

# 노동패널자료 연구(1)

- 패널자료 비교연구를 중심으로 -

홍민기 · 강신욱 · 성재민 · 이상호  
장지연 · 김재광 · 이지은





# 목 차

요 약 .....	i
제1장 노동패널 불평등지표와 소득 조사 .....	1
제1절 머리말 .....	1
제2절 노동패널과 가계동향조사 표본과 대표성의 비교 .....	1
제3절 소득 불평등지표의 비교 .....	3
1. 소득의 분류 .....	3
2. 지니계수 비교 .....	5
제4절 노동패널의 지니계수가 왜 더 높을까? .....	6
1. 연령 분포의 비교 .....	6
2. 표본탈락의 문제 .....	7
3. 소득 히핑의 문제 .....	8
4. 소득 조사의 차이 .....	8
제5절 결 론 .....	10
제2장 한국노동패널 자료를 통해 본 소득이동성 .....	12
제1절 머리말 .....	12
제2절 자료 및 분석방법 .....	13
1. 데이터 .....	13
2. 소득범주 .....	14
3. 자료의 결합 및 가중치 .....	14
제3절 소득지위 관련 자료 상태의 검토 .....	15
1. 소득이동성에 영향을 미치는 데이터의 특성 .....	15
제4절 소득이동성의 의미와 지표 .....	20

제5절 가구단위의 소득이동성 변화 추이 .....	22
1. 소득분위 이동성 .....	22
2. 전년도 소득과의 상관계수 .....	24
3. 전년도 소득과의 격차 .....	25
제6절 개인의 소득이동성 변화 추이 .....	27
1. 개인 소득이동성 분석 대상 및 방법 .....	27
2. 개인 소득분위의 이동성 .....	28
3. 개인소득의 전년도 소득 대비 상관계수 .....	30
4. 소득격차의 변화 추이 .....	30
제7절 소 결 .....	31

**제3장 한국노동패널조사에서 나타난 비정규직 현황과 추세 -  
경제활동인구조사와 비교를 중심으로** ..... (성재민) .... 33

제1절 서 론 .....	33
제2절 경제활동인구조사와의 비교·분석 .....	34
1. 추 이 .....	34
2. 인구분포 차이 .....	35
3. 설문구조, 조사방식의 차이 .....	39
제3절 정규직-비정규직 변화에 나타나는 비표본오차의 교정 .....	42
제4절 결론 및 정책 함의 .....	44

**제4장 노동패널 고용지표의 신뢰성 검토** ..... (이상호) .... 46

제1절 머리말 .....	46
제2절 고용률 및 실업률 추이 .....	48
제3절 청년층의 노동시장 참여가 낮은 원인 .....	52
1. 표집 오차와 가중치 .....	55
2. 개인 응답률 .....	58
3. 조사체계의 차이 .....	63
4. 소 결 .....	69

제4절 노동력 상태 이동성 분석 .....	70
제5절 정책적 함의 및 토론 .....	74
1. 고용지표 신뢰성 향상을 위한 방향 .....	74
2. 패널자료 신뢰성 연구의 확장 .....	75

## 제5장 노동패널 중장기 발전방안 - 해외 장기패널과의 비교를

중심으로 ..... (장지연 · 김재광 · 이지은) ..... 77

제1절 문제제기 .....	77
제2절 한국노동패널의 발전경로와 현재 상황 .....	78
1. 표본의 규모와 대표성 .....	78
2. 조사방법 .....	81
3. 연구범위 .....	81
4. 인력과 조직 .....	82
제3절 미국 PSID의 경험 .....	83
1. 표본의 규모와 대표성 .....	83
2. 조사 및 데이터의 배포 .....	84
3. 연구범위 .....	85
4. 인력과 조직 .....	86
제4절 영국 BHPS의 경험 .....	87
1. 표본의 규모와 대표성 .....	87
2. 조사방법 .....	88
3. 연구범위 .....	88
4. 인력과 조직 .....	89
제5절 독일 SOEP의 경험 .....	90
1. 표본의 규모와 대표성 .....	90
2. 조사방법 .....	92
3. 연구범위 .....	93
4. 인력과 조직 .....	95
제6절 소 결 .....	96

1. 주요 데이터 비교가 주는 시사점 .....	96
2. 더 많은 표본수 확보를 위한 고려사항 .....	99
3. 데이터 생산기관 간 협력 .....	104

참고문헌 .....	108
------------	-----

## 표 목 차

<표 1- 1> 노동패널과 가계동향조사 샘플링 비교 .....	2
<표 1- 2> 가구의 대표성(2010년) .....	2
<표 1- 3> 노동패널, 가계동향조사 소득 조사의 비교 .....	4
<표 1- 4> 지니계수의 비교 .....	6
<표 1- 5> 인구총조사와 노동패널의 인구 분포 비교 .....	7
<표 2- 1> 노동패널 자료의 가구 가치분소득 결측 표본의 비중 .....	16
<표 2- 2> 가구소득 결측치 처리방법별 분위 분포 (노동패널 9차 자료 이용) .....	17
<표 2- 3> 노동패널 자료의 가구 가치분소득 분위 분포 .....	18
<표 2- 4> 가구 결합자료 구축 시 분위별 탈락가구 비율 .....	19
<표 2- 5> 가구 소득분위의 전년 대비 변화 분포 .....	22
<표 2- 6> 전년도와의 가치분소득 상관계수 .....	24
<표 2- 7> 전년도와의 절대적 및 상대적 소득수준 변화 .....	26
<표 2- 8> 개인 소득분위의 이동성 .....	28
<표 2- 9> 전년도와의 개인노동소득 상관계수 .....	30
<표 2-10> 전년도와의 절대적 및 상대적 소득수준 변화 .....	31
<표 3- 1> 한국노동패널조사 통합 표본과 경찰 8월 부가조사의 비정규직 추이 .....	35
<표 3- 2> 비전형 근로 추이 .....	35
<표 3- 3> 연령별 분포 - 2014년 8월 경찰과 2014년 한국노동패널 조사(17차년도) .....	36
<표 3- 4> 학력별 분포 - 2014년 8월 경찰과 2014년 한국노동패널 조사(17차년도) .....	36

<표 3- 5> 임금근로자 중 한시근로 비중 추이 .....	37
<표 3- 6> 연령별 임금근로자 중 한시근로 비중 .....	38
<표 3- 7> 연령별 임금근로자 중 기간제 비중 .....	38
<표 3- 8> 임금수준별 비정규직 분포 .....	39
<표 3- 9> 임금수준별 한시근로 분포 .....	40
<표 3-10> 사유별 비기간제 한시근로자의 임금근로자 중 비중 .....	40
<표 3-11> 한시근로의 규모 변화 - 일일근로 통제 여부별 규모 추이 변화(임금근로자 중 비중) .....	41
<표 3-12> 노동패널에서 발견되는 정규직(=0), 비정규직(=1) 변화의 패턴 - 한 직장 내의 경우 .....	42
<표 3-13> 비정상적으로 판단되는 정규↔비정규 이동 교정 시 고용형태별 임금근로자 중 비중 추이 .....	43
<표 4- 1> 전체 고용률 및 실업률 .....	49
<표 4- 2> 2014년(17차) 성별·연령별 실업률 및 고용률 비교 .....	50
<표 4- 3> 청년층(15~29세) 고용률 및 실업률 .....	51
<표 4- 4> 고용률 과소 추정의 잠재적 요인들 .....	53
<표 4- 5> 경찰 수준의 청년층 고용률 달성을 위해 KLIPS에서 추가적으로 조사되어야 할 취업자수 .....	56
<표 4- 6> 청년층 가구원의 응답 여부 및 유형 .....	60
<표 4- 7> 2009년 KLIPS와 경제활동인구조사의 비경제활동인구 분포 .....	65
<표 4- 8> KLIPS의 월별 청년층 조사자수 및 고용률 .....	67
<표 4- 9> 고교 및 대학 재학생 아르바이트 취업 보정 시 고용률 증가 효과 .....	68
<표 4-10> 경제활동 상태의 동태적 변화 .....	72
<표 4-11> 미취업 상태의 동태적 변화 .....	73
<표 5- 1> 1~17차년도 조사 결과 .....	79
<표 5- 2> 통합표본 1~6차년도 조사 결과 .....	80

<표 5- 3> 1~17차년도 한국노동패널 자료의 구성 .....	82
<표 5- 4> SOEP 표본의 종류별 표집시점의 규모 .....	91
<표 5- 5> 주요 패널데이터 비교 .....	97

## 그림목차

[그림 1-1] 가구의 대표성 .....	3
[그림 1-2] 지니계수의 비교 .....	5
[그림 1-3] 노동패널, 가계동향조사, 가계금융복지조사 소득분포의 비교 .....	9
[그림 2-1] 노동패널 자료에서 개인 노동소득이 결측인 표본의 비중 변화 .....	17
[그림 2-2] 노동패널과 복지패널의 가구소득 분위별 탈락률 비교 (2005~12년간 평균) .....	20
[그림 2-3] 소득 10분위 변동가구 비율 비교 .....	23
[그림 2-4] 소득 10분위 상승 및 하락 가구 비율 비교 .....	23
[그림 2-5] 전년도 소득과의 상관계수 변화 추이 비교 .....	25
[그림 2-6] 전년 대비 소득격차 비교 .....	26
[그림 2-7] 소득이 결측인 관측치를 유지(0으로 대체)했을 때와 제외했을 때의 분위 변동비율 .....	28
[그림 2-8] 개인소득 분위(10분위 구분) 변동자의 비율 비교 .....	29
[그림 2-9] 개인소득의 전년 대비 격차 비교 .....	31
[그림 4-1] 가중치 적용여부에 따른 청년층 고용률 변화 .....	57
[그림 4-2] 노동패널 12차 조사의 연령집단별 응답률 .....	59
[그림 4-3] 노동패널 12차 조사의 무응답 사유 .....	61
[그림 4-4] 비동거 제외 시 개인 응답률(12차 조사) .....	62
[그림 4-5] KLIPS와 경찰조사의 취업자 판별 설문구조 차이 .....	64
[그림 4-6] 학교 재학 여부별 고용률 격차(12차 조사) .....	66
[그림 4-7] 경제활동인구조사의 월별 청년층(15~29세) 고용률 및	

실업률 추이 .....	67
[그림 4-8] KLIPS과 경찰조사 청년층 취업자의 고용형태 구성 비교 .....	68
[그림 5-1] 전체 표본의 크기(개인 응답자수 기준) .....	92
[그림 5-2] SOEP 표본과 관련 프로젝트의 관계 .....	95



## 요약

제1장에서는 노동패널 불평등지표와 소득 조사에 대하여 연구하였다. 특히 두 자료에서 지니계수가 얼마나 다른지, 추세는 어떠한지를 비교하고자 한다. 그리고 노동패널로 계산한 지니계수가 왜 더 높은지 원인을 규명하고자 한다.

