

Research in Brief

Building D, 370 Sicheong-daero, Sejong City 30147 KOREA **Korea Institute for Health & Social Affairs**

Issue No 2020-18

Publication Date October 14 2020

ISSN 2092-7117

Differences in Income Distribution Estimates Due to Using Administrative Records

Won-jin Lee

Associate Research Fellow, KIHASA



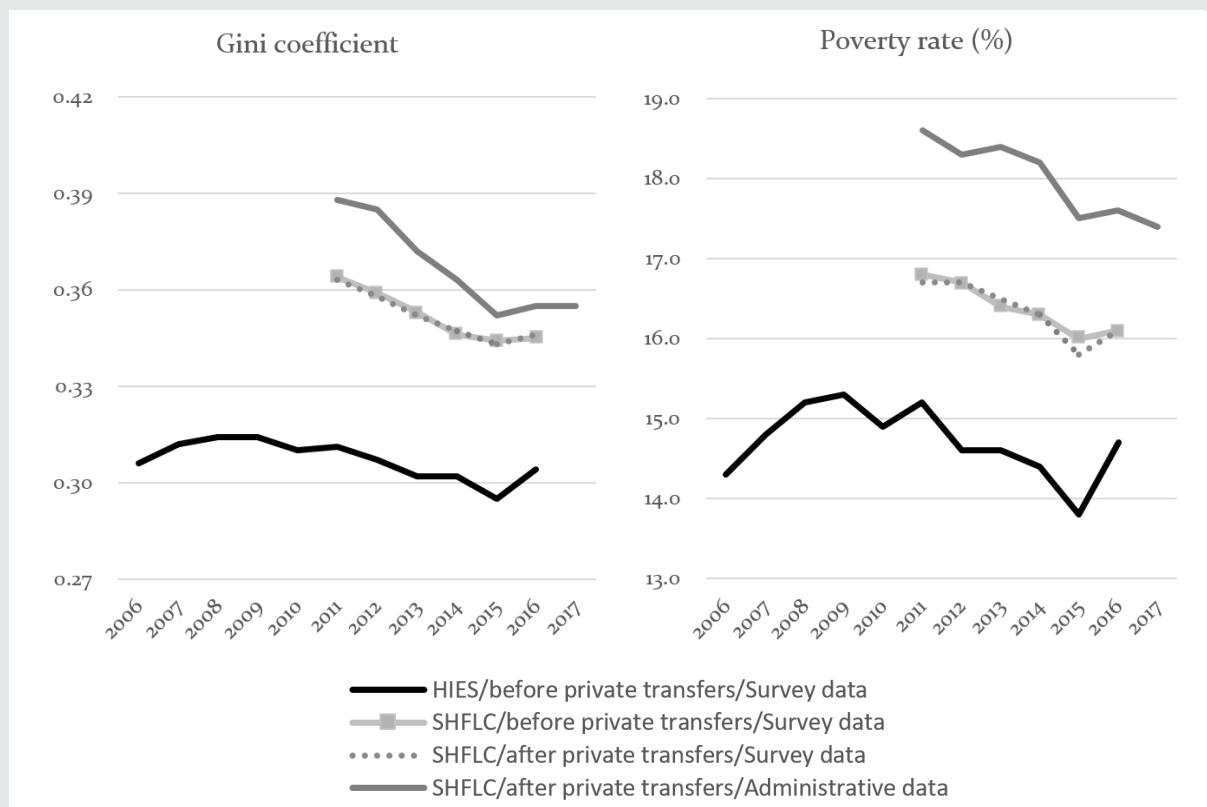
Introduction

Statistics Korea (KOSTAT) based the production of its income distribution measures on the Household Income and Expenditure Survey (HIES) until before 2016. HIES had been persistently criticized for having omitted high-income households while including cases of underreported income, thereby underestimating the actual state of income inequality. Considerable dispute was occasioned as the disposable income Gini turned out to be higher when calculated based on the Survey of Household Finances and Living Conditions (SHFLC), which was first conducted in 2012, than on HIES. In December 2017, KOSTAT began producing its income distribution measures based on the SHFLC data combined with administrative records. In what follows, income data drawn from SHFLC will be referred to as “survey data”, and “survey data supplemented with administrative information” as “combined data.”

