Naksungdae Institute of Economic Research

Working Paper Series

가계조사의 행정자료에 의한 보정:
2016년 가계금융복지조사를 중심으로

김낙년

Working Paper 2020-02

Jan, 2020

Naksungdae Institute of Economic Research

B1204, 164, Gwanak-ro, Gwanak-gu,

Seoul, 08788, South Korea

가계조사의 행정자료에 의한 보정:
2016 년 가계금융복지조사를 중심으로

김낙년*

<요약>

본고는 2016 년 가계금융복지조사를 대상으로 조사 결과와 행정자료로 보정된 결과를 비교하여 피조사자가 자신의 소득과 자산을 응답할 때 어떠한 편향이 있는지를 보였다. 첫째, 근로소득의 경우 소득이 높을수록 과소 보고하는 경향이 커졌다. 금융소득의 경우 아예 응답하지 않거나 응답한 경우에도 소득을 크게 줄여서 보고하였다. 그런데 이와 반대로 소득이 낮을수록 응답한 소득이 행정자료로 파악된 소득보다 커지는 경향이 나타났고, 양자의 괴리가 매우 큰 경우도 적지 않았다. 소득의 개념이나 분류에 관한 응답자의 오해나 혼동이 하나의 요인으로 작용한 것으로 보인다. 둘째, 행정자료를 이용함으로써 비 샘플링 오류는 대부분 제거되었지만, 그 결과를 다시 과세 자료와 비교해 보면 특히 금융소득의 경우 최상층에서 누락이 많았다. 이것은 가계조사가 실제의 분포를 제대로 반영하지 못하는 샘플링 오류라고 할 수 있다. 셋째, 자산의 경우 행정자료를 이용한 보정이 이루어지지 않아 조사결과를 직접 과세 자료와 비교하였다. 주택의 경우 두 자료의 자산 분포가 근접한 것으로 나타났다. 자산 편중이 더 심한 토지의 경우에는 최상층의 일부가 누락되었고, 금융자산의 경우 이 문제가 더욱 심각한 것으로 드러났다.

핵심 주제어: 소득, 자산, 샘플링 오류, 비 샘플링 오류, 가계금융복지조사

* 동국대학교(서울캠퍼스) 경제학과 교수(nnkim@dongguk.edu)

1. 머리말

가계조사와 같은 표본조사는 경제사회의 제 현상을 분석하는데 이미 불가결한 자료가 되었으며, 마이크로 데이터를 이용할 수 있다는 장점이 더해져서 수많은 연구에 널리 활용되고 있다. 그렇지만 거기에서 조사된 소득이나 자산이 실태를 얼마나 잘 반영하고 있을까라는 의문은 여전히 중요한 문제로 남아있다. 예컨대 통계청의 가계동향조사는 소득분배 실태를 보여주는 대표적인 자료로서 오랫동안 널리 이용되어 왔지만, 거기에서 파악된 소득이 특히 상층의 소득을 제대로 반영하지 못하고 있다는 문제가 제기되었다. 김낙년·김종일(2013)은 이 조사가 소득이 높아질수록 소득 파악률이 급속히 떨어지며 연소득이 2 억원을 넘는 샘플은 아예 존재하지 않는다든지, 이 조사에서 파악된 금융소득은 전체 금융소득의 5% 정도에 불과하다는 점들을 밝혀내었다. 최상층의 소득이 주로 금융소득으로 이루어져 있다는 점을 생각하면 이 조사는 소득격차를 심각하게 과소 평가한 것이 된다.

가계조사가 소득이나 자산을 파악하는데 한계가 많다는 점은 비단 우리나라만의 문제는 아니다. 정도의 차이는 있지만, 다른 나라에서도 크게 다르지 않다. 피조사자가 자신의 소득이나 자산을 드러내는 것을 꺼릴 수도 있으며, 예컨대 금융소득이나 자산과 같이 본인도 그 규모가 정확히 얼마인지 알기 어려운 경우도 있다. 가계동향조사의 경우 표본으로 선정된 가구가 매우 상세한 조사항목의 가계부를 작성하는 방식으로 이루어졌는데, 그만큼 피조사자의 부담이 커서 성실한 응답을 기대하기 어려운 점도 있다. 무응답 가구도 점차 크게 늘어났다. 통계청이 가계동향조사를 중단하기로 하고[1], 소득 등의 조사를 행정자료를 이용하여 보정하는 방식으로 이행하게 된 것은 그 때문이다. 가계금융복지조사가 이를 대체하였는데, 이 조사는 2011 년부터 시작되었으며 2016 년부터 조사결과를 행정자료로 보정한 결과를 발표하고 있다. 종래의 가계동향조사에 의거해 작성되었던 소득분배 지표도 이제는 가계금융복지조사에 의거한 것으로 변경되었다.

가계금융복지조사에서 행정자료에 의한 보정이란 다음과 같은 방식으로 이루어지고 있다. 먼저 통계청이 조사가구 구성원의 개인 정보[2] (이름 주소 등)를 국세청과 같이 소득이나 비소비지출(세금이나 사회보험 부담금 등)의 정보를 갖고 있는 행정기관에 보내면, 그 행정기관이 가지고 있는 정보와 매칭해서 해당 개인의 소득 정보 등을 회신해주고, 통계청은 이에 의거하여 조사결과를 보정하는 것이다. 행정 기관이 파악한 소득 정보가 조사결과보다 더 신뢰할 수 있는 경우에는 이 방식으로 가계조사의 정확성이 높아질 것으로 기대할 수 있다. 다만 자산의 경우에는 자료상의 제약으로 인해 현재 행정자료에 의한 보정이 이루지지 못했다.

그러면 행정자료로 보정된 가계금융복지조사는 기존의 조사결과에 비해 소득 등의 파악이 얼마나 개선되었을까? 통계청은 2016 년의 가계금융복지조사의 경우 행정자료로 보정되기 전의

---

[1] 가계동향조사를 중단하기로 하였지만, 문재인 정부에 들어와 소득분배 등의 분기별 통계가 필요하다는 이유로 부활하였다. 그렇지만 가계동향조사에 대해 제기되었던 소득의 과소 파악의 문제는 여전히 남아 있다.

[2] 개인정보호법의 시행으로 주민등록번호를 이용할 수 없어 개인을 식별할 수 있는 이름과 주소 등의 정보를 이용하고 있다. 그로 인해 매칭이 제대로 되지 않는 경우도 있지만, 그것이 큰 문제가 될 정도는 아니다.

조사결과를 제공하고 있으며, 이것을 보정 결과와 비교해 보면 피조사자가 소득을 얼마나 과소 또는 과대 보고하였는지를 드러낼 수 있다. 현재 대부분의 가계조사는 조사결과를 그대로 이용하고 있는데, 위 결과는 소득 조사에 대해 개인이 어떻게 응답하는가를 구체적으로 보여줌으로써 기존 가계조사의 소득 파악에 어떠한 문제가 있는지에 대해 많은 시사를 줄 수 있다.

나아가 행정자료로 보정된 결과는 과연 실제의 소득분포를 정확히 반영한다고 볼 수 있을까? 이를 위해서는 소득 분포에 관한 신뢰할 수 있는 정보가 필요한데, 근로소득과 금융소득의 경우 그러한 검토가 가능하다. 근로소득의 경우 급여를 지급하는 고용주가 소득의 신고 의무를 가지며 이를 통해 지급한 급여를 비용으로 인정받을 수 있기 때문에 신고를 누락할 유인이 없다. 따라서 국세청의 근로소득 연말정산 자료와 일용근로소득 자료를 이용하면 근로소득의 전모를 파악할 수 있다. 금융소득의 경우 금융기관을 통해 개인의 소득이 모두 파악되며, 금융소득종합과세 자료를 이용하면 금융소득이 2000 만원 이상의 과세 대상자의 소득을 모두 파악할 수 있다. 이에 대해 본인의 신고에 의거하여 과세되는 사업소득의 경우는 과소 보고의 유인이 크며, 국세청이 파악한 사업소득이 실태를 정확하게 반영한다고 보기 어렵다. 본고에서 사업소득이나 부동산 임대소득을 검토하지 못한 것은 그 때문이다.

한편, 자산의 경우는 행정자료에 의한 보정이 이루어지지 못했지만, 조사결과가 얼마나 실태에 근접한지를 검토해 볼 수 있다. 주택이나 토지의 경우는 재산세 부과 자료에서 공시가격에 의한 그 분포 정보를 얻을 수 있기 때문에 이를 이용해서 이 문제를 검토할 수 있다. 금융자산의 경우도 그 분포를 직접 보여주는 자료는 구할 수 없지만, 전술한 행정자료로 보정된 가계금융복지조사와 금융소득종합과세 자료를 이용하면 실태에 가까운 금융소득의 분포를 알 수 있다. 여기에 후술하는 소득 자본화 방법(income capitalization method)을 적용하면 실태에 가까운 금융자산의 분포를 구할 수 있다.

본고는 다음과 같이 구성되어 있다. 2 절에서는 가계금융복지조사의 근로소득과 금융소득을 대상으로 하여 조사결과와 행정자료에 의한 보정 결과를 비교한다. 이를 통해 가계조사에서 소득에 관한 설문에 대해 피조사자가 어떻게 응답하는지를 드러내 보이고자 한다. 나아가 이를 과세자료로 파악된 소득분포와 비교하여 실제의 소득분포에 얼마나 근접하였는지를 살펴본다. 만약 그렇지 못하다고 한다면 그것은 무엇 때문인지도 검토한다. 3 절에서는 가계금융복지조사의 자산 중에서 주택 및 토지, 금융자산을 대상으로 하여 그것이 얼마나 실제의 자산 분포에 접근하였는지를 검토한다. 4 절에서는 이상에서 드러난 사실에 의거하여 가계조사의 행정자료에 의한 보정이 갖는 의의와 한계를 논한다.

