

노동통계개선시리즈 5

한국노동패널조사(KLIPS)의 2003년 가중치

김기헌*

I. 2003년 가중치 개발 배경

- 표본조사자료의 결과가 대표성을 가지려면 적절한 가중치를 부여하는 것이 필요함. 한국노동패널조사(이하 ‘KLIPS’)의 경우도 표본자료에 대표성을 부여할 목적으로 1999년 가중치(이하 ‘구 가중치’)를 개발한 바 있음(강석훈, 1999).
- 구 가중치는 미국의 가구패널조사인 PSID(Panel Study of Income Dynamics)와 동일한 방법으로 가중치가 부여됨. PSID는 1차 조사(웨이브)에 대해서 표본추출확률과 응답률을 토대로 가중치를 산출하였으며, 2차 조사(웨이브)부터 표본 탈락(panel attrition)을 감안하여 가중치를 산출함. KLIPS의 구 가중치 역시 표본추출확률과 응답률을 토대로 1998년 1차 조사(웨이브)의 가중치로 산출된 것임.
- KLIPS의 구 가중치가 1999년 공개된 이후 추출확률계산과정에서 사용하지 않은 정보가 추가로 발견되었음.
 - KLIPS의 경우 1995년 인구센서스 10% 표본조사구(ED)를 모집단으로 하고 있으나 실제 표본추출은 1997년 고용구조특별조사(이하 ‘고특조사’)의 조사구를 통해 이루어짐. 구 가중치에서는 ED의 가구수와 고특조사 조사구의 가구수가 같다는 가정하에 표본추출확률을 계산하였지만, 실제로는 양자간에 차이가 있음이 발견되었음.

* 한국노동연구원 책임연구원(kihuns@kli.re.kr).

- 구 가중치를 계산할 당시에는 KLIPS 표본조사구가 도의 동부와 읍면부가 추출 확률이 상이하다는 정보를 이용하지 않았음.
- 1차년도에 이러한 문제점들을 보완하고, 2차년도 이후의 가중치를 개발하는 작업이 착수됨. 2003년도에 보완된 1차 조사(웨이브)의 가중치(이하 ‘신 가중치’)는 통계청의 해당 부서의 협조하에 인구센서스 및 고용구조특별조사의 조사구 내 가구수와 동부 및 읍면부 도시조사구 비중에 관한 정보를 토대로 구 가중치를 보완한 것임(강석훈, 2003).

II. 신 가중치 부여방법 및 결과

◆ 1차 조사(웨이브)의 가중치 부여방법 및 결과

- 1차 조사(웨이브)의 가구 가중치는 기본적으로 구 가중치와 동일하게 두 단계를 거쳐 계산됨.
 - 가중치를 계산하는 첫번째 단계는 표본이 추출될 확률(selection probability)을 계산하는 것임. 이것은 표본조사구의 추출확률과 표본가구의 추출확률로 구분됨. 먼저 표본조사구 추출확률은 인구센서스의 10% 표본조사구에 선정되고 다시 KLIPS의 표본조사구로 선정될 확률을 의미함. 신 가중치에서 표본조사구의 추출확률은 구 가중치와 마찬가지로 “10%의 표본조사구의 추출확률이 모든 지역에서 10%로 동일하다”는 가정을 그대로 사용함. 다만 지역별 표본조사구의 추출확률은 구 가중치와는 달리 도의 동부 및 도의 시에 속한 읍면부를 구분하여 계산함. 구 가중치에서는 도 지역의 경우 동부 및 읍면부를 구분하지 않고 동일한 추출확률로 계산함. 반면 신 가중치는 동부 및 읍면부의 도시가구수에 관한 자료를 토대로 도 지역의 동부 추출확률과 읍면부 추출확률을 분리해서 계산함. <표 1>은 신·구 가중치의 지역별 표본추출확률 계산의 차이를 보여주고 있음.
 - 다음으로 표본조사구에서 최종 표본가구로 추출될 확률은 구 가중치의 경우 조사구 내의 총접촉시도가구수를 고특조사에 응답한 총가구수로 나누어 계산함. 반면 신 가중치는 인구센서스 10% 자료의 조사구 내 가구수와 고특조사의 조사구 내 가구수에 관한 정보를 갖고 있기 때문에 고특조사의 응답가구수로 총접촉가

