,607	55,815	53,719	311,500	172,333	138,766	79,276	66,998
2000	1.397	3.850	6.176	17.849	29.510	210,761	120,632	104,713	58,400	55,410	338,448	186,605	149,671	86,505	71,511
2001	1.398	3.723	5.965	17.819	29.400	201,383	118,145	104,853	60,167	55,678	346,139	184,348	147,665	88,223	72,779
2002	1.430	3.880	6.157	18.276	30.011	220,232	125,446	111,732	62,818	59,143	367,645	199,503	158,279	93,970	77,154
2003	1.330	3.687	5.842	17.927	29.351	216,819	123,894	110,150	64,143	59,280	353,881	196,213	155,444	95,408	78,101
2004	1.470	3.873	6.113	19.017	31.507	214,405	137,530	110,537	75,935	57,835	398,815	210,095	165,826	103,171	85,466
2005	1.748	4.193	6.482	19.418	31.910	234,991	139,431	114,716	75,619	59,411	481,699	231,163	178,667	107,050	87,957
2006	1.886	4.468	6.872	20.413	33.729	267,810	160,749	128,664	77,018	67,990	523,944	248,272	190,945	113,433	93,711
2007	2.168	4.828	7.250	20.800	33.537	277,115	152,626	125,525	80,070	64,353	611,944	272,511	204,618	117,414	94,655
2008	2.133	4.846	7.271	20.989	33.751	285,826	150,883	124,274	80,410	63,878	594,898	270,297	202,811	117,085	94,136
2009	2.078	4.771	7.155	20.622	33.197	280,133	148,053	120,863	78,586	62,497	575,589	264,303	198,195	114,247	91,956
2010	2.138	4.971	7.442	21.279	33.882	301,104	155,582	126,403	80,466	63,014	598,414	278,296	208,347	119,137	94,850
2011	2.156	4.984	7.440	21.204	33.630	309,468	157,608	128,401	80,975	63,350	616,013	284,847	212,582	121,180	96,095
2012	2.050	4.850	7.253	20.739	32.961	312,141	160,188	129,497	82,142	64,357	605,186	286,292	214,090	122,429	97,291
2013	2.037	4.823	7.210	20.556	32.658	320,776	163,438	132,124	83,510	65,478	617,749	292,523	218,656	124,674	99,036
2014	2.021	4.792	7.152	20.343	32.279	328,891	165,254	133,405	84,193	66,007	626,266	296,978	221,619	126,077	100,026
2015	1.974	4.774	7.155	20.377	32.372	339,759	170,603	137,390	86,409	67,920	625,291	302,405	226,634	129,083	102,531
2016	1.981	4.761	7.126	20.167	32.012	348,612	175,216	141,162	87,912	69,496	649,067	311,929	233,462	132,134	104,870

Note: 1) The amount of income is converted into 2015 constant prices by CPI.

2) For example, P99.9–100 is the average income or income share of workers in the top 0.1%.

3) P99.9 is the threshold income to be included in the top 0.1%.

Appendix Table 2 Top income shares, threshold income, and average income: Total income