2013년 노동패널 09표본에서 시장소득 지니계수는 0.358로, 가계동향조사의 지니계수 0.336보다 약간 높다. 추세는 2009년 이후 비슷하다. 이 연구에서는 노동패널의 지니계수가 가계동향조사보다 높은 이유에 대해 1) 연령의 문제, 2) 표본탈락(attrition)의 문제, 3) 소득 히핑(heaping)의 문제, 4) 샘플링(sampling)의 차이 혹은 소득과 약 정도의 차이를 검토하였다. 그 결과 파악된 소득분포의 차이가 결정적인 이유로 파악되었다. 노동패널에서는 가계동향조사에 비해 저소득과 고소득이 많이 포착된다. 가장 크게 차이가 나는 부분은 고소득 가구이다. 노동패널에서는 1억 원 이상 고소득이 상대적으로 더 많이 포착되고 있다. 조사된 소득분포가 퍼진 정도는 가계금융복지조사, 노동패널, 가계동향조사 순이며, 지니계수도 이 순서대로이다.

소득 조사와 관련된 시사점은 다음과 같다. 첫째, 조사마다 소득에 대한 구분이 다른데, 특히 임대료를 어떻게 구분할 것인가의 문제에 대해 논의를 할 필요가 있고, 구분의 문제와는 별도로 최소한 부동산임대소득과 사업소득을 구분하여 조사하고 통계를 제시하여 연구자가 소득의 면모를 다 알 수 있도록 하는 것이 필요하다. 둘째, 국세 자료와 조사 자료를 결합하여 불평등지표를 개선하는 노력이 필요하다.

제2장에서는 노동패널 9~17차 자료를 이용하여 가구 및 개인의

소득이동성 추이를 분석하고, 그와 관련된 데이터의 특성을 한국복지패널 1~9차 자료에서 나타나는 특성과 비교하였다.

소득이동성을 분위 변동비율, 전년도 소득과의 상관관계, 소득 변화의 절대적 및 상대적 크기 등을 이용하여 분석한 결과, 가구 및 개인의 소득 이동성 추이는 복지패널을 이용한 결과와 크게 다르지 않다는 점을 확인할 수 있었다. 다만, 2012~13년간 소득지위 변동가구(및 개인)의 비율이 복지패널에서는 상승하는 반면 노동패널에서는 감소하는 것으로 나타났다. 인접한 두 해 사이의 가구소득 상관관계 변화는 노동패널과 복지패널에서 다소 상이한 양상으로 나타났다. 전년도 소득과의 격차는 노동패널과 복지패널에서 큰 차이가 없었으나, 변동폭은 노동패널의 변동폭이 복지패널보다 항상 높게 나타나고 있었다.

제3장에서는 노동패널조사에서 나타난 비정규직 현황과 추세를 경제활동인구조사와 비교하여 분석하였다. 한국노동패널조사는 10차년도부터 경제활동인구조사(이하 경활)의 근로형태별 부가조사와 동일한 비정규직 판별설문을 조사하였다. 이 설문항목 분석을 통해 경활과 비정규직 규모, 추이를 비교 분석하였다. 분석결과 비정규직 전체 규모가 2014년 다소 크게 조사된 것을 제외하면 경활과 비교해 크게 다르지 않은 모습이 나타났다. 다만, 세부로 구분해 보면 한시근로에서 차이가 보인다. 한시근로 비중과 추이가 경활과 반대이고, 과거엔 기간제가 지나치게 적었고, 기간제 추이가 경활과 반대이며, 비기간제 한시근로가 많은 차이가 나타난다.

연령, 성, 교육수준 등 인구분포의 경활과의 차이는 이와 같은 다른 한시근로 규모와 추세를 설명하는 데 전혀 도움이 되지 않는 것으로 나타났다. 기간제 추이가 반대이긴 하지만, 그로 인해 역설적으로 최근에는 노동패널과 경활 간 기간제 규모 차이가 2%포인트 내외로 좁혀졌다.

한편, 고연령대에서 한시근로, 특히 비기간제 한시근로 비중이 매우 높은 것으로 나타나는 한편, 저소득층에서 비기간제 한시근로 비

중이 월등히 높은 것으로 분석되었다. 고연령대나 저소득층의 일자리는 비공식 고용 특성을 갖는 경우가 많은데, 이로 인해 주관적 판단이 많이 개입되는 비기간제 한시근로 판별설문에서 응답 차이가 나타난 것이 아닌가 판단된다. 또한, 설문구조상 일일근로가 한시근로에 응답하게 되어 있는 노동패널 설문구조가 한시근로 규모를 증가시키는 원인으로 작용했다. 동일인을 추적 조사하는 노동패널 조사구조를 이용해 동일 기업 내 정규-비정규 전환을 분석해 본 결과 특별한 이유 없이 두 상태를 왔다갔다 하는 사람들이 포착되었다. 이들의 상태변화를 교정해주면 비정규직 규모가 감소하였다. 비정규직 설문이 하나의 설문이 아니고 여러 설문을 조합해 정의하는 특성상 비표본오차가 개입될 소지가 크다. 이는 조사 현장에서 과거 정보를 연동시켜 응답이 바뀐 사람들을 대상으로 정말 바뀐 것인지를 확인하는 루틴을 조사 시스템에 탑재하는 것의 필요성을 보여준다.

4장은 노동패널자료의 핵심적인 고용지표의 신뢰성을 검증하기 위해 고용률 및 실업률 지표를 중심으로 통계청의 경제활동인구조사와 비교를 실시하였다. 분석결과 KLIPS의 고용률은 경제활동인구조사보다 약 3~4%포인트 낮았는데, 주된 원인은 청년층 취업자가 과소 추정되었기 때문인 것으로 확인되었다.

이에 따라 청년층 취업자가 과소 추정된 원인을 표집 오차와 비표집 오차 영역으로 구분하여 분석한 결과 주요 원인은 크게 두 가지임이 확인되었다. 첫째, 청년층 비재학자와 비동거자의 낮은 응답률 문제이다. 비재학생(졸업, 수료, 중퇴 등)의 응답률은 76.6%로 재학생의 응답률 87.2%보다 10.6%포인트나 낮았다. 동거 형태와 관련하여 청년층 중 동거 가구원은 개인 응답률이 90.8%지만, 비동거 가구원인 경우 개인 응답률이 29.6%에 그쳤다. 군입대를 제외할 경우 청년층 응답률은 85.8%까지 증가하기 때문에 비동거 응답률이 심각하게 낮은 수준이라고 볼 수 없다.

둘째, 대학 재학생의 취업자 식별 문제이다. 비재학생의 고용률 격차는 3%포인트 이내로 다른 연령층에서 나타나는 두 조사 간 고용

를 격차와 비슷한 수준으로 확인되었다. 그러나 대학 이상 재학자의 고용률 격차는 경제활동인구조사(27.4%)와 KLIPS(11.5%) 간에 16%포인트나 되는 것으로 확인되었다. 대학 재학생의 취업자 과소 보고가 발생한 원인은 취업자 설문에 대한 응답 비용은 큰 반면, 이를 식별할 설문체계상의 제약이나 인센티브 부족 등으로 응답기피 현상, 즉 패널 컨디셔닝(Panel Conditioning)이 나타나기 때문인 것으로 추정된다.

고용지표의 신뢰성을 향상시키기 위해서는 청년층 응답률 제고를 위한 조사체계상의 개선, 취업자와 미취업자 간 응답 부담의 불균형을 완화시키기 위한 설문구조 변화, 표본 확대 등의 전략적 대응이 필요하다. 또한 다양한 분야에서 패널자료의 신뢰성을 높이기 위해 연구가 확대될 필요가 있음을 정책제언으로 제시하였다.

제5장에서는 2017년 20차를 맞이하는 노동패널의 문제점을 극복하기 위해 장기간의 가구패널자료를 구축해 온 사례로 미국의 Panel Study of Income Dynamics(이하 PSID), 영국의 British Household Panel Study(이하 BHPS), 독일의 Socio-Economic Panel Study(이하 SOEP)를 참고하여 한국노동패널조사의 중장기적인 발전방안을 세 가지 측면에서 마련해 보았다. 세계적으로 유명한 패널조사와의 비교가 주는 시사점, 더 많은 표본수 확보를 위한 고려사항, 데이터 생산기관 간 협력사항에 대해 논의하였다.

세 국가의 대표적인 가구패널조사 자료의 역사와 현황을 살펴본 결과, 다음과 같은 시사점을 얻게 된다. 첫째, 국가의 사회·경제 정책 수립과 다양한 연구주제를 소화하기 위해서는 표본 규모의 확대가 필수적이다. 노동패널의 경우도 최소한 2만 명 이상의 개인응답자를 조만간 확보하고, 필요에 따라 지속적으로 확대해나가는 방안을 강구해야 한다.

둘째, 조사의 내용을 확대하고 다른 조사들에서 도입한 혁신적인 설계를 노동패널도 도입할 때가 되었다. 조사내용의 폭을 넓히는 것은 데이터의 활용도를 높이는 데 기여하고, 데이터의 가치를 높이게

된다. 현실적으로 빠른 시일 내에 추구해 볼 수 있는 혁신은 행정데이터의 결합이다. 해외 패널조사들이 모두 행정데이터와의 연계를 추구하지만, 가장 성공적인 사례이자 우리가 벤치마킹 할 만한 사례는 SOEP를 독일 연방고용청의 고용보험 행정데이터와 연계한 사례이다.

셋째, 노동패널은 조사방법(survey mode)을 CAPI로 전환한 이래 다른 시도는 없는 실정이다. 따라서 가장 조사비용이 비싼 직접면접 방식만을 고수할 것이 아니라 다른 방법을 병행하는 안을 검토하여야 한다. 다만, 조사방법의 변화가 응답률이나 응답의 내용에 영향을 주지 않도록 면밀한 사전 검토와 준비가 있어야 할 것이다.

넷째, 데이터의 활용도를 높이기 위하여 데이터의 구조를 사용자가 이해하기에 좀 더 쉬운 방식으로 안내하고, 가중치도 보다 간단하게 제공하는 것이 바람직할 것이다. 데이터 추출방식도 지금보다 편리성을 제고할 필요가 있다.

노동패널조사의 장기 플랜에서 가장 핵심적인 부분은 표본추가와 종단면 가중치 작성 등과 관련된 통계학적 부분과 패널 표본 관리 등과 관련된 실사적 부분이다. 표본추가와 관련해서는 현재 표본을 분석하고 현재 표본에서 과소 집계된 인구학적 변수들을 파악하여 이를 보완한 후 표본을 보완하는 방법과, 어떤 계층에 대한 사회적 관심이 높아져서 그 계층(도메인)에 대한 표본가구를 과대 표집하는 방법을 고려할 수 있다. 표본이탈에 따른 가중치 처리 문제에서 고려할 사항은 현재 사용되고 있는 성별, 연령, 지역 등 인구학적 변수만을 반영한 방법에서 소득과 교육수준 같은 경제학적 변수를 반영하는 방법론의 개발이다. 실사 측면에서는 인센티브 시스템의 정비, 조사 방법, 설문지, 테스트 관련 사항들을 고려해야 한다.