<표 1> 신·구 가중치의 지역별 표본추출확률 비교

	인구센서스 10% 추출확률	지역별 표본조사구 추출확률	
		광역시	(시부 표본조사구수/시부 조사구수)
구 가중치(1999)	0.1	도	(도시조사구수/조사구수)× (표본조사구/도시조사구수)
		광역시	(시부 표본조사구수/시부 조사구수)
신 가중치(2003)	0.1	도의 동부	(동부 표본조사구수/동부 조사구수)
		도의 읍면부	(읍면부 표본조사구수/읍면부 조사구수)
		광역시	(시부 표본조사구수/시부 조사구수)

주: 표본조사구수는 KLIPS의 표본조사구수, 조사구수 및 도시조사구수는 10% 인구센서스의 조사구수 및 도시조사구수를 의미함.

구수를 나눈 값과 센서스 조사구 내의 전체 가구수로 곱해 조사의 응답가구수를 나눈 값을 곱하여 계산함. 최종적으로 표본의 추출확률은 조사구 추출확률과 최종 표본가구 추출확률의 곱으로 계산됨.

- 가중치를 계산하는 두번째 단계는 응답률을 계산하는 것임. 응답률을 고려하지 않는 경우는 전지역에서 응답가구와 비응답가구의 특성이 동일하다는 가정을 전제로 한 것임. 이러한 가정을 완화하기 위해 이루어지는 응답률 조정은 무응답가구에 관한 특성 파악을 전제로 함. 이에 따라 가능한 응답률 계산은 지역정보나 조사구(ED) 정보를 이용하여 계산하는 방법임. 지역정보를 이용하는 경우 동일 지역 내에서 응답가구와 무응답가구의 특성이 같다는 것을 가정하는 것이며, ED 정보를 이용하는 경우 ED 내의 응답가구와 무응답가구의 특성이 같다는 것을 가정하는 것임. 구 가중치는 지역 정보를 활용하였으나 신 가중치는 ED 정보를 활용하여 ED 내에서의 응답률을 면접확률로 이용하였음.
- 마지막 단계로 사후층화(poststratification)를 통해 패널자료의 가중치를 보완하는 경우가 있으나 신 가중치는 사후층화 방법을 구 가중치와 마찬가지로 사용하지 않음. 그 이유는 가구 자료이든, 개인 자료이든 사후층화에 사용될 적당한 자료가 없다고 판단했기 때문임. 먼저 KLIPS의 1차 조사(웨이브)가 1998년도임을 고려할 때 가구 단위의 사후층화 자료로 활용할 수 있는 가장 최근 자료는 통계청의 1995년 인구센서스 자료이나 이를 사용하여 사후층화를 하는 것은 시의성이 떨어지는 것으로 판단함. 개인 단위의 사후층화 자료로는 통계청의 장래인구추계 자료나 경제활동인구조사 자료, 고용구조특별조사 자료 등이 있으나 장래인

구주계 자료는 광역시를 제외하고 농촌 지역에 대한 연령별, 성별 인구분포의 추계를 제시하고 있지 않으며 경제활동인구조사 자료나 고용구조특별조사 자료는 KLIPS와 마찬가지로 전수조사 자료가 아닌 표본조사 자료이기 때문에 사후층화의 의미가 퇴색된다고 판단함.

