연구논문

노동패널 표본이탈과 불평등 지표*

The Attrition of KLIPS and Inequality Index

홍민기[a] · 최효미[b]
Minki Hong · Hyo-mi Choi

본 논문에서는 노동패널 자료에서 표본이탈의 형태와 표본이탈이 추정에 미치는 영향을 분석하였다. 노동패널 11차까지 표본유지율은 74.2%이며, 약 20%의 가구가 응답과 비응답을 오가고 있다. 소득 최상위 가구에서는 무응답을 이유로, 소득 최하위 가구에서는 소멸을 주된 이유로 하여 표본에서 탈락하는 확률이 높다. 가구는 소득분포의 양쪽에서 이탈이 많이 일어나고, 개인의 경우에는 저소득 쪽에서 이탈이 많이 일어난다. 표본이탈이 불평등 지표에 미치는 영향을 분석한 결과, 표본이탈을 고려하지 않고 응답한 가구만을 대상으로 추정하면 불평등도를 과소평가하게 된다는 것을 발견하였다. 패널 가중치를 설정할 때 표본이탈의 사유를 고려하고 응답 성향을 반영하는 것이 바람직하다.

주제어: 표본이탈, 노동패널, 무응답, 소멸, 불평등도

This paper examines the pattern of attrition and the effect on estimations of Korea Labor and Income Panel Study. The retention

* 이 논문은 노동연구원의 보고서 "패널자료 품질개선 연구(Ⅱ)" 중 제4장을 수정·보완한 것이다.

a) 교신저자(corresponding author): 한국노동연구원 연구위원 홍민기.
E-mail: hminki@kli.re.kr

b) 육아정책연구소 부연구위원

samples account for 74.2% up to 11 waves. About 20% of households belongs to in and out samples. The higher income households are likely to be out of sample due to non-responses, while the lower income households are likely due to extinction. The attrition is likely to happen in the both side of distribution of income for households and in the left side of distribution for individuals. When we do not consider the attrition correctly, the degree of inequality is underestimated.

Key words: attrition, KLIPS, non-response, extinction, inequality

Ⅰ. 머리말

패널조사는 동일한 표본을 계속 조사하는 것이므로 시간이 지남에 따라 조사대상이 어떻게 변화하는지를 추적할 수 있는 장점이 있다. 패널자료를 이용하면 상태변이 행렬(transition matrix)을 쉽게 계산할 수 있고, 패널분석을 통해 관측 불가능한 변수들을 효과적으로 통제할 수 있다는 장점도 있다. 이러한 이유에서 국내에서 지난 10여 년 동안 패널조사가 급격히 증가하였다(이희길 2008).

한편, 패널조사는 동일한 대상을 반복적으로 조사하기 때문에 패널유지와 관리에 많은 비용과 노력이 요구된다. 패널조사의 특성상 표본이탈(attrition, 혹은 패널유실)은 가장 핵심적인 문제로 부각된다. 표본이탈이 무작위적(random)으로 발생한다면 문제가 적을 수 있지만 표본이탈이 무작위적이지 않다면 표본의 대표성을 유지하기 어렵게 된다. 패널표본의 대표성이 의심되면 패널자료를 통한 분석의 대표성도 의심을 받게 된다.

표본이탈은 우선 가중치의 조정을 통해 해결할 수 있고, 표본이탈로 표본

의 대표성이 심각하게 훼손되면 표본을 추가하거나 표본을 개편하기도 한다. (신동균 1998) 그런데 표본이탈의 문제점을 해결하기 이전에 표본이탈의 형태를 먼저 파악하여, 만약 표본이탈을 고려하지 않으면 분석에 어떠한 문제점이 생기는지 파악할 필요가 있다. 문제점을 먼저 파악하여야 가중치를 어떻게 조정할 것인가, 어떻게 표본을 추가할 것인가도 정확하게 알 수 있기 때문이다. 이 논문에서는 한국노동패널조사(Korean Labor and Income Panel Study: KLPS, 이하 노동패널) 자료에서 이탈하는 표본의 특성을 분석하고, 표본이탈(attrition)이 불평등 지수에 미치는 영향을 살펴보고자 한다.

본 논문이 기존의 국내 연구와 다른 점은 크게 두 가지이다. 첫째, 본 논문에서는 표본이탈 이유를 소멸과 비응답으로 나누어 분석한다. 김대일 외(2000)와 이상호(2005) 등 기존 국내 연구에서는 비응답과 소멸을 구분하지 않고 한데 묶어서 표본이탈과 표본특성과의 관계를 살펴보았다. 소멸과 비응답가구의 특성은 서로 매우 달라서 표본이탈 사유를 나누어서 고려하는 것이 매우 중요하다는 점을 이 논문에서 밝히고자 한다.

둘째, 본 논문에서는 표본이탈이 분포에 미치는 영향을 분석한다. 이상호(2005)의 연구에서는 표본이탈이 회귀분석에는 영향을 거의 주지 않는다고 한 바 있다. 회귀분석은 평균에 대한 영향을 추정하는 것이다. 표본이탈이 평균에는 영향을 주지 않지만 분포에는 영향을 줄 수 있다. 소멸가구와 비응답가구는 소득 분포에서 매우 다른 곳에 위치해 있다. 소멸가구는 매우 불평등도가 높고 저소득 가구인 반면, 비응답가구는 평균보다 약간 고소득 가구이다. 소멸때문에 소득 분포의 왼쪽에서 이탈이 발생하고 비응답으로 소득 분포의 오른쪽에서 이탈이 발생하여, 전체적으로 소득분포가 가운데로 몰리게 된다. 표본이탈이 평균값의 측정이나 추정에는 상대적으로 적게 영향을 줄 수 있지만 분포의 측정에는 영향을 줄 가능성을 고려하여 본 논문에서는 표본이탈이 불평등 척도에 미치는 영향을 분석한다.

본 연구에서는 한국노동패널 11차 조사까지를 분석대상으로 하여 표본이탈 가구나 개인의 특성을 분석한다. 노동패널은 국내에서 가장 오래된 패널조사이다. 노동패널에서 발생하는 표본이탈의 현황을 보고하고 분석하는 것이 다른 패널조사에도 도움을 줄 것으로 기대한다.

다음 2장에서는 기존 연구를 개괄한다. 3장에서는 표본이탈을 정의하고, 4장에서는 가구 및 개인의 특성과 표본이탈과의 상관관계를 분석한다. 5장에서는 표본이탈이 임금식의 추정과 불평등도의 측정에 어떠한 영향을 미치는지 분석하고, 마지막 6장은 결론이다.

Ⅱ. 기존 연구

국내 패널조사의 현황에 대해서는 이희길(2008)의 연구에서 매우 상세하게 정리하였다. 이희길의 연구에서는 국내의 모든 패널자료를 대상으로 표본추출, 패널관리방법, 패널자료의 활용 현황을 보여주고 있다. 본조사가 2차 이상 진행된 9개 패널을 대상으로 패널유지율을 분석한 결과 개인패널의 경우 패널유지율이 높고 가구패널은 상대적으로 낮은 편임을 보여주었다. 패널유지의 경향에 대한 강석훈(1998)의 연구에서는 외국 패널조사의 경우 2차 조사에서 비교적 높은 탈락률을 보이지만 세 번째 조사부터는 안정화되는 경향이 있다고 하였다.

가구나 개인의 특성이 표본이탈에 미치는 영향을 분석한 기존 연구에서는 고소득층 가구의 이탈률이 높고, 경제 상태나 혼인 상태의 변화를 겪는 가구나 개인일수록 이탈율이 높다는 점을 보고하였다. 예를 들어, Hausman & Wise (1979)과 Fitzgerald et al.(1998)은 고소득층의 이탈률이 높다는 사실을 관찰하였다. 그리고 근로소득의 변동폭이 크거나 혼인상태의 변화(이혼, 별거, 배우자의 사망)를 겪으면 이탈할 확률이 높아진다는 점은 Lillard & Panis (1998)에 의해 보고되었다. 그리고 Zabel(1998)과 Hill & Wills(2001)은 면접원의 영향을 연구하여, 조사시간이 짧을수록 그리고 동일한 면접원이 계속 조사할수록 표본이탈이 줄어든다는 것을 보여주었다.

