

소득증빙서류와 조사자료 간 교차검증 : 재정패널조사를 중심으로



정재현
한국조세재정연구원
부연구위원
(jjung6@kipf.re.kr)

I. 서론

코로나19로 인한 막대한 경제피해에 대처하기 위해 조세·재정 정책의 역할이 그 어느 때보다 중요하다. 특히 한정된 자원하에 최대의 효과를 이끌어내기 위해서는 조세·재정 정책의 비효율성을 개선하려는 적극적인 노력이 요구되며, 이러한 노력이 결실을 맺기 위해서는 개별 정책이 기업과 가계에 미치는 영향을 분석하는 효과성 분석을 토대로 제도 개선 방안이 모색되어야 한다. 이처럼 정책효과의 실증적 검증의 필요성이 증가하며 최근 정부의 정책평가 및 제도 개선에 많은 경제학자들이 활발히 참여하고 있다. 실제로 2019년 노벨 경제학상 수상자인 에스터 뒤플로(Esther Duflo)는 정책의 평가와 설계에 참여하는 경제학자들에게 ‘배관공적인’ 태도가 요구된다고 주장하였는데, 이는 정책의 세세한 부분까지 고려하는, 다시 말해 정책의 일부를 세밀하게 조정하고 이에 따른 결과를 면밀하게 관찰하면서 최적의 정책을 조율해가는 자세를 의미한다(Duflo, 2017).

그렇다면, 현행 제도를 ‘배관공처럼’ 분석함에 있어서 가장 필요한 것은 무엇일까? 그것은 바로 신뢰성 있는 소득, 자산, 부채, 조세, 지출, 복지 등에 대한 포괄적인 통계자료의 확보이다. 우리나라의 정책효과 분석에 활용할 수 있는 대표적인 통계자료로는 국세 및 지방세 신고자료, 건강보험자료, 국민연금자료 등과 같은 행정자료를 들 수 있으며, 이러한 행정자료의 가장 큰 장점으로 대표성과 정확성을 꼽을 수 있다. 예를 들어, 의료보험 대상자의 전수 데이터를 수집하고 있는 건강보험자료는 개인의 성별, 나이, 진료기록 등의 정보뿐만 아니라 수가산정을 위해 타 행정기관과 연계하여 건강보험 가입자의 정확한 종합소득, 연금소득 등의 소득 정보를 제공하고 있다. 그러나 행정자료의 가장 큰 단점은 개인정보 보호 등의 이유로 접근이 제한되어 있어, 개별 정책의 효과를 세밀히 살피는 데 필요한 데이터 분석 작업에 상당한 제약이 존재한다는 점이다. 또한 정책효과를 보다 정확히 분

석하기 위해서는 개별 행정자료 외에 다른 행정 또는 설문 조사 간의 연계가 필요하나, 이미 구축된 행정자료 외에 다른 변수들, 가령 가구 단위의 조세-지출-복지에 관한 포괄적 데이터의 구축도 쉽지 않은 것이 현실이다. 가령 국세청 자료의 소득항목은 응답오차로부터 자유로워 신뢰성 있는 분석을 수행할 수 있으나, 기타 행정 및 설문 자료와의 연계가 쉽지 않아 신고된 소득, 조세부담 외에 가계가 정부로부터 받는 혜택을 종합적으로 파악하는 데는 불충분하다.

이에 한국조세재정연구원은 실증증거에 기반한 정책설계(Evidence-based policymaking)의 요구에 부응하기 위해 가구 단위의 조세-지출-복지에 관한 포괄적인 데이터를 수집하는 재정패널조사를 기획하고 재정패널 DB사업을 2008년부터 수행하고 있다. 2008년부터 1년의 조사주기로 실시하여 현재 12차 조사가 완료되었으며, 2008년 구축된 5,014가구와 2009년에 추가된 620가구를 포함하는 전체 원표본 가구수는 5,634가구이다. 재정패널 가구조사의 응답자는 가구주이며, 가구원 조사는 만 15세 이상의 소득이 있거나 소득¹⁾활동을 하는 가구원을 대상으로 하고 있다.

재정패널 가구원 조사의 소득 및 소득공제, 결정세액 현황 등에는 자가응답(self-report) 내용이 일차적으로 기록되는데 각 항목별로 복잡한 기준과 계산법이 작동되는 경우가 많아 응답자가 정확한 금액을 회상해내기 어렵다는 문제가 있었다. 이에 현장실사 기간 동안 가구원의 협조를 통해 소득공제 내역을 증빙할 수 있는 가구원의 세금신고 증빙서류 제출을 요청하고 있다. 근로소득자의 경우에는 근로소득 원천징수영수증을, 근로소득 외의 소득이 있는 경우에는 종합소득세 신고를 위한 일체의 신고서류(과세표준확정신고 및 자진납부계산서, 소득공제 명세서 등) 제출을 요청하고 있다. 따라서 재정패널조사의 소득 및 세액 공제, 결정세액 내역은 서류 제출자의 경우 제출된 증빙서류를 바탕으로 입력된 값이며, 서류 미제출자의 경우 해당 내역은 본인의 회상에 의한 응답값이 기록되고 있다. 이에 소득증빙서류 제출 여부에 따른 응답방식의 차이로 인해 소득 및 세액 공제 세부 내역 등 민감한 조사항목에 대한 응답 내용에 과소·과대 응답 등 오차가 발생할 우려가 있었다. 따라서 본고에서는 자가보고 내역과 행정자료와의 교차검증을 통해 재정패널조사에서 발생 가능한 조사편향을 논의하고 그 원인을 분석하여, 재정패널을 활용한 정책효과성 분석의 정확성을 제고할 수 있는 시사점을 제공하고자 한다.

**본고에서는
자가보고 내역과
행정자료와의
교차검증을 통해
재정패널조사에서
발생 가능한 조사편향을
논의하고 그 원인을
분석하여, 재정패널을
활용한 정책효과성
분석의 정확성을
제고할 수 있는 시사점을
제공하고자 한다.**

1) 가구원 조사의 대상은 만 15세 이상의 소득이 있거나 소득활동을 하는 가구원으로, 소득은 근로소득, 사업소득, 부동산임대소득, 이자 및 배당소득(100만원 이상), 연금/보험소득, 기타소득, 정부 현금보조금을 포함하고, 경제활동을 하지 않는 고령자라 할지라도 연금소득이나 재산소득이 있을 경우 조사 대상이 된다.

통계청의
가계금융복지조사는
조사에서 발생할 수
있는 응답오차를
줄이기 위해
국세청, 보건복지부,
한국신용정보원 등
30종의 행정자료를
연계하고 있다.

II . 조사자료와 행정자료 간 검증

본 장에서는 재정패널의 교차검증에 앞서 대표적인 조사자료인 가계금융복지 조사와 행정자료의 교차검증 사례를 먼저 살펴보고자 한다.