	Top income shares (%)					Threshold income (thousand KRW)					Average income (thousand KRW)				
	P99.9-100	P99.5-100	P99-100	P95-100	P90-100	P99.9	P99.5	P99	P95	P90	P99.9-100	P99.5-100	P99-100	P95-100	P90-100
1933	8.376	16.135	22.168			83,578	31,401	24,732			192,889	74,314	51,050		
1934	8.316	15.612	20.436			75,409	25,728	21,043			189,517	71,161	46,574		
1935	7.580	14.332	18.553			84,517	26,178	19,262			186,600	70,559	45,671		
1936	7.365	13.967	18.111			79,644	25,293	18,569			179,557	68,105	44,154		
1937	7.102	13.117	16.769			89,482	28,764	17,484			206,949	76,446	48,863		
1938	7.750	13.905	17.726			94,044	27,454	16,771			222,760	79,935	50,952		
1939	7.424	13.476	17.214			70,105	24,811	16,027			190,638	69,210	44,204		
1940	7.421	13.382	17.036			97,102	29,602	17,671			223,274	80,522	51,255		
1941	7.668	14.262	17.987			91,501	35,268	19,741			240,972	89,637	56,527		
1942	7.517	14.929	19.314			87,017	32,805	18,726			199,017	79,056	51,138		
1976	2.230	4.944	7.075	16.176	25.683	53,997	28,357	18,542	10,207	10,103	119,005	52,766	37,755	17,263	13,704
1977	2.318	5.001	7.181	16.578	25.982	56,363	29,067	21,091	10,944	10,243	130,577	56,342	40,449	18,675	14,634
1978	2.179	5.033	7.404	18.441	28.815	62,847	32,913	25,125	12,756	12,299	130,825	60,434	44,456	22,144	17,301
1979	2.043	4.773	7.148	18.608	28.078	62,463	33,835	26,416	12,930	11,411	126,606	59,144	44,287	23,058	17,396
1980	1.996	4.927	7.444	19.731	29.183	56,941	32,846	25,270	13,142	9,395	111,404	55,008	41,553	22,027	16,290
1981	1.881	4.698	7.144	19.243	29.306	61,367	32,369	25,699	13,343	11,101	109,523	54,718	41,609	22,414	17,068
1982	1.788	4.738	7.272	20.055	29.884	64,830	35,220	27,625	14,745	10,419	108,964	57,742	44,312	24,442	18,211
1983	1.838	4.866	7.469	20.440	30.198	67,449	38,378	29,781	15,591	10,191	117,277	62,084	47,647	26,078	19,264
1984	1.859	4.839	7.363	19.761	29.894	75,607	41,857	30,930	16,203	12,971	127,592	66,437	50,543	27,129	20,520
1985	1.847	4.694	7.148	19.400	29.236	73,654	40,297	31,472	16,759	12,920	132,165	67,194	51,157	27,768	20,924
1995	1.975	4.730	7.234	20.554	31.825	134,387	75,351	60,471	37,177	25,223	265,501	127,157	97,242	55,256	42,779
1996	2.147	5.101	7.811	21.217	35.030	160,021	89,716	68,109	42,631	36,895	309,696	147,133	112,663	61,203	50,524
1997	2.209	5.199	8.026	21.690	35.683	160,065	91,128	72,042	43,155	37,495	319,680	150,465	116,140	62,774	51,634
1998	1.822	4.564	7.080	20.109	33.817	141,264	81,296	61,157	40,801	35,484	253,816	127,158	98,625	56,020	47,105
1999	2.144	5.140	7.828	21.079	35.150	151,169	90,020	66,746	43,514	37,599	310,072	148,664	113,204	60,962	50,829
2000	2.347	5.459	8.295	22.107	35.963	183,553	94,844	78,755	45,662	39,569	362,015	168,428	127,949	68,202	55,476
2001	2.564	5.787	8.647	23.076	36.805	197,386	94,900	82,113	45,976	39,050	396,112	178,827	133,609	71,309	56,867
2002	2.743	6.174	9.086	24.144	37.751	206,431	100,255	86,407	46,891	38,855	429,211	193,253	142,189	75,572	59,081
2003	2.812	6.263	9.142	24.637	38.385	204,562	100,829	86,065	47,993	39,128	441,306	196,588	143,478	77,334	60,244
2004	3.055	6.513	9.523	25.428	38.561	195,414	111,490	86,155	53,058	35,377	490,572	209,144	152,907	81,655	61,913
2005	3.255	6.686	9.600	25.116	37.099	234,125	118,841	92,916	48,713	35,490	544,501	223,644	160,572	84,018	62,051
2006	3.589	7.381	10.637	28.067	41.828	261,679	125,945	97,654	62,839	43,853	630,726	259,394	186,923	98,644	73,504
2007	3.996	7.985	11.292	28.458	42.352	283,212	135,503	106,201	61,330	41,540	715,177	285,851	202,124	101,875	75,805
2008	4.006	8.049	11.385	28.798	42.874	289,162	136,312	107,351	62,353	41,357	714,742	287,193	203,110	102,753	76,490
2009	3.958	8.012	11.313	28.522	42.474	286,134	133,526	104,515	60,789	39,851	695,428	281,521	198,772	100,226	74,626
2010	4.175	8.355	11.745	29.074	43.032	306,488	140,143	109,186	61,407	40,660	750,188	300,224	211,015	104,470	77,314
2011	4.340	8.620	12.018	29.246	43.128	338,074	146,547	114,841	63,350	42,383	816,052	324,116	225,948	109,970	81,085
2012	4.179	8.394	11.740	28.851	42.742	340,099	149,031	117,492	64,952	44,128	812,794	326,474	228,304	112,215	83,121
2013	4.082	8.300	11.632	28.649	42.545	350,643	153,982	120,906	66,809	46,079	821,499	334,047	234,076	115,305	85,618
2014	4.128	8.430	11.779	28.831	42.695	368,975	159,295	123,936	68,645	46,970	849,762	347,108	242,480	118,708	87,895
2015	4.291	8.685	12.068	29.173	43.127	387,073	164,241	126,196	70,237	48,384	896,078	362,754	252,037	121,850	90,066
2016	4.306	8.766	12.164	29.271	43.317	397,014	167,332	127,198	71,291	49,301	906,220	368,976	256,002	123,208	91,165

Note: 1) Figures in this table use the population aged 20 years and above as a control total, while figures in Appendix Table 1 use number of employees as a control total.

2) The amount of income is converted into 2015 constant prices by the GDP deflator, following the WID. It is different from the figures in Appendix Table 2 of Korean version of this paper which used CPI to deflate income.