국내에서 김대일 외(2000)의 연구에서는 경제활동인구조사와 대우패널자료의 표본이탈을 분석하였다. 이들은 연령이 낮을수록, 학력이 높을수록, 이혼하였거나 미혼인 경우, 임금근로자인 경우, 실업자이거나 미취업인 경우 표

본이탈율이 높다는 것을 관찰하였다. 노동패널 7차 조사까지 분석한 이상호(2005)의 연구에서는 고소득층과 실업자에서 이탈율이 높았으며, 비무작위적 이탈에도 불구하고 표본이탈이 가구소득 추정에 심각한 영향을 주지는 않는다는 것을 보여주었다.

김세원・정익중(2012)은 서울아동패널 자료를 사용하여 단위무응답(unit nonresponse)과 항목무응답(item nonresponse) 간의 관계를 분석하였다. 이들의 연구에서는 1차 년도 조사에서 한 가지 이상의 설문문항에 응답을 하지 않았을 경우, 특히 인적사항에 대한 질문에 응답을 하지 않았을 경우 추적조사에 참여하지 않을 가능성이 높다는 사실을 보여주었다. 1차 조사에서 항목무응답을 감소시키는 것이 이후 패널유지에 영향을 준다는 것을 의미한다.

항목무응답이 단위무응답에 영향을 줄 수 있다는 점이 연구되었으므로 이와의 관련성을 살피는 것이 필요하지만 본 연구에서는 항목무응답과 표본이탈의 관련성은 검토하지 않는다. 본 연구의 분석 대상은 완전한 표본이탈이며 항목무응답은 제외된다.

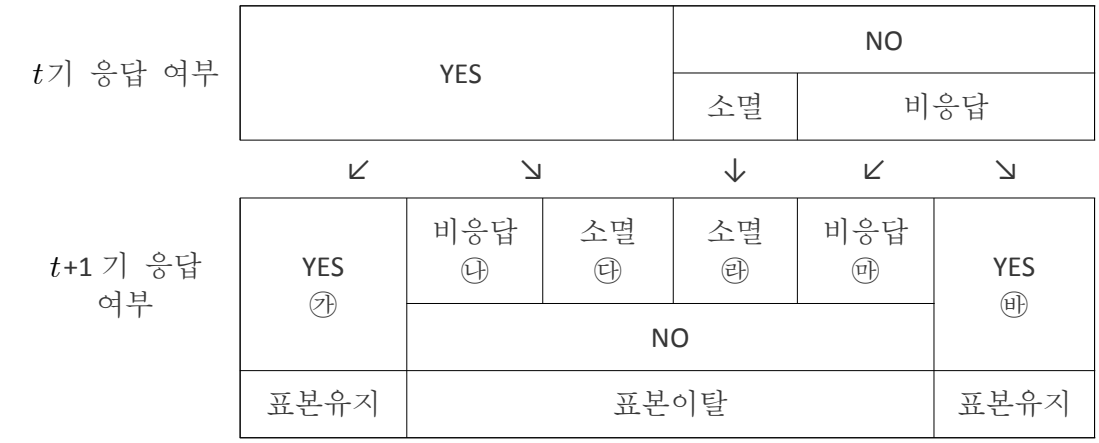
표본이탈을 보정하는 계량경제학적 방법과 통계학적 방법에 대해서는 Wooldridge(2010) 제19장과 홍민기 외(2014)의 제4장과 제5장을 참고할 수 있다.

Ⅲ. 자료와 표본이탈의 정의

이 연구에서 사용하는 자료는 한국노동패널이다. 노동패널 조사는 도시 지역에 거주하는 5,000가구와 가구원을 대상으로 1998년부터 가구의 특성, 가구원의 경제활동, 소득 및 소비활동 등을 추적・조사하는 패널조사이다. 본 연구에서는 11차 조사까지를 분석대상으로 한다.

특정 조사년도 t기에 가구는 응답형태에 따라 표본유지와 표본이탈로 나뉜다. 조사에 응하였으면 표본유지, 조사에 응하지 않았으면 표본이탈이 된다. 표본유지는 다시 계속응답과 다시응답으로 나뉜다. 계속응답 가구는 전년

도에도 응답하고 당해 년도에도 응답한 가구이다. 다시응답 가구는 전년도에는 응답하지 않았지만 당해 년도에 응답한 가구를 말한다. 표본이탈 가구는 이탈 사유에 따라 소멸과 비응답으로 나뉜다. 소멸가구는 사망, 합가, 이민 등으로 조사가 불가능한 가구이다. 비응답가구는 당해 년도에 조사가 가능하였지만 응답하지 않은 가구를 말한다. 응답형태에 따른 가구의 구분과 설명은 <그림 1>과 <표 1>에 정리하였다.



〈그림 1〉 응답형태에 따른 가구의 구분

<표 1> 응답형태에 따라 구분된 가구에 대한 설명

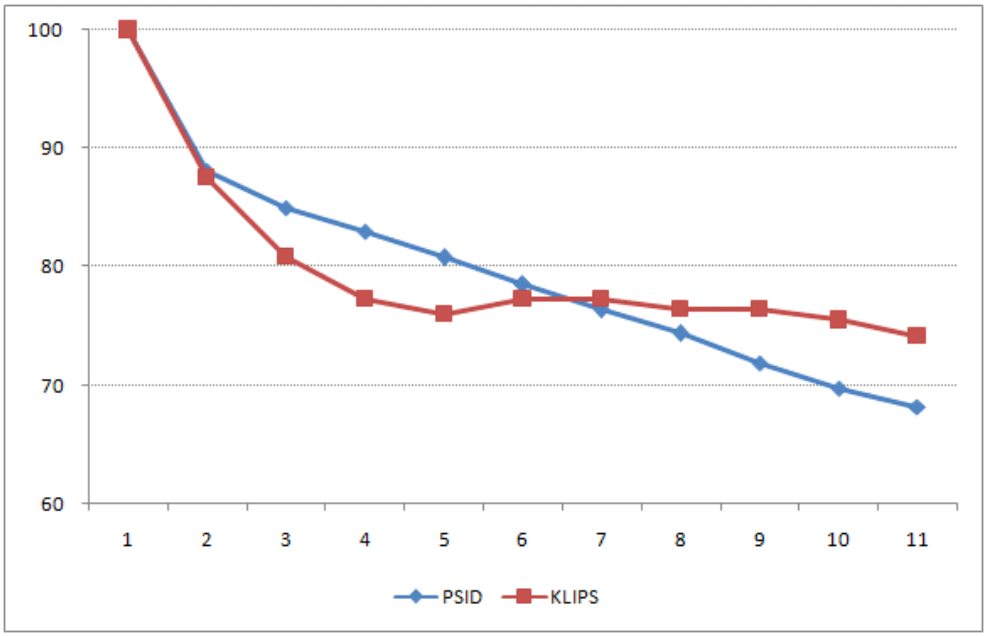
가구 명칭	설명	그림 1
표본유지	당해 년도에 응답한 가구. 전년도에 비응답하였지만 당해 년도에 응답한 가구를 포함함.	㉮+㉳
계속응답	전년도에 응답하고 당해 년도에도 응답한 가구	㉮
소멸	사망, 합가, 이민 등으로 조사가 불가능한 가구	㉰+㉱
비응답	당해 년도에 조사가 가능하지만 응답하지 않은 가구	㉯+㉲

<표 2> 노동패널의 표본유지 및 이탈 가구 분석

차수	년도	표본 유지		표본 이탈				전년 비응답 →응답	이탈가구 중 다시응답 비율(%)
		표본 유지	유지율 (%)	표본이탈 전체(누적)	전년응답 →비응답	전년 비응답 →비응답	소멸 (당해년)		
		㉮+㉲		㉯+㉰+㉱+㉳	㉯	㉳	㉰+㉱	㉲	
1 차	1998 년	5,000	100.0	-	-	-	-	-	
2 차	1999 년	4,378	87.6	622	622	-	0	-	
3 차	2000 년	4,044	80.9	956	483	473	0	149	24.0
4 차	2001 년	3,866	77.3	1,134	430	704	0	252	26.4
5 차	2002 년	3,798	76.0	1,202	290	912	0	222	19.6
6 차	2003 년	3,862	77.2	1,138	271	867	0	335	27.9
7 차	2004 년	3,863	77.3	1,137	222	915	50	223	19.6
8 차	2005 년	3,822	76.4	1,178	219	959	15	178	15.7
9 차	2006 년	3,820	76.4	1,180	185	995	9	183	15.5
10 차	2007 년	3,775	75.5	1,225	163	1,062	27	118	10.0
11 차	2008 년	3,709	74.2	1,291	182	1,109	37	116	9.5

주: '표본유지' 가구는 1차 년도 조사 당시 원표본이었던 5,000가구 중 해당 연도에 조사가 성공적으로 이루어진 가구를 의미함. '표본유지율'은 1차 년도에 조사에 성공한 원표본 5,000가구 중에서 해당 연도에 조사 성공한 원표본가구 비율임.