1. 가계금융복지조사의 조사자료와 행정자료의 연계

통계청의 가계금융복지조사는 민감한 조사항목에 대한 가구의 응답 거부나 과소·과대 응답 등 조사에서 발생할 수 있는 응답오차를 줄이기 위해 국세청, 보건복지부, 한국신용정보원 등 30종의 행정자료를 연계하고 있다. 행정자료와 조사자료의 개인고유정보(성별, 생년월일, 주소, 성명 등)를 활용한 식별키를 통해 표본 대상(가구, 가구원)을 연계하고 있으며, 2018년에는 9개 기관 24종의 행정자료를, 2019년에는 12개 기관의 30종 행정자료를 활용하였다. 자료 보완 기준으로는, 근로소득과 사업·임대소득은 조사자료를 기본으로 사용하나 금융소득, 공적이전소득, 비소비지출은 행정자료를 기준으로 대체 보완하고 있다.

<표 1> 가계금융복지조사의 행정자료 활용 현황(소득항목)

조사항목		행정자료	행정자료 보유기관
근로소득		근로소득(종합소득세, 원천세),	국세청
		일용근로소득, 자활급여	보건복지부
사업소득		사업소득(종합소득세, 원천세)	국세청
재산소득	금융소득	금융소득(종합소득세, 원천세)	국세청
	임대소득	사업소득 중 부동산 임대소득(종합소득세)	국세청
공적 이전 소득	공적연금	연금지급액	각 연금공단 등
	기초연금	기초연금	보건복지부
	양육수당 (출산장려금 포함)	양육수당, 양육보조금 아동수당, 한부모가족 아동양육수당 등	보건복지부, 여성가족부
	장애수당 (장애아동수당, 장애인연금 등)	장애인연금, 장애수당, 장애아동수당·장애인급여(지자체지원) 등	보건복지부
	기초생활보장 지원금	생계급여, 주거급여, 교육급여, 해산급여, 장제급여	보건복지부, 국토교통부, 교육부
	근로·자녀장려금	근로·자녀장려금	국세청

출처: 통계청(2019), pp. 69~70

<표 2> 행정자료 보완 전·후 소득 평균 및 증감값

(단위: 만원, %, %p)

구분	소득연도	평균			증감값		
		보완	조사	차이	보완	조사	차이
가구소득	2016년	5,478	5,020	459	4,300	4,040	260
	2017년	5,705	5,195	509	4,457	4,200	257
	증감률	4.1	3.5	0.6	3.7	4.0	-0.3
처분가능소득	2016년	4,520	4,128	391	3,616	3,374	242
	2017년	4,668	4,251	417	3,719	3,471	248
	증감률	3.3	3.0	0.3	2.8	2.9	0.0

출처: 통계청(2018. 12. 20.), p. 3

2017년 기준 표본가구의 가구원 48,590명의 소득통계를 행정자료로 보완한 결과 44,887명(92%)을 연계하였으며, 2017년 가구소득 평균은 5,705만원으로 기존 조사 결과(5,195만원)보다 509만원 증가하고, 처분가능소득 평균은 4,668만원으로 조사 결과(4,251만원)보다 417만원 증가하였다.

2. 행정자료와 조사자료 간 교차검증에 관한 선행연구

가. 가계금융복지조사

2016년 가계금융복지조사를 활용한 김낙년(2020)²⁾의 연구에서 가계금융복지 조사의 근로소득, 금융소득, 자산을 행정자료와 비교하여 응답편향을 분석한 결과, 근로소득은 소득이 높을수록 과소 보고하고, 금융소득과 자산은 가계금융복지 조사의 조사자료에서 누락된 경우가 많은 것으로 분석되었다. 같은 2016년의 가계금융복지조사를 활용한 또 다른 연구에서 이원진 외(2019)³⁾는 2016년(귀속 연도) 조사데이터와 행정보완데이터의 소득과 가처분소득 분포의 차이를 분석한 결과, 행정자료로 보완하면 소득과 가처분소득의 불평등 정도가 더 증가하는 것으로 나타났다. 이 연구는 가계금융복지조사와 행정보완데이터의 소득분포에 주목하여 두 자료를 비교분석하였으며, 국세청 행정데이터와 비교할 때 조사자료의 근로소득이 과소 보고되는 경향이 있음을 밝혔다. 구체적으로는, 행정자료 보완 후 근로소득이 존재하는 집단 비율이 77.1%에서 82.1%로, 근로소득은 연평균 2,748만원에서 2,760만원으로 증가하였다. 사업소득은 주로 무응답 사례를 행정

2017년 기준

표본가구의 가구원 48,590명의 소득통계를 행정자료로 보완한 결과, 가구소득 평균은 기존 조사 결과보다 509만원 증가하고, 처분가능소득 평균은 조사 결과보다 417만원 증가하였다.

2) 김낙년, 「가계조사의 행정자료에 의한 보정: 2016년 가계금융복지조사를 중심으로」, 『한국사회정책』, 27(1), 2020, pp. 39-61.

3) 이원진·정해식·전지현, 「소득조사 마이크로데이터 비교분석: 가계동향조사와 가계금융복지조사를 중심으로」, 한국보건사회연구원, 2019.

행정자료 보완 시
소득과 가처분소득의
불평등 정도가
더 증가하는 것으로
나타났다. 특히
금융소득이 존재하는
집단 비율이 기존
12.4%에서
행정데이터로 보완 후
83.3%로 크게
증가했으며, 금융소득
크기 또한 연평균
17만원에서 94만원으로
대폭 증가하였다.

<표 3> 조사데이터와 행정보완데이터 간 소득 현황 비교

(단위: 만원/년, %, %p)

구분	평균			0 초과비율			0 제외평균		
	조사(A)	행정(B)	B-A	조사(A)	행정(B)	B-A	조사(A)	행정(B)	B-A
근로소득	2,118	2,267	149	77.1	82.1	5.0	2,748	2,760	13
사업소득	738	782	44	32.2	45.3	13.1	2,292	1,727	-565
재산소득	125	213	88	23.8	88.7	64.9	526	240	-287
금융소득	17	94	77	12.4	88.3	75.9	140	107	-33
임대소득	101	111	11	13.7	14.5	0.9	737	766	29
근로·자녀장려금	2	7	5	2.8	12.0	9.1	53	55	2

주: '조사'는 조사데이터, '행정'은 행정보완데이터를 의미하며, '0 초과 비율'은 해당 항목이 0원 초과 존재하는 비율을, '0 제외 평균'은 해당 항목이 0원 초과 존재하는 집단의 평균을 의미함
출처: 통계청, 「2017년 가계금융복지조사」; 이원진 외(2019), p.132 <표 4-9> 재인용

데이터로 대체한 결과 사업소득이 존재하는 집단 비율이 32.2%에서 45.3%로 증가하였으며, 소득의 크기를 비교하면, 사업소득 연평균은 738만원에서 782만원으로 증가하였다. 금융소득은 조사자료에서는 금융소득이 존재하는 집단 비율이 12.4%에 불과하였으나 행정데이터로 보완한 후 83.3%로 크게 증가한 것으로 분석되었으며, 신고자의 금융소득의 크기 또한 연평균 17만원에서 94만원으로 대폭 증가하였다. 근로·자녀장려금 항목의 경우 조사자료에서는 수령했다는 자가 보고 비율은 2.8%였으나 행정자료로 보완한 후에는 12.0%로 나타났으며, 연평균 소득도 5만원이 증가하였다.