2. 소득

1) 근로소득

가계금융복지조사는 약 2 만개의 샘플 가구를 대상으로 매년 4 월에 가구의 자산과 부채, 소득 등을 조사한다. 자산과 부채는 3 월 31 일 기준으로 조사하며, 소득의 경우 소득이 있는 모든

가구원을 대상으로 지난 해(1 월 1 일~12 월 31 일)의 연간 소득을 조사한다[3]. 이 때 예금통장이나 원천징수영수증(연말정산), 종합소득세신고서와 같은 자료를 참조해서 정확한 수치를 기입하도록 안내하고 있다. 만약 모든 조사 가구가 이 안내에 따랐다면 보다 정확한 소득이 파악될 수 있었겠지만, 후술하듯이 많은 가구가 실제 그렇게 한 것으로 보기는 어렵다.

통계청의 마이크로 데이터 서비스(MDIS)에서 다운로드 받을 수 있는 가계금융복지조사는 가구 단위로 되어 있기 때문에 개인의 소득을 알 수 없다. 개인별 소득에 관해서는 MDIS 의 원격접근 서비스로 접근할 수 있는 "가구원" 파일을 이용하였다. 여기서는 조사결과와 행정자료로 보정된 결과를 모두 이용할 수 있는 2016 년을 대상으로 분석하기로 한다. 이들 두 마이크로 데이터를 가구원 개인의 ID 로 통합(merge)하면, 개인별로 근로소득의 조사결과와 보정결과를 매칭시켜 양자의 비율을 구할 수 있다. 이 비율(=조사/보정)이 1 이라면 양자가 일치한 것을, 1 보다 작다면 조사결과가 실제(보정된 결과)보다 과소 보고된 것을, 1 보다 크다면 반대로 과대 보고된 것을 뜻한다.

표 1 은 그 결과를 근로소득의 소득 구간별로 나누어 제시한 것이다. 여기서 근로자란 상용근로자와 임시 및 일용 근로자의 합계(종사상의 지위가 1~3)이며, 20 세 이상의 성인으로 한정하였다. 이들의 표본 수는 16,742 명이며, 1,907 만명을 대표한다. 여기서 소득구간은 보정된 근로소득을 기준으로 설정된 것이다. 근로소득이 연 5-6 천만원인 근로자의 경우를 예로 들어 설명하면, 조사/보정 비율의 평균(mean)은 0.909 로 나왔음을 알 수 있다. 즉 이 구간의 근로자는 평균적으로 실제 근로소득보다 9.1%(=1-0.909) 과소하게 보고하였음을 말한다. 다만 동 비율의 분포는 해당 소득구간에서 0(min)에서부터 2.469(max)까지로 퍼져 있다. min 이 0 으로 나온 것은 보정된 근로소득은 5-6 천만원 구간에 들지만 조사결과는 0 인 근로자가 적어도 1 명 이상 존재한다는 뜻이다. 해당 구간의 근로자 수(1,064,411 명)를 동 비율의 순서로 정렬했을 경우 중앙에 위치하는 근로자(p50)의 동 비율은 0.939 이며, 하위 10%나 25%의 경우(즉 p10 과 p25)는 각각 0.664 와 0.821 이며, 상위 75%와 90%(즉 p75 와 p90)는 각각 1.004 와 1.079 임을 알 수 있다.

이를 소득구간별로 보면 연소득 3 천만원 이상에서는 mean 의 경우 1 을 밑돌기 시작해서 점차 하락하며 10 억원 이상 구간에서는 0.429 로까지 떨어졌음을 알 수 있다. 중앙값(p50)을 기준으로 보아도 소득이 높아질수록 소득을 과소 보고하는 경향이 강해지는 것을 마찬가지로 확인할 수 있다. 종래 가계조사의 소득이 근로소득의 경우는 실태에 가깝게 파악했을 것으로 생각되어 왔지만, 과소 보고가 상당히 일반적이며 특히 소득이 높아짐에 따라 과소 보고 비율이 높아진다는 점은 주목할만하다.

그런데 3 천만원 미만 그룹의 경우는 거꾸로 소득이 과대 보고되고 있으며, 소득이 낮을수록 과대 보고의 정도가 높아져 1 천만원 미만 구간에서는 동 비율의 평균이 5 배를 넘었다. 동 비율이 500 배(max)인 경우마저 있었다. 표준편차(sd)도 소득이 낮을수록 커져서 최하위 소득구간에서는 이상치가 많음을 보이고 있다. 이를 어떻게 해석해야 할까? 소득이 높을수록 과소 보고 비율이 높아지는 것은 자신의 소득이 노출되는 것을 꺼리기 때문으로 해석할 수 있다. 그러면 반대의 경우에는 자신의 소득을 과장하는 경향이 존재하는 것일까? 그러한 가능성을 배제할 수 없지만,

[3] 통계청은 2017 년에 조사한 것을 2017 년 조사라고 칭한다. 그렇지만 그 때 조사된 소득의 경우는 2016 년의 소득이므로 본고에서는 조사 대상 연도에 따라 이를 2016 년 조사로 부르기로 한다.

그것이 어느 정도 영향을 미쳤는지를 확인할 수는 없다. 여기서는 다른 가능성을 생각해 보고자 한다.

<표 1> 근로소득의 소득구간별 과소(과대) 보고 비율(=조사/보정)의 분포
(2016 년 가계금융복지조사)

	표본수	근로자수	mean	sd	min	p10	p25	p50	p75	p90	max
-1천만원	2,696	2,961,252	5.073	22.988	-	0.643	1.000	1.000	2.675	7.333	500
1-1.5천만원	2,398	2,685,859	1.273	0.765	-	0.869	1.000	1.000	1.390	2.080	15.714
1.5-2천만원	2,056	2,367,994	1.156	0.523	-	0.778	0.957	1.000	1.265	1.702	6.667
2-3천만원	3,101	3,519,781	1.043	0.394	-	0.713	0.886	1.000	1.114	1.414	5.512
3-4천만원	1,859	2,150,635	0.951	0.261	-	0.653	0.826	0.993	1.030	1.176	3.183
4-4.5천만원	702	808,873	0.929	0.249	-	0.642	0.831	0.966	1.009	1.135	2.499
4.5-5천만원	626	736,008	0.936	0.228	-	0.654	0.841	0.983	1.018	1.090	3.602
5-6천만원	909	1,064,411	0.909	0.212	-	0.664	0.821	0.939	1.004	1.079	2.469
6-8천만원	1,292	1,492,683	0.898	0.218	-	0.647	0.810	0.939	1.000	1.067	3.721
8-10천만원	557	641,062	0.897	0.181	-	0.669	0.826	0.943	1.000	1.030	1.862
1-2억원	498	591,123	0.867	0.216	-	0.589	0.766	0.916	0.988	1.031	3.158
2-3억원	32	34,096	0.793	0.235	0.195	0.473	0.559	0.826	1.000	1.034	1.086
3-5억원	11	11,269	0.640	0.216	0.304	0.388	0.529	0.567	0.825	0.918	1.093
5-10억원	3	2,012	0.431	0.237	0.176	0.176	0.176	0.441	0.651	0.651	0.651
10억원-	2	2,216	0.429	0.026	0.419	0.419	0.419	0.419	0.419	0.462	0.462
합계/평균	16,742	19,069,274	1.666	9.184	-	0.691	0.887	1.000	1.080	1.698	500

주: 1) 20 세 이상인 근로자(상용, 임시, 일용 근로자)를 대상으로 한 것이다.

2) 소득구간은 보정된 근로소득을 기준으로 설정된 것이다. 이 비율이 0 인 것은 보정된 근로소득은 0 보다 큰데 조사된 근로소득이 0 임을 뜻한다.

국세청은 근로소득을 근로소득 연말정산 자료(비과세자 포함)와 일용근로소득 자료의 두 가지로 파악하는데, 세법상의 차이로 인해 두 데이터는 통합되지 않고 있다. 따라서 양쪽에 모두 포함되는 경우가 존재한다. 국세청은 2016 년의 일용근로소득자를 816 만명으로 파악하였는데, 그 중에서 상용근로소득이 있는 자가 219 만명, 사업소득이 있는 자가 135 만명으로 나왔다[4]. 이들 정보가 통계청으로 보내지면, 통계청에서는 동일인이 복수의 소득원을 가지고 있는 경우 이들 소득을 개인별로 통합해서 이용한다. 그런데 여기서 주목하고자 하는 것은 사업소득과 일용근로소득을 모두 얻고 있는 경우가 적지 않게 존재한다는 것이다. 예컨대 낮에는 편의점 아르바이트를 하면서 밤에는 대리운전을 하는 경우를 상정해 보자. 그는 국세청의 일용근로소득 자료와 사업소득 자료에 모두 포착되지만, 이들 자료가 통계청으로 보내지면 개인별로 통합해서 동일인이 근로소득과 사업소득을 가진 것으로 파악된다. 통계청의 행정자료로 보정된 그의 근로소득은 편의점 아르바이트 수입만으로 잡히지만, 당사자는 두 수입을 구분하지 않고 모두 근로소득으로 인식할 수 있다. 즉 국세청과 피조사자가 파악하거나 인식하는 근로소득의 범위에

[4] 이 수치는 유승희 의원의 요구로 국세청이 제공한 순수 및 기타 일용근로자 지급명세서 자료에 의거한 것이다.

차이가 있을 수 있으며, 이러한 문제는 소득이 낮은 구간으로 갈수록 더욱 심해질 것으로 생각된다[5].