마지막으로 데이터 생산기관들과 협력을 통해 노하우를 공유하는 것이 바람직하다. 국제적인 가구패널조사 연구팀들의 네트워크는 두 가지 방식으로 진행되고 있다. 하나는 데이터 자체를 국가 간 비교가 가능하도록 한 Cross National Equivalent Files(CNEF) 사업이

추진되고 있는데, 노동패널도 이 사업에 참여하고 있으며, 최근 데이터를 업데이트 해서 업로드 할 계획이다. 다른 협력 네트워크는 2년에 한 번씩 각 나라의 가구패널조사 연구진들이 모여서 조사방법과 관련된 워크숍을 개최하는 방식으로 활발하게 진행되고 있다(패널조사방법론 워크숍: Panel Survey Method Workshop). 이 워크숍은 각국의 데이터 생산기관이 돌아가면서 개최하는데, 2016년에는 SOEP를 생산하는 독일의 DIW 연구소가 개최한다. 우리 노동패널 팀도 한 차례 워크숍에 참석한 적이 있으나 향후 더욱 활발한 참여로 협력사업을 떠나가는 것이 바람직할 것이다. 마지막으로 중국의 China Family Panel Studies(CFPS)와 일본의 가구패널도 아직은 국제적으로 유명하지는 않지만 협력사업의 대상으로 검토되어야 할 것이다.

제 1 장

# 노동패널 불평등지표와 소득 조사

## 제1절 머리말

노동패널로 계산한 불평등지표(예: 지니계수)가 가계동향조사로 계산한 공식적인 불평등지표보다 높고, 추세도 다르다는 지적이 있었다. 이 장에서는 두 자료에서 지니계수가 얼마나 다른지, 추세는 어떠한지를 비교하고자 한다. 그리고 노동패널로 계산한 지니계수가 왜 더 높은지 그 원인을 규명하고자 한다.

## 제2절 노동패널과 가계동향조사 표본과 대표성의 비교

‘가계동향조사 분배지표 자료’는 가계동향조사와 농가경제조사로 이루어져 있다. ‘가계동향조사’는 비농가가구를, 농가경제조사는 농가가구를 대표한다. 이 두 자료를 합쳐서 ‘가계동향조사 분배지표 자료’를 만든다. 본 연구에서는 특별한 언급이 없는 한 불평등지표와 관련되어 ‘가계동향조사’를 언급할 때에는 두 자료를 합친 분배지표 자료를 가리킨다.

노동패널에는 98표본과 09표본이 있다. 노동패널 98표본은 제주도를

## 2 노동패널자료 연구(1)

제외한 도시지역을 대표하고, 표본을 확대한 노동패널 09표본은 전국을 대표한다.

두 조사의 표본과 대표가구를 비교한 것이 <표 1-1>에 나와 있다. 가계동향조사는 약 10,000가구를 표본으로 하며 약 1,730가구를 대표한다. 노동패널 09표본은 6,721가구를 표본으로 하며, 1,820만 가구를 대표한다.

두 조사의 가구 대표성을 비교한 것이 <표 1-2>와 [그림 1-1]에 나와 있다. 2010년 인구총조사의 총가구(1,734만 가구)와 비교하여 노동패널은 100%, 가계동향조사는 97.7%를 대표하고 있다. 공통적으로, 노동패널과 가계동향조사 모두 4인 가구와 5인 가구를 과대 대표하고 있으며 나머지 가구는 과소 대표한다. 4인이나 5인 가구를 과대 대표하는 것은 조사자료

<표 1-1> 노동패널과 가계동향조사 샘플링 비교

	대표성	샘플수	대표가구수
노동패널 98표본	제주도 제외 도시지역	5,000가구	
노동패널 09표본	전국	6,721가구	1,821만 가구
가계동향조사	전국 비농가가구	약 7,100가구	1,600만 가구
농가경제조사	전국 농가가구	약 2,800가구	126만 가구
가계금융복지조사	전국	약 20,000가구	1,814만 가구

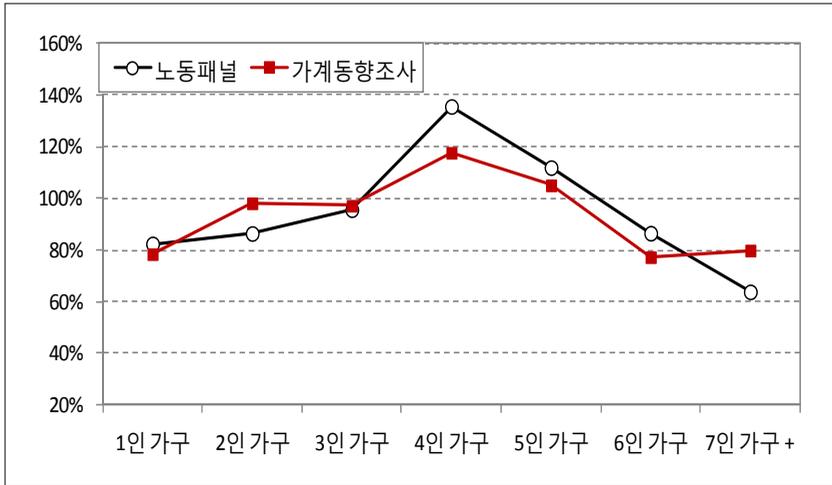
주: 가계동향조사, 농가경제조사는 2010년 기준.

<표 1-2> 가구의 대표성(2010년)

(단위: 가구)

	인구총조사 (A)	노동패널 (B)	가계동향조사 (C)	노동패널 (B)/(A)	가계동향조사 (C)/(A)
총계	17,339,422	17,339,423	16,939,220	1.000	0.977
1인 가구	4,142,165	3,414,844	3,249,134	0.824	0.784
2인 가구	4,205,052	3,636,255	4,127,858	0.865	0.982
3인 가구	3,695,765	3,535,986	3,592,529	0.957	0.972
4인 가구	3,898,039	5,285,830	4,587,052	1.356	1.177
5인 가구	1,078,444	1,207,404	1,133,341	1.120	1.051
6인 가구	241,063	208,699	186,282	0.866	0.773
7인 가구+	78,894	50,404	63,024	0.639	0.799

(그림 1-1) 가구의 대표성



의 특성이다.

가계동향조사에 비해 노동패널이 4인과 5인 가구를 과대 대표하는 경향이 더 크다. 예를 들어 4인 가구의 경우, 인구총조사 대비 가계동향조사는 17.7%를 과대 대표하는 반면, 노동패널은 35.6%를 과대 대표한다.

### 제3절 소득 불평등지표의 비교

#### 1. 소득의 분류

불평등지표를 계산할 때 사용하는 소득에는 총소득, 시장소득, 가처분소득 등이 있다. 각 소득의 정의는 다음과 같다.

$$\begin{aligned} \text{총소득} &= \text{임금} + \text{사업소득} + \text{재산소득(이자, 배당, 임대료)} \\ &\quad + \text{양도소득} + \text{이전소득} + \text{기타소득} \end{aligned}$$

$$\text{시장소득} = \text{임금(근로소득)} + \text{사업소득} + \text{재산소득} + \text{사전이전소득}$$

$$\text{가처분소득} = \text{시장소득} + \text{공적이전소득} - \text{공적 비소비지출<sup>1)</sup>}$$

#### 4 노동패널자료 연구(1)

〈표 1-3〉 노동패널, 가계동향조사 소득 조사의 비교

		노동패널	가계금융복지조사	가계동향조사
근로 소득		임금	임금 (기본급, 수당, 보너스)	임금
사업 소득		자영업자 소득 (세후)	사업소득(세전)	사업소득(세후) 주택 등 임대소득 의제자가주거소득
재산 소득	금융 소득	이자, 배당, 주식매매차익	금융소득(이자, 배당) 주택·건물·토지임대수입 개인연금, 퇴직연금	이자소득
	부동산 소득	집세, 토지임대료, 부동산매매차익		배당소득 의제이자소득 기타재산소득
기타 소득		보험금 퇴직금 증여/상속 축의금/조의금 당첨상금/상금 재해보상금		저축·보험 탄 금액 보증금 회수 유가증권 매각 부동산 매각 기타자산변동수입 벌린 돈 자산이전수입

가계동향조사에서는 지니계수를 계산할 때 ‘시장소득’을 사용한다. 참고로, 시장소득에는 매매차익(자본이득, capital gain)이 포함되지 않는다.

소득을 조사할 때 조사마다 소득의 하위구분이나 분류방식이 달라서 이에 주의할 필요가 있다. <표 1-3>에서는 노동패널과 가계동향조사에서 소득의 구분을 비교하여 보여주고 있다. 부동산 임대소득의 경우 노동패널에서는 재산소득으로 분류되고 있는 반면, 가계동향조사에서는 사업소득으로 분류된다. 노동패널에서 주식매매차익은 금융소득으로 조사되고 있고, 부동산매매차익은 부동산소득으로 조사되고 있어서 시장소득을 계산할 때에는 이 둘을 제외하여야 한다. 가계동향조사에서 유가증권 매각, 부동산 매각은 기타소득으로 분류되고 있으며 매매차익이 아니라 매각금액이 조사된다는 것에 주의해야 한다.

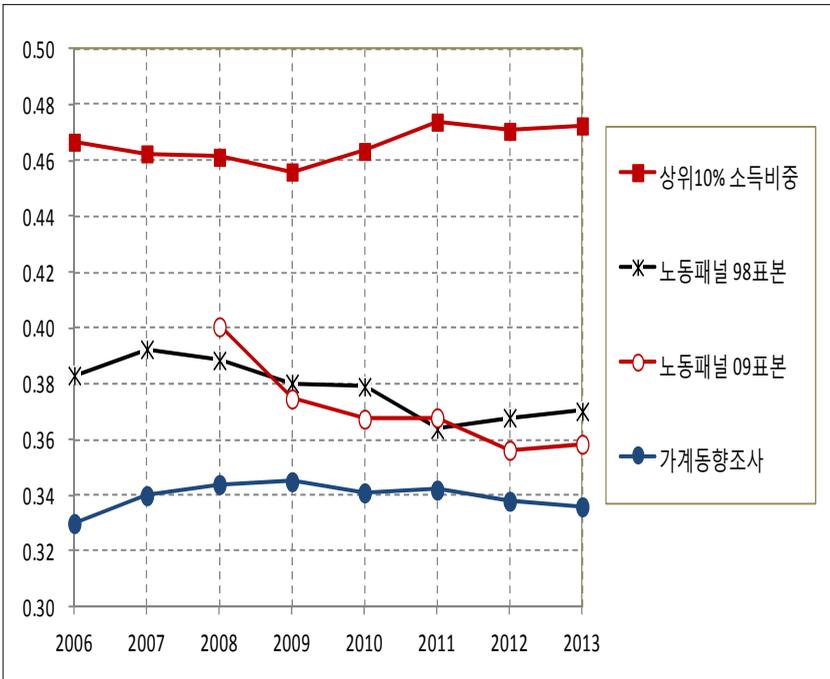
1) 참고로, 공적이전소득에는 공적연금, 기초노령연금, 사회수혜금, 세금환급금 등이 있다. 공적 비소비지출에는 경상조세, 연금, 사회복지장 등이 있다.