이어서 2016년(귀속연도) 조사데이터와 행정보완데이터의 소득과 가처분소득 분포의 차이를 분석한 결과, 행정자료로 보완하면 소득과 가처분소득의 불평등 정도가 더 심화되는 것으로 나타났다. 조사데이터와 행정보완데이터의 소득과 가처분소득 분포의 차이를 분석한 결과 첫째로, 행정자료 보완으로 가처분소득 10분위 경계값이 모두 증가하는데, 월평균 소득 하위 10% 가구의 경계값을 의미하는 P10은 4.5% 증가한 반면, P50~P90은 9.1~11.4% 증가한 것으로 분석하였다. 이에 따라 불평등 정도(P90/P10, P90/P50, P50/P10, 5분위배율, 지니계수, 빈곤율) 지표들이 모두 악화되었다.

가계금융복지조사는 또한 비소비지출(세금 등)의 항목에 대해 행정자료로 조사 자료를 전면 대체하여 보완하고 있으며, 해당 항목으로는 소득세, 공적연금기여금(국민연금, 공무원연금 등 5개 연금) 등을 포함하고 있다. 소득세의 경우도 <표 5>

<표 4> 조사데이터와 행정보완데이터 시장소득 및 가처분소득 분포

(단위: 만원/년, %, %p)

구분	균등화 시장소득				균등화 가처분소득			
	조사(A)	행정(B)	C=B-A	C/A(%)	조사(A)	행정(B)	C=B-A	C/A(%)
P10	624	696	72	11.6	882	922	40	4.5
P20	1,270	1,286	16	1.2	1,316	1,369	53	4.0
P30	1,700	1,772	72	4.3	1,684	1,783	100	5.9
P40	2,079	2,214	135	6.5	2,007	2,155	148	7.4
P50	2,477	2,652	175	7.1	2,336	2,550	214	9.1
P60	2,900	3,165	265	9.1	2,689	2,971	282	10.5
P70	3,464	3,745	281	8.1	3,173	3,485	312	9.8
P80	4,158	4,568	410	9.9	3,766	4,175	409	10.9
P90	5,400	5,911	511	9.5	4,742	5,281	540	11.4
P90/P10	8.66	8.50	-0.17	-1.9	5.37	5.73	0.35	6.6
P90/P50	2.18	2.23	0.05	2.2	2.03	2.07	0.04	2.1
P50/P10	3.97	3.81	-0.16	-4.1	2.65	2.77	0.12	4.5

주: '조사'는 조사데이터, '행정'은 행정보완데이터를 의미하며, 소득을 가구원수의 제곱근으로 나누어 균등화한 가구소득의 개인단위 분포임. P90/P10은 월평균 소득 상위 10% 가구의 경계값을 하위 10% 가구 경계값으로 나눈 배율을 의미하며, P90/P50은 (월평균 소득 상위 10% 가구의 경계값/월평균 소득 중위값)을, P50/P10은 (월평균 소득 중위값/월평균 소득 하위 10% 가구 경계값)을 각각 의미함
출처: 통계청, 「2017년 가계금융복지조사」 이원진 외(2019), p. 125 <표 4-4> 재인용

<표 4>를 보면, 행정자료 보완으로 가처분소득 10분위 경계값이 모두 증가하는데, 월평균 소득 하위 10% 가구의 경계값을 의미하는 P10은 4.5% 증가한 반면, P50~P90은 9.1~11.4% 증가한 것으로 분석하였다. 이에 따라 불평등 정도 지표들이 모두 악화되었다.

<표 5> 조사데이터와 행정보완데이터 간 세금 현황 비교

(단위: 만원/년, %, %p)

구분	평균			0 초과비율			0 제외평균		
	조사(A)	행정(B)	B-A	조사(A)	행정(B)	B-A	조사(A)	행정(B)	B-A
세금	154	199	45	98.8	98.9	0.1	156	201	45
소득세	116	160	45	75.6	80.2	4.6	153	200	47
재산세	20	20	-	72	72	-	28	28	-
자동차세	17	17	-	75.4	75.4	-	23	23	-
기타 세금	1	1	-	95.9	95.9	-	1	1	-
사회보험료	200	199	-1	89.3	90.2	0.9	224	220	-3
국민연금기여금	82	78	-5	72.4	75.6	3.2	114	103	-11
기타 연금기여금	18	21	4	9.3	8.7	-0.6	189	243	54
건강보험료	89	89	-	88	88	-	101	101	-
고용보험료	11	11	-	60.7	60.7	-	18	18	-

주: '조사'는 조사데이터, '행정'은 행정보완데이터를 의미하며, 가구원 수 제곱근으로 나누어 균등화한 가구소득의 개인 단위 분포임. '0 초과 비율'은 해당 항목이 0원 초과 존재하는 비율을, '0 제외 평균'은 해당 항목이 0원 초과 존재하는 집단의 평균을 의미함
출처: 통계청, 「2017년 가계금융복지조사」 이원진 외(2019), p.132 <표 4-9> 재인용

『국세통계연보』에
제시된 소득분포를
모집단이라 가정하고
재정패널조사에서
자가보고된 소득의
분포가 모집단의 소득
분포와 비교했을 때
어느 정도로 차이가
나타나는지,
소득증빙서류 수집률이
증가하면서 그 차이가
증가 또는 감소하는지를
살펴본다.

와 같이 소득자료와 마찬가지로 조사자료가 과소 보고되는 경향을 발견할 수 있다.