그리고 표 1 에 따르면 조사/보정 비율의 전체 평균이 1.666 으로 나왔는데, 이것은 하위 소득구간으로 갈수록 이상치가 많아져서 과장된 것이다. 표의 max 에서 보듯이 낮은 소득구간으로 갈수록 동 비율의 이상치가 더욱 커질 뿐만 아니라 그 숫자도 늘어남을 알 수 있다. 만약 소득의 크기를 감안해서 동 비율의 평균을 구해보면 0.984 로 낮아지는 것은 그 때문이다. 따라서 표 1 의 결과를 해석할 때 특히 하위 소득구간의 수치에는 이러한 자료상의 문제가 포함되어 있음을 감안해서 볼 필요가 있다.

그러면 가계금융복지조사의 보정된 근로소득은 실제의 근로소득 분포를 얼마나 잘 반영하고 있을까? 이를 검토하기 위해 표 2 를 제시하였다. 거기에서 조사 결과와 보정 결과는 전술한 데이터에서 구한 것이다. 그와 함께 소득세 자료에서 구한 소득구간별 인원수 분포를 제시하였다. 소득세 자료는 마이크로 데이터로 제공되지 않고 소득구간별 정보만 이용할 수 있을 뿐이며, 표 2 의 소득구간을 소득세 자료의 소득구간에 따라 작성되었다. 여기서 이용된 것은 국세청의 『국세통계연보』에서 근로소득 연말정산 자료와 일용근로소득 자료이다. 두 자료의 근로소득자 수를 합하되 전체 근로자 수에 맞추어 조정한 것이다[6]. 이로 인해 하위 소득구간에서는 소득세 자료가 실태와 다소 괴리가 있을 수 있지만, 그 이상의 소득구간에서는 전체 근로소득자를 비교적 정확히 파악한 것으로 볼 수 있다. 따라서 이를 기준으로 해서 가계금융복지조사의 조사결과와 보정 결과를 평가할 수 있다.

먼저 조사결과를 소득세 자료와 비교해 보면(표 2 의 a/c 참조), 3-4.5 천만원의 중위 소득구간의 인원수가 실제보다 21% 정도 많게 파악되었으며, 이 구간을 경계로 하여 하위 또는 상위 쪽으로 멀어질수록 파악률이 떨어지는 역 U 자형을 보이고 있다. 최상위로 갈수록 과소 보고되거나 과소 대표되는 경향이 더 심해져서 3-10 억원 구간에는 실제 소득자 수의 43%로 떨어졌고, 10 억원이 넘는 샘플은 아예 존재하지 않음을 알 수 있다. 표 1 에서 3 천만원이 넘는 경우에는 소득이 높아질수록 조사/보정 비율이 떨어진 데에서 알 수 있듯이 조사결과의 인원수는 실제 자신이 속하는 소득구간보다 아래 쪽 구간에 들어가 있는 경우가 많았을 것으로 생각된다. 그에 비해

---

[5] 본문에서 든 사례로 설명하면 편의점 아르바이트를 잠깐 하다(수입 50 만원)가 그만두고 대리운전이 한 해 동안의 주된 수입원(950 만원)이었다고 해 보자. 그 경우 국세청이 파악한 그의 근로소득은 50 만원이 된다. 만약 그가 자신의 근로소득을 1000 만원으로 인식하고 그렇게 응답했다고 한다면 조사/보정의 비율은 20 배(=1000/50)가 된다. 근로소득이 낮을수록 이러한 이상치가 나올 가능성이 높아짐을 알 수 있다.

[6] 2016 년의 근로소득 연말정산 신고자(비과세자 포함)는 1,774 만명이며, 일용근로소득자수는 816 만명인데, 양자를 합치면 통계청의 경제활동인구조사에서 파악된 전체 근로자수 1,967 만명을 넘는다. 일용근로소득자 중에는 연소득이 100 만원에도 미치지 못하는 자가 229 만명, 100-300 만원인 경우는 210 만명, 300-600 만원은 125 만명, 600-800 만원은 45 만명에 이른다. 이들 중에는 한 해 동안에 일시적으로 근로소득이 있었지만 경제활동인구조사의 취업자 정의(지난 1 주간 수입을 목적으로 일을 했는가)에 따르면 근로자로 파악되지 않는 자가 상당수 포함되어 있을 것으로 생각된다. 여기서는 양자의 합계가 경제활동인구조사의 근로자수와 일치하도록 일용근로소득자 중에서 소득이 낮은 순으로 제외하였다.

하위 소득구간으로 갈수록 반대로 조사/보정의 비율이 1 보다 높았기 때문에 실제 자신이 속하는 소득구간보다 위쪽 구간에 들어간 경우가 많았다. 조사결과에서 중위 소득구간이 실제보다 비대해진 것은 그 때문이라 생각된다.

<표 2> 근로소득 구간별 인원수 분포(2016 년)

	조사결과 a	보정결과 b	소득세자료 c	a/c	b/c
-1천만원	2,119,929	3,002,861	3,573,072	59.3%	84.0%
1-1.5천만원	1,931,009	2,703,250	2,402,133	80.4%	112.5%
1.5-2천만원	2,549,182	2,373,676	2,443,489	104.3%	97.1%
2-3천만원	3,851,907	3,528,810	3,540,749	108.8%	99.7%
3-4천만원	2,860,525	2,151,523	2,362,130	121.1%	91.1%
4-4.5천만원	1,026,893	809,966	845,725	121.4%	95.8%
4.5-5천만원	732,638	736,008	698,277	104.9%	105.4%
5-6천만원	1,111,909	1,064,411	1,065,028	104.4%	99.9%
6-8천만원	1,433,665	1,492,683	1,415,396	101.3%	105.5%
8-10천만원	617,034	641,062	654,436	94.3%	98.0%
1-2억원	512,935	591,123	598,398	85.7%	98.8%
2-3억원	27,373	34,096	42,572	64.3%	80.1%
3-5억원	8,172	11,269	18,728	43.6%	60.2%
5-10억원	2,917	2,012	6,722	43.4%	29.9%
10억원-		2,216	2,144	0.0%	103.4%
합계	18,786,090	19,144,965	19,669,000	95.5%	97.3%

주: 1) 조사결과와 보정결과는 2016 년 가계금융복지조사의 행정자료로 보정되기 전과 보정된 후의 결과를 말한다.

2) 소득세 자료는 근로소득 연말정산 신고자(비 과세자 포함)와 일용근로소득자를 합한 것이다. 다만 양자의 합계가 경제활동인구조사의 근로자수를 넘는데, 그 초과 분은 일용근로소득자 중에서 소득이 낮은 순으로 제외한 것이다.

자료: 통계청(MDIS), 가계금융복지조사(마이크로 데이터), 2016 년; 국세청, 국세통계연보.

이에 대해 보정결과를 소득세 자료와 비교하면(표 2 의 b/c 를 참조), 중위 소득구간이 과대하게 파악되는 현상은 보정되어 더 이상 나타나지 않게 되었다. 그렇지만 2-10 억원의 소득구간에서는 소득이 높아지면 파악률이 30%까지 떨어지는 양상을 보여 여전히 보정되지 못한 경우가 있음을 알 수 있다.

표본조사에서 발생하게 되는 오류를 샘플링 오류(sampling error)와 비 샘플링 오류(non-sampling error)로 구분할 수 있다. 전자는 표본이 모집단을 제대로 대표하지 못해서 발생하는 오류이며, 이것은 표본 규모를 늘리거나 표본의 설계를 개선해서 줄일 수 있다. 이에 대해 후자는 그 외의 모든 요인을 포괄하는 것인데, 조사 대상자가 응답을 기피하거나 과소 또는 과대하게 보고하는 경우가 대표적인 이유가 된다. 후자는 표본 수를 늘린다고 해결되지는 않는다. 이 개념 구분에 따르면, 조사결과를 행정자료에 의해 보정한 것은 후자의 비 샘플링 오류를 줄이는데

기여했다고 생각된다. 그렇지만 보정된 결과도 일부 소득구간에서 포착률이 크게 떨어진 데에서 드러나듯이 여전히 실제와 괴리가 남아 있는데, 이것은 샘플링 오류라고 할 수 있다.

2) 금융소득

가계금융복지조사에서 금융소득은 지난 한 해 동안에 이자와 배당으로 얻은 수입으로 조사된다. 금융소득은 통장에 잔고가 있으면 이자 수입이 발생하므로 누구나 가질 수 있는 매우 일반적인 소득이라 할 수 있다. 그 반면 금융소득 중 특히 배당이 그러한데 소수에게로 편중이 심하고 최상층 소득의 주된 형태가 금융소득으로 이루어져 있기도 하다. 이러한 특성으로 인해 금융소득의 조사는 근로소득에 비해 양상이 크게 다르다.

표 3 은 표 1 과 같은 요령으로 금융소득의 조사/보정 비율의 분포를 보인 것이다. 20 세 이상의 성인인구 중에서 행정자료로 파악된 금융소득이 0 보다 큰 경우로 한정하였다. 여기에 해당하는 표본수가 26.850 개이며, 이들이 2,865 만명을 대표하고 있다. 금융소득이 조금이라도 있는 자들은 근로소득자보다 훨씬 많음을 알 수 있다. 금융소득의 규모가 미미한 경우가 많기 때문에 금융소득 구간을 표 3 과 같이 세분해서 조사/보정의 비율의 분포를 제시하였다. 예컨대 실제의 금융소득이 2-5 백만원에 속한 소득자는 205 만명이지만, 대부분의 금융소득이 0 으로 조사되었고, 이 비율이 높은 순으로 상위 25%와 10%(즉 p75 와 p90)에 해당 하는 자는 실제 금융소득의 5.3%와 57.4%로, 가장 높은 경우(max)는 실제 소득보다 5 배가 넘게 조사되었다.

