

조세재정 브리프

소득증빙서류와 조사자료 간 교차검증: 재정패널조사를 중심으로

정재현 한국조세재정연구원 부연구위원 / jjung6@kipf.re.kr

- 1 재정패널조사의 개요와 행정자료와의 교차검증의 필요성
- 2 행정자료와 조사자료 간 교차검증에 관한 다른 사례
- 3 재정패널조사의 소득증빙서류와 조사자료 간 교차검증
- 4 맺음말

참고문헌





요약

- ▶ 재정패널조사는 납세정보와 함께 납세자들의 자산, 소득, 소비지출 등 포괄적인 정보수집을 통해 조세부담과 복지수혜의 연계성 분석, 조세 모의실험 등을 가능하게 하였음
- ▶ 재정패널조사의 가구원 설문 중 소득 및 소득공제 내역, 결정세액의 경우, 가구원의 자발적인 관련 증빙서류 제출을 통해 상세항목을 보완하고 있는바, 서류 미제출자의 신고소득액, 소득공제 등의 항목에 응답오차가 발생할 가능성이 있음
- ▶ 증빙서류의 제출·미제출 여부에 따라 근로소득액이나 세액공제 여부 등에 차이가 존재하며, 따라서 이러한 조사편향(bias)은 재정패널을 활용한 연구를 수행하거나 분석결과를 이해하는 데에 있어 필수적으로 고려해야 할 사항 중 하나임

1 재정패널조사의 개요와 행정자료와의 교차검증의 필요성

- ▶ 재정패널조사는 실증증거에 기반한 정책설계(Evidence-based policymaking)의 요구에 부응하기 위해 가구 단위의 조세-지출-복지에 관한 포괄적인 데이터를 제공하고자 2008년부터 시작됨
 - 재정패널조사는 납세정보와 함께 납세자들의 자산, 소득, 소비지출, 복지수혜 정보 등을 포괄하여 이를 바탕으로 조세부담과 복지수혜의 연계성 분석 및 조세 모의실험이 가능함
 - 2008년부터 1년의 조사주기로 실시되어 현재 제12차 조사가 완료되었으며, 2008년 구축된 5,014가구와 2009년에 추가된 620가구를 포함하는 전체 원표본가구수는 5,634가구임
- ▶ 재정패널조사의 가구원의 소득¹⁾ 및 소득공제, 결정세액 현황 등에는 자가응답(self-report) 내용이 일차적으로 기록되며, 보다 정확한 내역을 조사하여 보완하기 위해 가구원의 세금신고 증빙서류 제출을 요청함
 - 가구원 중 근로소득자의 경우 근로소득 원천징수영수증을, 근로소득 외의 소득이 있는 경우 종합소득세 신고를 위한 일체의 신고서류(과세표준확정신고 및 자진납부계산서, 소득공제 명세서 등)의 제출을 요청함

1) 가구원 조사의 대상은 소득이 있거나 소득활동을 하는 만 15세 이상의 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 조사대상이 됨

- 따라서 서류 제출자의 소득 및 세액공제, 결정세액 내역은 제출된 증빙서류를 바탕으로 입력된 값이며, 서류 미제출자의 경우 해당 내역은 본인의 회상에 의한 응답값이 기록됨
- ▶ 소득증빙서류 제출 여부에 따라 소득 및 세액공제 세부내역 등 민감한 조사항목에 대한 응답 내용에 과소·과대 응답 등 오차가 발생할 우려가 상존하므로, 자가보고 내역과 행정자료와의 교차검증이 필요함
- 국세청 자료와 같은 행정자료는 응답오차로부터 자유로워 신뢰성 있는 분석을 수행할 수 있으나, 기타 행정 및 통계자료와의 연계가 쉽지 않아 신고된 소득, 조세부담 외에 가계가 정부로부터 받는 혜택을 종합적으로 파악하기에는 불충분함