나. 재정패널조사

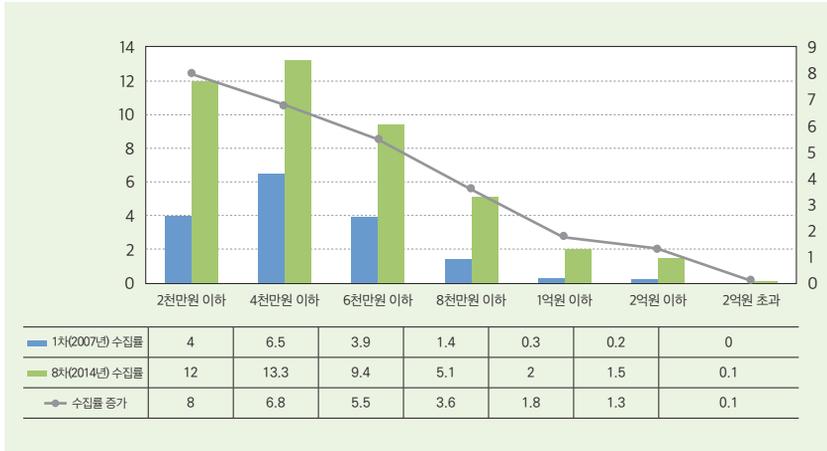
본고의 분석 대상인 재정패널조사를 대상으로 2017년 한국조세재정연구원의 세수추계패널센터에서 소득증빙서류 수집률에 따라 재정패널의 근로소득 구간별 인원분포 차이를 분석한 바 있다.⁴⁾ 구체적으로는, 『국세통계연보』에 제시된 소득분포를 모집단이라 가정하고 재정패널조사에서 자가보고된 소득의 분포가 모집단의 소득분포와 비교했을 때 어느 정도로 차이가 나타나는지, 그리고 소득증빙서류 수집률이 증가하면서 그 차이가 증가 또는 감소하는지를 살펴보았다. 변화를 분석하기에 앞서 근로소득자의 증빙서류 제출 비중을 소득구간별로 살펴보면, 1차 연도(2007년) 기준 전체 근로소득자 중 16.3%가 소득증빙서류를 제출한 가운데, 6.5%에 해당하는 가구원이 근로소득 2천만원 초과 4천만원 이하 구간에 속하는 것으로 조사되었다. 8차 연도(2014년)에는 제출자 중 13.3%의 가구원들이 이 구간에 속하는 사람들로 가장 높은 수집률을 보이고 있다. 조사차수가 증가하면서 수집률의 증가 폭이 가장 큰 구간은 근로소득 2천만원 이하 구간으로 1차 연도에는 4%가 서류를 제출하여 6천만원 이하 구간과 비슷한 수집률을 보였으나, 7년 후인 8차 연도에는 수집률이 8%p 증가하였다.

조사차수가 증가하면서 『국세통계연보』와 재정패널자료 간 소득규모별 인원분포의 차이가 증가 또는 감소하는지를 살펴보면, 소득규모별 인원분포 차이는 대부분의 구간에서 0에 가까워지는 모습을 보이고 있다. 특히, 증빙서류 수집률이 가장 높게 나타난 구간인 2천만원 이하 구간과 4천만원 이하 구간에서 그 차이가 감소하고 있다. 하지만, 소득증빙서류 수집에 따라 응답 오차가 감소하고 있다고 단언하기에는 주의가 필요하다. 표집변동(sampling variation), 그리고 자가보고 편의(self-report bias)를 구분할 수 없기 때문이다. 예를 들어, 2천만원 이하 소득자의 경우 전체 근로소득자의 약 44.9%(『국세통계연보』, 2014년 귀속 기준)를 차지하는 높은 비중으로, 조사차수가 증가할수록 2천만원 이하 그룹의 인원분포 비중의 차이가 감소하는 것이 실질소득 증가로 인한 상위 소득구간으로의 이동 때문인지, 소득증빙서류 제출에 따른 보정에 의한 소득구간 이동인지를 정확하게 판단할 수 없기 때문이다. 따라서 증빙서류 제출 여부에 따른 소득 크기의 차이뿐 아니라 개인 특성의 차이까지 따져보아야 조사편향의 여부를 평가할 수 있을 것으로 판단된다.

4) 세수추계패널센터, 「소득증빙서류 수집률이 소득분포에 미치는 영향」, 『재정포럼』, 제249호, 2017. 3., pp. 52-64.

[그림 1] (전체 근로소득자 기준) 근로소득 규모별 수집률 변화(2007년, 2014년)

(단위: %, %p)

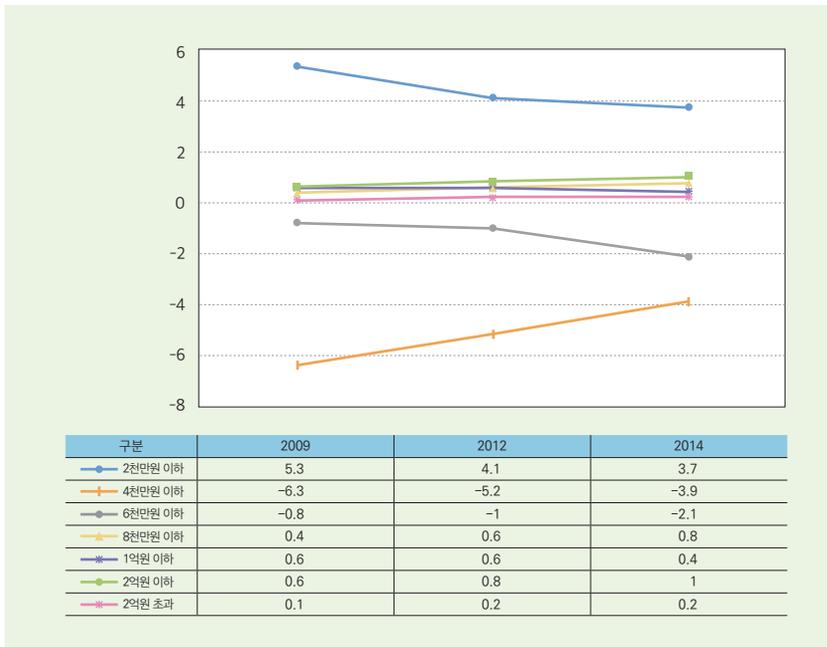


출처: 세수추계패널센터(2017), p.58 [그림 4]

소득규모별 인원분포 차이는 대부분 구간에서 0에 가까워지는 모습을 보인다. 특히, 증빙서류 수집률이 가장 높게 나타난 2천만원 이하 구간과 4천만원 이하 구간에서 그 차이가 감소하고 있다. 다만, 개인 특성의 차이까지 따져보아야 조사편향의 여부를 판가름할 수 있을 것으로 판단된다.

[그림 2] 근로소득 구간별 인원분포 차이(국세통계연보-재정패널)

(단위: %p)



출처: 세수추계패널센터(2017), p. 61 [그림 7]

근로소득증빙서류
수집률은 재정패널
1차 연도(2007년 귀속)에
26.8%를 기록한 후,
6차 연도(2012년 귀속)
까지 지속적으로
증가하여 그 이후로는
60%대에 머물고
있으며, 종합소득의
경우 6차 연도 이후
수집률은 평균
41%대에 머물고 있다.

Ⅲ. 소득증빙서류와 조사자료 간 교차검증 : 재정패널을 중심으로

본 장에서는 재정패널조사의 자가보고 내역과 행정자료와의 교차검증을 통해 자가보고에 따른 조사편향을 논의하고자 한다.

1. 소득증빙서류 제출 여부

<표 6>은 증빙서류 수집 대상이 되는 가구와 가구원 중 실제로 증빙서류를 제출한 가구와 가구원의 현황을 나타낸 것이다. 근로소득자나 사업소득자 중 소득 신고 대상자의 수를 나타낸 것으로 근로소득자의 경우에는 연말정산을 하지 않았다고 응답한 근로자는 수집 대상에 포함되지 않는다. 마찬가지로 사업소득자나 금융소득자 중에서도 소득신고를 한 가구원만 수집 대상에 포함된다.