<표 3> 금융소득의 소득구간별 과소(과대) 보고 비율(=조사/보정)의 분포
(2016 년 가계금융복지조사)

	표본수	인구수	mean	sd	min	p10	p25	p50	p75	p90	max
-2만원	4,449	4,987,592	4.363	56.188	-	-	-	-	-	-	1,800
2-5만원	3,596	3,973,103	1.459	18.501	-	-	-	-	-	-	600
5-10만원	2,221	2,475,542	0.678	8.037	-	-	-	-	-	-	240
10-20만원	3,452	3,657,608	0.210	2.010	-	-	-	-	-	-	50
20-50만원	4,027	4,268,912	0.228	3.755	-	-	-	-	-	-	233
50-100만원	3,309	3,397,979	0.207	1.103	-	-	-	-	-	0.488	30.769
1-2백만원	2,789	2,813,735	0.169	0.663	-	-	-	-	-	0.524	11.009
2-5백만원	1,995	2,052,653	0.148	0.388	-	-	-	-	0.053	0.570	5.114
5-10백만원	610	613,365	0.194	0.516	-	-	-	-	0.156	0.658	5.848
1-2천만원	270	286,310	0.154	0.311	-	-	-	-	0.142	0.574	1.973
2-5천만원	91	95,105	0.278	0.433	-	-	-	-	0.383	0.968	1.770
5-10천만원	28	23,387	0.144	0.271	-	-	-	-	0.266	0.352	1.000
1-3억원	9	6,378	0.136	0.221	-	-	-	-	0.374	0.374	0.609
3억원-	4	2,992	0.070	0.133	-	-	-	-	0.264	0.264	0.264
합계/평균	26,850	28,654,662	1.140	24.651	-	-	-	-	-	0.038	1,800

주: 소득구간은 보정된 금융소득을 기준으로 설정한 것이다. 조사/보정의 비율이 0 인 것은 보정된 금융소득이 0 보다 크지만 조사된 금융소득은 0 인 것을 뜻한다.

금융소득의 조사/보정의 평균 비율(mean)을 보면, 근로소득에 비해 훨씬 낮지만, 소득이 높아질수록 동 비율이 떨어지는 경향이 강해진다는 점은 마찬가지다. 그리고 하위의 소득구간으로 갈수록 동 비율이 높아지는 경향도 그러하다. 예컨대 금융소득이 2 천만원 이상인

경우 모두 금융소득종합과세 신고를 하였을 것으로 생각되지만, 조사 설문에는 아예 소득을 기입하지 않거나 매우 줄여서 기입하였음을 알 수 있다. 그 이유 중의 하나는 자신의 금융소득이 정확히 얼마인지 알기 어렵기 때문이다. 금융소득종합과세 신고 때에는 국세청이 금융기관을 통해 이미 파악한 금융소득 합계가 제시되지만, 가계조사에서는 그러한 정보가 제공되지 않는다. 따라서 피조사자가 자신의 금융소득을 어림 짐작으로 기입하거나 공란으로 두게 된다. 후자의 경우 소득이 0 으로 처리되었다고 생각된다. 표 3 에서 p50 에서 조사/보정 비율이 모두 0 이라는 것은 과반수가 공란으로 비워두었고, p75 나 p90 의 동 비율이 1 보다 훨씬 낮은 것은 어림 짐작으로 기입하는 경우에도 실제보다 크게 줄여서 기입했음을 보여준다.

그런데 하위구간으로 가면 근로소득에서 나타난 것처럼 동 비율이 급격히 높아지는 현상이 나타났다. 그리고 하위 소득구간이 아니더라도 동 비율의 최대값(max)이 실제 소득의 몇 배에서 몇 백 배로 나오고 심지어는 1800 배까지 나왔다. 이러한 이상치(outlier)를 어떻게 해석할 수 있을까? 그 이유를 밝힐 수 있는 정보를 얻기 어렵지만, 오식의 가능성 이외에 몇 가지 가능성을 추정해 볼 수 있다. 하나는 피조사자가 금융소득과 금융자산을 혼동했을 가능성을 들 수 있다. 예컨대 자신의 통장에 1000 만원의 예금 잔고가 있고, 한 해 동안의 이자 소득이 20 만원인 경우 자신의 금융소득을 20 만원이 아니라 1000 만원으로 응답했을 가능성이다. 이 경우 조사/보정 비율은 50 배(=1000/20)의 이상치가 된다. 만약 그렇다면, 금융소득 조사는 조사당국의 의도가 피조사자에게 충분히 전달되지 못한 셈이 되는데, 이것은 비 샘플링 오류의 전형적인 한 형태이다. 통계청에 문의한 바에 따르면, 피조사자가 금융소득이 얼마인지 응답하지 못할 경우 금융자산에 일정한 이율을 곱해서 구한 값을 기입한 경우도 있다고 한다. 그런데 금융소득과 금융자산은 조사 시점의 차이로 인해 이 방식이 오히려 오차를 낳았을 가능성을 생각해 볼 수 있다[7]. 그리고 표 3 에서 조사/보정의 평균 비율이 1.140 으로 나왔지만, 소득의 크기를 감안한 평균은 0.188 로 나온다. 이것은 그러한 이상치가 금융소득이 낮을수록 더 많았음을 뜻한다.

그러면 이렇게 행정자료로 보정된 금융소득은 과연 실제의 소득분포를 얼마나 잘 반영하고 있을까? 근로소득에서는 연말정산과 일용근로소득 자료를 이용하여 그 소득분포의 전모를 파악할 수 있었지만, 금융소득의 경우 그에 대응하는 과세 자료가 없다. 다만 금융소득이 2 천만원을 넘는 자는 금융소득종합과세의 대상이 되기 때문에 그들은 모두 파악된다. 이 자료는 특히 최상층의 금융소득을 정확히 보여준다는 점에서 중요하다. 그리고 국세청은 개인에게 귀속되는 전체 이자

[7] 금융소득의 조사/보정 비율이 이상치를 보인 샘플을 대상으로 금융소득/금융자산의 비율, 즉 수익률을 구해 보았다. 대체로 100% 전후 또는 그것을 넘는 경우가 있는가 하면, 2% 전후로 나온 경우로 나뉜다. 전자는 금융 소득과 자산을 혼동한 경우라고 생각된다. 후자의 경우는 금융자산에 수익률을 곱해 금융소득을 구한 경우가 아닐까 생각된다. 그런데 행정자료로 보정된 금융소득을 보면 대체로 1-2 만원에 불과한 것으로 나오는데 이를 어떻게 이해할 수 있을까? 이와 관련하여 먼저 자산의 조사 시점은 3 월 말이고 소득의 조사는 작년 1 년 동안을 대상으로 하므로 양자의 조사 기준에 괴리가 있다는 점에 유의할 필요가 있다. 만약 1 월에 금융자산을 새롭게 취득한 경우라면 작년의 금융소득은 0 이 될 수 있다. 또는 예컨대 3 년 만기 정기예금의 이자가 만기 때 일괄 지급되는 경우라면 금융자산은 있지만 금융소득은 없는 해가 나올 수 있다. 이러한 경우에는 금융자산으로부터 금융소득을 추정해서 기입하는 방식이 오히려 조사/보정 비율을 이상치로 만드는 결과를 낳게 된다.

및 배당 소득의 정보를 준다. 따라서 과세 자료를 이용하면 2 천만원 미만의 금융소득에 관해 그 전체 규모는 알 수 있지만 그 분포에 관한 정보는 얻을 수 없다.

표 4 에는 먼저 가계금융복지조사에서 행정자료로 보정되기 전의 조사결과(a)와 보정된 결과(b)를 제시하였다[8]. 한국은행의 자금순환표에서 확인할 수 있는 전체 금융소득이 41 조원이므로 조사결과는 그 13.9%를 파악한 반면, 행정자료에 의한 보정으로 그 비율을 70.6%로 끌어올렸음을 알 수 있다. 행정자료의 보정으로 소득 파악의 정확도를 크게 올렸다고 평가할 수 있지만, 그럼에도 불구하여 여전히 30%에 가까운 금융소득은 누락되고 있음에 유의할 필요가 있다.

<표 4> 소득구간별 금융소득의 조사 결과와 누락된 금융소득의 보정(단위: 10 억원)

	가금복		소득세 자료		가금복 추가보정		전체		
	조사결과	보정결과	2000만원 이상		2000만원 미만		금융소득		
	a	b	이자 c	배당 d	e	f=e/(A/B)	g=c+d+f	a/g	b/g
0~1백만원	171	1,432	-	-	1,432	1,825	1,825	9.4%	78.5%
1~5백만원	561	3,584	-	-	3,584	4,568	4,568	12.3%	78.5%
5~10백만원	437	2,435	1	0	2,435	3,103	3,105	14.1%	78.4%
1~2천만원	679	3,828	7	5	3,828	4,878	4,890	13.9%	78.3%
2~3천만원	494	2,631	113	123	2,304	2,936	3,172	15.6%	82.9%
3~4천만원	434	1,897	87	94	1,677	2,137	2,318	18.7%	81.8%
4~5천만원	296	1,668	76	111	1,269	1,618	1,804	16.4%	92.5%
5~6천만원	239	1,228	68	100	999	1,274	1,442	16.6%	85.2%
6~7천만원	488	1,171	60	112	955	1,217	1,389	35.1%	84.3%
7~8천만원	192	930	57	107	721	919	1,083	17.7%	85.9%
8~9천만원	349	1,083	50	119	694	885	1,054	33.1%	102.8%
9~10천만원	223	552	49	118	498	635	803	27.8%	68.8%
1~2억원	891	3,324	319	1,044	1,699	2,165	3,528	25.3%	94.2%
2~3억원	123	891	167	693	232	296	1,157	10.7%	77.0%
3~5억원	30	1,452	184	947	86	109	1,240	2.5%	117.2%
5억원~	78	788	770	6,713	60	77	7,561	1.0%	10.4%
합계 A	5,684	28,894	2,009	10,287	22,474	28,642	40,937	13.9%	70.6%
자금순환표 B	40,937	40,937	23,015	17,922	28,642	28,642	40,937		
A/B	13.9%	70.6%	8.7%	57.4%	78.5%	100.0%	100.0%		

주: 1) 소득구간은 전체소득(=근로소득+사업소득+금융소득+부동산임대소득)으로 설정되었다.