The differences between HIES-based estimates and the current, SHFLC-based income distribution estimates are attributed to three things, as illustrated in Figure 1. Firstly, the HIES data are different from the SHFLC data. In 2016, the Gini coefficient and poverty rate for disposable income (before private transfers) were 0.345 and 16.1 percent, respectively, in the SHFLC data, compared to 0.304 and

14.7 percent in the HIES data. The concept of disposable income differs between the two surveys. HIES defines disposable income as current gross income minus tax and social insurance contributions. Disposable income in SHFLC means HIES disposable income after private transfers. The difference in disposable income is found to make little difference in income distribution indicators (that is, before data combination). Thirdly, a difference exists also between the SHFLC data (survey data) and the SHFLC data combined with administrative data (combined data). After data combination, the Gini coefficient increased from 0.346 to 0.355, and the poverty rate from 16.1 percent to 17.6 percent. Thus, the differences between past and current income distribution indicators were due in the main to the shift in data source and the use of administrative data.

[Figure 1] Trends in the distribution of disposable income (2006~2017)



Note: The distribution of equivalised income across individuals. Equivalised income is household income divided by the square root of the number of household members. Poverty rate is based on the cutoff line of 50 percent of the population-wide median disposable income.

Source: Statistics Korea. Survey of Household Finances and Living Conditions for years 2012~2017
 Statistics Korea. Income Distribution Indicators based on the Survey of Household Finances and Living Conditions. <http://kosis.kr>
 Statistics Korea. Income Distributions Indicators based on the Household Income and Expenditure Survey. <http://kosis.kr>



Combining SHFCL data with administrative data

Statistics Korea combines, at both the household and individual levels, its survey data with administrative data on income and non-consumption expenditures that are generated by national authorities such as the National Tax Service, the Ministry of Health and Welfare and various public pension corporations. In principle, administrative data are used in place of their conceptually compatible survey items. In detail, however, separate rules are employed for different income items, as described in Table 1.

[Table 1] Rules employed in supplementing SHFCL data with administrative data

	Rules on supplementing SHFCL data with administrative data
Earned income	<ul style="list-style-type: none"> - (Basic rule) Replace or supplement with administrative data - (Exceptions) If the amount identified in the administrative data is less than the non-taxable ceiling, take whichever is larger from either the survey data or the administrative data from the National Tax Service; in cases where an overlap is assumed, as in the case of households whose administrative data have earnings information but no business income records or whose survey data have business income records but no earnings information, use survey data to ensure the accuracy of household unit income estimates.
Business income; rental income	<ul style="list-style-type: none"> - (Basic rule) In principle, use survey data; for unreported cases, use administrative records - (Point of consideration) In cases where, in the same household, the nominee is not the person who runs the business, make corrections at the household level
Financial income	<ul style="list-style-type: none"> - (Basic rule) Replace survey data overall with administrative data - (Exceptions) In the case of the amount lent out, use whichever is larger, from either datasets
Public transfer income	<ul style="list-style-type: none"> - (Basic rule) Replace survey data overall with administrative data - (Relevant items) public pension, basic pension, disability benefits, basic living security benefits, in-work and child benefits, childcare allowance (including childbirth promotion allowance) - (Exceptions) For childcare allowance, use the amount whichever is larger, from either survey data or administrative data
Non-consumption transfers	<ul style="list-style-type: none"> - (Basic rule) Replace survey data overall with administrative data - (Relevant items) Earned income tax, public pension contributions

Source: Statistics Korea. Report on the Household Finance and Consumption Survey for 2018 (<http://kosis.kr>). p. 34



The impact of administrative records on income distribution measures

In our comparison below of the survey and combined data, market income is defined as the sum of earned income, business income and private transfer income less private transfers (any negative value is taken as zero). Disposable income is defined as market income less public transfer expenditures (a negative value is treated as zero).