- 1차 조사(웨이브)의 개인 가중치는 KLIPS의 경우 가구와 그 가구에 속한 가구원의 추출확률이 동일하므로 동일한 가구 내의 가구원은 모두 동일한 가구 가중치를 부여함. 이는 PSID에서 사용하는 가중치 부여방법과 동일한 것임.
- <표 2>는 1999년도에 개발된 1차 조사(웨이브)에 대한 구 가중치와 2003년도에 보완된 1차 조사(웨이브)에 대한 신 가중치의 기초 통계치를 보여주고 있음. 구 가중치의 가구 가중치의 평균값은 2457.5이며 개인 가중치는 2707.0인데 비해서 신 가중치는 각각 2331.7과 2332.1임. 신 가중치의 경우 가구와 개인 가중치간의 차이가 없으나 구 가중치는 가구와 개인 가중치간에 차이가 존재함을 알 수 있음. 표준편차를 살펴보면 구 가중치의 경우 신 가중치에 비해서 편차가 2배 가량 높음을 알 수 있음. 한편 신 가중치에서 가구 가중치의 합은 모집단의 추정 가구수에 근사하게 scale 조정을 한 값(11,276,899가구)이나 개인 가중치의 경우 가구원들의 총수에 관한 모수값을 알 수 없기 때문에 scale 조정을 하지 않은 값임.

<표 2> 신·구 가중치의 기초통계치 비교

		사례수	평균	표준편차	최소값	최대값
구 가중치(1999)	가구 가중치	10,316,198	2457.5	1040.8	551.2	5540.6
	개인 가중치	28,785,217	2707.0	1355.3	551.2	7259.8
신 가중치(2003)	가구 가중치	11,276,899	2331.7	453.5	1513.2	4514.6
	개인 가중치	30,039,371	2332.1	455.8	1513.2	4514.6

주: 구 가중치는 표본추출확률만을 부여한 가중치(변수명: cwt01h1)와 응답률까지 고려한 가중치(변수명: cwt01h2)를 동시에 제공하고 있으나 응답률까지 고려한 가중치가 보다 개선된 가중치이므로 이것을 사용함.

자료: 1998년 노동패널조사(KLIPS) 자료.

<표 3> 지역별 표본가구수 및 신·구 가중치 부여 결과

	Un-weighted		구 가중치(1999)		신 가중치(2003)	
	가구수	비중	가구수	비중	가구수	비중
서울	1,353	27.1	1,780,613	17.3	2,961,684	26.3
부산	486	9.7	831,845	8.1	1,105,238	9.8
대구	314	6.3	536,727	5.2	700,459	6.2
인천	153	3.1	309,257	3.0	372,062	3.3
광주	298	6.0	511,429	5.0	646,022	5.7
대전	158	3.2	310,040	3.0	352,060	3.1
울산	116	2.3	228,655	2.2	264,968	2.3
경기	856	17.1	1,946,958	18.9	1,976,455	17.5
강원	130	2.6	407,572	4.0	312,451	2.8
충북	113	2.3	386,406	3.8	255,802	2.3
충남	131	2.6	543,413	5.3	301,386	2.7
전북	199	4.0	513,186	5.0	451,777	4.0
전남	127	2.5	526,848	5.1	289,056	2.6
경북	274	5.5	738,764	7.2	619,352	5.5
경남	292	5.8	744,484	7.2	668,127	5.9
전 체	5,000	100.0	10,316,198	100.0	11,276,899	100.0

자료 : 1998년 노동패널조사(KLIPS) 자료.

○ <표 3>은 지역별로 표본가구수와 신·구 가중치를 부여한 결과를 보여주고 있음. 구 가중치의 경우 서울지역의 표본 구성비가 가중치를 사용하지 않은 경우에 비해 9.5% 포인트 낮아지고 있으나 신 가중치를 부여했을 경우 가중치를 부여하지 않은 경우와 큰 차이를 보이지 않음. 경기 지역의 경우 구 가중치를 부여했을 때의 표본 구성비는 가중치를 부여하지 않은 경우에 비해 약 2% 포인트 상회하고 있으나 신 가중치를 부여했을 경우 가중치를 부여하지 않았을 때보다 0.4%포인트 상회하는데 그침. 전반적으로 구 가중치에 비해서 신 가중치에 의한 지역별 표본 구성비의 변화는 크지 않다는 점을 알 수 있음.