재정패널조사의 경우 7차 연도 이후로 근로소득자의 약 60%가 소득증빙서류를 제출하고 있으며, 종합소득 신고 대상자의 경우 평균 41%가 관련 소득증빙서류를 제출하고 있다. 연도별로 살펴보면, 근로소득 증빙서류 수집률은 재정패널 1차 연도(2007년 귀속)에 26.8%(근로소득 연말정산 대상자 2,387명 중 639명 제출)를 기록한 후, 6차 연도(2012년 귀속)까지 지속적으로 증가하여 그 이후로는 60%대에 머물고 있으며, 종합소득의 경우 6차 연도 이후 수집률은 평균 41%대에 머물고 있다. 근로소득 증빙서류의 경우 원천징수영수증 1개 항목을 수집하는 반면, 종합소득의 경우 5월 소득세 확정신고를 완료한 과세표준확정신고 및 자진납부계산서, 소득공제명세서 등을 수집하기 때문에 근로소득 증빙서류 제출에 비해 수집률이 낮다. 따라서 본고에서는 증빙자료 제출의 용이성과, 이후 진행될 결정세액 및 공제내역 분석의 편의상 근로소득자를 대상으로 분석을 수행하기로 한다.

이어서 근로소득자를 대상으로 연말정산 서류를 얼마나 계속해서 제출하는지를 살펴보았다. 귀속연도 2013년(7차)부터 2017년(11차)까지 연속으로 재정패널 조사에 참여한 근로소득 연말정산 대상자 1,939명 중 1,137명(58.6%)이 매조사 시 증빙서류를 제출한 반면, 같은 기간 근로소득 연말정산 서류를 한 번도 제출하지 않은 경우는 21.4%로 조사되었다.

<표 6> 재정패널 차수별 근로소득, 사업소득 신고 대상자 중 신고서류 제출 여부

(단위: 명, %)

구분	조사 가구원 수	근로소득				종합소득		
		근로소득 보유자	연말정산 대상자	정산서류 제출자	수집률	연말정산 대상자	정산서류 제출자	수집률
1차 연도 (2007년 귀속)	6,912	3,928	2,387	639	26.8	721	97	13.5
2차 연도 (2008년 귀속)	7,095	4,106	2,576	1,223	47.5	806	213	26.4
3차 연도 (2009년 귀속)	6,981	3,998	2,558	1,345	52.6	879	252	28.7
4차 연도 (2010년 귀속)	7,216	3,984	2,546	1,516	59.5	874	289	33.1
5차 연도 (2011년 귀속)	7,298	4,078	2,645	1,544	58.4	852	316	37.1
6차 연도 (2012년 귀속)	7,380	4,118	2,786	1,651	59.3	920	364	39.6
7차 연도 (2013년 귀속)	7,559	4,294	2,954	1,785	60.4	894	378	42.3
8차 연도 (2014년 귀속)	7,585	4,328	3,064	1,876	61.2	901	377	41.8
9차 연도 (2015년 귀속)	7,806	4,475	3,137	1,921	61.2	920	376	40.9
10차 연도 (2016년 귀속)	7,865	4,583	3,196	1,939	60.7	937	401	42.8
11차 연도 (2017년 귀속)	7,965	4,645	3,266	2,001	61.3	974	415	42.6

출처: 재정패널조사 11차 자료를 이용하여 저자 작성

<표 7> 재정패널 근로소득 연말정산 서류 계속 제출 여부

(단위: 명, %)

2013~2017년 귀속연도 간 제출 횟수	해당자	비중
0	415	21.40
1	87	4.49
2	56	2.89
3	75	3.87
4	169	8.72
5	1,137	58.64
총합	1,939	100.00

출처: 재정패널조사 11차 자료를 이용하여 저자 작성

이를 통해 제출자는 계속해서 제출하고 미제출자는 반대의 경향을 보여 소득 수준, 납세 성향 등 응답자 개인 특성에 따라 증빙서류 제출 여부가 좌우될 수 있음을 가정할 수 있다.

<표 7>을 보면, 귀속연도 2013년(7차)부터 2017년(11차)까지 연속으로 재정패널 조사에 참여하고, 근로소득 연말정산 대상자 중 58.6%가 매조사 시 증빙서류를 제출한 반면, 같은 기간 근로소득 연말정산 서류를 한 번도 제출하지 않은 경우는 21.4%로 조사되었다.

최신 11차 조사
(2017년 귀속자료)
기준 근로소득을
지니면서 연말정산
대상자로 분류된
가구원 중 근로소득
연말정산 서류 제출자와
미제출자의 개인 특성별
평균의 차이를 가늠하고
그 차이의 통계적
유의성을 측정하는
분석을 진행하였다.

2. 제출자 및 미제출자 간 개인 특성 비교

최신 11차 조사(2017년 귀속자료) 기준 근로소득을 지니면서 연말정산 대상자로 분류된 가구원 3,253명 중 근로소득 연말정산 서류 제출자(1,992명)를 미제출자(1,261명)와 각 개인 특성별 평균의 차이를 가늠하고 그 차이의 통계적 유의성을 측정하는 분석을 진행하였다. 여기서 분석에 포함된 개인의 특성은 나이, 성별, 결혼 여부, 이혼 여부, 가구주 여부, 대리 응답 여부, 대학교 이상의 학력 여부, 전문직 사무직의 근무 여부, 정규직 여부 등이며 여기에 근로소득의 크기를 측정하였다. 분석 결과로는 첫째로, 증빙서류 제출자의 증빙서류상 근로소득(총급여)이 미제출자의 자가보고에 비해 평균 429만원 더 높은 것으로 조사되었다. 근로소득 차이 이외에도, 제출자는 미제출자에 비해 나이가 평균 3.6세 더 많으며, 가구주일 확률은 13% 더 높고, 본인이 직접 설문에 응했을 경우가 미제출자에 비해 26.5% 더 높은 것으로 조사되었다. 이는 제출자가 가구주 본인 혹은 배우자며 설문에 직접 참여하여 증빙서류도 함께 제출한 경우가 많은 반면, 미제출자는 <표 9>에서 정리된 바와 같이 가구주의 자녀(즉, 가구주의 대리응답)인 경우가 많은 것을 의미한다. 반면, 학력, 성별의 차이는 크지 않은 것으로 나타났다.