[Table 2] The distribution of market and disposable income

	Market income				Disposable income				
	Survey data (A)	Combined Data (B)	C=B-A	C/A (%)	Survey data (A)	Combined data (B)	C=B-A	C/A (%)	
Average (in KRW 10 thousand/year)	2915	3195	+2.80	+9.6	2723	2981	+2.58	+9.5	
Income decile upper bounds	P10	624	696	+0.74	+11.6	882	922	+0.40	+4.5
	P20	1270	1286	+0.16	+1.2	1316	1369	+0.53	+4.0
	P30	1700	1772	+0.72	+4.3	1684	1783	+1.00	+5.9
	P40	2079	2214	+1.35	+6.5	2007	2155	+1.48	+7.4
	P50	2477	2652	+1.75	+7.1	2336	2550	+2.14	+9.1
	P60	2900	3165	+2.65	+9.1	2689	2971	+2.82	+10.5
	P70	3464	3745	+2.81	+8.1	3173	3485	+3.12	+9.8
	P80	4158	4568	+4.10	+9.9	3766	4175	+4.09	+10.9
	P90	5400	5911	+5.11	+9.5	4742	5281	+5.40	+11.4
Gini coefficient	0.390	0.402	+0.012	+3.2	0.346	0.355	+0.009	+2.7	
Poverty rate (%), %p)	17.9	19.8	+1.9	+10.7	16.1	17.6	+1.5	+9.1	

Note: The distribution of equivalised income across individuals. Equivalised income is household income divided by the square root of the number of household members. Both market income poverty rate and disposable income poverty rate are based on the cutoff line of 50 percent of the population-wide median disposable income.

Source: Statistics Korea. Survey of Household Finances and Living Conditions for 2017

Market income and disposable income were higher by 9.6 percent and 9.5 percent, respectively, in the combined data, where income inequality was also higher than in the survey data. Disposable income was higher by 4.5 percent for the bottom decile (P10), and by on the order of 9.1~11.4 percent for the upper five deciles. The disposable income Gini was 0.355, compared to 0.346 in the survey data. The poverty rate was 17.6 percent, compared to 16.1 percent in the survey data. As is the case with disposable income, the increase in market income is more pronounced in deciles 6 to 9 than in deciles 2 to 5. However, the sharpest increase in market income (11.6 percent) was observed in the first decile, unlike the case of disposable income. It is likely that some of those in the bottom income bracket who in the survey reported having no market income were consistently identified through administrative data as having a certain market income. The market income Gini increased from 0.390 to 0.402, and the market income poverty rate from 17.9 percent to 19.8 percent.

Table 3 shows differences in income amounts between the survey and combined data linked to personal identifiers. Earned income estimates were lower in the combined data than in the survey data for 24.8 percent of respondent households, but higher for 41.1 percent. The average amount of earned income was KRW 34.99 million in the combined data, higher than KRW 32.81 million in the survey data, as administrative records reflected the earnings of some (5.4 percent) of the respondents who had reported no earnings. The amount of business income remained unchanged for most of those households, except for the 11.3 percent newly identified in the combined data as having a certain level of

business income. The average property income was KRW 3.53 million in the combined data, compared to KRW 2.12 million in the survey data. Those whose financial income went unreported in the survey, but was identified through administrative data, accounted for as much as 72.9 percent of the households. The average amount of financial income was KRW 1.54 million in the combined data, while it was KRW 0.29 million in the survey data.

After correcting for the data on the 9.4 percent whose benefit income was unreported, an increase in the amount of public transfers was observed for 35.1 percent of the surveyed households. The average transfer income was higher (KRW 3.28 million) in the combined data than in the survey data (KRW 2.97 million). The category with the highest rate of undeclared public transfer income detected was “in-work and child” benefit. The average amount of in-work and child benefits received was larger by KRW 80 thousand in the combined data. The impact of combining the two data sources was substantial also on public pension estimates. The average amount per household of public pension was larger by KRW 160 thousand in the combined data. The impact was relatively insignificant on the other categories of public transfer.