<표 2>에서는 노동패널 조사가구의 표본 유지 및 이탈 형태를 분석한 결과를 보여주고 있다. 노동패널 원표본 가구 중 11차 년도에 응답한 가구는 3,709가구로 표본유지율은 74.2%이다. 참고로 미국 PSID(Panel Study of Income Dynamics)의 11차 조사에서 표본유지율은 68.2%로, 노동패널의 표본유지율이 약간 높다. PSID 자료 분석의 단위는 개인이기 때문에 표본단위가 가구인 노동패널과 직접적으로 비교하기는 어려우나 표본이 유지되는 패턴이 달라서 그 이유를 살펴볼 필요가 있다.



〈그림 2〉 노동패널(KLIPS)과 PSID의 표본유지율 비교

(주) 가로축은 조사차수, 세로축 단위는 %임.

<그림 2>에서는 노동패널과 PSID의 표본유지율을 비교하여 보여주고 있다. PSID의 표본유지율은 2차 년도에 급속히 감소하였고 이후에도 거의 선형적으로 꾸준히 감소하였다. 반면 노동패널의 표본유지율은 2차와 3차에 급격히 감소한 후 이후 감소세가 둔화되면서 4차 조사 이후에는 거의 변동하지 않았다. 그 결과 6차 조사까지는 노동패널의 표본유지율이 PSID보다 낮았지만 7차 조사부터는 오히려 더 높아졌다.

두 조사에서 이러한 형태가 나타나게 된 것은 표본에서 이탈한 가구(개인) 중 다시 응답을 하여 표본에 들어온 가구(In from Nonresponse) 비율에서 크게 차이가 나기 때문이다. PSID에서는 한 번 이탈한 가구를 추적하여 다시 조사하려는 노력을 거의 하지 않았기 때문에, 이탈하였다가 다시 들어온 가구의 비율이 매 조사마다 약1% 정도로 매우 낮다. 반면 노동패널에서는 이탈한 가구가 다시 표본에 들어온 비율이 약 20%로 매우 높다. 이러한 결과로 노동패널의 표본유지율이 7차 조사부터 PSID보다 높게 되었다.

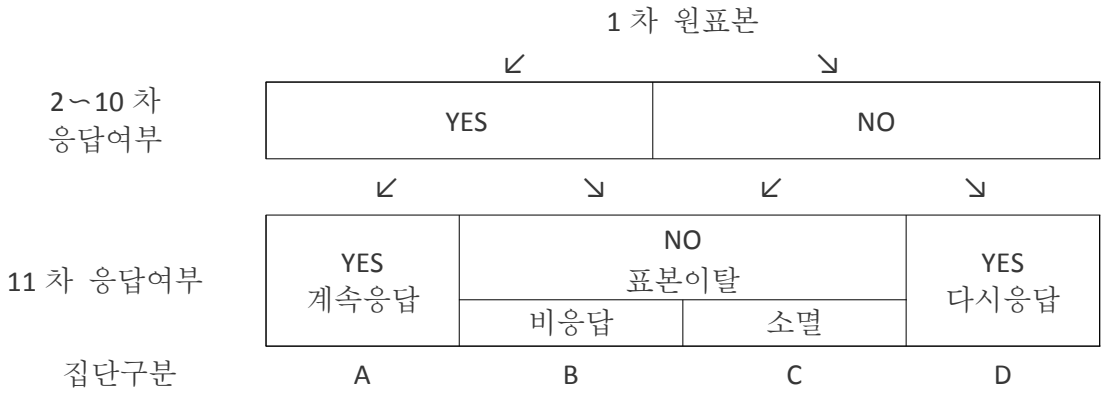
한편, 노동패널에서 1차 년도부터 11차 년도까지 계속응답 가구는 2,607가구로 원가구 표본의 52.1%를 차지하는데, 계속응답 비율은 조사 초반에 급격히 줄어드는 경향을 보이고 이후에는 감소세가 둔화되는 양상을 보여주고 있다. 조사 초반에 계속응답을 한 가구는 이후에 표본 이탈 가능성이 매우 낮다는 것을 의미한다.

앞서 살펴본 바와 같이 원표본유지율은 75% 내외를 유지하고 있는데, 계속응답가구 비중은 52.1%이므로, 약 20%의 가구가 응답과 비응답을 오가고 있음을 알 수 있다. 응답과 비응답을 오가는 가구의 비중이 높다고 할 수 있다.

Ⅳ. 가구 및 개인의 특성과 표본이탈과의 관계

1. 가구의 특성과 표본이탈과의 관계

이 장에서는 어떠한 가구들이 표본에서 이탈을 하는지 그리고 이탈한 가구들의 특성이 남아 있는 가구들의 특성과 얼마나 다른가를 파악하고자 한다. 이를 위해 1차 년도 자료와 11차 년도 자료를 비교하여 11차 년도의 응답여부를 기준으로 가구를 구분하고 이러한 구분을 이용하여 분석을 한다. 이러한 기준으로 가구를 (A) 계속응답 (B) 비응답 (C) 소멸 (D) 다시응답의 네 가지 유형으로 나눈다.



〈그림 3〉 응답방식에 따른 가구 구분 (1차와 11차 년도)

<표 3> 표본유지 및 이탈 가구의 분포 (1차와 11차 자료를 비교)

	계속응답	비연속응답						
			비응답	소멸				다시응답
	A	B+C+D	B	소멸전체 C	사망	합가	원가구 없음	D
가구수	2,607	2,393	1,153	138	99	3	36	1,102
%	52.1	47.9	23.1	2.8	(71.7)	(2.2)	(26.1)	22.0

주: 괄호 안은 'C 집단'을 100으로 보고 산출한 값임.

'계속응답' 가구는 1차 년도부터 11차 년도까지 한 번도 빠짐없이 모두 응답한 가구이다. 11차 년도에 응답을 하였지만 2차부터 10차 년도 사이에 한 번이라도 비응답한 적이 있는 가구는 '다시응답' 가구라고 부를 것이다. 그리고 11차 년도에 비응답한 가구는 표본이탈 가구에 속한다. 표본이탈 가구는 이탈한 이유에 따라 다시 비응답(B)과 소멸(C) 가구로 나뉜다.

<표 3>에는 이러한 기준에 따라 분류된 노동패널 원표본 가구의 분포가 나와 있다. 1차 년도부터 11차 년도까지 계속응답한 가구는 2,607가구로 원표본의 52.1%를 차지한다. 2차부터 10차 년도 사이에 한번이라도 비응답한 적이 있다가 11차 년도에 다시 응답한 가구는 1,102가구로 22.0%에 달한다.

11차 년도에 표본에서 이탈한 원가구는 총 1,291가구(25.8%)이며, 이 중에서 비응답가구가 1,153가구(23.1%), 소멸가구가 138가구(2.8%)이다. 소멸사유별로는 원가구원 사망이 99가구(71.7%)로 압도적으로 많으며, 원가구원 전원이 이민가거나 더 이상 응답할 수 없게 됨에 따라 소멸된 경우가 36가구(26.1%), 다른 원가구로 합가함으로써 소멸된 경우가 3가구이다.

1차 년도 소득별로 응답 및 이탈 가구의 분포가 어떻게 다른지는 <표 4>에 나타나 있다. 이 표는 가구총소득과 근로소득의 분위별로 응답 및 이탈 가구의 분포를 계산한 결과이다. 여기서 가구총소득은 가구근로소득, 금융소득, 부동산소득, 이전소득, 사회보험소득, 기타소득을 모두 합한 값이며, 가구근로소득은 가구원들의 근로소득을 모두 합한 것이다.