이어서 머신러닝 기법 중 랜덤포레스트를 활용하여 어떤 특성변수가 제출/미

<표 8> 근로소득 연말정산 대상자 중 증빙서류 제출자의 특성(11차 기준)

특성변수	미제출자		제출자와의 차이 (통계적 유의성)
	평균	표준편차	
나이	42.013	11.930	3.645***
성별	0.619	0.486	-0.033*
결혼 여부	0.630	0.483	0.161***
이혼 여부	0.023	0.150	0.017***
가구주 여부	0.494	0.500	0.132***
대리 응답 여부	0.550	0.498	-0.265***
대학교 이상 학력	0.699	0.459	-0.017
전문직 사무직	0.504	0.500	0.017
정규직 여부	0.929	0.256	0.018**
근로소득(만원)	3,834.434	2,800.890	429.045***
해당자 수	1,261		3,253(제출자+미제출자)

주: *, **, *** 은 각 10%, 5%, 1% 수준에서 통계적으로 유의함
출처: 재정패널조사 11차 자료를 이용하여 저자 작성

제출 여부를 결정(예측)하는 데에 가장 큰 영향을 끼치는지 분석(Random forest, variable importance)⁵⁾해 보았다. 분석 결과 소득(income)이 가장 큰 영향을 끼치고, 그다음으로는 나이, 대리응답 여부, 가구주 여부 순으로 영향을 끼치는 것으로 나타났다. 물론 소득은 미제출자의 경우 자가보고 값이므로 증빙서류 제출 여부에 영향을 받는 내생변수일 가능성도 상존한다. 또한 가구원의 나이변수도 소득의 크기 및 가구주 여부와 높은 상관관계를 지닌다는 점 또한 고려해야 한다.

3. 제출자 및 미제출자 간 근로소득 분포 비교

이어서 제출자 및 미제출자 간 근로소득의 분포를 비교해보았다. 앞서 <표 8>에서 제출자의 평균 소득이 미제출자에 비해 높은 것으로 조사되었다. 소득분포의 경우, <표 10>에서 제시된 바와 같이 미제출자 소득의 표준편차(2,800만원)가 제출자(3,079만원)보다 작은 것을 발견할 수 있다. 단, 미제출자의 실제 소득분포가 제출자의 소득분포와 다른 것인지 아니면 자가보고 과정에서 소득을 개략적으로 올림 또는 내림하는 경향에 의해 좀 더 편차가 작게 나온 것인지는 개별 가구원의 행정자료와 자가응답 자료가 모두 있어야만 판단이 가능하다.

<표 9> 근로소득 연말정산 대상자 중 증빙서류 제출자와 미제출자의 가구주와의 관계(11차 기준)

구분	제출자의 가구주와의 관계		미제출자의 가구주와의 관계		
	Freq.	Percent	Freq.	Percent	비중 차이
가구주	1,248	62.68	623	49.37	13.31
가구주의 배우자	546	27.42	248	19.65	7.77
가구주의 부모	1	0.05	0	0.00	0.05
형제자매, 그 배우자	5	0.25	8	0.63	-0.38
형제자매의 자녀, 그 배우자	1	0.05	1	0.08	-0.03
기타 친·인척	2	0.10	0	0.00	0.10
가구주의 첫째 자녀	96	4.82	202	16.01	-11.19
가구주의 둘째 자녀	73	3.67	145	11.49	-7.82
가구주의 셋째 자녀	10	0.50	20	1.58	-1.08
가구주의 넷째 자녀	1	0.05	2	0.16	-0.11
가구주의 다섯째 자녀	0	0	1	0.08	-0.08
가구주의 일곱째 자녀	0	0	1	0.08	-0.08
가구주 첫째 자녀의 배우자	6	0.30	4	0.32	-0.02

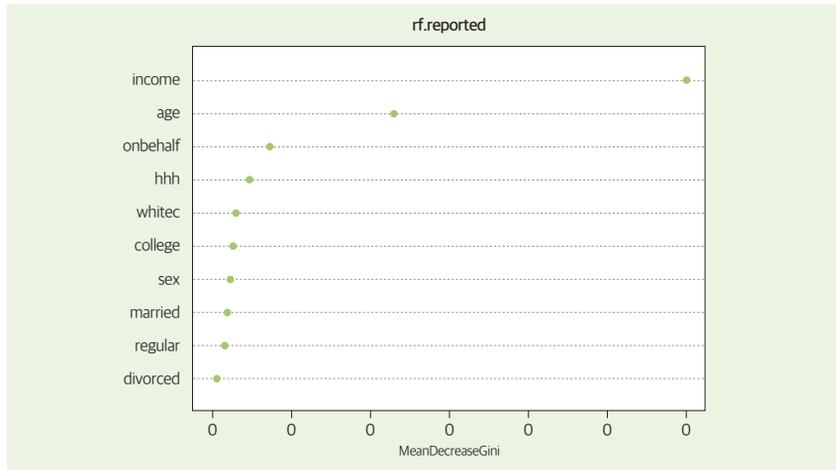
출처: 재정패널조사 11차 자료를 이용하여 저자 작성

분석 결과, 증빙서류 제출자의 증빙서류상 근로소득(총급여)이 미제출자의 자가보고에 비해 더 높은 것으로 조사되었다. 또한 제출자는 미제출자에 비해 나이가 평균 3.6세 더 많으며, 가구주일 확률은 13% 더 높고, 본인이 직접 설문에 응했을 경우가 미제출자에 비해 26.5% 더 높은 것으로 조사되었다.

5) 랜덤포레스트의 특성 중요도는 어떤 특성을 사용한 노드가 랜덤포레스트 내 모든 트리에 걸쳐서 평균적으로 불순도를 얼마나 감소시키는지를 파악하여 측정한다. 큰 특성 중요도 점수를 가지는 변수는 작은 값을 갖는 변수보다 반응변수 예측(제출 여부)에 높은 중요성을 갖고 있음을 의미한다.

증빙서류 제출자와
미제출자 간 근로소득
분포를 비교해보면,
<표 10>과 같이
미제출자 소득의
표준편차(2,800만원)가
제출자(3,079만원)보다
작은 것을
발견할 수 있다.

[그림 3] Random forest variable importance 분석 결과



출처: 재정패널조사 11차 자료를 이용하여 저자 작성

<표 10> 근로소득 연말정산 대상자 중 증빙서류 제출자의 소득분포 특성(11차 기준)

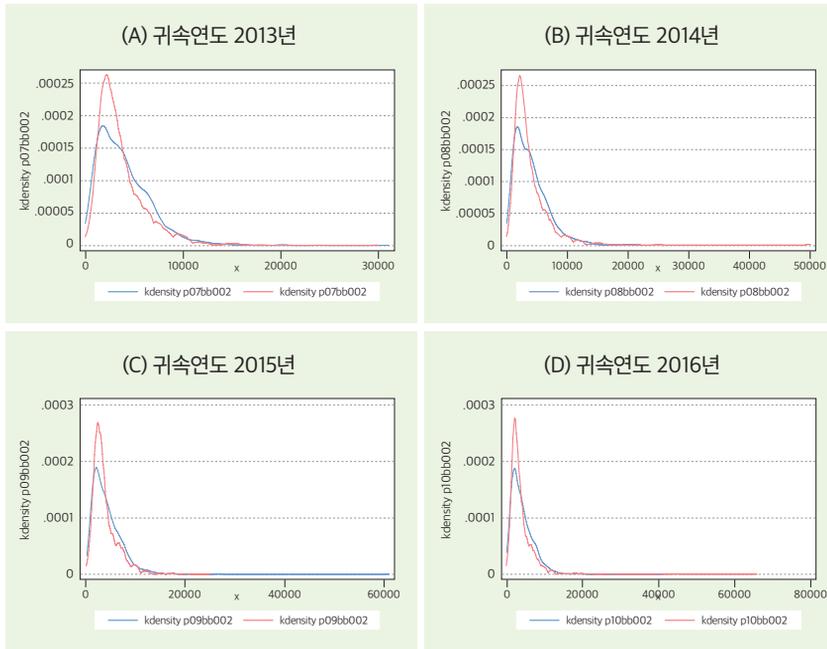
(단위: 만원)