[Table 3] The impact of using administrative data on income estimates, by income source

	Average				From 0 with survey data to greater than 0 with combined data (%)	% of those with an increase data (B)	% of those with a decrease	% of those with the same amount
	Survey data	Combined Data	C=B-A	C/A (%)				
Earned income	3281	3499	+218	+6.6	5.4	41.1	24.8	34.1
Business income	1149	1217	+68	+5.9	11.3	14.1	0.0	85.9
Property income	212	353	+142	+66.9	62.0	82.0	3.1	14.8
Financial income	29	154	+125	+426.9	72.9	82.0	3.2	14.8
Rental income	170	187	+17	+9.9	0.8	1.1	0.0	98.9
Public transfer income	292	328	+36	+12.2	9.4	35.1	9.9	55.1
Public pension	176	192	+16	+9.2	4.8	16.6	6.7	76.6
Basic pension	55	55	-0	-0.0	1.2	11.1	2.8	86.1
Child allowance	9	13	+4	+41.7	2.9	4.6	0.0	95.4
Disability allowance	7	5	-2	-29.2	0.5	1.6	1.3	97.1
Basic living security	20	24	+4	+22.7	0.9	3.2	1.4	95.4
In-work and child	2	10	+8	+359.8	9.0	10.2	0.7	89.1
Public transfer expenditure	547	614	+67	+12.3	0.2	42.2	40.6	17.3
Income tax	176	246	+70	+39.5	15.6	41.6	37.5	20.9
National pension contributions	127	120	-8	-6.0	5.6	28.1	36.8	35.1
Other pension contributions	27	32	+5	+19.1	1.0	4.7	3.2	92.0

Source: Statistics Korea. Survey of Household Finances and Living Conditions for 2017

Of the sources of income, earned income was the largest and one most affected by the use of administrative data. Table 4 shows the extent to which earned income estimates are affected by the use of administrative data. The proportion of households with no earned income was 24.5 percent in the combined data, 5.4 points lower than the 30.0 percent in the survey data. The proportion of households with an earned income of less than KRW 10 million and those in the income bracket between KRW 10 million and KRW 20 million were larger in the combined data by 7.6 percentage points and 2.4 percentage points, respectively, as many of the households who in the survey reported having no earned income were identified through administrative data as having some. The proportion of households in the income brackets of above KRW 10 million and below KRW 80 million was lower in the combined data, but the proportion of those with an income of KRW 80 million or more was higher. A similar pattern was observed in earnings at the individual level. In the combined data, the proportion of individuals with no earned income was lower by 8.5 percentage points, while the share of those whose income was more than zero but less than KRW 10 million was larger by 7.7 percent points. The share of individuals in the income brackets of above KRW 20 million and below KRW 60 million was smaller in the combined data than in the survey data, but the percentage of individuals with an earned income of more than KRW 60 million was larger.

[Table 4] Differences in earned income estimates between the survey data and the combined data for households and individuals

	Household earned income			Individual earned income		
	Survey data (A)	Combined data (B)	B-A	Survey data (A)	Combined data (B)	B-A
No earned income	30.0	24.5	-5.4	63.5	55.0	-8.5
~ less than KRW 10m	5.9	9.9	+4.1	4.9	12.5	+7.6
~ less than KRW 20m	10.1	12.3	+2.2	8.8	11.2	+2.4
~ less than KRW 40m	19.9	19.0	-0.8	12.7	11.3	-1.4
~ less than KRW 60m	14.3	13.2	-1.1	5.3	4.9	-0.4
~ less than KRW 80m	8.9	8.8	-0.1	2.6	2.8	+0.1
~ less than KRW 100m	5.0	5.2	+0.1	1.1	1.2	+0.2
~ less than KRW 200m	5.6	6.6	+0.9	0.9	1.1	+0.2
More than KRW 200m	0.3	0.5	+0.2	0.1	0.1	+0.0
Total	100.0	100.0		100.0	100.0	

Note: Household here means the distribution of non-equivalised earned income across households; individual earned income means the distribution of non-equivalised earned income across individuals.