KIPF ISSUE PAPER

2 행정자료와 조사자료 간 교차검증에 관한 다른 사례-가계금융복지조사의 경우

- ▶ 통계청의 가계금융복지조사는 민감한 조사항목에 대한 가구의 응답 거부나 과소·과대 응답 등 응답오차를 줄이기 위해 30여 종의 행정자료를 연계해 보완함
 - 행정자료와 조사자료의 개인고유정보(성별, 생년월일, 주소, 성명 등)를 활용한 식별키를 통해 표본대상(가구, 가구원)을 연계하는데, 2017년 기준 표본가구의 가구원 4만 8,590명 중 4만 4,887명(92%)을 연계함
 - 근로소득과 사업·임대소득은 조사자료를 기본으로 사용하나 금융소득, 공적이전소득, 비소비지출은 행정자료를 기준으로 대체 보완함
- ▶ 2017년 기준 소득통계를 행정자료로 보완한 결과, 2017년 가구소득 평균은 5,705만원으로 조사결과(5,195만원)보다 509만원 증가하였으며, 처분가능소득 평균은 4,668만원으로 조사결과(4,251만원)보다 417만원 증가함

〈표 1〉 가계금융복지조사 행정자료 보완 전·후 소득 평균 및 증감값(2017년 기준)

(단위: 만원, %, %p)

구분	평균			증감값		
	보완	조사	차이	보완	조사	차이
가구소득	5,478	5,020	459	4,300	4,040	260
	5,705	5,195	509	4,457	4,200	257
처분가능소득	4,520	4,128	391	3,616	3,374	242
	4,668	4,251	417	3,719	3,471	248

주: 2017년 기준

자료: 통계청 보도자료, 2018. 12. 20, p. 3.



▶ 이원진 외(2019)²⁾는 2016년(귀속연도) 가계금융복지조사의 조사데이터와 행정보완데이터의 소득의 차이를 분석한 결과, 근로·사업소득 등 소득원에 대한 자가보고 내역이 행정자료에 비해 과소 보고된 경향을 찾아냄

- 행정자료 보완 후 근로소득이 존재하는 집단의 비율이 77.1%에서 82.1%로, 근로소득은 연평균 2,748만원에서 2,760만원으로 증가함
- 사업소득은 무응답 사례를 행정데이터로 대체한 결과 사업소득이 존재하는 집단의 비율이 32.2%에서 45.3%로 증가하였고, 연평균 사업소득은 738만원에서 782만원으로 증가함
- 금융소득 또한 연평균 17만원에서 94만원으로 크게 증가하였으며, 조사자료에서는 금융소득이 존재하는 집단의 비율이 12.4%에 불과하였으나 행정데이터로 보완한 후 83.3%로 크게 증가함

▶ 또한 행정데이터로 보완 시 가처분소득의 불평등 정도가 심화됨³⁾

- 행정자료 보완으로 가처분소득 10분위 경계값이 모두 증가하는데, p10은 4.5% 증가하는 반면 p50~p90은 9.1~11.4% 증가함
- 이에 따라 불평등 정도(p90/p10, p90/p50, p50/p10, 5분위배율, 지니계수, 빈곤율) 지표들이 모두 악화됨

〈표 2〉 조사데이터/행정보완데이터 시장소득 및 가처분소득 분포

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	1270	1286	16	1.2	1316	1369	53	4
p30	1700	1772	72	4.3	1684	1783	100	5.9
p40	2079	2214	135	6.5	2007	2155	148	7.4
p50	2477	2652	175	7.1	2336	2550	214	9.1
p60	2900	3165	265	9.1	2689	2971	282	10.5
p70	3464	3745	281	8.1	3173	3485	312	9.8
p80	4158	4568	410	9.9	3766	4175	409	10.9
p90	5400	5911	511	9.5	4742	5281	540	11.4
P90/P10	8.66	8.5	-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 의미, 소득을 가구원 수의 제곱근으로 나누어 균등화한 가구소득의 개인 단위 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

원자료: 통계청, 「2017년 가계금융복지조사」

자료: 이원진 외, 2019, p. 125, 〈표4-4〉 자료 재인용

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

3) 이원진·정해식·전지현, 2019, p. 126.