소득분포	제출자	미제출자	차이(제출자-미제출자)
1%	348	600	-252.1
5%	853	1,420	-567.1
10%	1,306	1,800	-494.0
25%	2,114	2,160	-46.1
50%	3,503	3,000	502.6
75%	5,719	4,800	918.9
90%	8,155	7,200	955.2
95%	9,833	9,000	832.8
99%	14,113	13,000	1,112.8
평균	4,263	3,834	429
표준편차	3,079	2,800	279
해당자 수	1,985	1,260	

출처: 재정패널조사 11차 자료를 이용하여 저자 작성

이러한 제출자와 미제출자 간 소득분포의 차이는 [그림 4]에서처럼 11차뿐만 아니라 다른 귀속연도에도 유사하게 발견할 수 있다. 이러한 경향은 소득을 자가 보고할 경우 발견할 수 있는 일반적인 패턴으로도 이해할 수 있으나, 소득 양극단의 경우, 즉 저소득층은 소득을 높게 신고하려는 경향을, 고소득층은 (분포 99% 이상) 소득을 낮게 자가신고하려는 경향을 반영한 결과일 수도 있다.

[그림 4] 근로소득 연말정산 대상자 중 증빙서류 제출자의 소득분포



주: 푸른색이 제출자, 붉은색이 미제출자의 소득분포임
출처: 재정패널조사 11차 자료를 이용하여 저자 작성

증빙서류 제출자와 미제출자 간 소득분포의 차이는 다른 귀속연도에서도 발견할 수 있다. 이는 소득을 자가보고할 때의 일반적인 패턴일수도 있으나, 저소득층은 소득을 높게 신고하고, 고소득층은 소득을 낮게 신고하려는 소득 양극단의 경향을 반영한 결과일 수도 있다.

4. 제출자 및 미제출자 간 신고세액의 여부 및 신고세액의 차이

재정패널조사를 활용한 조사 중 특히 소득공제내역 분석 시 증빙서류 미제출자를 분석에 포함시키느냐, 포함시키지 않느냐는 분석의 신뢰성을 제고하기 위해 가장 중요하게 고려해야 할 문제 중 하나이다. 제11차 조사의 경우, 근로소득 신고 대상이지만 증빙서류를 제출하지 않은 미제출자 1,262명 중 534명(42.3%)이 상세 결정세액을 자가보고하고, 나머지 728명은 실제 결정세액이 있을 가능성이 있음에도 불구하고 기억나지 않거나 모름이라고 응답하였다. 더 상세한 분석을 위해, 증빙서류 제출 여부를 관심변수로, 기타 개인 특성변수를 통제변수로 포함하여 서류 제출 여부에 따른 결정세액, 세액 유무 여부 그리고 신용카드 공제내역 보고 여부의 선형회귀 분석을 실시하였다. 분석 결과 <표 11>과 같이 ① 작년 한 해 근로소득금액에 대한 납부 세금(결정세액)의 경우 제출자가 미제출자에 비해 평균 약 40만원 더 많은 결정세액 금액을 보고하였고, ② 증빙서류 제출자가 결정세

<표 11>을 보면, 정확하게 기억해내기 쉽지 않은 결정세액의 크기나 결정세액 여부, 소득공제 세부사항 등을 분석하는 데 있어 제출/미제출 여부를 필수적으로 고려하여 분석에 임해야 할 필요성을 제시하고 있다.

<표 11> 서류 제출 여부에 따른 결정세액, 세액 유무 여부, 신용카드 공제내역 보고 여부 (귀속연도 2017년 기준)

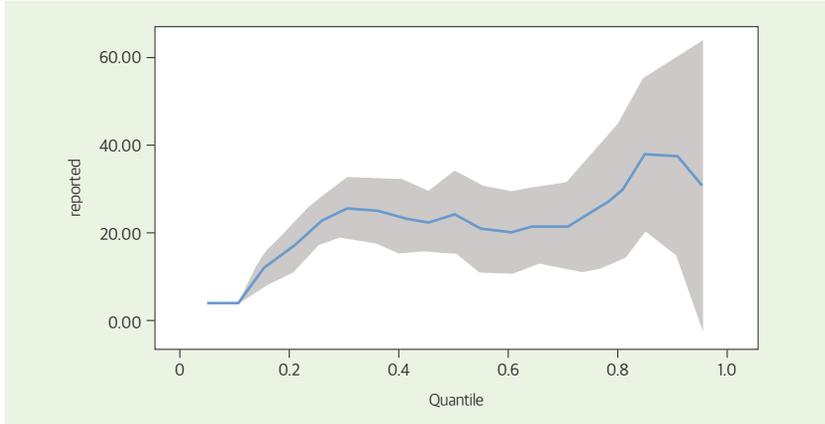
변수명	(1) 근로소득 결정세액	(2) 결정세액(=0) 여부	(3) 신용카드 공제 여부
제출 여부(제출=1)	39.674*** (11.144)	-0.183*** (0.022)	0.197*** (0.023)
나이	1.726*** (0.449)	-0.001 (0.001)	-0.005*** (0.001)
성별	-73.535*** (19.476)	-0.046* (0.025)	0.027 (0.025)
결혼 여부	-88.482*** (20.128)	0.119*** (0.025)	0.04 (0.025)
이혼 여부	-15.085 (18.519)	0.091* (0.051)	-0.006 (0.054)
가구주 여부	-66.735*** (19.145)	0.005 (0.026)	0.006 (0.025)
대졸 이상	-32.556** (16.066)	-0.017 (0.023)	0.012 (0.023)
전문직 사무직 여부	-43.329*** (15.816)	-0.03 (0.020)	0.077*** (0.02)
정규직 여부	-178.812*** (67.693)	-0.127*** (0.042)	0.319*** (0.038)
근로소득(만원)	0.138*** (0.018)	-0.000*** (0.000)	0.000*** (0.000)
상수항	-149.332*** (48.172)	1.009*** (0.059)	0.213*** (0.06)
Observations	2,518	2,518	2,518
R-squared	0.686	0.269	0.177

주: *, **, *** 은 각 10%, 5%, 1% 수준에서 통계적으로 유의함
출처: 재정패널조사 11차 자료를 이용하여 저자 작성

6) 본고의 분석에서는 재정패널 조사의 가중치를 고려하지 않았으나, 가구원 가중치를 고려한 분석에서도 결과는 대동소이 하였다. 참고로 재정패널조사에서는 가구원 가중치를 따로 구하지 않고 가구 가중치를 먼저 산출하고, 이를 가구원 가중치로 사용하는 방식을 따르고 있다.