## 2. 지니계수 비교

가계동향조사를 이용하여 지니계수를 계산할 때에는 균등화 시장소득에 인구가중치를 가중치로 하여 계산한다. 여기서 균등화 시장소득=시장소득/가구원수의 제곱근이며, 인구가중치=가구원수\*가구가중치이다.

[그림 1-2]와 <표 1-4>에서는 지니계수를 계산한 결과를 비교하여 보여주고 있다. 2013년 노동패널 09표본에서 시장소득 지니계수는 0.358로, 가계동향조사의 지니계수 0.336보다 약간(6.7%) 높다. 추세를 보면, 2009년 이후 노동패널, 가계동향조사 지니계수 모두 약간 하락하고 있다. 상위 10% 소득비중은 2009년 약간 상승하였고 2011년 이후 정체하고 있어서(홍민기, 2015) 조사자료와 약간 차이가 난다. 간략히 말해, 노동패널의 지니계수가 약간 높고, 추세는 가계동향조사와 비슷하다.

[그림 1-2] 지니계수의 비교



## 6 노동패널자료 연구(1)

〈표 1-4〉 지니계수의 비교

	가계동향조사	노동패널 09표본	노동패널 98표본	가계금융 복지조사	상위 10% 소득비중
2006	0.330		0.383		0.467
2007	0.340		0.392		0.463
2008	0.344	0.401	0.388		0.461
2009	0.345	0.375	0.380		0.456
2010	0.341	0.367	0.379		0.463
2011	0.342	0.368	0.364		0.474
2012	0.338	0.356	0.368	0.406	0.471
2013	0.336	0.358	0.370	0.389	0.473

### 제4절 노동패널의 지니계수가 왜 더 높을까?

그러면 노동패널의 지니계수가 가계동향조사보다 왜 더 높을까? 가능한 설명은 1) 연령의 문제, 2) 표본탈락(attrition)의 문제, 3) 소득 히핑(heaping)의 문제, 4) 샘플링(sampling)의 차이 혹은 소득과악 정도의 차이이다. 이하에서는 각각의 요소가 설명력을 가지고 있는지 검토한다.

#### 1. 연령 분포의 비교

노동패널은 패널조사이기 때문에 시간이 흐르면 자연스럽게 가구들이 고령화된다. 일반적으로 고령층의 소득불평등도가 청년, 장년에 비해 높기 때문에 고령화된 가구를 유지한 패널조사에서 불평등도가 높게 측정될 가능성이 있다. 이러한 가능성을 살펴보기 위해 노동패널과 인구총조사에서의 인구 분포를 비교하였고, 그 결과가 <표 1-5>에 나와 있다.

인구총조사와 노동패널을 비교하여 보면, 노동패널이 대체로 인구 분포의 대표성을 유지하고 있음을 알 수 있다. 이는 가중치를 부여할 때 연령 집단을 명시적으로 고려하기 때문이다. 따라서, 노동패널의 고령화 때문

〈표 1-5〉 인구총조사와 노동패널의 인구 분포 비교

	인구총조사 (A)	노동패널 (B)	비율 (B)/(A)
15~19세	2,922,281	3,438,414	1.00
20~24세	2,640,489	2,753,259	0.90
25~29세	3,080,113	3,844,038	1.09
30~34세	3,167,638	3,872,820	1.05
35~39세	3,483,148	3,918,746	0.96
40~44세	3,508,366	3,971,653	0.96
45~49세	3,426,557	4,233,128	1.04
50~54세	3,131,150	3,720,326	0.98
55~59세	2,209,244	2,842,406	1.03
60~64세	1,675,619	2,267,581	1.04
65~69세	1,307,441	1,727,934	0.95
70~74세	1,035,369	1,630,519	1.04
75~79세	676,567	1,033,774	0.95
80~84세	370,966	513,376	0.86
85~89세	168,706	318,556	1.17
90~94세	48,720	82,676	1.06
95~99세	9,425	25,800	1.69

에 지니계수가 높은 것은 아니라고 할 수 있다. 패널조사이기 때문에 고령화되더라도 가중치를 잘 조정하면 인구 대표성을 유지할 수 있다.

## 2. 표본탈락의 문제

패널조사의 특성상 시간이 지나면 일부 표본이 탈락하게 된다. 표본이탈의 영향을 분석한 홍민기·최효미(2014)의 연구에 따르면, 노동패널에서의 표본이탈로 지니계수가 약간 감소하게 된다. 표본이탈은 소득분포의 양쪽 끝에서 상대적으로 많이 일어나고, 그 결과 남아 있는 가구의 불평등도는 전체 가구에 비해 약간 감소한다. 표본이탈을 고려하지 않고 응답한 가구만 대상으로 계산하면 불평등도를 약간 과소 평가하게 된다. 따

## 8 노동패널자료 연구(1)

라서 표본이탈 때문에 가계동향조사에 비해 노동패널의 지니계수가 높은 것은 아니다.

### 3. 소득 히핑의 문제

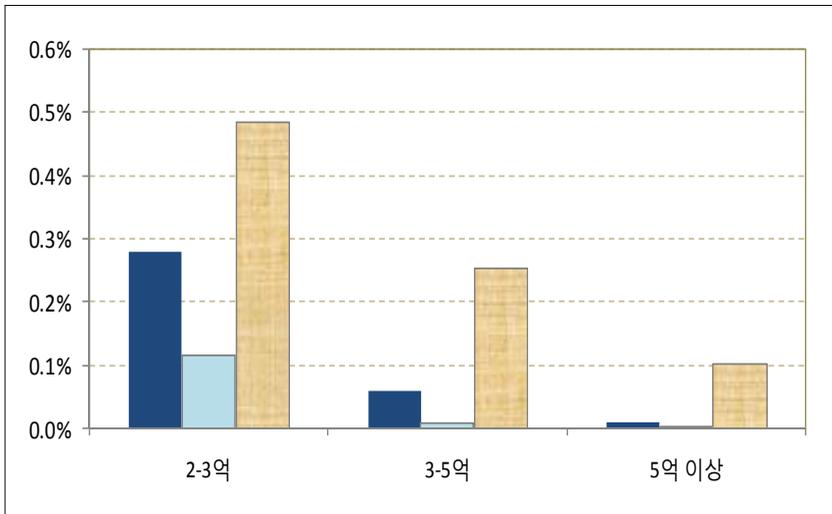
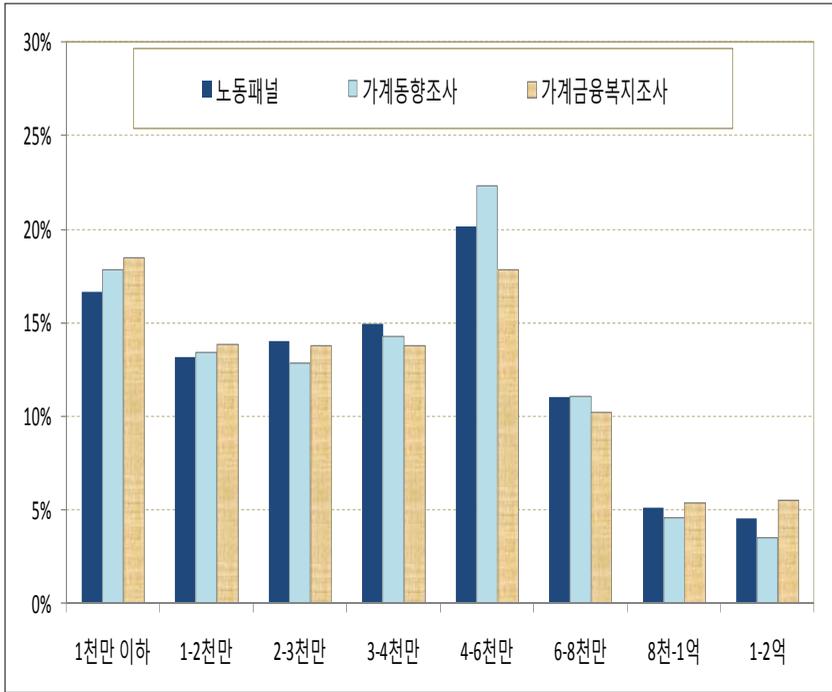
소득 히핑(heaping)이란 소득에 대해 응답자가 대략적인 값을 응답하여 특정 구간에 응답값이 집중되는 현상을 말한다. 소득 히핑이 불평등 척도에 미친 영향에 대해서는 홍민기(2015)의 연구가 있다. 이 연구에서는 노동패널 조사의 소득 히핑 문제를 MCEM(Monte Carlo Expectation Maximization)의 방법으로 보정하고, 원자료와 히핑을 보정한 자료를 이용하여 양극화 지수를 계산한 결과를 비교한 결과, 소득 히핑은 지니계수에 영향을 주지 않는다고 하였다. 그리고 양극화 지수 중 EGR(Esteban, Gradin, and Ray) 지수와 FW(Foster and Wolfson) 지수도 소득 히핑의 영향을 크게 받지 않았다. 반면, DER(Duclos, Esteban, and Ray) 지수는 소득 히핑에 민감해서 소득 히핑이 심한 경우 DER 지수는 실제의 극화 정도를 과대 평가하게 된다.

### 4. 소득 조사의 차이

마지막으로 노동패널과 가계동향조사에서 파악된 소득분포를 비교하여 본다. 가구가중치를 사용하여 소득분포를 그린 결과가 [그림 1-3]에 나와 있다. 노동패널에서는 가계동향조사에 비해 저소득과 고소득이 많이 포착된다. 가계동향조사와 비교하여 노동패널에서는 2천만~4천만 원이 상대적으로 많이 포착되고, 중간소득인 4천만~6천만 원은 적게 포착된다.

가장 크게 차이가 나는 부분은 고소득 가구이다. 노동패널에서는 1억 원 이상 고소득이 상대적으로 더 많이 포착되고 있다. 가계동향조사보다 표본의 크기가 더 큰 가계금융복지조사는 노동패널에 비해 고소득을 더 많이 포착하고 있고, 저소득도 더 많이 포착하고 있어서 지니계수도 높게 나온다. 조사된 소득분포가 퍼진 정도는 가계금융복지조사, 노동패널, 가계동향조사 순서며, 지니계수도 이 순서대로이다.

[그림 1-3] 노동패널, 가계동향조사, 가계금융복지조사 소득분포의 비교



## 제5절 결 론

이 장에서는 노동패널과 가계동향조사의 소득 불평등지표가 왜 다른지를 검토하였고 그 결과 조사된 소득분포의 차이로부터 비롯된 것임을 보여주었다. 소득 조사와 관련하여 두 가지 시사점을 밝히고자 한다.