3 재정패널조사의 소득증빙서류와 조사자료 간 교차검증

가. 소득증빙서류 제출 여부

- ▶ 근로소득자의 약 60%가 소득증빙서류를 제출하고 있으며, 종합소득 신고대상자의 경우 평균 41%가 관련 소득증빙서류를 제출하고 있음
 - 근로소득증빙서류 수집률은 재정패널 1차년도(2007년 귀속)에 26.8%(근로소득 연말정산 대상자 2,387명 중 639명 제출)를 기록한 후, 6차년도(2012년 귀속)까지 지속적으로 증가하여 그 이후로는 60%대에 머물고 있으며, 종합소득의 경우 6차년도 이후 수집률은 평균 41%대에 머물고 있음
 - 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 낮음
- ▶ 제출자는 계속해서 제출하고 미제출자는 반대의 경향을 보여, 소득 수준, 납세 성향 등 응답자 개인 특성에 따라 증빙서류 제출 여부가 좌우될 수 있음을 가정할 수 있음
 - 귀속연도 2013년(제7차)부터 2017년(제11차)까지 연속으로 재정패널조사에 참여하고, 근로소득 연말정산 대상자인 1,939명 중 1,137명(58.6%)이 매 조사 시 증빙서류를 제출함
 - 반면 같은 기간 근로소득 연말정산 서류를 한 번도 제출하지 않은 경우는 21.4%로 조사됨

나. 제출자·미제출자의 개인별 특성 비교

- ▶ 증빙서류 제출자의 증빙서류상 근로소득(총급여)이 미제출자의 자가보고에 비해 평균 429만원 더 높은 것으로 조사됨
 - 최신 제11차 조사(2017년 귀속자료) 기준 근로소득을 지니면서 연말소득 대상으로 분류된 가구원 3,253명 중 근로소득 연말정산 서류 제출자(1,992명)를 미제출자(1,261명)와 비교함
 - 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 높음
 - 이는 제출자가 가구주 본인 혹은 배우자이며 설문에 직접 참여하여 증빙서류도 함께 제출한 경우가 많은 반면, 미제출자는 가구주의 자녀(즉, 가구주의 대리응답)인 경우가 많은 것을 의미함
 - 반면 학력, 성별의 차이는 크지 않은 것으로 나타남
- ▶ 소득 분포의 경우, 미제출자 소득의 표준편차(2,800만원)가 제출자(3,079만원)보다 작음
 - 미제출자의 실제 소득분포가 다른 것인지 아니면 자가보고 경향에 의해 좀 더 편차가 작게 나온 것인지는 행정자료와 자가응답 자료를 모두 갖고 있어야만 판단 가능함



〈표 3〉 근로소득 연말정산 대상자 중 증빙서류 제출자의 특성

특성 변수	미제출자		제출자와의 차이(통계적 유의성)
	평균	표준편차	
나이	42.013	(11.93)	3.645***
성별	0.619	(0.486)	-0.033*
결혼 여부	0.63	(0.483)	0.161***
이혼 여부	0.023	(0.15)	0.017***
가구주 여부	0.494	(0.5)	0.132***
대리응답 여부	0.55	(0.498)	-0.265***
대학교 이상 학력	0.699	(0.459)	-0.017
전문직 사무직	0.504	(0.5)	0.017
정규직 여부	0.929	(0.256)	0.018**
근로소득(만원)	3,834.434	(2,800.89)	429.045***
해당자 수	1,261		3,253

주: 1. 제11차 기준

2. *, **, *** 은 각 10%, 5%, 1% 수준에서 통계적으로 유의함

자료: 재정패널조사 제11차를 이용하여 저자 작성

▶ 조세부담에 대한 인식 정도와 증빙서류 제출 여부 간 직접적인 상관관계는 없는 것으로 판단됨

- 개인의 조세부담에 대한 인식이 증빙서류 제출 여부에 영향을 끼칠 수 있다고 가정할 수 있으나, 경제적 능력이 본인과 비슷한 납세자들과 비교하여 세금 부담이 높고 낮음을 질문한 설문내역을 활용하여 분석한 결과, 직접적인 상관관계를 판단할 수 있는 통계적 유의성을 발견하지 못함