Source: Statistics Korea. Survey of Household Finances and Living Conditions for 2017

Table 5 displays a hypothetical distribution of disposable income, with the estimates of various disposable income components derived from linking the survey data with administrative records

through personal identifiers.

The combining of the survey data and administrative records is found to result in a more unequal income distribution. Annual disposable income per household was KRW 28.73 million in the combined data, higher than KRW 27.23 million in the survey data. The Gini coefficient was 0.346 in the survey data and 0.367 in the combined data. The poverty rate was 16.1 percent in the survey data and 18.9 percent in the combined data. Business income and property income on average were larger by KRW 440 thousand and KRW 870 thousand, respectively, in the combined data. There was little if any change in the disposable income Gini. In the combined data relative to the survey data, the disposable income of those in the bottom two deciles was considerably higher, with the Gini coefficient and poverty rate lower respectively by 0.003 and 0.7 percentage points. The level of disposable income inequality was lower in the combined data than in the survey data, but not to a significant extent.

[Table 5] Differences in the distribution of disposable income between the survey data and the combined data, by income decile, by income source

		Disposable income in the survey data	Disposable income in the combined data					
			Earned income	Business income	Property income	Public transfer income	Public transfer expenditure	Total
Average (in KRW 10 thousand/year)		2723	2873	2767	2811	2742	2686	2981
			+149	+44	+87	+19	-38	
Income decile upper bounds	P10	882	820	897	923	915	853	922
			-62	+15	+40	+33	-29	+40
	P20	1316	1272	1338	1359	1342	1294	1369
			-44	+21	+42	+25	-22	+53
	P30	1684	1678	1708	1726	1700	1664	1783
			-6	+24	+43	+16	-20	+100
	P40	2007	2054	2035	2051	2023	1987	2155
			+47	+28	+44	+16	-20	+148
	P50	2336	2439	2366	2395	2703	2671	2971
			+103	+30	+59	+14	-18	+282
	P60	2689	2855	2731	2748	2703	2671	2971
			+166	+42	+59	+14	-18	+282
	P70	3173	3364	3207	3237	3178	3152	3485
			+191	+35	+65	+6	-21	+312
	P80	3766	4067	3812	3847	3776	3743	4175
			+300	+45	+81	+9	-24	+409
	P90	4742	5163	4819	4888	4768	4734	5281
			+421	+77	+146	+27	-8	+540
Gini coefficient		0.346	0.367	0.347	0.348	0.342	0.344	0.355
			+0.021	+0.001	+0.002	-0.003	-0.001	+0.009
Poverty rate (% , %p)		16.1	18.9	16.0	15.8	15.4	16.5	17.6
			+2.8	-0.1	-0.3	-0.7	+0.4	+1.5

Note: The distribution of equivalised income across individuals. Equivalised income is household income divided by the square root of the number of household members. Poverty rate is based on the cutoff line of 50 percent of the population-wide median disposable income.

Source: Statistics Korea. Survey of Household Finances and Living Conditions for 2017

Table 6 shows between-dataset differences in income distribution estimates for different age groups. The impact of administrative data use on disposable income distribution varied across the age groups. The disposable income Gini coefficient for those 17 and younger was higher by 0.032 in the combined data than in the survey data. The Gini for working-age people (aged 18 to 64) was also higher, by 0.021, than in the survey data. For those 65 and older, the difference in the Gini coefficient was as small as 0.007. The poverty rates for the two younger groups were higher in the combined data, by 5.6 percentage points and 2.6 percentage points, respectively. The poverty rate for people 65 and older was lower by 0.4 percentage points in the combined data.