첫째, 조사마다 소득에 대한 분류가 약간씩 다른데, 왜 다른지를 검토하면서 개념에 대한 정의, 공표방식을 공론화할 필요가 있다. 소득은 일반적으로 임금, 사업소득, 재산소득으로 구분된다. 조사마다 차이가 나는 것은 부동산임대소득, 부동산매매차익, 주식매매차익에 대한 분류이다. 부동산임대소득의 경우 국제청과 통계청의 조사에서는 사업소득으로 분류하고 있고, 노동패널에서는 재산소득으로 분류하고 있다.

전통적으로 임대료는 노동의 댓가로 얻은 것이 아니라 재산을 보유한 댓가로 얻는 불로소득으로 보았기 때문에 임대료를 재산소득으로 구분하여 왔다. 그런데, 점차 부동산 임대료를 주거서비스 제공에 대한 댓가로 보는 경향이 생기면서 사업소득으로 구분하고 있다. 국제청 자료에서 1960~70년대에는 부동산소득과 사업소득을 따로 구분하여 통계를 제시하였는데, 최근에는 부동산임대소득과 사업소득을 ‘사업(부동산임대)소득’으로 합쳐서 통계를 제시하면서 부동산임대소득에 대한 정보를 찾기 어려워졌다.

먼저, 임대료를 어떻게 구분할 것인가에 대한 논의를 하는 것이 바람직하다. 구분의 문제와는 별도로, 최소한 부동산임대소득과 사업소득을 구분하여 조사하고 통계를 제시하여 연구자가 소득의 면모를 다 알 수 있도록 하는 것이 바람직하다.

두 번째, 불평등지표를 개선하기 위한 노력이 필요하다. 공식 지니계수를 산정하기 위해 사용되는 가계동향조사는 고소득을 충분히 포착하지 못하고 있어서 결과적으로 실제 불평등도를 낮게 측정하게 된다. 공식 지니계수를 보면 한국의 불평등 정도가 크지 않은 것으로 보이지만 최상위 소득 집중도로 보면 국제적으로도 매우 높은 수준에 와 있다. 궁극적으로

는 국세자료와 조사자료를 결합하여야 소득분포, 불평등도의 실제 모습을 알 수 있을 것이다.

## 제 2 장

# 한국노동패널 자료를 통해 본 소득이동성

### 제1절 머리말

소득이동성은 시간이 경과함에 따라 개인이나 가구의 소득지위가 얼마나 빈번히, 어느 방향으로 변화하는지를 의미하는 개념이다. 일반적인 소득 불평등지표가 특정 시점에서 파악한 소득분배의 문제를 보여준다면, 소득이동성 지표는 그 소득분포 안에서 특정 개인이나 가구의 위치가 얼마나 역동적으로 변화하는지를 보여준다. 계층간 이동이 둔화되고 소득 지위가 고착화된 사회일수록 사회통합의 정도가 낮아질 수 있음을 감안할 때 소득이동성은 사회통합의 중장기적 양상을 파악할 수 있는 중요한 지표이다.

이 장의 목적은 한국노동패널자료(이하 노동패널)를 통해 가구 및 개인의 소득이동성 변화 추이를 분석하고, 이를 다른 패널자료에서 나타나는 경향성과 비교 분석하는 것이다. 소득이동성에 대한 분석은 패널자료를 통해서만 가능하며, 비교적 긴 시계열의 자료를 제공하는 국내 패널자료가 많지 않다는 점에서 노동패널 자료와 타 자료와의 비교를 통해 노동패널 자료의 품질을 점검하는 것은 불가피한 방법이라 할 수 있다. 이 장에서는 이러한 목적하에 구체적으로 다음과 같은 분석을 수행할 것이다.

먼저, 소득이동성의 분석 결과에 영향을 미칠 수 있는 노동패널 데이터상의 기본적 특징들에 대해 점검하고, 소득이동성 변화 추이를 분석하며,

또한 소득이동성 변화 추이상 특이한 사항이 관측되는 연도의 자료에 대해 데이터의 품질 관련 사항을 분석하고자 한다. 이 과정에서 한국복지패널자료에서 나타나는 소득이동성 양상과 비교를 통해 노동패널 데이터의 품질 관련 특이점을 살펴볼 것이다.

소득이동성 분석은 패널조사자료의 품질 점검 및 개선이란 맥락에서 특별히 의미를 지닌다고 할 수 있다. 먼저 소득이동성을 분석하기 위해서는 가구 및 개인의 각 연도별 자료를 결합하여 사용하여야 한다. 따라서 소득이동성은 횡단면적 분석에 비해 패널자료 고유의 특성을 살릴 수 있는 분석 주제이다. 또한 소득이동성 분석은 각 연도별 자료의 소득분포와 밀접한 연관이 있다는 점에서 연도별 소득분포의 특징을 간접적으로 검토할 수 있다.

한편 본 연구의 목적이 소득이동성과 관련된 자료(주로 소득 관련 변수의)의 품질을 검토하는 데 있는 만큼 소득이동성에 영향을 미치는 다양한 요인들에 대한 분석은 연구 내용에 포함되지 않을 것이다. 또한 소득이동성의 크기 또는 추세적 변화와 관련하여 규범적 해석을 부가하는 것은 본 연구의 범위를 넘어서는 것으로 보아, 이하에서는 이동성 변화 추이에 대한 별도의 정책적 제안이나 시사점을 언급하지는 않을 것이다.

## 제2절 자료 및 분석방법

### 1. 데이터

이 연구에서는 한국노동패널(이하 노동패널) 9~17차 자료와 한국복지패널(이하 복지패널) 1~9차 자료를 비교하여 분석할 것이다. 노동패널 자료를 9차 자료부터 이용하는 것은 복지패널의 1차년도 자료와 비교하기 위해서이다. 소득이 발생한 시점을 기준으로 비교하자면 노동패널은 9차(2005년)부터 17차(2013년), 복지패널은 1차(2005년)부터 9차(2013년)까지 동일한 기간의 소득을 보고하고 있는 것이다.

분석에 사용되는 자료의 경우, 노동패널은 가구 및 개인용 데이터셋을, 복지패널은 가구 데이터셋과 가구-개인 결합 데이터셋을 이용하였다. 복지패널 자료에서 7차 조사에 새롭게 추가된 표본은 분석에서 제외하였다

## 2. 소득범주

소득이동성을 분석할 때, 소득획득의 단위를 무엇으로 볼 것인가와 어떤 소득범주를 이용할 것인가가 문제가 된다. 이 장에서는 소득을 가구단위와 개인단위 소득으로 각각 구분하여 비교할 것이다. 이때 가구소득은 가처분소득을, 개인소득은 노동소득을 이용하도록 할 것이다. 가구 소득 이동성 비교에서 가처분소득을 기준 범주로 선택하는 가장 큰 이유는 노동패널의 경우 가구 세전소득을 보고하지 않기 때문이다. 따라서 세전소득 범주를 별도로 보고하는 복지패널과 비교하기 위해서는 가처분소득을 분석 대상 소득범주로 채택하는 것이 불가피하다. 노동패널 자료에서 가구 가처분소득은 다음과 같은 식으로 표현된다.

$$\text{가구 가처분소득} = \text{세후 총소득} - \text{국민연금보험료} - \text{건강보험료}$$

한편 개인 노동소득에 대해서는 노동패널이 세전, 세후 소득 정보를 모두 제공하는 데 반해 복지패널의 경우 세전/후 소득을 구분하지 않고 있다. 따라서 개인소득의 이동성을 비교할 때에는 세전 노동소득을 비교 범주로 삼을 것이다.

가구단위의 소득 비교 시 가구원수의 차이를 감안해주기 위해 균등화 지수를 적용하는 것이 일반적이다. 이 장의 분석에서 균등화지수는 가구 소득을 가구원수의 제곱근으로 나누는 방식을 공통적으로 적용하였다.

## 3. 자료의 결합 및 가중치

소득이동성이 두 해에 걸친 소득지위의 변동을 의미하는 것인 만큼 개인이나 가구의 소득이동성을 살펴보기 위해서는 인접한 두 해의 데이터를 결합하여 사용하여야 한다. 이를 위해 노동패널의 경우 9~10차, 10~

11차 자료를 결합하는 방식으로 모두 8개의 데이터셋을 구성하였다. 9~17차 자료를 모두 결합하여 하나의 데이터셋으로 만들지 않은 이유는 모든 자료를 결합하여 사용할 때 발생할 수 있는 정보의 손실을 막기 위해서이다.

그런데 각 연도별 소득지위를 파악할 경우 소득분위 등의 구분은 당해 연도 소득분포를 기준으로 이루어진 것이다. 즉 결합데이터에서 표본이 유지되는 관측치만을 대상으로 구분한 것이 아님을 유의할 필요가 있다.

노동패널 자료의 가중치는 1998년 표본 가중치를 사용하였다. 결합된 데이터에서 소득 및 기타 소득지위 관련 정보는 당해 연도의 것을 사용하며, 가중치는 전년도의 값을 이용하였다.

### 제3절 소득지위 관련 자료 상태의 검토

#### 1. 소득이동성에 영향을 미치는 데이터의 특성

##### 가. 소득이 결측인 관측치의 수

이후 제4절에서 살펴보겠지만 소득이동성은 크게 두 가지 방식으로 파악된다. 하나는 기준연도와 비교연도의 소득값의 관계(두 해 소득 간 상관관계수, 소득변동비율 등)를 이용하여 분석하는 방법이고, 다른 하나는 각 가구 또는 개인이 당해 연도의 소득분포에서 차지하는 상대적 지위(소득분위)의 변동으로 파악하는 방식이다.

이 가운데 소득분위의 변동으로 소득이동성을 파악하기 위해서는 소득분위의 구분이 적절한 방식으로 이루어지는지를 판단할 필요가 있다. 즉, 개념적으로 각 소득분위에는 동일한 비중의 가구 또는 개인이 포함되어야 하나, 데이터 분포의 특성상 특정 분위가 타 분위에 비해 더 많은 가구 또는 개인을 포함할 수 있는데, 실제 데이터상에서 이런 문제가 나타나지는 않는지 점검해야 하는 것이다.

16 노동패널자료 연구(1)

이러한 경향은 소득 하위분위에서 쉽게 나타나는데, 소득이 0인 관측치가 지나치게 많거나, 분위수(quantile), 즉 분위 간 경계값에 동일한 소득을 갖는 관측치가 다수 몰려 있는 경우 나타나게 된다.

먼저 가구소득이 관측되지 않은(missing) 표본의 수와 비중은 다음 <표 2-1>과 같다.

가처분소득이 결측인 가구의 비중은 가중치를 적용하지 않은 경우 전체 표본 가운데 약 1~2%, 가중치를 적용할 경우 약 1~3%로 나타났다. 2008년 소득을 보고하고 있는 12차 자료의 경우 소득이 결측인 관측치의 비중이 예외적으로 높았다.

[그림 2-1]은 개인 노동소득이 결측치인 표본의 비중이 어떻게 변했는지를 보여주는데, 가구 가처분소득 결측치의 비중에 비해 높게 나타나는 것이 특징이다. 특히 2009년 소득을 보고하는 13차 표본까지는 결측치의 비중(가중치를 적용한)이 10%를 넘는데, 이는 소득을 10개 분위로 구분할 때 1분위가 정확히 정의되지 않을 수 있음을 의미하는 것이다. 이하의 분석에서는 이를 고려하여 1~2분위를 하나의 분위로 간주하여 분석하였다.