◆ 2차 조사(웨이브) 이후의 가중치 부여방법

○ 그동안 KLIPS는 2차 조사(웨이브) 이후의 가중치를 제시하지 않았으나 이번에 신 가중치가 공개되면서 1차년도 자료 가중치와 함께 2~4차년도 자료에 대한 가중치도 동시에 제공하고 있음. 2차 조사(웨이브) 이후부터 가중치는 종단면 조사(longitudinal survey) 자료의 특성상 1차 조사(웨이브)에서 발견되지 않는 다음과 같

은 상황을 반영해야 함.

- 첫번째는 1차 조사(웨이브)에 응답한 가구 및 가구원이 2차 조사(웨이브) 이후에 응답을 거절할 수 있기 때문에 각 조사(웨이브)간 표본 탈락(panel attrition)이 발생하게 됨.
- 두번째는 표본가구에 결혼, 동거, 입양, 출생 등의 이유로 새롭게 진입한 가구원이 발생하게 되는데 이러한 비표본가구원(non-sample household)으로 인해서 조사(웨이브)가 진행됨에 따라 가구구성이 달라지게 됨.

○ KLIPS의 신 가중치는 2차 조사(웨이브) 이후에 발생할 할 수 있는 상황을 고려한 던컨(Duncan, 1995)의 방법론을 이용하여 가중치를 계산함. 이 방법에 따른 가중치 부여절차는 다음과 같음.

- 먼저 이미 산출된 1차 조사(웨이브)의 개인 가중치를 가구원들의 상이한 응답률을 이용하여 조정함. 이 단계에서는 가구와 가구원의 정보를 이용하여 응답여부에 관한 로짓(logit) 모형을 설정한 후 도출된 계수를 사용하여 모든 가구원들의 응답확률을 추정함. 이 응답확률의 역수를 최초의 개인 가중치에 곱하여 무응답 조정 가중치(non-response adjusted weights)를 산출함.
- 다음으로 새롭게 산출된 해당 조사(웨이브)의 개인 가중치의 평균을 이용하여 해당 조사(웨이브)의 가구 가중치를 산출함. 이 때 결혼, 동거 등의 사유로 새롭게 진입한 비표본가구원의 경우에는 0의 가중치를 부여함. 이렇게 해서 가구 차원의 무응답 조정 가중치가 산출됨.
- 이런 절차를 통해 산출된 개인 가중치는 비표본가구원의 경우 0의 가중치를 주고 표본가구원의 경우 응답률로 조정한 추출확률의 역수를 가중치로 부여한 것이며 가구 가중치는 개인 가중치의 평균으로 산출한 것이라고 할 수 있음.

<표 4> 분석 단위에 따른 1-4차년도 가중치 변수

		1차년도	2차년도	3차년도	4차년도
횡단면 분석	가구 분석	w1_h	w2_h	w3_h	w4_h
	개인 분석	w1_p	w2_pc	w3_pc	w4_pc
종단면 분석	가구 분석	w1_h	w2_h	w3_h	w4_h
	개인 분석	w1_p	w2_pl	w3_pl	w4_pl

주: 신 가중치는 한국노동패널 홈페이지(www.kli.re.kr/klips)에서 제공되고 있음. 1-4차년도 가중치는 가구번호(HHID), 개인고유번호(PID) 등 기준 변수가 포함된 엑셀 파일(.xls), SPSS 파일(.sav), SAS 파일(.sd2) 등 세 가지 형식의 데이터 파일로 제공됨.