〈표 4〉 소득 분위별 가구의 분포 (단위: %)

가구형태		5분위	4분위	3분위	2분위	1분위	전체
가구총소득							
계속응답	A	44.6 (0.024)	52.2 (0.005)	50.5 (0.006)	59.1 (0.011)	54.3 (0.033)	52.1 (0.017)
비응답	B	27.4 (0.028)	25.3 (0.007)	24.2 (0.008)	18.9 (0.017)	19.7 (0.073)	23.1 (0.024)
소멸	C	2.0 (0.090)	0.5 (0.060)	0.9 (0.040)	1.9 (0.056)	8.6 (0.099)	2.8 (0.138)
다시응답	D	26.1 (0.027)	22.1 (0.008)	24.5 (0.008)	20.2 (0.016)	17.4 (0.079)	22.0 (0.025)
응답	A+D	70.7 (0.018)	74.3 (0.004)	75.0 (0.005)	79.3 (0.009)	71.7 (0.031)	74.1 (0.014)
이탈	B+C	29.4 (0.027)	25.8 (0.007)	25.1 (0.008)	20.8 (0.016)	28.3 (0.060)	25.9 (0.024)
가구근로소득							
가구형태		5분위	4분위	3분위	2분위	1분위	전체
계속응답	A	46.7 (0.015)	53.1 (0.005)	51.9 (0.005)	56.2 (0.013)	52.8 (0.040)	52.1 (0.016)
비응답	B	26.6 (0.020)	25.2 (0.007)	23.8 (0.008)	20.0 (0.021)	19.9 (0.092)	23.1 (0.021)
소멸	C	1.7 (0.062)	0.7 (0.057)	0.7 (0.037)	2.1 (0.059)	8.7 (0.111)	2.8 (0.121)
다시응답	D	25.1 (0.018)	21.1 (0.008)	23.7 (0.008)	21.8 (0.018)	18.7 (0.096)	22.0 (0.021)
응답	A+D	71.8 (0.012)	74.2 (0.004)	75.6 (0.004)	78.0 (0.011)	71.5 (0.037)	74.1 (0.013)
이탈	B+C	28.3 (0.019)	25.9 (0.007)	24.5 (0.007)	22.1 (0.020)	28.6 (0.070)	25.9 (0.021)

(주) 괄호안은 표준오차임. 가구총소득은 가구근로소득, 금융소득, 부동산소득, 이전소득, 사회보험소득, 기타소득을 모두 합한 금액에 로그를 취한 값임. 가구근로소득은 가구근로소득(가구원 전체)의 합에 로그를 취한 값임.

소득분위별로 살펴보면, 소득 최상층인 5분위와 최하층인 1분위에서 표본이탈 비율(각각 29.4%, 28.3%)이 다른 분위에 비해 높다. 5분위에서 표본이탈 비율이 높은 이유는 비응답의 비율이 높기 때문이다(27.4%). 반면, 1분위에서 표본이탈 비율이 높은 이유는 소멸의 비율이 상대적으로 높기 때문이다(8.6%). 이는 소득 최하위층인 1분위에 고령 1인 가구가 많이 포함되어 있음이 반영된 것이다. 가구근로소득만을 기준으로 하여 소득분위를 나눈 경우에도 유사한 표본이탈 패턴을 보인다.

<표 5>는 가구원 수에 따라 응답의 형태가 어떻게 다른가를 나타낸 것이다. 표본이탈 가구의 비중은 1인 가구에서 매우 높으며 그 이유는 소멸로 인한 표본이탈의 비중이 다른 가구에 비해 높다는 것이다.

한편, 계속응답 가구의 가구원 수 구성을 살펴보면, 6인, 2인, 5인 가구 순으로 응답 비중이 높다. 이러한 현상은 원가구원 중 한명이라도 응답하면, 원가구의 조사가 성공한 것으로 간주하는 노동패널 표본 관리의 특징이 반영된 결과이다. 특히 2인 가구의 계속응답 확률이 높다는 점이 눈에 띄는데, 이는 2인 가구의 경우 자녀와 함께 살지 않는 노인 가구일 가능성이 높고 노인 가구의 경우 주거지 이주가 많지 않아 접촉이 용이할 뿐 아니라 응답 확률이 높은 특징이 있기 때문이다.

〈표 5〉 가구원 수별 응답 분포 (단위: %, 명)

응답형태 /가구원 수	1인	2인	3인	4인	5인	6인	평균
계속응답 A	40.0	58.2	50.8	51.6	56.0	59.3	3.6
비응답 B	24.4	19.9	25.4	25.1	19.1	16.4	3.4
소멸 C	13.4	4.3	1.4	1.0	1.3	0.4	2.1
다시응답 D	22.2	17.7	22.5	22.4	23.6	24.0	3.5
응답 A+D	62.2	75.9	73.3	74.0	79.6	83.3	3.6
이탈 B+C	37.8	24.2	26.8	26.1	20.4	16.8	3.3

<표 6> 가구의 특성과 표본이탈의 관계

		계속응답(A)	비응답(B)	소멸(C)	다시응답(D)
가구특성					
평균 가구원 수(명)		3.6	3.4	2.1	3.5
가구원 수(%)	1 명	7.9	10.9	50.0	10.4
	2 명	14.6	11.3	20.3	10.5
	3 명	18.8	21.3	9.4	19.7
	4 명	37.4	41.1	13.0	38.5
	5 명	15.0	11.5	6.5	14.9
	6 명 이상	6.3	3.9	0.7	6.0
평균가구총소득(만원)		1,830.7	1,976.7	965.0	1,973.5
평균가구근로소득(만원)		1,501.5	1,690.4	762.3	1,614.6
가구주 특성					
혼인 상태(%)	미혼	4.4	9.3	10.9	8.6
	기혼유배우	81.8	79.1	39.9	78.5
	별거	1.8	2.3	3.6	2.2
	이혼	2.0	2.7	5.1	2.9
	사별	9.9	6.5	39.1	7.6
성별(%)	남성	86.8	86.3	60.9	86.3
	여성	13.2	13.7	39.1	13.7
경제활동 상태(%)	임금근로자	42.3	49.3	28.3	47.3
	비임금근로자	29.7	27.1	10.9	27.2
	무급가족종사자	0.9	1.3	0.0	0.5
	실업자	7.8	7.4	5.1	8.0
	비경제활동인구	19.3	14.9	55.8	17.0
평균연령(년)		48.4	43.2	60.3	45.3
평균교육기간(년)		10.0	11.7	7.1	11.3
교육기간 범주(%)	< 12	46.1	28.1	63.8	34.5
	12	34.4	38.9	15.9	35.4
	12～15	6.1	9.6	2.2	9.3
	16+	13.5	23.4	18.1	20.9
평균임금(만원)		129.6	150.5	120.4	136.3
평균근로시간(시간)		54.0	54.0	46.5	54.7

(주) 괄호 안은 단위를 나타냄. 가구 소득은 연평균 소득, 가구주의 임금은 월평균 임금임.

<표 6>에서는 응답 형태와 가구 특성과의 관계를 보여주고 있다. 평균 가구원 수는 계속응답 가구가 3.6명, 다시응답 가구는 3.5명, 비응답가구는 3.4명이다. 노동패널조사의 가구용 설문은 가구주뿐 아니라 가구를 대표할 수 있는 누구나 응답자가 될 수 있기 때문에, 가구원 수가 많을수록 접촉 가능한 가구원이 많아져 응답의 가능성도 높아진다.

가구 총소득과 가구 근로소득은 비응답(B), 다시응답(D), 계속응답(A), 소멸가구(C) 순으로 높았다. 비응답, 다시응답, 계속응답 가구의 소득은 큰 차이를 보이지 않은데 반해, 소멸가구의 소득은 다른 집단에 비해 매우 낮다.

응답형태별 가구주 특성을 살펴보면, 계속응답가구(A)는 비응답가구(B)에 비해 기혼자 비중이 높고, 비경제활동인구 비중이 높다. 비응답가구(B)의 가구주는 계속응답가구(A)의 가구주에 비해 평균 연령은 낮고, 교육수준은 높은 것으로 나타났다. 또한, 가구주의 주된 일자리의 평균 임금은 비응답가구(B)가 높은 것으로 나타났다. 이러한 현상은 젊고 소득이 높은 가구주일수록 접촉이 어렵고, 응답거부가 많기 때문이다.