<표 2-1> 노동패널 자료의 가구 가처분소득 결측 표본의 비중

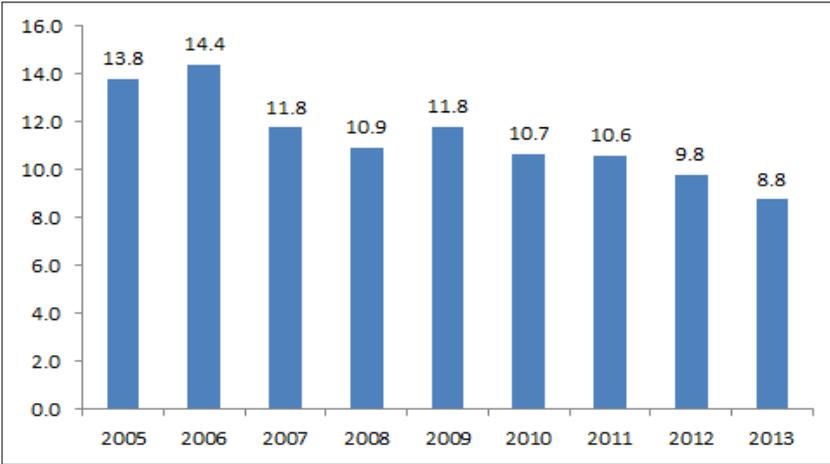
(단위: 가구, %)

	소득기준 연도	결측표본수	전체표본수	표본비율 (%)	가중치적용비율 (%)
9차	2005	83	5,001	1.66	1.74
10차	2006	79	5,069	1.56	1.56
11차	2007	55	5,116	1.08	1.07
12차	2008	288	6,721	4.29	2.59
13차	2009	120	6,683	1.80	1.79
14차	2010	117	6,686	1.75	2.05
15차	2011	138	6,753	2.04	2.69
16차	2012	133	6,785	1.96	2.28
17차	2013	89	6,838	1.30	1.52

자료: 한국노동패널 9~17차 연도별 연결자료.

<표 2-2>는 결측치를 어떻게 처리하는가에 따라 소득분위별 비중이 달라지는 양상을 9차 자료를 이용하여 예시하고 있다. 결측치를 제거하고

(그림 2-1) 노동패널 자료에서 개인 노동소득이 결측인 표본의 비중 변화 (단위: %)



자료: 한국노동패널 9~17차 연도별 연결자료.

<표 2-2> 가구소득 결측치 처리방법별 분위 분포(노동패널 9차 자료 이용) (단위: %)

결측치 처리방식 분위	결측제거		결측=0		결측=최하위	
	비중	누적비중	비중	누적비중	비중	누적비중
1	9.91	9.91	10.00	10.00	11.48	11.48
2	10.08	19.99	9.99	19.99	9.90	21.38
3	10.01	30.00	10.01	30.00	9.83	31.22
4	9.97	39.97	10.00	40.00	9.80	41.01
5	10.03	49.99	9.86	49.86	9.85	50.86
6	9.99	59.98	10.11	59.97	9.82	60.68
7	9.95	69.94	10.01	69.98	9.78	70.46
8	10.06	79.99	9.99	79.97	9.88	80.34
9	10.00	90.00	10.03	90.00	9.83	90.17
10	10.00	100.00	10.00	100.00	9.83	100.00

자료: <노동패널> 가구데이터 각 연도.

18 노동패널자료 연구(1)

소득분위를 구분할 경우 각 분위별 가구비중은 9.95~10.08%의 값을 지닌다(분위별 비중의 표준편차는 약 0.05). 반면 결측치를 0으로 간주할 경우 분위별 비중은 9.86~10.03%(표준편차 약 0.06)가 된다.

한편 분위수(분위간 경계값)를 도출하는 단계에서는 결측치를 사용하지 않되, 소득이 결측된 가구를 최하위 분위로 간주할 경우 분위별 비중은 9.82~11.48%(표준편차 0.52)로 나타난다.

분석의 목적에 따라 결측치를 어떻게 간주하고, 어떻게 처리할 것인가가 달라질 수 있겠으나, 이하에서는 소득분위 구분 시 결측가구를 제거하고 분석할 것이다.

<표 2-3>은 노동패널 가구데이터의 각 연도별 가구 가처분소득의 소

<표 2-3> 노동패널 자료의 가구 가처분소득 분위 분포

(단위: %)

	자료 차수	9차	10차	11차	12차	13차	14차	15차	16차	17차
	소득 년도	2005	2006	2007	2008	2009	2010	2011	2012	2013
10분위	1	9.9	10.0	10.0	9.9	10.0	10.0	10.0	10.0	10.0
	2	10.1	10.0	9.9	10.1	10.0	10.0	10.0	10.0	10.0
	3	10.0	10.0	10.1	10.0	10.0	10.0	10.0	10.0	10.0
	4	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	5	10.0	10.0	10.0	10.0	10.0	10.0	10.0	9.9	10.0
	6	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.1	10.0
	7	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	8	10.1	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	9	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	10	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
5분위	1	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.1
	2	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	19.9
	3	20.0	20.0	19.9	20.1	20.0	20.0	20.0	20.0	20.0
	4	19.9	20.0	20.0	19.9	20.0	20.0	20.0	20.0	20.0
	5	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0

자료: <노동패널> 가구데이터 각 연도.

득분위별 분포 변화를 보여준다. 결측치를 제외하고 분위를 구분했을 때 대부분의 소득분위는 균등한 비중을 차지하고 있으며, 균등 비중을 초과하거나 그 미만인 분위의 경우는 많지 않았다.

#### 나. 탈락가구의 분위별 분포의 차이

앞서 설명한 바와 같이 인접한 두 해 사이의 소득분위 변화를 파악할 때, 두 해 모두 소득분위가 관측된 가구만 포함시킴에 따라 비교대상에서 탈락한 가구가 발생하게 된다. 이때 표본에서 탈락하는 가구가 어느 소득분위에서 발생하는가에 따라 소득이동성 결과에도 영향을 줄 수 있다.

이 문제를 검토하기 위해 인접한 두 해 모두에서 소득분위가 정의되지 않는(즉 소득이동성을 파악할 수 없는) 가구가 어느 분위에서 발생하는지를 전년도 소득분위를 기준으로 비교하였는데, <표 2-4>는 이에 대한 결과를 보여준다.

<표 2-4>에서 알 수 있듯이 최하위 소득분위에서 평균 약 4.6%의 가구가 결함데이터 구축과정에서 탈락되었고 다음으로 9분위, 10분위의

<표 2-4> 가구 결함자료 구축 시 분위별 탈락가구 비율

(단위: %)

분위	9차	10차	11차	12차	13차	14차	15차	16차	평균
	2005	2006	2007	2008	2009	2010	2011	2012	
1	4.3	5.2	4.8	5.8	3.7	4.3	3.6	4.8	4.6
2	3.9	4.1	5.0	1.8	1.9	2.3	2.8	2.2	3.0
3	2.9	5.1	3.1	2.6	2.9	2.0	1.6	0.4	2.6
4	3.4	4.2	4.1	5.7	3.2	2.5	2.0	1.3	3.3
5	2.0	3.4	3.0	4.1	2.7	1.3	2.1	1.5	2.5
6	3.7	3.6	3.4	4.7	3.2	1.9	2.5	1.5	3.1
7	2.8	2.7	3.7	1.9	2.6	2.9	1.7	1.2	2.4
8	2.6	2.9	3.1	3.5	2.7	2.9	1.3	0.0	2.4
9	4.8	4.7	5.3	2.0	4.4	3.3	3.1	1.1	3.6
10	4.1	3.6	4.3	5.3	4.5	2.8	2.8	0.9	3.5

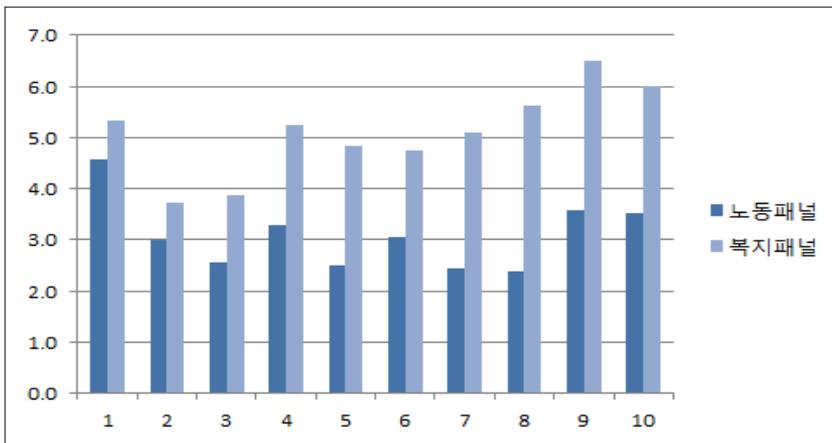
자료: <노동패널> 가구데이터 각 연도.

순으로 탈락가구 비율이 많은 것으로 나타난다.

한편 노동패널과 복지패널의 가구데이터에서 가구소득 분위별 탈락률을 비교하여 보면 복지패널 자료의 탈락률이 더 높음을 알 수 있다. [그림 2-2]는 이를 비교하여 보여준다. 즉 연결데이터를 구축할 때 각 소득분위 별로 어느 정도씩의 가구가 탈락하는지를 비교한 것인데, 노동패널의 경우 9~17차 기간 동안 1분위에서 매년 평균 4.6%의 가구가 탈락한 반면 한국복지패널은 1~9차 자료 기간 동안 1분위에서 매년 5.3%의 가구가 탈락하였다. 그런데 이러한 표본 탈락률의 격차는 각 자료의 조사 차수가 얼마나 오래되었는지에 대한 차이 때문에 비롯되었을 수 있다는 점에 유의할 필요가 있다.

(그림 2-2) 노동패널과 복지패널의 가구소득 분위별 탈락률 비교(2005~12년간 평균)

(단위: %)



자료: <노동패널> 9~17차 가구 연결데이터, <복지패널> 1~9차 가구연결데이터.

## 제4절 소득이동성의 의미와 지표

소득 및 소득지위의 이동성은 다양한 방법으로 정의될 수 있으며, 각

방법에 따라 특정한 소득이동성 지표가 해당된다(Fields, 2008; 이병희 외, 2013). 소득지위의 변동을 나타내는 지표들을 간략히 설명하면 다음과 같다.<sup>2)</sup>

첫째는 소득분위의 변동률이다. 이는 전년도 대비 금년의 소득분위가 변화한 가구 또는 개인의 비율을 의미한다. 소득분위 변동률을 측정하기 위해서는 특정 가구(개인)가 소득분포상에서 어떤 분위(10분위 또는 5분위로 구분)에 속하는지를 파악한 후, 기준연도와 비교연도 사이에 분위가 상승/하락/불변인지를 분석하면 된다. 분위 불변가구의 비율이 높을수록 소득이동이 활발하지 않은 것이다. 다만 분위 변동비율에 대한 규범적 해석에 유의할 필요가 있다. 즉 분위 변동비율이 높은 것이 바람직한지, 낮은 것이 바람직한지는 보는 관점에 따라 다른 해석을 할 수 있다.