액이 0인 경우를 보고할 확률이 미제출자에 비해 약 18% 더 낮으며, ③ 증빙서류 제출자가 신용카드 사용금액 소득공제 내역을 보고할 확률이 미제출자에 비해 약 20% 높음을 발견해내었다.⁶⁾ 이는 정확하게 기억해내기 쉽지 않은 결정세액의 크기나 결정세액 여부, 소득공제 세부 사항 등을 분석하는 데에 있어 제출/미제출 여부를 필수적으로 고려하여 분석에 임해야 할 필요성을 제시하고 있다.

[그림 5] 결정세액분포별 제출 및 미제출자 간 결정세액의 차이
(Quantile regression) 결과



출처: 재정패널조사 11차 자료를 이용하여 저자 작성

더 나아가 증빙서류 제출/미제출 여부에 소득의 크기가 영향을 미치는 경우, 소득분포별 제출자/미제출자 간 결정세액 차이, 소득공제 내역에서 회상의 오류 등은 재정패널을 활용한 연구를 진행하거나 이해할 때 더욱 주의를 기울여야 할 사항이 된다. 이에 참고로, 결정세액과 서류제출 여부가 내생적 관계를 지니고 있지만, 분위회귀분석(Quantile regression)을 활용하여 결정세액 분포별로 제출자와 미제출자 간 결정세액 차이의 비선형성을 체크해보았다. 분석 결과, 결정세액이 큰 구간에서는 서류 제출 여부에 따라 결정세액의 차이가 커지나, 결정세액의 사분위 범위(0.2~0.8)에서 증빙서류 제출 여부 차이에 따른 결정세액의 차이는 어느 정도 일정하게 유지됨을 발견할 수 있다.

5. 제출자 및 미제출자 간 조세부담에 대한 인식 차이

마지막으로 조세부담에 대한 인식 정도와 증빙서류 제출 여부 간 상관관계를 분석해보았다. 개인의 조세 부담에 대한 인식이 증빙서류 제출 여부에 영향을 끼칠 수 있다고 가정할 수 있으나, 본인의 경제적 능력과 비슷한 납세자들과 비교하여 세금부담이 높고 낮음을 질문한 설문내역⁷⁾을 활용하여 분석한 결과 직접적인 상관관계를 단정할 수 있는 통계적 유의성을 발견하지 못하였다.

분위회귀분석 결과, 결정세액이 큰 구간에서는 서류 제출 여부에 따라 결정세액 차이가 커지나, 결정세액의 사분위 범위(0.2~0.8)에서는 결정세액의 차이가 어느 정도 일정하게 유지됨을 발견할 수 있다.

7) “귀하는 본인의 경제적 능력과 비슷한 납세자들과 비교하여 세금부담이 높다고 생각하십니까? 아니면 낮다고 생각하십니까?”의 질문에 대해 “매우 많이 납부한다” 혹은 “약간 납부한다”고 응답한 경우를 1로 설정한다.

재정패널조사는
가계금융복지조사와
같은 행정자료 보완이
아닌 설문 참여자의
자발적인 소득증빙서류
제출에 의지하여
상세 항목의 정확성을
보완하고 있는바,
재정패널조사의
신고소득액,
소득공제항목 등에는
응답오차의 가능성이
상존해왔다.

<표 12> 서류 제출 여부에 따른 조세부담 인식의 차이(귀속연도 2017년 기준)

변수명	(1) 세부담이 높다(=1)
제출 여부(제출=1)	0.030* (0.017)
나이	0.001** (0.001)
성별	-0.030 (0.019)
결혼 여부	0.069*** (0.019)
이혼 여부	0.007 (0.041)
가구주 여부	0.073*** (0.019)
대졸 이상	0.042** (0.017)
전문직 사무직 여부	0.021 (0.014)
정규직 여부	-0.043* (0.026)
근로소득(만원)	0.000** (0.000)
상수항	0.713*** (0.045)
Observations	2,518
R-squared	0.041

주: *, **, ***은 각 10%, 5%, 1% 수준에서 통계적으로 유의함
출처: 재정패널조사 11차 자료를 이용하여 저자 작성

IV. 맺음말

재정패널조사는 그동안 납세정보와 함께 납세자들의 자산, 소득, 소비지출 등 종합적이고 포괄적인 정보수집을 통해 가계금융복지조사, 국세청 데이터로는 불가능하였던 조세부담과 복지수혜의 연계성 분석, 조세모의실험 등을 가능하게 하였다. 하지만 가계금융복지조사와 같은 행정자료 보완이 아닌 설문 참여자의 자발적인 소득증빙서류 제출에 의지하여 상세 항목의 정확성을 보완하고 있는바,

재정패널조사의 신고소득액, 소득공제 항목 등에는 응답오차의 가능성이 상존해 왔다. 특히, 증빙서류의 제출/미제출 여부가 가구원의 소득뿐 아니라 개인 특성 과도 체계적인 상관관계를 갖고 있다 판단되며, 따라서 이러한 조사편향은 재정패널을 활용한 연구를 수행하거나 분석 결과를 이해하는 데에 있어 필수적으로 고려해야 할 사항 중 하나이다. 한국조세재정연구원의 재정패널조사팀에서는 향후 증빙서류 제출자의 자가보고 내역 또한 추가 분석변수로 제공할 예정으로, 제출자의 신고내역과 자가보고 내용을 비교하여 미제출자의 소득분포와 세금 및 공제내역에 대한 분석에 시사점을 제공할 수 있을 것으로 기대한다. **KIF**

<참고문헌>

김낙년, 「가계조사의 행정자료에 의한 보정: 2016년 가계금융복지조사를 중심으로」, 『한국사회정책』, 제27권 제1호, 2020, pp. 39~61.

세수추계패널센터, 「소득증빙서류 수집률이 소득분포에 미치는 영향」, 『재정포럼』, 제249호, 2017. 3., pp. 52~64.

이원진·정해식·전지현, 『소득조사 마이크로데이터 비교분석: 가계동향조사와 가계금융복지조사를 중심으로』, 한국보건사회연구원, 2019.

통계청, 「2018년 가계금융복지조사 행정자료 활용」, 보도참고자료, 2018. 12. 20.

_____, 『가계금융복지조사』 통계정보 보고서, 2019., https://kostat.go.kr/portal/korea/kor_pi/8/7/index.board?bmode=read&bSeq=&aSeq=351214&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=srch&sTarget=title&sTxt=%EA%B0%80%EA%B3%84%EA%B8%88%EC%9C%B5, 검색일자: 2020. 5. 7.

Duflo, E. “Richard T. Ely lecture: The economist as plumber,” *American Economic Review*, 107(5), 2017, pp. 1~26.

한국조세재정연구원의 재정패널조사팀에서는 향후 증빙서류 제출자의 자가보고 내역 또한 추가 분석변수로 제공할 예정으로, 제출자의 신고내역과 자가보고 내용을 비교하여 미제출자의 소득분포와 세금 및 공제내역에 대한 분석에 시사점을 제공할 수 있을 것으로 기대한다.