2) e 는 가계금융복지조사에서 행정자료로 보정된 금융소득 중에서 2 천만원 미만인 금융소득을 말하여, f=e/(A/B)에서 A/B 란 e 열의 값인 78.5%를 말한다.

3) 2 천만원 미만의 소득구간에 금융소득이 2000 만원 이상(c 와 d)인 경우가 나와 모순으로 보인다. 그것은 국내에서 원천징수 되지 않는 금융소득 중 종합 과세되는 소득이 있기 때문이다.

자료: 통계청(MDIS), 가계금융복지조사(마이크로 데이터), 2016 년; 국세청, 국세통계연보; 한국은행, ECOS.

[8] 표 4 의 소득구간은 종합소득세 자료의 소득구간과 동일하게 설정되어 있다. 그것은 금융소득종합과세 자료가 그와 같은 소득구간별로 제시되기 때문에 그 자료를 이용하기 위한 것이다. 그 소득구간은 근로소득, 사업소득, 금융소득, 부동산임대소득을 합한 전체소득 기준으로 설정되어 있다.

이 누락된 30%의 금융소득의 실태를 보여주는 것이 2 천만원 이상의 금융소득에 부과되는 금융소득종합과세 자료이며, 표 4의 소득세 자료로 제시하였다. 이자(c)와 배당(d)이 각각 2조원과 10 조원으로 전체 이자와 배당의 각각 8.7%와 57.4%에 해당한다. 신고대상자가 2016 년의 경우 94,129 명으로 금융소득이 조금이라도 있는 자(2,890 만명)의 0.33%에 불과한 이들이 전체 금융소득의 30%를 차지하고 셈이다. 그런데 전체 금융소득 41 조원에서 금융소득종합과세 대상자의 몫을 뺀 29 조원이 2 천만원 미만의 금융소득에 해당한다.

그런데 이들의 분포가 어떻게 되어 있는가가 문제이다. 여기서는 그 실마리로서 가계금융복지조사의 보정결과에서 2 천만원 미만인 자들의 분포를 구해 본 것이 표 4 의 e 이다. 그런데 그 합계가 22 조원이며 전체의 78.5%에 불과하다. 여기서는 누락된 6 조원의 금융소득도 2 천만원 미만의 금융소득(표 4 의 e)의 분포와 다르지 않다는 가정으로 구한 것이 표 4 의 f 이다. 그러면 2 천만원 이상의 금융소득은 소득세 자료(c+d)에서 가져오고, 2 천만원 미만은 여기서 추정된 f 를 추가하면 전체 금융소득(g)이 추정된다. 이 추정에는 가정이 들어가 있기 때문에 실제와 다소 괴리가 있을 것으로 생각하지만, 금융소득종합과세 자료는 전수 조사된 수치이기 때문에 특히 최상층의 금융소득은 정확하게 반영되었다고 볼 수 있다.

이를 기준으로 가계금융복지조사의 조사결과가 실제보다 얼마나 과소 보고되어 있는가를 보면 표 4 의 a/g 와 같다. 많아야 35% 정도이고 대체로 10% 대의 파악율을 보이고 있다. 5 억 이상의 최상위 구간은 1%에 불과함을 알 수 있다. 이에 대해 보정결과는 파악율(b/g)을 크게 높였음을 알 수 있다. 금융소득 조사가 대부분 무응답 또는 크게 줄여서 보고하는 것이 보통이고 금융소득과 금융자산의 혼동에 의한 오류도 포함되어 있을 것으로 생각하는데, 이들 비 샘플링 오류는 행정자료에 의한 보정으로 상당히 개선되었음을 알 수 있다. 다만 그럼에도 불구하고 최상위 소득구간에서는 보정된 결과마저도 10.4%에 불과하다는 점이 주목된다. 그것은 가계금융복지조사가 최상층의 금융소득자가 조사대상으로 많이 빠져 있기 때문이다[9]. 조사대상 가구로 선정된 경우에는 행정자료에 의한 보정이 유효하지만, 누락이 많아 대표성을 갖지 못한 경우(이를 샘플링 오류라고 한다)에는 행정자료로 보정해도 정확성을 높이지 못한다.

3. 자산

가계금융복지조사는 소득과 함께 각종 자산과 부채를 조사하였다. 다만 자산의 경우는 소득과 달리 행정자료에 의한 보정이 이루어지지 않았다. 행정자료도 자산 가치의 전모를 제대로

[9] 가계금융복지조사에서 행정자료로 보정된 2 천만원 이상의 분포는 표 4 의 b 에서 e 를 빼서 구할 수 있다. 5 억원 이상의 소득구간에서 2 천만원 이상인 자의 금융소득은 7,280 억원(=7880-600)으로 나왔는데, 이것은 소득세 자료로 파악된 금융소득(c+d)인 7 조 4840 억원(=7700+67130)의 10%에도 미치지 못함을 알 수 있다. 이것은 최상위 금융소득자가 샘플에서 빠져 실태를 제대로 반영하지 못했기 때문으로 생각된다. 예컨대 금융소득이 1 억 이상인 자들이 18,585 명이며 전체 가구수의 0.1%에도 미치지 못한다는 점을 감안하면, 전체 가구수의 0.1% 정도를 표본으로 선정하는 가계금융복지조사에서 그들이 표본으로 선정되지 못했을 가능성이 높다.

파악했다고 보기 어렵기 때문이다. 소득의 경우 일부 지하경제가 남아 있지만 과세제도에 의해 소득의 발생 원천이 거의 파악되었다고 할 수 있다. 이에 대해 자산의 경우에는 거래 자체가 없거나 빈번하지 않기 때문에 그 가치를 객관적으로 정하기 어려운 경우가 많다. 여기서는 자산의 전체 분포를 개략적으로 보여주거나 추정할 수 있는 자료가 있는 경우로 한정해서 검토하지 않을 수 없다.

1) 주택 및 토지

가계금융복지조사에서 주택은 현재 살고 있는 주택과 그 외에 소유하고 있는 다른 부동산으로 나누어 조사하고 있다. 먼저 현재 살고 있는 주택에는 자가 소유한 경우와 전세나 월세 등으로 임차해서 살고 있는 경우로 나뉜다. 이들은 가구 단위로 조사되어 누구의 소유인지를 알 수가 없다. 본고에서는 개인별 자산의 분포를 보기 때문에 이를 가구주의 소유로 가정하기로 한다. 이에 대해 그 외에 소유하고 있는 부동산은 각 물건별로 소유자가 조사되었다. 개인이 복수의 물건을 가지고 있는 경우 이를 개인별로 합산하였다. 이들 부동산은 주거용 건물(단독주택, 아파트, 그 외의 주거용 건물), 비 주거용 건물(상가 및 빌딩 등), 토지(논밭, 임야, 대지 등), 기타로 나누어 조사되었다. 여기서 주택은 현재 소유주가 살고 있는 자가 주택과 임대해 준 주거용 건물의 합계로 구했다.

그런데 이들 주택 가액의 합계는 표 5 에서 보듯이 3,836 조원으로 나왔다. 이를 한국은행의 국민대차대조표에서 주택 자산의 합계액(3,749 조원) [10]과 비교하면 2.3% 더 많았지만 양자가 거의 근접한 것으로 볼 수 있다. 통상 자산을 조사하면 실제보다 과소하게 보고되는 것이 일반적이라는 점에 비추어 보면 예외적이라 할 수 있다. 주택의 현재의 시장가격을 물은 것인데, 실제 시장에서 거래되는 가격이라기보다는 피조사자가 생각하는 시장가격이 조사되었다고 할 수 있다. 표 5 는 성인 인구를 대상으로 보유하는 주택 가액 구간별 인원수 분포를 제시하였다. 복수의 주택을 소유한 경우는 합산되었다. 그에 따르면 전체 주택의 소유자는 1.228 만명으로 나왔으며, 1-5 억원 구간에 794 만명(전체의 64.6%)이 몰려 있다. 10 억원이 넘는 주택 소유자는 39 만명(전체 소유자의 3.2%)으로 나왔으며, 1 억원에 미달되는 경우는 259 만명(전체 소유자의 21%)에 해당하였다.

한편, 토지의 경우 가계금융복지조사에서는 665 조원으로 조사되었지만, 그에 대응하는 국민대차대조표의 토지 자산(1,332 조원)에 비해 49.9%에 불과한 것으로 나왔다. 주택은 미미하지만 과대 파악된 반면, 토지는 절반 정도로 과소 파악된 셈이다. 표 5 는 양자를 비교하기 위해 과소(과대) 보고된 것을 다음과 같이 조정하였다. 즉 주택은 조사된 가액을 1.023 으로, 토지는 0.499 로 각각 나누어 주어 각 자산의 합계가 국민대차대조표의 수치와 일치하도록 하였다. 이것은 과소 평가되거나 누락된 자산의 분포가 가계금융복지조사에서 파악된 자산의 분포와

[10] 한국은행의 국민대차대조표는 주택을 건설자산(주거용)과 토지(주거용 건물의 부속 토지)로 나누어 파악된다. 이에 대해 가계금융복지조사가 현재 살고 있는 자가 주택과 그 외에 임대한 주택으로 나누어 조사하되 주택의 가액에는 건물과 토지가 합산되어 있다는 점에서 차이가 있다. 여기서 주택이라 함은 주거용 건물과 부속 토지를 합친 것을 말한다.