2. 개인의 특성과 표본이탈과의 관계

이 절에서는 가구를 구분한 것과 비슷하게 응답형태에 따라 개인의 특성을 살펴 본다. 앞에서와 마찬가지로 1차 년도 원가구원이었던 개인을 기준으로 ㉠ 계속응답, ㉡ 비응답, ㉢ 소멸, ㉣ 다시응답의 4가지 유형으로 구분하였다. 개인의 응답 여부는 개인용 설문지의 응답 여부를 기준으로 하였으며, 2~10차 년도 사이에 15세 이상이 되어 개인용에 응답한 원가구원은 제외된다. 노동패널조사는 1차 년도 조사 당시 원표본 가구에서 응답된 모든 가구원을 원가구원으로 보기 때문에, 1차 년도 조사 당시 원표본 가구에 속한 15세 미만의 모든 개인도 원가구원이 된다. 따라서 이들이 15세가 되는 시점부터 신규용 설문조사를 시작으로 이후 지속적으로 개인 조사가 이루어진다. 본 논문에서는 개인 표본의 특성 변화 및 이탈 형태를 분석하고자 하는 것이 주요 목적이므로, 1차 년도 조사 당시 15세 이상이었던 원가구원만을 대상으로 분석을 실시하였다.

〈표 7〉 응답형태별 개인 분포 (단위: 명, %)

1차 년도 원가구원	11차 년도 조사성공 원가구원	원가구원 조사성공률				
			계속응답 (㉠)	비응답 (㉡)	소멸 (㉢)	다시응답 (㉣)
13,321	8,621	64.7	5,289 (39.7)	4,494 (33.7)	206 (1.6)	3,332 (25.0)

(주)괄호 안은 비중을 나타냄. 소멸 원가구원은 사망, 합가 등으로 소멸된 가구에 속한 원가구원으로 11차 년도에 비응답한 원가구원을 의미함.

<표 7>에 나타나듯이, 1차 년도 당시 15세 이상 원가구원이었던 13,321명 중에서 11차 년도에 응답한 원가구원은 8,621명이며, 원가구원 조사 성공률은 64.7%이다. 이는 11차 년도 원가구 조사 성공률인 74.2%에 비해 상당히 낮은 수치로, 원가구원(개인) 중 한 명이라도 응답을 하면 원가구도 조사에 성공하여 표본이 유지되는 가구패널 조사의 특성에 따른 것이다.

1차 년도부터 11차 년도까지 계속응답한 원가구원은 5,289명(39.7%)이며, 중간에 응답하지 않은 적이 있지만 11차 년도에 응답한 개인은 3,332명(25.0%)이다. 11차 년도에 비응답한 원가구원은 4,700명(35.3%)으로, 비응답 원가구원이 4,494명이었으며 소멸가구에 속한 비응답 원가구원은 206명이다. 원가구는 소멸되었으나 원가구가 소멸되기 전에 분가하여 11차 년도에 응답한 원가구원이 33명이 있었으며, 이들은 원가구원 조사성공률에 포함된다. 이와 비슷한 형태로 11차 년도에 원가구는 응답하지 않았으나 원가구원은 분가하여 응답된 경우가 82명 존재한다(<표 8> 참조).

<표 9>에서 나타나듯이, 1차 년도 원가구원의 주된 일자리의 근로소득을 기준으로 계속응답자 비중이 가장 높은 소득집단은 소득이 가장 낮은 1분위이다. 반면 소득 최상층인 5분위에는 비응답자와 소멸가구원의 비중이 다른 분위에 비해 높아서 계속응답자의 비중이 41.6%이다.

〈표 8〉 가구 및 개인 집단간 표본이탈 (단위: 명)

		원가구 응답여부				
		계속응답 A	비응답 B	소멸 C	다시응답 D	계
원가구원 (개인) 응답여부	계속응답㉠	5,225	23	14	27	5,289
	비응답 ㉡	1,082	2,833	0	579	4,494
	소멸 ㉢	0	0	206	0	206
	다시응답㉣	860	59	19	2,394	3,332
	계	7,167	,2915	239	3,000	13,321

〈표 9〉 개인근로소득(주된 일자리) 분위별 분포(개인) (단위: %)

	5분위	4분위	3분위	2분위	1분위	전체
계속응답㉠	41.6 (0.014)	46.5 (0.005)	42.9 (0.003)	40.5 (0.005)	49.5 (0.021)	44.2 (0.014)
비응답 ㉡	33.2 (0.016)	29.9 (0.006)	30.3 (0.003)	31.9 (0.006)	25.9 (0.028)	30.2 (0.016)
소멸 ㉢	1.6 (0.065)	0.6 (0.020)	0.8 (0.018)	0.6 (0.042)	1.5 (0.178)	1.0 (0.130)
다시응답㉣	23.6 (0.017)	23.0 (0.006)	25.9 (0.003)	26.9 (0.007)	23.2 (0.025)	24.5 (0.017)
응답 ㉠+㉣	65.2 (0.01)1	69.5 (0.00)4	68.8 (0.00)2	67.4 (0.00)4	72.7 (0.01)6	68.7 (0.01)1
탈락 ㉡+㉢	34.8 (0.016)	30.5 (0.006)	31.1 (0.003)	31.5 (0.006)	27.4 (0.029)	31.2 (0.016)

<표 10> 기초통계량(개인)

		계속응답 (㉠)	비응답 (㉡)	소멸 (㉢)	다시응답 (㉣)
혼인상태(%)	미혼	15.5	36.5	14.1	40.8
	기혼유배우	75.5	54.7	49.0	52.7
	별거	1.0	1.1	2.4	0.8
	이혼	1.0	1.0	3.9	1.2
	사별	6.8	6.5	29.6	4.4
성별(%)	남성	44.3	50.8	44.7	52.6
	여성	55.7	49.2	55.3	47.5
가구주 여부	가구주	42.9	34.1	63.1	31.8
	비가구주	57.1	65.9	36.9	68.2
경제활동 상태(%)	임금근로자	31.7	27.5	20.9	31.7
	비임금근로자	17.5	12.5	7.8	11.5
	무급가족종사자	6.0	2.8	1.9	2.5
	실업자	7.5	7.0	5.3	8.6
	비경제활동인구	37.4	50.2	64.1	45.7
평균연령(년)		43.0	38.1	57.0	35.5
평균교육기간(년)		10.0	11.0	7.3	11.4
교육기간 범주(%)	< 12	45.8	35.1	62.6	33.4
	12	34.8	35.0	16.5	35.1
	12～15	9.0	15.2	6.3	16.5
	16+	10.4	14.7	14.6	15.0
평균임금 > 0(만원)		114.6	122.86	124.5	114.7
근로소득	20 분위 수/ 50 분위	0.50	0.60	0.40	0.60
	40 분위 수/ 50 분위	0.80	0.90	0.80	0.80
	60 분위 수/ 50 분위	1.10	1.20	1.40	1.00
	80 분위 수/ 50 분위	1.58	1.70	2.00	1.60
평균근로시간 > 0(시간)		53.8	53.9	45.2	52.4
응답자 수(명)		5,289	4,494	206	3,332

소득 2분위에서 계속응답자 비중은 40.5%로 응답자 비중이 가장 낮다. 2분위의 응답자들은 응답-비응답을 오가는 응답자(In and Out)의 비중이 높다. 상대적으로 저소득층에 속하는 2분위 응답자의 표본 이탈이 심각한 이유는 접촉의 어려움과 낮은 소득을 알리기 싫어하는 응답자들의 성향때문인 것으로 보인다. 그리고 실업, 이혼 등 신변에 부정적인 변화가 발생하면 설문응답을 꺼려하는 경향도 작용한 것으로 보인다. 2분위 응답자들의 계속응답 비중은 가장 낮지만, 응답-비응답을 오가는 응답자는 가장 많다는 점에서도 이를 짐작케 한다.