The estimates of public transfer income increased with the use of administrative data, reducing both inequality and poverty estimates for all age groups. Those 65 and older were the age group that saw the most increase in the mean and median disposable income after data combination. The extents to which the Gini and poverty estimates declined due to data combination were similar across all age groups. The use of administrative data increased the Gini and poverty rate by 0.019 and 4.4 percentage points, respectively, for children aged 17 and younger, and by 0.011 and 1.6 percentage points for those aged 18 to 64. The poverty rate for those 65 and older declined by 2.9 percentage points. This suggests that the survey may have underestimated the Gini coefficients and poverty rates for children aged 17 and younger and the working-age population, while overestimating the poverty rate for those 65 and older.

[Table 5] Differences in the distribution of disposable income between the survey data and the combined data, by income decile, by income source

		Indicators of disposable income distribution in survey data	Difference in disposable income distribution indicators for individuals, between survey data and combined data					
			Earned income in combined data	Business income in CD	Property income in CD	Public transfer income in CD	Public transfer expenditure in CD	Total in CD
Children 0~17 years of age	Mean (in KRW 10,000/year)	2706	+98	+48	+62	+19	-33	+188
	Median (in KRW 10,000/year)	2311	+105	+39	+30	+9	+7	+213
	Gini coefficient	0.31	+0.032	+0.002	+0.005	-0.004	-0.003	+0.019
	Poverty rate (%, %p)	10.8	+5.6	-0.1	+0.4	-0.9	+0.2	+4.4
Individuals aged 18~64	Mean (in KRW 10,000/year)	2943	+171	+46	+86	+16	-41	+274
	Median (in KRW 10,000/year)	2526	+133	+31	+55	+11	-10	+240
	Gini coefficient	0.327	+0.021	+0.001	+0.004	-0.003	-0.001	+0.011
	Poverty rate (%, %p)	11.2	+2.6	-0.2	-0.1	-0.6	+0.5	+1.6
Individuals aged 65 and older	Mean (in KRW 10,000/year)	1723	+120	+32	+125	+33	-30	+277
	Median (in KRW 10,000/year)	1263	+79	+13	+89	+42	-21	+209
	Gini coefficient	0.425	+0.007	+0.004	-0.001	-0.004	-0.002	-0.003
	Poverty rate (%, %p)	46.6	-0.4	-0.1	-1.9	-0.8	+0.2	-2.9

Note: The distribution of equivalised income across individuals. Equivalised income is household income divided by the square root of the number of household members. Poverty rate is based on the cutoff line of 50 percent of the population-wide median disposable income.

Source: Statistics Korea. Survey of Household Finances and Living Conditions for 2017



Concluding remarks

As household disposable income and its Gini coefficient were both higher in the combined data than in the survey data, the supplementation of survey data with administrative data was found to reduce to some extent the problem income underreporting among those in the higher end of the income distribution. The impact of data supplementation was significant on the distribution earned income. In the combined data, there was a substantial rise in the proportion of those with higher earned income, with a share of aggregate earned income newly captured by administrative records. Administrative records uncovered instances of unreported business income and accordingly increased household financial income markedly. The proportion of households in receipt of in-work and child benefits and the average amount of public pension turned to be higher in the combined data than in the survey data, although not to the extent shown in the case of primary income sources. In the combined data, income distribution improved for those 65 and older more than those younger.

The findings suggest that the sample design may not have been properly representative of the population. Such a problem may arise from various stochastic and non-stochastic errors in the process of sample selection, or as a result of selective nonresponse among those with very high or low income. Also, the income estimates as computed based on the survey data may have been affected by underreporting of income among those surveyed.

The supplementation of SHFLC with administrative records does little to resolve the problem of sampling error arising from disproportionate omission of high-income earners, but it can correct to some extent the problem of income underreporting. Policymakers may consider further use of administrative data on transfer receipt. A proper assessment of the impact of social security programs targeting the working-age population will require increased use of administrative data, especially on unemployment insurance benefit receipt. Challenges remain regarding the combining of survey data with administrative records on business income and rental income and also on various taxes including health insurance payments, social security contributions and unemployment insurance contributions.