다르지 않다고 가정한 것과 동일하다. 토지의 소유자 수를 보면, 342 만명으로 나와 주택 소유자의 28%에 불과하였다. 그렇지만 상위 자산 가액 구간에서 주택보다 토지 소유자가 더 많아 토지 자산의 집중도가 더 높았음을 알 수 있다.

그런데 가계금융복지조사가 파악한 이러한 자산 분포는 실제의 분포에 얼마나 근접해 있을까? 이를 검토하기 위한 기준을 종합부동산세법에 의거한 재산세 부과 자료에서 구했다. 이 자료는 박원석 의원의 요청으로 국세청이 제출한 것인데, 개인과 법인이 각각 소유하고 있는 주택과 토지를 공시가액을 기준으로 100 분위 분포 현황을 보여주는 것이다. 표 6 은 2013 년의 개인을 대상으로 이를 10 분위 분포로 간략하게 만든 것이다. 그에 따르면, 주택 과 토지 소유자수는 각각 1,300 만명과 767 만명으로 나왔고, 주택과 토지 가액은 각각 1,957 조원과 1,194 조원으로 나왔다. 먼저 소유자수를 표 5와 비교하면, 주택의 경우는 근접한 반면 토지 소유자수는 절반에도 미치지 못했다. 가계금융복지조사에서 파악된 토지 가액이 국민대차대조표에서 파악된 것의 절반 정도였음을 지적했는데, 토지 가액이 과소 평가 이외에도 조사에서 누락된 토지가 상당히 많았음을 알 수 있다. 주택과 토지의 가액 기준이 현재가격(표 5)와 공시가격(표 6)으로 차이가 있고, 또 두 조사의 시점이 다르기 때문에 직접 비교하기는 어렵다.

<표 5> 주택과 토지 가액 구간별 분포(2016 년)

	주택		토지	
	가액	인원수	가액	인원수
	(10억원)	(천명)	(10억원)	(천명)
0	-	28,940	-	37,805
-1천만원	922	117	299	63
1-2천만원	2,911	168	1,256	97
2-4천만원	12,560	382	5,810	226
4-6천만원	28,037	529	9,240	211
6-8천만원	46,043	626	13,198	208
8-10천만원	72,922	769	14,763	174
1-2억원	490,512	3,183	88,043	676
2-3억원	610,606	2,391	111,210	496
3-5억원	937,423	2,361	182,823	491
5-7억원	518,874	868	154,642	266
7-10억원	420,042	494	164,008	200
10-15억원	309,371	247	188,263	161
15-20억원	186,590	107	80,141	48
20-30억원	64,871	27	110,583	50
30-50억원	28,542	8	107,326	32
50억원-	18,367	3	100,280	17
합계 A	3,748,592	41,221	1,331,884	41,221
가금복 B	3,836,036	12,281	664,572	3,416
B/A	102.3%	29.8%	49.9%	8.3%

주: 1) 가액 항목의 B 는 가계금융복지조사가 파악한 주택 또는 토지의 가액을 말한다.

2) 인원수 항목의 A 는 20 세 이상의 성인인구 수이고, B 는 자산 가액이 0 보다 큰 소유자 수를 말한다.

그런데 표 5 와 표 6 의 자산 분포를 어떻게 비교할 수 있을까? 표 6 의 자료는 백분위 분포를 제시하고 있지만, 그것은 주택이나 토지의 소유자수를 기준으로 상위 1% 또는 10%의 비중을 구한 것이다. 이것은 그 나름으로는 유용한 정보를 제공하지만, 여기서는 소득이나 자산 분포의 국제비교에 널리 이용되고 있는 20 세 이상 성인 인구를 기준으로 산출하는 자산 집중도(top wealth shares)를 구하고자 한다. 즉 성인인구를 보유하는 주택(또는 토지) 자산의 가액 순으로 정렬한 다음, 예컨대 성인 인구의 상위 1% 또는 10%에 속하는 그룹이 차지하는 자산 가액이 전체 가액의 몇 %에 해당하는지를 추정하였다[11]. 그림 1 은 이렇게 추정된 결과를 그래프로 비교하였다. 이 때 자료상의 제약으로 두 자료의 연도뿐만 아니라 자산 가액의 기준도 공시가격과 현재가격으로 차이가 있다는 점에 유의할 필요가 있다.

<표 6> 재산세 부과 자료에 의한 개인의 주택 및 토지 보유실태(2013 년)

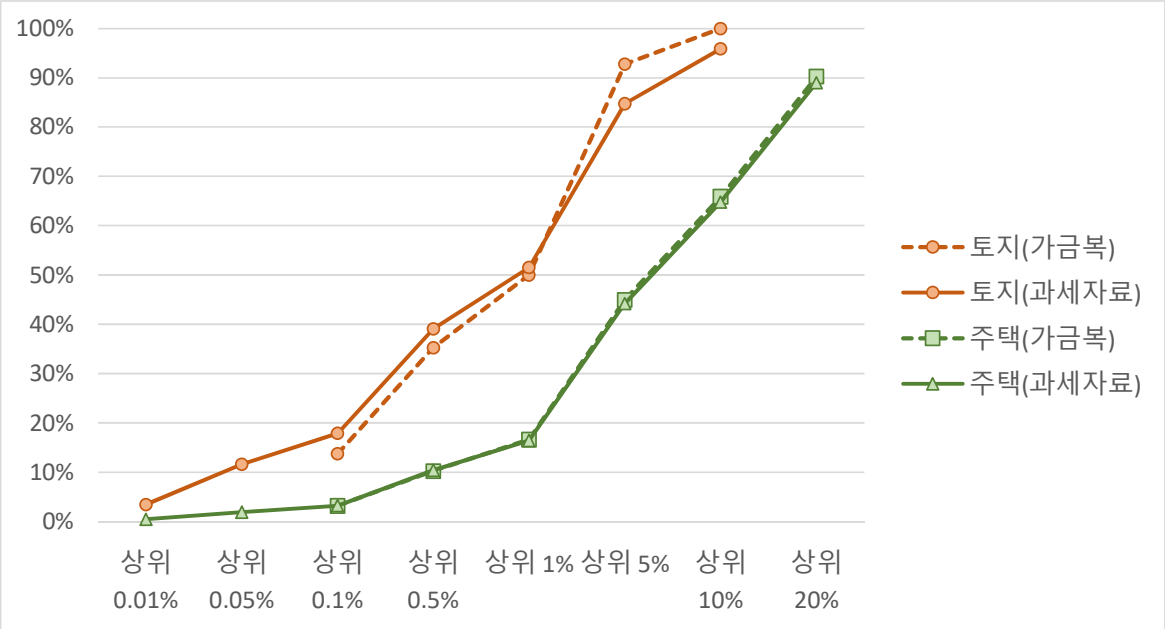
	주택			토지		
	인원수 (명)	가액 (10억원)	평균 (100만원)	인원수 (명)	가액 (10억원)	평균 (100만원)
1분위	1,300,448	11,162	9	767,515	840	1
2분위	1,300,360	44,274	34	767,460	3,744	5
3분위	1,300,360	70,977	55	767,460	8,416	11
4분위	1,300,360	96,705	74	767,460	15,096	20
5분위	1,300,360	124,420	96	767,460	24,313	32
6분위	1,300,360	155,697	120	767,460	37,583	49
7분위	1,300,360	194,411	150	767,460	58,088	76
8분위	1,300,360	247,448	190	767,460	94,239	123
9분위	1,300,360	335,642	258	767,460	174,861	228
10분위	1,300,360	676,309	520	767,460	776,500	1,012
상위 6-10%	650,180	234,414	361	383,730	168,420	439
상위 2-5%	520,144	287,618	553	306,984	300,057	977
상위 1%	130,036	154,277	1,186	76,746	308,023	4,014
전체	13,003,688	1,957,045	150	7,674,655	1,193,682	156

[11] 표 6 의 소유자수 기준 기준의 백분위 분포를 성인 인구 기준으로 재편할 때 평균분할 히스토그램(mean split histogram)이라는 보간법(interpolation)을 적용하였다. 예컨대 성인 인구(2013 년에 3,950 만명) 기준의 상위 1%에 들어가기 위한 경계(threshold)는 소유자수 기준으로는 상위 3%와 4% 사이의 어느 곳에 위치하는데, 이를 확정하는데 이 보간법을 이용하였다. 평균분할 히스토그램에 관해서는 Atkinson(2005: 333-334)이나 김낙년(2014: 63-73)을 참조하기 바란다. 그리고 가계금융복지조사의 자산 분포는 마이크로 데이터를 이용할 수 있기 때문에 그로부터 직접 계산할 수 있다.

자료: 국세청 "2013 년 보유부동산 가격기준 100 분위 현황(개인)", 박원석 의원의 요청의 의한 국세청 제출 자료.

과세 자료를 이용한 경우에는 전수가 조사되었기 때문에 상위 0.01%까지 구할 수 있지만, 표본조사인 가계금융복지조사의 경우는 상위 0.1%까지만 구했다. 먼저 주택의 경우를 보면, 두 자료에 각각 의거한 자산 분포가 거의 차이가 없음을 알 수 있다. 과세자료의 상위 0.1%, 1%, 10%의 자산 비중은 각각 3.2%, 16.4%, 64.8%였는데, 가계금융복지조사의 동 비율은 3.2%, 16.6%, 65.9%로 나왔다. 가계금융복지조사가 주택의 가액을 다소 과대하게 파악했다는 점을 지적했지만, 그 분포는 실제와 다르지 않았을 것으로 판단된다.