- 이렇게 해서 산출된 KLIPS의 신 가중치는 <표 4>에서처럼 분석 단위에 따라서 다르게 사용될 수 있음.
 - 먼저 개인 차원의 분석은 개인 가중치와 가구 가중치가 사용될 수 있음. 개인 가중치는 비표본가구원을 제외하고 원표본가구원만을 대상으로 분석할 때 사용할 수 있으며(의미상 ‘개인 종단면 가중치’로 명명), 가구 가중치는 비표본가구원의 추출확률이 표본가구원과 동일하다는 가정하에 비표본가구원을 포함하여 해당 연도의 전체 가구원을 대상으로 분석할 때 사용할 수 있음(의미상 ‘개인 횡단면 가중치’로 명명).
 - 가구 차원의 분석에는 가구 가중치가 사용될 수 있음. 다만 종단면 분석시에 사용할 경우 해당 연도의 가구가 1차 조사(웨이브)의 가구와 동일하다는 가정을 해야 하는데 앞서 살펴보았듯이 비표본가구원으로 인해 가구구성이 바뀌므로 과연 동일한 가구인가에 대해 논란이 있음. 이에 따라 영국의 가구패널조사인 BHPS나 독일의 가구패널조사인 GSEP의 경우에는 미국의 PSID와는 달리 종단면가구 가중치를 산출하고 있지 않음.

Ⅲ. 가중치 부여에 따른 표본 대표성 비교

◆ 가구 자료의 대표성 비교

- 1998년 1차 년도 KLIPS 자료를 중심으로 가중치 부여에 따른 표본 대표성을 살펴보기 위하여 먼저 KLIPS의 가구 소득과 통계청의 도시가계연보의 가구 소득을 비교 검토하고자 함. 두 자료간의 표본설계 및 조사방법 등의 차이로 인해 엄밀한 비교분석은 가능하지 않으나 도시가계연보는 가중치 부여에 따른 표본 값의 변화에 대한 준거자료로서 가치가 있다고 판단함. KLIPS와 도시가계조사 자료를 비교하기 위해서는 다음과 같은 조정작업이 필요함.
 - 도시가계조사 자료가 단독가구 및 농·어가 등을 제외한 도시근로자 가구를 대상으로 하고 있기 때문에 도시 가구를 대상으로 하는 KLIPS의 비교 자료는 2인 이상 가구이고 근로소득이 있는 가구로 국한하여 계산하였음.
 - 도시가계조사 자료의 월평균 가구소득은 경상소득과 비경상소득의 합으로 정의되며 KLIPS에서 제공하고 있는 가구소득은 근로소득, 금융소득, 부동산소득, 이

전소득, 기타 소득 등으로 구분됨. 소득원천별로는 도시가계조사와 KLIPS의 개념정의가 다르기 때문에 KLIPS의 원천별 가구소득을 월평균으로 환산하여 합산한 가구 총소득과 도시가계조사의 월평균 가구소득을 비교함.

- 이러한 조정에도 불구하고 두 자료간의 기본적인 차이가 존재함. 도시가계조사는 1998년 당해 연도의 월평균 가구소득인데 비해서 KLIPS의 1998년 월평균 가구총소득은 응답시점을 기준으로 1년간의 평균소득임.
- <표 5>는 두 자료의 월평균 가구소득 및 소득계층별 가구의 분포를 보여주고 있음.
 - 전반적으로 도시가계조사에 비해서 KLIPS의 월평균 가구소득이 적게 추정되었으며 소득계층별 분포에 있어서 KLIPS의 저소득 계층의 비중이 상대적으로 높게 나타남.
 - 가중치를 부여하지 않은 KLIPS의 월평균 가구총소득은 도시가계조사의 월평균 근로자소득의 80.8%에 머물고 있음. 구 가중치를 부여했을 경우 도시가계조사의 77.9%에 머물러 그 격차가 더 벌어지고 있음을 알 수 있음. 반면 신 가중치를 부여했을 경우 가중치를 부여하지 않은 경우와 큰 차이가 없는 것으로 나타남.
 - 소득계층별 분포를 살펴보면 KLIPS의 경우 월평균 가구소득이 95만원 미만인 저소득계층은 22.7%로 도시가계조사보다 9.9%포인트 높게 나타나고 있음. 반면 365만원 이상의 고소득계층은 KLIPS가 4.2%포인트 낮게 나타나고 있음. 이러한 현상은 구 가중치를 부여했을 때 더 강화되는 양상을 보여줌. 구 가중치를 부여했을 때 격차는 각각 11.7%포인트, 4.8%포인트로 나타나고 있음.