둘째는 두 해의 소득 간 상관계수이다. 이는 기준연도의 소득과 비교연도의 소득 사이의 상관계수를 계산함으로써 측정된다. 상관관계가 높다는 것은 소득지위의 이동성이 그만큼 낮음을 의미한다.

셋째, 소득의 절대적 (또는 상대적) 수준의 변동이다. 기준연도 대비 비교연도의 소득수준이 어느 정도인지(소득의 증가분)를 파악함으로써 소득의 이동성을 표현할 수 있는데, 이때 사회 전체의 소득증가분은 다음과 같이 집계된다.

$$M_d = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)$$

한편, 위의  $M_d$ 에서 소득이 늘어난 가구와 소득이 감소한 가구의 변화가 서로를 상쇄하는 것을 방지하기 위해 소득변화분의 절대값을 이용할 수 있다. 이때 소득변화분은 다음과 같이 표현된다.

$$M_{nd} = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

한편, 평균소득이 증가함에 따라 각 가구 소득의 절대값도 증가하게 되고, 따라서 두 해 사이의 소득격차값도 커지게 된다. 이러한 문제를 통제

2) 소득이동성 지표에 대한 설명은 이병희 외(2013)의 5장의 내용을 일부 인용한 것이다.

해주기 위해 소득의 절대적 수준 대신에 평균소득 대비 개별 가구소득의 비율(상대소득)을 사용하기도 한다. 이때 소득이동성을 나타내는 상대소득의 변화분은 다음과 같이 표현된다.

$$Ms = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i}{\mu_y} - \frac{x_i}{\mu_x} \right|$$

이하에서는 소득분위의 변동과 함께 두 해의 소득 상관계수, 그리고 소득값 변동지표를 소득이동성 지표로 사용할 것이다.

## 제5절 가구단위의 소득이동성 변화 추이

### 1. 소득분위 이동성

<표 2-5>는 노동패널 자료에서 가구단위의 소득분위 변동 양상을 보여준다. 10분위로 파악한 가구 소득분위가 전년 대비 불변인 가구의 비중이 약 33~40%(평균 약 35.3%)이며, 5분위로 구분시에는 평균 약 53.8%

<표 2-5> 가구 소득분위의 전년 대비 변화 분포

(단위: %)

		2006	2007	2008	2009	2010	2011	2012	2013
10분위 구분시	불변	34.3	34.2	32.9	33.0	35.7	36.9	35.6	39.4
	변동	65.7	65.8	67.1	67.0	64.3	63.1	64.4	60.6
	상승	32.5	32.4	34.9	32.9	32.8	31.0	32.9	29.7
	하락	33.2	33.4	32.3	34.2	31.5	32.1	31.5	30.9
5분위 구분시	불변	52.9	52.1	52.4	51.0	53.3	55.2	55.4	57.9
	변동	47.1	47.9	47.6	49.0	46.7	44.9	44.6	42.2
	상승	23.5	23.8	24.4	24.2	24.0	22.3	22.5	20.8
	하락	23.6	24.1	23.2	24.8	22.8	22.6	22.2	21.3

자료: <노동패널> 가구데이터 각 연도.

수준이다. 소득분위의 변동성은 2009년을 정점으로 이후 감소하는 추이이며, 특히 13차 자료(2013년 소득)에서 변동성이 급히 하락하는 모습을 보이고 있다.

(그림 2-3) 소득 10분위 변동가구 비율 비교

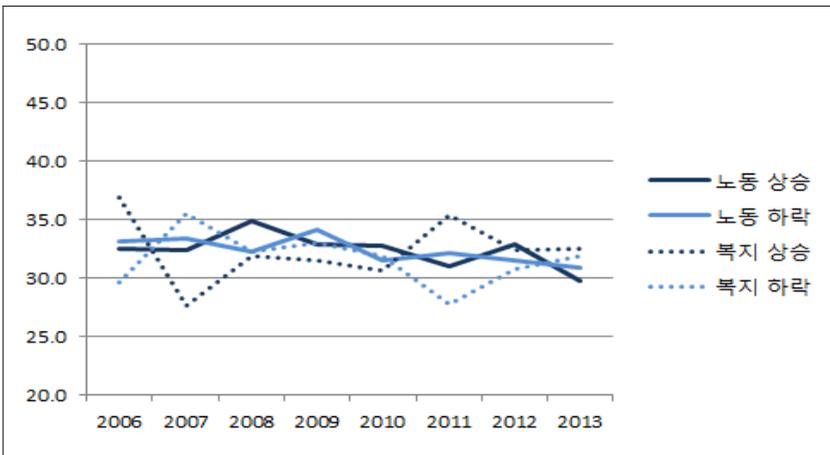
(단위: %)



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

(그림 2-4) 소득 10분위 상승 및 하락 가구 비율 비교

(단위: %)



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

이상의 결과를 가구 소득분위 변동에 대한 복지패널 자료상의 추이와 비교하여 보면, [그림 2-3]과 같이 전체적인 변화 방향은 유사한 것으로 나타난다. 두 자료 모두 2009년이 변동비율의 최고점을 보여주는 해이다.

한 가지 특징적인 점은 2013년의 전년 대비 분위 변동 가구 비율이 노동패널에서 큰 폭으로 떨어진다는 점이다. 2006~13년간 전년 대비 10분위 변동가구의 비율은 복지패널에서는 약 64.0%, 노동패널에서는 65.1%로 나타나 평균적으로 1.1%포인트의 차이를 보인다.

## 2. 전년도 소득과의 상관계수

소득이동성은 전년도 소득과의 상관계수를 통해서도 측정할 수 있다. 상관계수가 높을수록 소득이동성이 낮을 가능성이 높으며 반대로 상관계수가 낮으면 소득이동성이 높을 가능성이 높아진다.

<표 2-6>은 전년 대비 금년 소득의 상관계수가 어떻게 변화했는지를 보여준다. 전년도 소득과 상관계수의 평균값은 약 0.399로 복지패널의 같은 기간 상관계수 평균값 0.5026에 비해 약 0.1 이상 낮은 수치이다. 즉 상관계수를 통해 보았을 때 노동패널의 소득이동이 좀 더 활발한 것으로 나타난다.

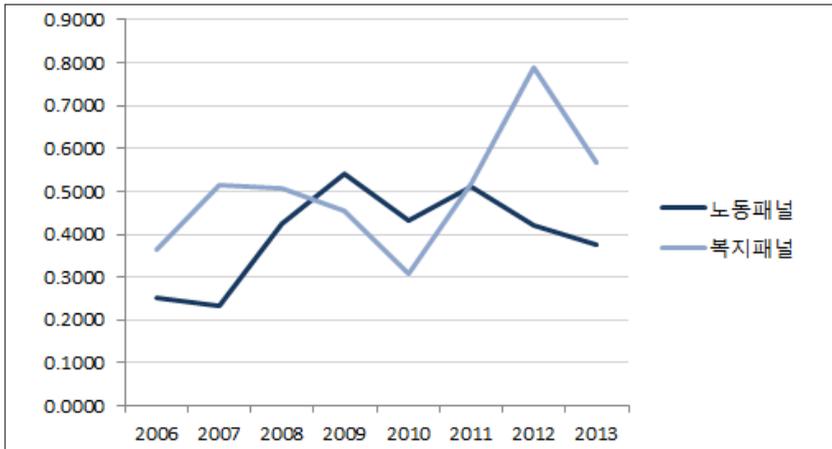
상관계수의 연도별 변동성을 비교할 때 상관계수의 연도별 표준편차는 노동패널 자료는 약 0.135, 복지패널은 약 0.145로 나타났다. 즉 노동패널의 소득 상관계수가 좀 더 안정적으로 변화하고 있다고 할 수 있다. 한편 연도별 상관계수의 변화 추이는 두 데이터에서 상이하게 나타나고 있다(그림 2-5 참조).

<표 2-6> 전년도와의 가처분소득 상관계수

	2006	2007	2008	2009	2010	2011	2012	2013
상관계수	0.2532	0.2322	0.4243	0.5406	0.4329	0.5097	0.4208	0.3763

주: 상관계수는 모두 오차확률 1% 수준에서 유의함.  
자료: <노동패널> 가구데이터 각 연도.

(그림 2-5) 전년도 소득과의 상관계수 변화 추이 비교



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

### 3. 전년도 소득과의 격차

마지막으로 <표 2-7>은 전년도 소득과의 절대적 및 상대적 소득격차 변화 추이를 보여준다. 2008년의 경우 전년 대비 소득값이 평균 약 61.8만 원 하락한 것으로 나타났다.

2008년의 전년 대비 소득이 감소하는 것은 노동패널 자료에서 나타나는 특이한 모습이다. 복지패널 자료의 경우 소득증가폭이 둔화되긴 하지만 소득값이 감소하지는 않았다. 2008년이 글로벌 경제위기의 해였다는 점에서 소득감소가 나타나는 것은 불가능한 현상은 아닐 것이다. 다만, 2008년의 경제성장률이 (+)였다는 점에서 가구소득 증가율이 (-)였다는 점에서 이례적인 것은 사실이다.

복지패널 자료의 경우 2008년의 소득이 전년 대비 증가한 것은 소득조사항목의 추가와 무관하지 않을 것으로 추정된다. 2008년 소득에 대한 조사에서 복지패널 자료는 일시적으로 생겨난 각종 공적이전소득(예, 유류 환급금 등)에 대한 질문을 추가함으로써 소득이 전년 대비 증가한 것으로 나타날 요인을 갖고 있었기 때문이다.