<그림 1> 주택과 토지의 상위 집중도 비교: 가계금융복지조사 vs. 과세 자료



자료: 통계청, 가계금융복지조사 2016; 국세청 "2013 년 보유부동산 가격기준 100 분위 현황(개인)", 박원석 의원의 요청의 의한 국세청 제출 자료를 이용하여 산출하였다.

이에 대해 토지의 경우는 과세자료의 상위 0.1%, 1%, 10%의 자산 비중은 각각 17.8%, 51.5%, 95.9%였는데, 가계금융복지조사의 동 비율은 13.7%, 50.0%, 100.0%로 나왔다. 상위 1%에서는 양자가 근접하였지만 그 전후에서는 적지 않은 괴리가 보인다. 전술한 소득에서는 조사결과와 행정자료에 의한 보정 결과를 비교할 수 있었기 때문에 양자의 차이를 비 샘플링 오류에 의한 것으로, 그리고 이 보정결과와 실제에 가까운 과세자료와의 차이를 샘플링 오류에 의한 것으로 구분할 수 있었다. 그렇지만 자산에서는 행정자료에 의한 보정 결과가 없기 때문에 그러한 구분은 어렵고 두 가지 오류가 혼입되어 있다고 할 수 있다. 다만 최상위로 갈수록 토지 자산이 과소 파악되는 경향이 커지고 있다는 점이 주목된다. 앞 절에서 금융소득과 같이 소득의 편중이 심한 경우에는 최상위 소득자가 샘플에 제대로 반영되지 못해 샘플링 오류가 커진다는 점을

지적한 바 있다. 토지 자산은 금융소득 정도로 편중된 것은 아니지만, 주택 자산에 비하면 상당히 편중되어 있는 편이고, 따라서 금융소득 정도는 아니지만 주택에서는 나타나지 않았던 샘플링 오류가 나타난 것이 아닌가 생각된다.

2) 금융자산

금융자산의 경우는 실제의 분포를 직접 보여주는 믿을만한 자료가 없어 이를 검토할 수 있는 자료적 상황이 더욱 열악하다. 먼저 가계금융복지조사에서 파악한 금융자산은 표 7 과 같다. 이들은 주로 이자와 배당을 낳는 금융자산이라 할 수 있다. 현금은 이자를 낳지 않지만 여기에 포함해서 다룬다. 그 외에도 보험이나 연금과 같은 자산이나 금융자산에 대응하는 부채도 있지만, 자료상의 제약으로 여기서 다루지 않는다[12].

가계금융복지조사가 파악한 금융자산을 한국은행의 자금순환표에서 구한 전체 금융자산과 비교해 보면, 현금을 포함해서 입출금이 자유로운 예금은 조사결과가 더 많았지만, 저축성 예금(펀드 포함)은 1/3 정도, 주식 및 채권은 1/5 정도에 불과하였다. 이들 합계는 831 조원이며 전체 금융자산(2,296 조원)의 36.2%를 파악한 데 불과하였다.

전체 금융자산의 분포는 알 수 없지만, 금융소득으로부터 금융자산을 추정하는 방식으로 접근하고자 한다. 금융소득은 금융자산의 수익으로 얻어지는 것이므로 양자는 밀접한 관련을 갖는다. 다만 각 개인의 금융자산 수익률은 천차만별이므로 개인의 금융소득으로부터 곧 금융자산을 추정하기는 어렵다. 그렇지만 개인이 아니라 상위 1%나 상위 10%와 같이 어떤 그룹의 평균을 구할 때에는 개인별로 들쑥날쑥 한 수익률이 서로 상쇄되어 평균 수준에 접근할 것으로 예상할 수 있다. 이 아이디어를 소득 자본화 방법(income capitalization method)이라 하며, 이를 이용하면 소득으로부터 자산을 추정할 수 있다[13]. 구체적으로는 금융자산의 평균 수익률(=금융소득/금융자산)의 역수를 자본화 승수(capitalization factor)라고 하는데, 이 승수를 개인에 금융소득에 곱해서 금융자산을 구한다.

<표 7> 소득구간별 금융자산의 비교: 조사 결과와 추정 결과

[12] 김낙년(2019)은 2017 년을 대상으로 본고에서 다루지 않은 금융자산과 부채를 포함하여 개인의 순자산 분포를 추정하였다.

[13] 이 방법을 적용하여 미국의 자산의 분포를 추정한 대표적인 연구로는 Saez and Zucman(2014)을 들 수 있다.

	금융자산의 조사결과					금융소득	추정 결과	비율
	현금 등	저축	주식 등	빌려준 돈	합계	표4의 g	금융자산	
	a	b	c	d	e=a+b+c+d	f	g=f*승수	e/g
0	6,405	9,707	2,386	735	19,232			
0~1백만원	18,201	33,619	8,481	2,624	62,925	1,825	102,344	61.5%
1~5백만원	13,399	42,163	11,560	2,410	69,532	4,568	256,173	27.1%
5~10백만원	13,086	34,879	8,911	2,312	59,188	3,105	174,121	34.0%
1~2천만원	24,385	54,683	24,085	6,322	109,475	4,890	274,238	39.9%
2~3천만원	19,755	42,701	15,448	2,963	80,867	3,172	177,930	45.4%
3~4천만원	16,587	32,211	8,268	4,643	61,709	2,318	130,007	47.5%
4~5천만원	14,690	33,035	10,571	2,483	60,779	1,804	101,198	60.1%
5~6천만원	11,485	21,201	9,099	2,672	44,457	1,442	80,880	55.0%
6~7천만원	8,600	19,427	8,945	3,001	39,973	1,389	77,888	51.3%
7~8천만원	7,876	17,602	7,803	2,170	35,452	1,083	60,740	58.4%
8~9천만원	7,636	16,885	9,530	1,899	35,950	1,054	59,121	60.8%
9~10천만원	3,685	9,020	6,442	1,392	20,539	803	45,009	45.6%
1~2억원	20,414	36,009	32,926	3,669	93,018	3,528	197,880	47.0%
2~3억원	5,504	8,181	8,804	116	22,606	1,157	64,879	34.8%
3~5억원	1,924	3,454	1,707	135	7,220	1,240	69,532	10.4%
5억원~	4,021	1,522	2,837	18	8,398	7,561	424,035	2.0%
합계 A	197,655	416,298	177,804	39,564	831,320	40,937	2,295,975	36.2%
자금순환표 B	170,182	1,311,225	814,568		2,295,975		자본화 승수	
A/B	116.1%	31.7%	21.8%		36.2%		56.09	

주: 1) 금융자산을 추정하는데 사용된 자본화 승수는 전체 금융자산(2,295,975)을 금융소득(40,937)으로 나누어 구한 56.09 이다.

2) 누락을 보정한 금융소득은 표 4 의 g 를 가져온 것이다.

앞 절에서 추정한 금융소득(표 4 의 g)이 실태에 가까울 것으로 생각할 수 있으므로 이를 이용하기로 한다. 표 7 에서 전체 금융소득은 41 조원이고, 전체 금융자산이 2,296 조원이므로 양자의 비율로 구한 자본화 승수는 56.09 가 된다[14]. 금융자산은 소득구간별 금융소득에 이 자본화 승수를 곱해서 추정할 수 있다(표 7 의 g). 이것이 실제의 금융자산에 가깝다고 볼 수 있기 때문에 이를 기준으로 가계금융복지조사의 금융자산이 얼마나 실태로부터 벗어나 있는지를 보일 수 있다(표 7 의 e/g). 그에 따르면 연소득이 4-9 천만원 구간에서는 금융자산의 50-60% 정도가 파악되었지만, 연소득이 그보다 높거나 낮을수록 파악률이 떨어지는 역 U 자형의 양상을 보이고 있다. 특히 연소득이 5 억원인 최상위 소득구간에서는 가계금융복지조사가 파악한 금융자산은

[14] 56.09 라는 승수의 의미를 생각해 보자. 만약 승수가 50 이라면 그 자산의 평균 수익률은 2%(=1/50)가 되므로 우리나라 금융자산의 수익률은 2%에도 미치지 못한다는 뜻이 된다. 다만 여기에는 이자를 거의 없는 현금이나 요구불예금이 포함되었기 때문에 이를 제외할 필요가 있다. 그리고 법인기업이 올린 수익 중에서 배당으로 분배되지 않은 사내유보가 여기에 빠져 있다. 2016 년의 법인의 사내유보(177 조원)에서 전체 지분증권 중에서 가계의 몫인 21%를 적용하면 37 조원이 되는데, 이를 포함하면 자본화 승수는 29.4 가 되고, 수익률은 3.4%로 높아진다.

전체의 2%에 불과하였다. 앞의 금융소득에서도 가계금융복지조사는 5 억 이상의 최상위 소득구간에서 전체의 1%밖에 파악하지 못했음(표 4 의 a/g 참조)을 밝힌 바 있는데, 금융자산에서도 그와 다르지 않은 결과가 나왔다. 최상층의 자산이 주식을 비롯한 금융자산 위주로 이루어져 있음을 감안하면, 가계금융복지조사의 자산 통계를 보정하지 않고 그대로 이용할 경우 자산 분포를 심각하게 왜곡할 수 있음을 알 수 있다.

4. 맺음말

여기서는 이상에서 밝혀진 사실을 정리하고, 가계조사의 행정자료에 의한 보정이 어떤 의의와 한계를 갖는지에 관해 논하기로 한다.