<표 5> 가구 소득 및 소득계층별 분포 비교

(단위: 천원, %)

	KLIPS			도시가계조사
	Un-weighted	구 가중치(1999)	신 가중치(2003)	
가구 월평균 소득	1,725	1,661	1,720	2,133
소득계층별 분포				
950원 미만	22.7	24.5	22.8	12.8
950~1,550원 미만	34.2	33.9	34.2	25.5
1,550~2,150원 미만	20.5	20.3	20.5	23.5
2,150~2,750원 미만	9.0	8.9	9.0	16.1
2,750~3,650원 미만	8.1	7.4	8.0	12.6
3,650원 이상	5.5	4.9	5.4	9.7

자료: 1998년 노동패널조사(KLIPS) 자료, 1998년 도시가계연보 자료

◆ 개인 자료의 대표성 비교 검토

- 이번에는 KLIPS의 개인 자료의 대표성을 비교검토해 보고자 함. 개인 자료의 비교 대상 자료는 장래인구추계 자료와 경제활동인구조사 자료임. 전자는 인구학적 특성을 비교검토할 때 사용했으며, 후자는 경제활동상태를 비교검토할 때 사용하였음.
- 먼저 기본적인 인구구성과 관련 장래인구추계 자료와 KLIPS의 결과를 비교해 보고자 함. KLIPS의 경우 15세 이상을 대상으로 하고 있기 때문에 장래인구추계 자료의 15세 이상 인구의 성별 및 연령별 인구 자료를 활용하여 비교하였음. 앞서 언급한 도시가계조사 자료와 마찬가지로 두 자료간의 비교는 다음과 같은 차이를 염두에 두는 것이 필요함.
 - 먼저 장래인구추계는 1995년 인구센서스의 결과를 기초로 해서 추정된 것으로 KLIPS와 같이 도시지역에 국한된 정보를 제공하고 있지 않음. 곧 KLIPS와는 달리 장래인구추계의 결과는 농촌지역을 포괄한 결과임.
 - 다음으로 KLIPS는 군 복무자를 제외하고 있기 때문에 20대 초반의 남성 인구가 적게 추정될 가능성이 존재함.
- <표 6>은 두 자료의 15세 이상 인구의 성별 및 연령별 분포를 보여주고 있음.
 - 대체로 두 자료간의 성별 및 연령별 분포는 유사한 양상을 보여주고 있음. 남성에 비해서 여성 인구가 높게 나타나고 있으며 대략적인 인구 분포의 양상도 유사함. 다만 KLIPS의 여성 비중이 약간 높게 나타나고 있고 연령별 분포에서도 청·장년층 인구의 비중이 다소 낮게 나타나고 있음.
 - 먼저 성별 비중을 살펴보면 장래인구추계의 경우 여성 비중이 50.3%인데 비해서 가중치를 부여하지 않은 KLIPS의 여성 비중은 51.5%로 약간 높게 나타남. 큰 차이는 아니지만 구 가중치를 부여했을 때 여성 비중은 51.7%로 약간 높아지며 신 가중치를 부여했을 때에는 가중치를 부여하지 않은 경우와 동일하게 나타남.
 - 연령별 분포를 살펴보면 장래인구추계의 경우 30세 중반까지 10%의 비중을 보여주고 있으나 KLIPS는 40대 중반까지 10%대의 비중을 보여주고 있음. KLIPS의 20~24세 연령층의 인구구성비가 상대적으로 적게 추정된 것은 군입대 남성 인구가 제외된 결과로 풀이됨. 구 가중치를 부여할 경우 미미한 수준이기는 하지만 청·장년층 인구는 감소하고 중·장년층 인구는 증가하는 양상을 보여주고 있음. 반면 신 가중치를 부여했을 경우 성별 구성과 마찬가지로 가중치를 부여하지 않은 결과와 거의 유사한 양상을 보여주고 있음.