▶ 제출·미제출 여부를 결정하는 데에 어떤 특성 변수가 가장 큰 영향을 끼치는지 머신러닝의 Random Forest로 분석해 본 결과 소득이 제일 큰 영향을, 그다음으로는 나이, 대리응답 여부, 가구주 여부 순으로 영향을 끼치는 것으로 판단됨

- 물론 소득은 미제출자의 경우 자가보고한 내역이므로 증빙서류 제출 여부에 영향을 받는 내생변수일 가능성도 상존함

다. 제출자·미제출자 간 신고 세액 및 공제내역 차이

▶ 특히 소득공제 내역 분석 시, 증빙서류 미제출자를 분석에 포함시키느냐 마느냐는 재정패널을 활용한 분석의 신뢰성을 제고하기 위해 중요하게 고려해야 할 문제 중 하나임

- 가령 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 응답함

▶ 또한 소득 분포별 제출자·미제출자 간 결정세액 차이, 소득공제 내역에서 회상의 오류 등 또한 재정패널을 활용한 연구를 진행하거나 이해하는 데 있어 신중하게 고려해야 할 사항임

- 작년 한해 근로소득금액에 대한 납부 세금(결정세액)의 경우, 제출자가 미제출자에 비해 평균 약 40만원 더 많은 결정세액 금액을 보고함(표 4)의 (1)

- 증빙서류 제출자가 결정세액이 0인 경우를 보고할 확률이 미제출자에 비해 약 18% 더 낮음(〈표 4〉의 (2))
- 증빙서류 제출자가 신용카드 사용금액 소득공제 내역을 보고할 확률이 미제출자에 비해 약 20% 높음(〈표 4〉의 (3))

〈표 4〉 서류 제출 여부에 따른 결정세액, 세액 유무 여부, 신용카드 공제내역 보고 여부

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)
대출 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: 1. 귀속연도 2017년 기준

2. *, **, ***은 각 10%, 5%, 1% 수준에서 통계적으로 유의함

자료: 재정패널조사 제11차를 이용하여 저자 작성

4 맺음말

- ▶ 재정패널조사는 그동안 납세정보와 함께 납세자들의 자산, 소득, 소비지출 등 종합적이고 포괄적인 정보를 수집해 가계금융복지조사나 국세청 데이터로는 불가능하였던 조세 모의실험 등을 가능하게 하였음



- ▶ 하지만 행정자료 보완이 아닌 설문 참여자의 자발적인 소득증빙서류 제출에 의지하여 상세항목의 정확성을 보완하고 있는바, 재정패널조사의 신고소득액, 소득공제항목 등에서 발생 가능한 응답 오차에 주의를 기울여야 함
- ▶ 특히 증빙서류의 제출·미제출 여부가 가구원의 소득, 지출 등 개인 특성과 체계적인 상관관계를 갖고 있다고 판단되며, 따라서 이러한 조사편향(bias)은 재정패널을 활용한 연구를 수행하거나 분석결과를 이해하는 데에 있어 필수적으로 고려해야 할 사항 중 하나임
- ▶ 재정패널조사팀에서는 향후 증빙서류 제출자의 자가보고 내역 또한 추가 분석 변수로 제공할 예정으로, 제출자의 신고내역과 자가보고 내역을 비교하여 미제출자의 소득분포와 세금 및 공제내역 분석에 시사점을 제공할 수 있을 것이라 판단됨

[참고문헌]

- 이원진 · 정해식 · 전지현, 『소득조사 마이크로데이터 비교분석: 가계동향조사와 가계금융복지조사를 중심으로』, 한국보건사회연구원, 2019.
- 통계청, 「2018년 가계금융·복지조사 행정자료 활용」, 보도참고자료, 2018. 12. 20.