응답 여부별로 원가구원 특성 차이는 <표 10>과 같다. 첫째, 성별로는 계속응답한 경우 여성의 비중(55.7%)이 다소 높고, 비응답 원가구원은 남성의 비중(50.8%)이 약간 높다. 반면, 응답-비응답을 반복하는 원가구원도 남성이 좀 더 많다. 이는 남성이 여성에 비해 인터뷰 시간을 내기 어렵다는 점에 기인한 것으로 보인다.

둘째, 기혼유배자의 비중을 보면, 계속응답 원가구원은 75.5%로 매우 높은데 반해, 비응답 원가구원은 54.7%, 응답-비응답을 오가는 원가구원은 52.7%인 것으로 나타났다. 한편 소멸가구에 속한 원가구원의 혼인상태는 사별이 29.6%로 매우 높게 나타나는데, 이는 가구 소멸의 주요 원인이 원가구원의 사망이기 때문이다.

셋째, 가구주일수록 계속응답 가능성이 높다. 노동패널의 경우 가계를 책임지는 주소득원이 아니더라도 가구주가 될 수 있고, 가구에서 한명의 개인만이 응답하면 그 개인을 가구주로 보기 때문에 가구주 비중이 높게 나타나는 것은 당연하다.

넷째, 경제활동 상태별로는 응답자인 경우엔 취업자 비중이 높고, 비응답인 경우 비경제활동인구의 비중이 높다. 하지만, 계속응답자는 비응답자에 비해서 연령이 높고 교육수준이 낮으며, 비응답자의 소득분포는 응답자의 소득분포에 비해 오른쪽으로 치우쳐 있다.

3. 가구 및 개인의 특성과 표본이탈

표본이탈 가구와 계속응답 가구의 특성 차이를 비교해 보았다 (<표 11>

<표 11> 응답형태에 따른 집단간 특성의 차이(가구)

	계속응답(A)		표본이탈(B+C)		다시응답(D)	
	평균	표준오차	평균	표준오차	평균	표준오차
가구특성						
가구원 수	3.56	0.026	3.28 *	0.037	3.55	0.040
가구총소득	1830.7	98.1	1868.6	129.9	1973.5	154.7
가구근로소득	1501.5	27.5	1591.2	42.4	1614.6 *	38.9
가구주특성						
혼인상태[기혼유배우=1]	0.82	0.008	0.75 *	0.012	0.78 *	0.013
성별[남성=1]	0.87	0.007	0.84	0.010	0.86	0.010
경제활동상태[경활=1]	0.81	0.008	0.81	0.011	0.83	0.011
연령	48.40	0.252	4501 *	0.381	45.29 *	0.363
교육년 수	10.01	0.087	11.24 *	0.128	11.32 *	0.123
임금 > 0	129.60	2.521	148.70 *	3.493	136.30 *	2.959
근로시간 >0	54.03	0.461	53.60	0.621	54.66	0.661

(주) *는 계속응답 집단과 비교해서 평균값이 5% 수준에서 유의한 차이가 있음을 나타냄.

<표 12> 응답형태에 따른 집단간 특성의 차이(개인)

	계속응답(㉠)		표본이탈(㉡+㉢)		다시응답(㉣)	
	평균	표준오차	평균	표준오차	평균	표준오차
혼인상태 [기혼유배우=1]	0.76	0.006	0.54*	0.007	0.53*	0.009
성별 [남성=1]	0.44	0.007	0.51*	0.007	0.53*	0.009
가구주 여부	0.43	0.007	0.35*	0.007	0.32*	0.008
경제활동상태 [경활=1]	0.63	0.007	0.49*	0.007	0.54*	0.009
연령	43.0	0.199	38.9*	0.263	35.5*	0.258
교육년수	10.0	0.059	10.8*	0.065	11.4*	0.063
임금 > 0	114.6	1.913	122.8*	2.152	114.7	2.097
근로시간 >0	53.8	0.385	53.6	0.433	52.4	0.473

(주) * 표시는 계속응답 집단과 비교해서 5% 수준에서 유의한 차이가 있음을 나타냄.

과 <표 12> 참조). 평균근로시간을 제외하고 가구원 수와 같은 가구 특성에서도 집단 간의 차이가 포착되며, 혼인상태, 성별, 경제활동상태, 연령, 교육, 임금 등 가구주 특성의 평균값 차이도 통계적으로 유의미하였다. 요약하자면, 계속응답 가구주가 표본이탈 가구주에 비해 기혼자 비중이 높고, 나이가 많으며, 평균 교육기간이 짧고, 임금이 적은 것으로 분석됐다.

한편, 가구를 대상으로 하였을 때와 개인을 대상으로 하였을 때 표본이탈 패턴은 약간의 차이가 있었다. 비응답 집단은 개인 임금과 가구소득이 다른 집단에 비해 모두 높다. 소멸집단의 가구소득은 다른 집단에 비해 아주 낮은 반면, 개인임금은 약간 높다. 즉, 개인의 표본이탈은 고임금일 때 높은 반면, 가구의 이탈은 고임금과 저임금 양쪽에서 표본이탈이 발생하는 특징이 관찰된다.

비응답과 소멸에 영향을 주는 요인을 파악하기 위해 프로빗(probit)모형으로 추정한 결과가 <표 13>에 나와 있다. 이 표에서 두 번째와 세 번째 열은 가구단위의 비응답과 소멸 여부를 종속변수로 하여 추정을 한 것이고 네 번째와 다섯 번째 열은 개인단위의 비응답과 소멸여부를 종속변수로 하여 추정을 한 것이다. 추정결과를 요약하면 다음과 같다.

먼저 가구단위의 경우, 가구원 수가 적을수록, 소득이 낮을수록, 나이가 적을수록, 학력이 낮을수록 비응답 확률이 늘어난다. 그리고 가구원 수가 적을수록, 소득이 낮을수록, 나이가 많을수록, 학력이 높을수록 소멸확률이 늘어난다. 가구원 수가 적을수록, 소득이 낮을수록 비응답과 소멸이 같이 늘어난다. 반면 학력과 나이는 비응답과 소멸에 대해 반대방향으로 작용을 한다. 개인의 경우 혼인 상태가 비응답이나 소멸에 큰 영향을 준다는 것이 큰 특징이다. 노동패널은 가구를 단위로 조사되고 있기 때문에 개인단위의 표본이탈은 가구단위의 표본이탈에 종속되어 있음을 보여주고 있다. 프로빗모형으로 표본이탈과 관련된 요인을 분석한 결과는 앞에서 기초통계량으로 분석한 결과와 다르지 않다.

<표 13> 비응답과 소멸에 영향을 주는 요인(probit 모형 추정)

분석단위	가구		개인	
설명변수 / 종속변수	비응답=1	소멸=1	비응답=1	소멸=1
가구원 수(기준=1 명)				
2 명	-0.111(0.105)	-0.609*(0.160)		
3 명	-0.085(0.107)	-0.766*(0.185)		
4 명	-0.140(0.109)	-0.753*(0.196)		
5 명	-0.278*(0.117)	-0.714*(0.220)		
6 명 이상	-0.326*(0.138)	-1.276*(0.397)		
가구소득(기준=1 분위)				
2 분위	-0.086(0.063)	-0.528*(0.200)	-0.109(0.082)	0.157(0.258)
3 분위	-0.115(0.065)	-0.332*(0.178)	-0.124(0.082)	0.235(0.252)
4 분위	-0.234*(0.070)	-0.231(0.166)	-0.121(0.084)	-0.056(0.265)
5 분위	-0.140(0.076)	0.083(0.163)	-0.312*(0.084)	0.002(0.252)
혼인상태(기준=미혼)				
기혼유배우	-0.002(0.116)	-0.429(0.239)	0.243*(0.041)	-0.501*(0.105)
별거	0.180(0.171)	-0.101(0.307)	0.416*(0.128)	-0.085(0.238)
이혼	0.138(0.155)	-0.096(0.278)	0.338*(0.125)	0.222(0.208)
사별	0.038(0.127)	-0.146(0.226)	0.360*(0.080)	-0.120(0.154)
경제활동(기준=임금근로)				
비임금근로자	0.019(0.050)	-0.275*(0.140)	0.002(0.044)	-0.205(0.118)
무급가족종사자	0.572*(0.206)	NA	-0.193(0.100)	-0.104(0.311)
실업자	0.049(0.083)	-0.221(0.211)	-0.175*(0.088)	0.140(0.262)
비경제활동인구	0.101(0.070)	0.041(0.136)	-0.087(0.079)	0.208(0.245)
나이	-0.013*(0.002)	0.023*(0.005)	-0.011*(0.001)	0.023*(0.003)
교육(기준=고졸 미만)				
고졸	0.155*(0.053)	0.072(0.131)	0.168*(0.032)	0.079(0.083)
대졸 미만	0.206*(0.084)	-0.050(0.257)	0.138*(0.043)	0.092(0.117)
대졸 이상	0.295*(0.066)	0.537*(0.145)	0.259*(0.043)	0.346*(0.100)
여성			0.065*(0.028)	-0.036(0.067)
관측치	3,804		10,397	

(주) *는 5% 수준에서 유의함을 나타냄. 계속응답과 다시응답을 응답가구로 봄.