<표 6> 성별 및 연령별 분포 비교

(단위: %)

	KLIPS			장래 인구추계
	Un-weighted	구 가중치(1999)	신 가중치(2003)	
성비(15세 이상)	94.35	93.54	94.36	98.96
연령별 분포				
15~19세	11.8	11.7	11.8	14.6
20~24세	9.5	9.7	9.4	13.5
25~29세	11.2	10.9	11.2	13.9
30~34세	10.7	10.5	10.8	12.7
35~39세	12.2	11.8	12.2	9.0
40~44세	10.7	10.1	10.7	7.8
45~49세	8.4	8.2	8.4	7.1
50~54세	6.7	6.8	6.6	6.4
55~59세	6.1	6.2	6.1	4.9
60~64세	4.7	5.0	4.7	3.6
65~69세	3.5	3.9	3.4	2.8
70~74세	2.4	2.6	2.4	1.8
75~79세	1.3	1.5	1.3	1.1
80세 이상	0.9	1.0	0.9	0.8

주: 성비=남자/여자×100.

자료: 1998년 노동패널조사(KLIPS) 자료, 1998년 장래인구추계 자료.

○ 다음으로 경제활동인구조사(이하 ‘경활조사’) 자료를 이용하여 KLIPS와 경제활동상태 및 취업형태를 비교하고자 함. 비교가 되는 경활조사 자료는 전체 결과치와 도시 지역에 국한된 결과치를 동시에 비교하고자 함(김대일 외, 2000). 두 자료는 가구조사라는 공통점 외에 비교 기준이 되는 1주간 경제활동상태에 관한 문항이 동일하기 때문에 비교상 장점이 존재함. 반면 이 역시 두 자료상의 다음과 같은 차이를 고려하여 해석하는 것이 필요함.

- 먼저 경활조사는 매월 진행되는 조사인데 비해서 KLIPS는 1년 중 4~5개월간 진행되는 조사라는 차이가 존재함. KLIPS는 응답자마다 답변하는 시기가 다를 수 있으므로 월간 변동에 따른 차이를 통제할 수 없음.
- 다음으로 장래인구추계와 마찬가지로 KLIPS는 군 복무자를 제외하고 있으나 경활조사는 그렇지 않기 때문에 이에 따른 차이가 존재할 수 있음.

<표 7> 경제활동 및 취업형태 분포 비교

(단위: %)

	KLIPS			경찰조사
	Un-weighted	구 가중치(1999)	신 가중치(2003)	
경제활동				
취업 및 일시휴직	48.7	49.2	48.7	56.5(55.0)
실업	6.9	6.6	6.9	4.2(4.6)
비경제활동	44.4	44.2	44.4	39.3(40.4)
취업형태				
임금근로자	62.4	60.0	62.4	61.0(67.8)
비임금근로자	29.4	30.2	29.3	28.8(25.4)
무급가족종사자	8.2	9.8	8.3	10.1(6.8)

자료: 1) 1998년 노동패널조사(KLIPS), 경제활동인구조사 자료.

2) 경찰조사의 괄호 안 수치는 도시지역 통계치로 김대일 외(2000) 참조.

○ <표 7>은 경찰조사와 KLIPS의 경제활동상태 및 취업형태의 분포를 보여주고 있음.