먼저 2016 년의 가계금융복지조사의 소득에 관한 조사결과와 행정자료에 의한 보정결과를 비교하여 피조사자의 응답에 어떤 문제가 있는지를 드러낼 수 있었다. 근로소득의 경우 3 천만원을 경계로 해서 소득이 높을수록 과소 보고의 경향이 커짐을 확인하였다. 금융소득의 경우 피조사자가 자신의 금융소득이 얼마인지 정확히 알기 어려운 점도 있어 응답하지 않은 경우가 많고 응답한 경우에도 과소 보고의 경향이 뚜렷하다. 어느 경우이든 소득이 높아지면 드러나는 것을 꺼리기 때문으로 생각된다. 그런데 소득이 낮을수록 거꾸로 행정자료로 파악된 소득보다 응답한 소득이 커지는 경향이 나타났고, 양자의 비율(=조사/보정)이 이상치라 할 정도로 높은 경우가 적지 않았다. 이것은 피조사자가 근로소득으로 인식하고 있지만 행정자료에서는 그렇게 분류되지 않는 경우라든지, 금융소득과 금융자산을 혼동한 경우가 적지 않았을 가능성을 시사하고 있다.

가계조사에서 나타나는 이러한 무응답, 과소 또는 과대 보고, 피조사자의 오해와 혼동, 또는 소득에 따라 그 양상이 달라지는 것들은 모두 비 샘플링 오류의 여러 유형이라 할 수 있다. 이들은 조사결과와 보정결과의 차이로 나타났다고 할 수 있다. 2016 년 가계금융복지조사에서 확인된 바에 따르면 금융소득의 경우 무응답을 비롯하여 비 샘플링 오류가 매우 컸고, 조사결과가 실태를 거의 반영하지 못했다. 근로소득의 경우는 금융소득보다 훨씬 낫지만 여전히 무시할 수 없는 정도로 비 샘플링 오류가 남아 있음을 알 수 있다. 여기서 확인된 문제는 다른 가계조사에서도 크게 다르지 않을 것으로 생각된다.

둘째, 소득 조사가 행정자료에 의해 보정됨으로써 비 샘플링 오류는 상당히 제거되었다고 할 수 있지만, 그렇다고 해서 그 결과가 실제의 분포에 얼마나 접근하였는지는 추가로 검토할 문제이다. 이를 위해서 전수 조사라고 할 수 있는 과세자료의 소득구간별 분포와 비교하였다. 근로소득의 경우 보정결과와 과세자료를 비교하면 일부 소득구간(2-10 억원)에서 괴리가 보이지만 양자의 분포가 대체로 근접한 것으로 나타났다. 금융소득의 경우 과세자료로 전수가 파악된 것은 2 천만원 이상으로 한정되지만, 이 정보만으로도 최상위 소득구간에서 금융소득의 파악률은 10%에 불과한 것을 알 수 있다. 금융소득은 예컨대 상위 0.1%가 전체의 1/3 이상을 차지할 정도로 최상층으로의 편중이 워낙 심한데, 이들이 샘플에서 빠질 확률이 높아 실태를 제대로 반영하지 못하기 때문이다. 이것은 샘플링 오류라고 할 수 있으며, 샘플 수를 늘리거나 샘플 디자인을 개선할 필요가 있음을 시사하고 있다.

셋째, 자산의 경우 행정자료에 의한 보정이 이루어지지 않아 조사결과를 직접 과세자료(금융자산의 경우 과세자료 등을 이용한 추정 결과)와 비교하였다. 그로 인해 두 자료의 차이는 전술한 비 샘플링 오류와 샘플링 오류가 분리되지 않아 이들이 혼입된 것으로 볼 수 있다. 주택의 경우 두 자료의 차이가 미미한 것으로 나타났고, 토지 자산에서는 차이가 좀더 벌어졌고, 금융자산에서는 괴리가 큰 것으로 나타났다. 주택보다는 토지의 가치를 파악하기 어렵고, 부동산보다는 금융자산이 금융소득에서 나타난 바와 같은 무응답이 많았기 때문에 이들 자산 중 뒤에 나열된 것일수록 비 샘플링 오류가 컸을 것으로 생각된다. 거기다가 금융자산과 그보다 덜하지만 토지 자산이 최상층으로의 편중이 심하다는 점을 감안하면, 그로 인한 샘플링 오류가 더해졌을 것으로 생각된다.

여기서 가계조사를 보정하는데 행정자료의 활용이 어떤 의의와 한계를 갖는지를 살펴보았지만, 거기에 그치지 않고 보다 본격적인 행정자료의 활용 방식을 생각해 볼 수 있다. 즉, 소득이나 자산 또는 과세 등의 정보를 갖고 있는 행정기관들이 이들 정보를 개인별로 통합하고, 거기에 개인의 사회인구학적 정보를 더해서 빅 데이터(big data)를 만들고, 이를 통계 목적으로 제공하는 것이다. 국세청의 소득세 정보를 사례로 들면, 종합소득세 신고자료 이외에도 종합소득을 신고하지 않고 원천세 징수로 과세가 종결되는 경우가 많다. 근로소득이나 사업소득의 연말정산, 일용근로소득이나 기타소득 등의 원천징수 자료 등이 그러하다. 개인의 소득이 이들 각 자료에 흩어져 있기도 하고 중복되어 있기도 한데, 이들 정보를 모두 개인별로 통합할 필요가 있다. 국세청을 넘어 다른 행정기관의 정보를 통합할 필요가 있다. 공적 연금은 연금관리공단, 정부의 사회부조금의 지출은 보건복지부 등의 정보를 이용할 수 있다. 개인의 사회인구학적 정보는 통계청의 인구센서스 등의 정보를 통합하는 방식이다. 이러한 빅 데이터의 구축은 현재 기술적으로 가능하지만, 그 실현에는 넘어야 할 많은 장벽이 적지 않다. 이 과정에서 제기될 수 있는 개인정보 보호와 행정정보 활용을 양립시키는 것뿐만 아니라 각 부처의 이기주의를 넘어 그들간의 정보 공유와 협력을 이끌어내는 것이 과제라고 할 수 있다.

이것이 앞으로 실현해야 할 목표라 할 수 있지만, 거기에 이르기까지 갈 길이 멀다. 이 점을 감안하면 본고에서 검토한 가계조사를 행정자료로 보정하는 방식은 과도적이지만 현실적인 방안이라고 생각한다. 소득의 종류에 따라 정도의 차이가 있지만 가계조사에 포함되어 있는 비 샘플링 오류를 크게 줄여 통계의 정확도를 크게 높이기 때문이다. 다만 샘플링 오류는 여전히 남기 때문에 전술한 바와 같이 샘플을 늘리거나 디자인을 개선하는 것이 과제라고 할 수 있다. 자산에 대해서도 행정자료에 의한 보정이 추가될 필요가 있다. 나아가 가계지출 중에서 비소비지출의 경우는 행정자료로 보정되었지만, 소비지출의 경우에는 실태와의 상당히 괴리가 커서 이를 보완하는 것이 또 하나의 과제라 할 수 있다. 소비지출은 가구 단위로 이루어지고 있으며, 아직은 행정자료로 대체하기 어렵다는 점에서 가계조사가 아니면 그 실태를 드러내기 어렵다는 점에서도 그러하다.

<참고문헌>

국세청,『국세통계연보』, 2017.

국세청, "순수 및 기타 일용근로자 현황 2016-17년", 유승희 의원의 요청에 의한 국세청 제공 자료

국세청, "부동산 보유실태 현황 2013년", 박원석 의원의 요청에 의한 국세청 제공 자료

김낙년 (2019),「우리나라 개인 자산 분포의 추정」,『경제사학』, 43(3), pp. 437-482.

김낙년 (2014),「2013년 소득세제 개편과 계층별 소득세 부담률」,『재정학연구』, 7(2), pp. 59-93.

김낙년·김종일(2013),「한국 소득분배 지표의 재검토」,『한국경제의 분석』, 19(2), pp. 1-64.

통계청,『가계금융복지조사』(마이크로 데이터), 2017.

통계청,『가계동향조사』(마이크로 데이터), 각 연도.

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

한국은행, 자금순환표 (http://ecos.bok.or.kr/).

한국은행, 국민대차대조표 (http://ecos.bok.or.kr/).

Atkinson, Anthony B.(2005), "Top Incomes in the UK over the 20th Century," *Journal of the Royal Statistical Society*, 168(2), 325-343.

Saez E. and G. Zucman (2014), "Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data", NBER Working Paper 20625.

<abstract>

Correction of Household Survey Using Administrative Data: Focusing on 2016 *Survey of Household Finance and Living Conditions*

Nak Nyeon Kim*

This paper compares survey results of the 2016 *Survey of Household Finance and Living Conditions* with the revised results using administrative data to reveal what biases respondents have in reporting their income and assets. First, in the case of earned income, the higher the income, the greater the tendency to underreport. In the case of financial income, many respondents did not report at all, and even if they responded, the income was greatly reduced. On the other hand, when income is low, the reported income tends to be larger than the income found in administrative data, and the gap between them is often very large. Factors such as misunderstanding or confusion among respondents regarding the concept or classification of income may have been at work. Second, non-sampling errors were mostly corrected by using administrative data. However, when compared with the tax data again, there were many omissions at the top, especially in the case of financial income. This is a sampling error where household surveys do not properly reflect the actual distribution. Third, in the case of assets, corrections were not made using administrative data, so the survey results were directly compared with the tax data. In the case of housing assets, the asset distributions of the two data were very close. In the case of land where the asset is more concentrated on the top floor, some of the top floors were missing, and in financial assets the problem was even worse.

Keywords: income, asset, sampling error, non-sampling error, *Survey of Household Finance and Living Conditions*

* Professor, Department of Economics, Dongguk University (nnkim@dongguk.edu)