Ⅴ. 표본이탈이 불평등 지표에 미치는 영향

표본이탈이 분포에 대한 추정에 어떠한 영향을 끼치는지 파악하기 위해 1차 년도 자료를 이용하여 응답형태 집단별로 지니계수와 임금 불평등도를 추정하였다. 가구 총소득의 불평등 지표를 계산한 결과가 <표 14>에 있다.

비응답가구는 평균 소득이 다른 집단에 비해 높고, 네 집단 중 가장 불평등도가 낮은 집단이다. 불평등도가 낮은 비응답가구가 이탈하면(전체 표본에 비해), 남아 있는 표본의 불평등도는 높아진다. 반면 소멸가구는 평균소득이 낮고 네 집단 중 가장 불평등도가 높은 집단이다. 소멸가구가 이탈한 것은 남아 있는 표본의 불평등도를 낮추는 방향으로 작용한다. 따라서 표본이탈이 불평등 척도에 미치는 영향은 이 두 가지 힘의 크기에 달려 있다.

응답과 이탈가구를 합친 전체 가구의 p90/p50은 2.37인 반면 응답집단의 경우에는 2.32로서 표본이탈의 결과 고소득에서의 불평등도가 약간 감소하였다. 표본이탈 가구의 p90/p50값은 2.31로 응답가구에 비해 약간 낮다. 소득이 높은 쪽에서 불평등도가 감소한 것은 비응답가구의 영향이 소멸가구의 영향보다 크기 때문이다.

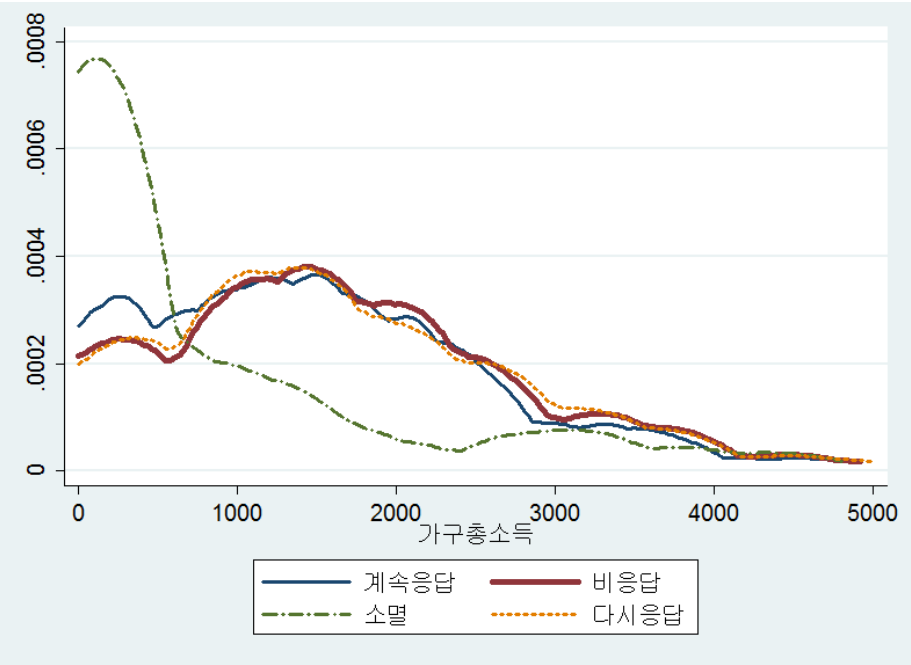
〈표 14〉 가구총소득 불평등 지표 (단위: 만원)

	전체	계속응답 (㉠)	비응답 (㉡)	소멸 (㉢)	다시응답 (㉣)	11차응답 (㉠+㉣)	표본이탈 (㉡+㉢)
평균	1871.9	1830.7	1976.7	965.0	1973.5	1873.1	1868.6
편차	4951.5	5009.2	4894.3	1710.4	5135.2	5046.7	4669.1
p90/p50	2.37	2.33	2.14	12.48	2.31	2.32	2.31
p50/p10	6.2	6.0	3.5	22.7	2.9	4.5	16.1
Gini	0.473	0.473	0.451	0.714	0.463	0.471	0.480
비중(%)	100	52.1	23.1	2.8	22.0	74.1	25.9

(주) 가구 횡단면 가중치를 사용하였음.

전체가구의 p50/p10은 6.2인 반면 응답집단의 경우에는 4.5로서 표본이탈의 결과 저소득에서의 불평등도도 감소하였다. 소득이 낮은 쪽에서는 소멸가구의 영향이 비응답가구의 영향을 압도하였다. 응답형태별로 가구 총소득의 분포를 나타낸 <그림 3>에서 볼 수 있듯이 소멸가구는 저소득 쪽에 몰려 있다. 비응답가구의 이탈로 불평등도가 증가한 요인이 있었지만 저소득 쪽에 몰려 있는 소멸가구가 이탈하면서 저소득에서 불평등도가 감소하였다.

소득분포의 양쪽 끝에서 표본이탈이 상대적으로 많이 일어난 결과 응답한 가구 즉 남아 있는 가구의 불평등도는 전체 가구에 비해 감소하였다. 전체 표본을 대상으로 지니계수를 계산하면 0.473인 반면, 11차 년도에 응답한 표본(계속응답+다시응답 가구)를 대상으로 지니계수를 계산하면 0.471이다. 표본이탈을 고려하지 않고 응답한 가구만을 대상으로 계산하면 불평등도를 약간 과소평가하게 된다.



〈그림 3〉 응답형태별 가구총소득의 분포

표본이탈을 적절히 고려하지 않았을 경우 불평등도가 과소평가될 수 있지만 그 정도가 매우 적은 것은 두 가지 이유 때문이다. 첫째는 비응답가구와 소멸가구가 작용하는 방향이 서로 반대이기 때문이다. 비응답가구의 이탈은 불평등도를 높이고 소멸가구의 이탈은 불평등도를 낮춘다. 이 중 소멸가구의 영향이 조금 더 커서 그 결과 표본이탈이 전반적으로 불평등도를 낮추게 된다.

두 번째 이유는 불평등도에 더 큰 영향을 준 소멸가구의 비중이 매우 낮기 때문이다. 소멸집단은 소득이 매우 낮으면서 불평등도도 매우 심한 집단이다(지니계수 0.714). 하지만 전체 표본에서 차지하는 비중이 매우 낮아서(2.8%) 전체 분포에 미치는 영향이 적었다. 소멸집단은 소득이 낮고 매우 불평등도가 높은 반면, 비응답집단은 소득이 높고 불평등도가 낮다. 매우 이질적인 두 집단의 합으로 이루어진 표본이탈 가구집단의 평균소득과 불평등도가 전체 분포에 근접하게 되었다.[1)]

일반적으로 가구소득의 불평등도를 계산할 때에는 가구원 수에 제곱근을 취하여 이를 가구소득에 곱하는 방식으로 균등화된 가구소득을 비교한다. 이렇게 균등화된 가구소득으로 불평등도를 계산해도 위 결과와 다름이 없다. 가구 가중치를 사용하지 않고 계산하여도 결론은 동일하다. 이 결과들은 따로 보고하지 않는다.