- 전반적으로 볼 때 KLIPS가 경찰조사에 비해서 경제활동상태의 경우 취업자의 비중이 상대적으로 낮고 미취업자의 비중이 상대적으로 높게 나타나고 있으며 취업형태별 비중은 유사한 양상을 보여주고 있음.
- 먼저 경제활동상태를 살펴보면 취업비중은 KLIPS가 7.8%포인트 낮게 추정되고 있으며 실업과 비경제활동 비중은 각각 2.7%포인트, 5.1%포인트 높게 추정되고 있음. 다만 도시지역에 국한된 수치를 살펴보면 두 자료간의 격차가 다소 완화되고 있을 알 수 있음. 구 가중치를 부여할 경우 경찰조사와의 차이는 약간 보정되고 있으나 신 가중치의 경우 가중치를 부여하지 않은 경우와 경제활동상태의 비중에 있어서 차이가 없는 것으로 나타남.
- 취업형태에 관한 결과를 살펴보면 두 자료 모두 임금근로자의 비중은 60%대를, 비임금근로자의 비중은 30% 미만을, 무급가족종사자의 경우 10% 미만의 비중을 보여주고 있음. 다만 KLIPS의 경우 임금근로자의 비중이 약간 높고 무급가족종사자의 비중이 약간 낮은 특성을 보여줌. 경찰조사를 도시지역으로 국한해 볼 때 KLIPS의 임금근로자 비중은 상대적으로 낮게 추정된 것으로 나타나며 비임금근로자와 무급가족종사자의 비중은 약간 높게 추정된 것으로 나타나고 있음. 이것은 구 가중치를 사용할 경우 임금근로자의 비중을 상대적으로 낮추고 있는 것과 대비됨. 신 가중치의 경우 가중치를 부여하지 않은 경우와 큰 차이가 없는 것으로 나타남.

IV. 결 론

- KLIPS는 대표적인 노동관련 패널조사로 자리잡아 왔으며 이번에 자료의 대표성을 강화할 수 있는 새로운 가중치가 제공되면서 보다 완성도 높은 가구패널조사로 자리매김할 것으로 기대됨.
- 다만, 도시가계조사나 경찰조사 등 다른 표본자료와의 비교검토를 통해서 알 수 있듯이 그동안 제기되어 왔던 KLIPS의 표본 대표성 문제가 신 가중치를 통해 완전히 보정되었다고 말하기 어려움. 물론 비교대상이 되는 자료간의 차이는 실질적인 내용상의 차이라기보다는 비교대상 자료간의 표본설계 및 조사대상, 조사방법상의 차이일 수 있음. 이와 함께 패널조사는 일반 공식 통계자료와는 달리 집계자료의 산출을 목적으로 한 것이 아니기 때문에 완전히 동일한 결과가 산출되어야 할 이유가 없다고 볼 수도 있음. 그렇다고 하더라도 표본 대표성의 문제는 계속 제기될 수밖에 없기 때문에 이를 보완하고 개선하는 노력이 필요하다고 할 수 있음.
- KLIPS가 진행중인 프로젝트라는 점에서 기존에 발견된 사실들과 함께 표본 이탈 등으로 인해 발생할 수 있는 추가적인 대표성의 문제를 해명하는 다각적인 노력이 향후 전개될 것으로 보임. 신 가중치의 개발은 이러한 작업의 밑거름이라고 할 수 있음.

참고문헌

- 강석훈(1999), 「KLIPS 1차 웨이브에서 가중치 부여방법에 관한 연구」, 『제1회 노동패널학술대회 자료집』.
- _____ (2000), 「KLIPS 2차 웨이브의 가중치 부여방법에 관한 연구」, 『제2회 노동패널학술대회 자료집』.
- _____ (2003), 「KLIPS의 가중치 부여방안 연구」, 『한국노동패널연구』, Working Paper Series 2003-04.
- 김대일 · 남재량 · 류근관(2000). 「한국노동패널 표본의 대표성과 패널조사 표본 이탈자의 특성 연구」, 『노동경제논집』, 3.
- Duncan, G. J(2003), “A Simple Methods for Weighting in Household Panel Surveys,” in Wolfgang Voges eds, *Dynamic Approaches to Comparative Social Research : Recent Developments and Applications*, Ashgate Publishing Company, London.