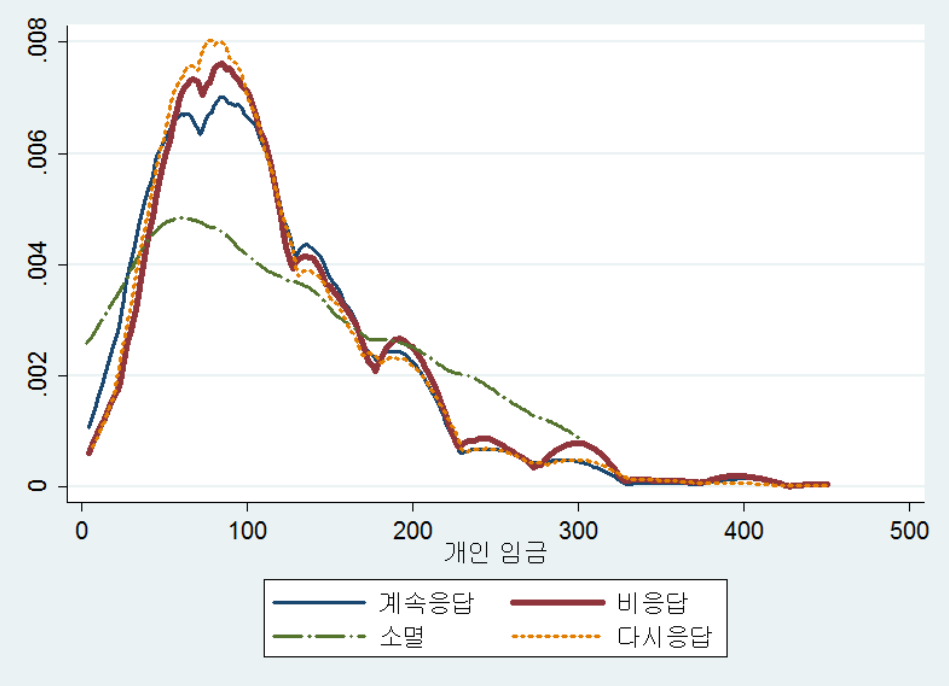
마지막으로 개인임금의 불평등도를 계산한 결과가 <표 15>에 나와 있다. 가구총소득에서와 마찬가지로 11차 년도에 응답한 개인을 대상으로 계산하면 전체 집단을 대상으로 하였을 때보다 지니계수의 값이 약간 작다(0.341 vs. 0.342).

표본이탈 집단의 대부분을 차지하는 비응답집단의 임금분포는 응답집단과 매우 비슷하다. 그리고 소멸집단이 다른 집단에 비해 이질적이기는 하지만

---

1) 참고로 응답유형별 소득분포가 통계적으로 다르다고 할 수 있는지 파악하기 위해 Kolmogorov-Smirnov 동일분포 검증을 해보았다. 검증 결과 가구소득의 경우 비응답과 다시응답 가구의 소득분포는 통계적으로 다르지 않았다 (D통계량 p값 0.882). 반면 계속응답과 소멸가구의 소득분포는 비응답이나 다시응답 가구의 소득분포와 통계적으로 다르다. 예를 들어 계속응답과 비응답 및 다시응답가구와의 D통계량 p값은 각각 0.003, 0.006이다. 이러한 결과는 [그림 3]에서 시각적으로 확인할 수 있는 것이다.

이질적인 정도가 가구총소득에서보다는 낮고, 전체 집단에서 차지하는 비중도 적어서(1.6%), 표본이탈이 임금불평등도에 미치는 영향이 적다.



〈그림 4〉 응답형태별 개인 임금 분포

〈표 15〉 개인 임금 불평등 지표 (단위: 만원)

	전체	계속응답 (㉠)	비응답 (㉡)	소멸 (㉢)	다시응답 (㉣)	11차응답 (㉠+㉣)	표본이탈 (㉡+㉢)
평균	117.2	114.6	122.8	124.5	114.7	114.7	122.8
편차	88.3	93.8	88.5	95.4	76.6	88.0	88.7
p90/p50	2.0	2.0	2.3	2.5	2.0	2.0	2.3
p50/p10	2.3	2.5	2.0	3.3	2.1	2.5	2.0
Gini	0.342	0.345	0.337	0.369	0.332	0.341	0.339
비중(%)	100	39.7	33.7	1.6	25.0	64.7	35.3

(주) 개인 횡단면 가중치를 사용하였음.

Ⅵ. 결론

본 논문에서는 노동패널 자료에서 표본이탈의 형태와 표본이탈이 추정에 미치는 영향을 분석하였다. 노동패널 11차까지 표본유지율은 74.2%이며, 약 20%의 가구가 응답과 비응답을 오가고 있다. 소득 최상위 가구에서는 무응답을 이유로, 소득 최하위 가구에서는 소멸을 주된 이유로 하여 표본에서 탈락하는 확률이 높다. 가구는 소득분포의 양쪽에서 이탈이 많이 일어나고, 개인의 경우에는 저소득 쪽에서 이탈이 많이 일어난다.

표본이탈이 지니계수와 임금불평등도에 미치는 영향을 분석한 결과, 표본이탈을 고려하지 않고 응답한 가구만을 대상으로 추정하면 불평등도를 과소평가하게 된다는 것을 발견하였다. 하지만 편의의 크기는 작았는데, 그 이유는 소멸집단은 소득이 낮고 매우 불평등도가 높은 반면 비응답집단은 소득이 높고 불평등도가 낮은데 이 두 집단으로 이루어진 표본이탈 가구의 불평등도가 전체 분포와 비슷하였기 때문이다. 개인소득의 분포의 경우 소멸집단의 비중이 매우 낮아서 표본이탈이 개인소득의 불평등도 측정에 미치는 영향이 적었다.

패널조사에서 표본이탈은 패널 가중치의 조정을 통해 일차적으로 해결할 수 있는데, 이 논문에서 나타나듯이 표본이탈을 이루고 있는 비응답과 소멸가구의 특성이 매우 다르고 이들이 불평등도에 미치는 방향이 다르기 때문에, 가능하면 표본이탈의 사유를 최소한 비응답과 소멸로 나누어서 패널 가중치를 설정하는 것이 필요하다. 더 나아가 가구의 응답성향을 적절한 모형으로 만들어 이를 가중치에 반영하는 것이 바람직하다.

참고문헌

강석훈. 2001. “유럽의 패널조사 현황과 시사점: 가구패널조사와 국제비교학적인 관점을 중심으로.” 워킹페이퍼. 한국노동연구원.

김대일・남재량・류근관. 2000. “한국노동패널 표본의 대표성과 패널조사 표본이탈자의 특성연구.” ≪노동경제논집≫ 23: 1-33.

김세원・정익중. 2012. “아동청소년패널조사의 항목무응답과 단위무응답에 영향을 미치는 요인.” ≪한국청소년연구≫ 23(3): 27-50.

신동균. 1998. “미국 패널 데이터의 현황과 시시점 - PSID, NLSY, KHPS, KLIPS를 중심으로.” 한국노동패널연구, 워킹페이퍼. 한국노동연구원.

윤윤규・홍민기・이상호・배기준・최효미・김기민・장숙랑・조성일・천희란. 2012. ≪패널자료 품질개선 연구(Ⅱ)≫. 한국노동연구원.

이상호. 2005. “한국노동패널(KLIPS)의 표본이탈 분석.” ≪노동리뷰≫ 11: 66-80. 한국노동연구원.

이희길. 2008. “국내 패널조사의 현황 분석.” 이희길 외, ≪통계조사의 정확성과 효율성 제고방안: 현장조사 및 표본추출을 중심으로≫. 통계개발원.

홍민기・김재광・한치록・김기민. 2013. ≪패널자료 품질개선 연구(Ⅲ)≫. 한국노동연구원.

Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics.” *The Journal of Human Resources* 33(2): 251-299.

Hausman, J. and D. Wise. 1979, “Attrition Bias in Experimental and Panel Data.” *Econometrica* 47(2): 445-473.

Hill, D. and R. Willis. 2001. “Reducing Panel Attrition: A Search for Effective Policy Instruments.” *The Journal of Human Resources* 36(3): 416-438.

Lillard, L. and C. Panis. 1998. “Panel Attrition from the PSID.” *The Journal of Human Resources* 33(2): 437-457.

Zabel, J. 1998. “An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior.” *The Journal of Human Resources* 33(2): 479-506.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*(2nd Ed.). The MIT Press.

<접수 2014/7/8, 수정 2014/9/12, 게재확정 2014/9/30>