다음의 [그림 2-6]에서 비교되고 있듯이 노동패널자료의 전년 대비 소

〈표 2-7〉 전년도와의 절대적 및 상대적 소득수준 변화

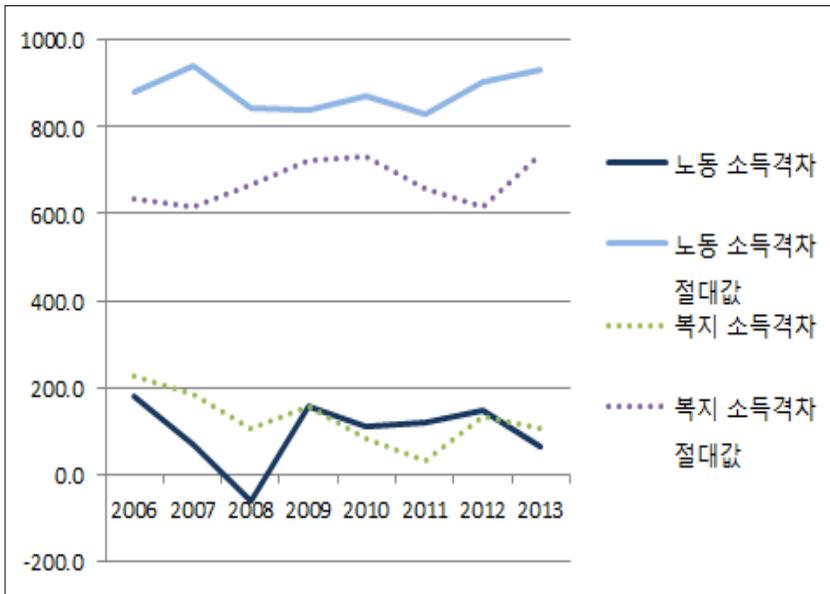
(단위: 만 원/연, %포인트)

	2006	2007	2008	2009	2010	2011	2012	2013
소득격차	178.6	67.9	-61.8	155.6	107.9	119.2	147.4	62.6
소득격차(절대값)	878.4	942.5	844.1	838.1	872.4	831.4	904.0	933.5
평균대비 변화 비율	0.2	-0.9	-0.2	-0.4	0.4	0.3	-0.1	-1.0
평균대비 변화(절대값) 비율	44.0	44.4	40.0	38.2	37.6	34.3	35.3	35.1

자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

(그림 2-6) 전년 대비 소득격차 비교

(단위: 만 원/년)



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

소득격차의 값은 복지패널보다 낮은 것으로 나타난다. 연도별 소득격차 평균값을 2006~13년 전 기간에 대해 평균했을 때 노동패널은 약 97.2만 원인 반면 복지패널은 126.9만 원이다. 그러나 연도별 변동치를 절대값으로 계산할 경우 2006~13년간 소득변동값의 평균은 노동패널에서는 약 881만 원이고 복지패널에서는 약 672만 원으로 나타난다.

## 제6절 개인의 소득이동성 변화 추이

### 1. 개인 소득이동성 분석 대상 및 방법

개인의 소득지위 변동에 대한 분석에서는 15세 이상 64세 이하의 개인 중 비경제활동인구를 제외한 개인에 대해 소득지위의 이동을 분석하였다. 비경제활동인구를 포함할 경우 노동소득이 0인 집단이 많고 이것이 소득이동성을 둔화시키는 방향으로 강력한 영향을 미칠 것이기 때문에 분석에서 제외하였다. 소득은 앞서 언급하였듯이 개인의 세전 노동소득을 기준으로 분석하였다. 세전소득을 비교하는 것은 복지패널의 개인 소득이동성과 비교하기 위한 것이다(복지패널에서는 개인의 세후소득을 별도로 보고하지 않음).

한편 노동패널은 노동소득이 없는 가구에 대해서 소득을 대부분 결측(missing)으로 처리하고 있다. 0의 소득값이 보고된 경우는 10차 조사(7개 관측치)와 14차(1개 관측치) 조사에 불과하다. 그런데 경제활동에 참여하는 개인의 노동소득이 없는 경우를 모두 결측치로 처리한 것으로 보아, 이 장의 분석에서는 결측값을 다시 변환하였다.

이와 같은 소득처리 방법의 결과 경제활동인구 가운데 소득값이 0인 개인의 비중이 최근 두 해를 제외하고는 10%를 넘고 있다(그림 2-2 참조). 이에 따라 소득분위를 10분위로 나눌 경우 하위 1, 2분위의 구분이 불가능해진다. 참고로 이러한 특성은 복지패널자료와 상이한 점이다. 복지패널은 1차 자료(2006년 소득 기준)에서 0 소득 개인이 약 14.9%인 경우를 제외하면 다른 연도에서는 모두 10% 미만이다. 따라서 이하의 분석에서는 1, 2분위를 구분하지 않고 하나의 분위로 간주하였다. 이러한 방법은 복지패널과의 비교에서도 그대로 유지된다.

## 2. 개인 소득분위의 이동성

<표 2-8>은 개인 소득분위의 연도별 변동 실태를 보여준다.

한편 노동소득이 관측되지 않은 경우를 어떻게 처리하는가에 따라 이동성이 다소 다르게 나타날 수 있다. [그림 2-7]은 소득이 결측된 관측치

<표 2-8> 개인 소득분위의 이동성

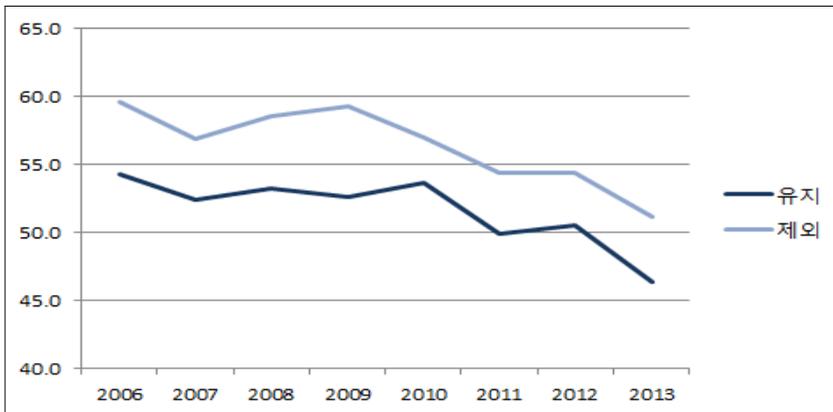
(단위: %)

		2006	2007	2008	2009	2010	2011	2012	2013
10분위	불변	45.7	47.6	46.8	47.4	46.4	50.1	49.5	53.6
	변동	54.3	52.4	53.2	52.6	53.6	49.9	50.5	46.4
	상승	30.5	27.1	30.7	31.4	27.6	29.5	28.3	23.9
	하락	23.8	25.3	22.5	21.2	26.0	20.5	22.2	22.4
5분위	불변	62.5	63.8	62.3	62.9	63.3	64.6	63.8	67.5
	변동	37.5	36.2	37.7	37.1	36.7	35.4	36.2	32.5
	상승	22.5	19.6	23.0	21.6	20.2	22.3	20.1	16.2
	하락	15.0	16.5	14.7	15.5	16.5	13.1	16.2	16.3

자료: 한국노동패널 9~17차 자료.

(그림 2-7) 소득이 결측인 관측치를 유지(0으로 대체)했을 때와 제외했을 때의 분위 변동비율

(단위: %)



자료: 한국노동패널 9~17차 자료.

를 앞서 설명한 바와 같이 0으로 대체했을 경우와 분석에서 제외했을 경우 어떤 차이가 나타나는지를 보여준다.

결측치를 제외하는 경우 소득의 이동성은 더 활발하게 나타난다. 이는 0 소득을 유지할 경우 최하위분위(실제로는 1~2분위)의 비중이 커지고, 지속적으로 최하위분위를 유지하는 개인의 비중이 높아짐에 따라 개인 소득이동성이 둔하게 나타나기 때문일 것으로 추정된다.

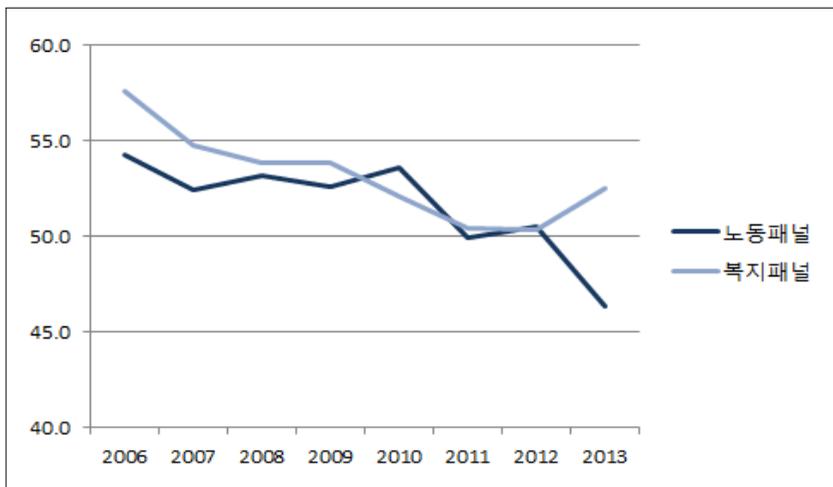
하지만 분위 변동성의 이동 추이는 결측치의 처리방법과 무관하게 유사하게 나타나며, 최근 조사연도로 올수록 점차 감소하는 모습을 보이고 있다.

분석 결과를 제시하지는 않았으나, 전년도 소득과의 상관계수에서도 결측치를 0으로 대체할 경우의 상관계수가 다소 높게 나타난다. 이는 두 해 모두 소득이 0인 개인의 영향일 것이다. 다만 이 경우도 연도별 변화 추이는 결측치 처리방법과 무관하게 유사한 것으로 나타난다.

개인소득의 분위 변동을 복지패널과 비교할 때 [그림 2-8]에서 보듯이 상당히 유사한 추이가 발견된다. 소득분위 변동자의 비율은 지속적으로 감소하는 추이를 보이고 있다. 2006~13년 기간 동안 소득분위 변동자의

(그림 2-8) 개인소득 분위(10분위 구분) 변동자의 비율 비교

(단위: %)



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

### 30 노동패널자료 연구(1)

비율 평균은 노동패널의 경우 51.6%, 복지패널의 경우 53.2%로 복지패널이 다소 높았다.

2012~13년간 소득분위 변동비율 변화가 상이한 점은 특징적인 현상이다. 노동패널의 분위 변동자 비율이 감소하는 반면 복지패널은 이전과 달리 증가 추이를 보이고 있다.

### 3. 개인소득의 전년도 소득 대비 상관계수

인접한 두 해의 개인소득을 비교했을 때, 전년 대비 당해 연도 소득의 상관계수 변화는 다음 <표 2-9>와 같다. 개인소득의 상관계수는 가구소득의 상관계수에 비해 평균적으로 더 낮은 것을 알 수 있다. 노동패널 자료의 상관계수 변화는 복지패널에 비해 더 안정적인 모습을 보이고 있다.

<표 2-9> 전년도와의 개인노동소득 상관계수

	2006	2007	2008	2009	2010	2011	2012	2013
상관계수	0.7611	0.6990	0.7017	0.8071	0.6403	0.8212	0.8324	0.8338

주: 상관계수는 모두 오차확률 1% 수준에서 유의함.  
자료: 한국노동패널 9~17차 자료.

### 4. 소득격차의 변화 추이

마지막으로 전년 대비 소득수준의 절대적, 상대적 증감은 다음 <표 2-10>과 같다. 복지패널 자료의 경우와 비교했을 때 노동패널의 소득변동액이 연평균 약 30만 원 정도 큰 것으로 나타난다. 반면 소득변동에 절대값을 이용하여 비교하면 오히려 복지패널 자료의 소득변동이 더 크게 나타나고 있다. 최근 연도에 이르러 개인소득의 전년 대비 격차는 두 데이터에서 매우 유사한 양상을 보이며 변화하고 있다.

〈표 2-10〉 전년도와의 절대적 및 상대적 소득수준 변화

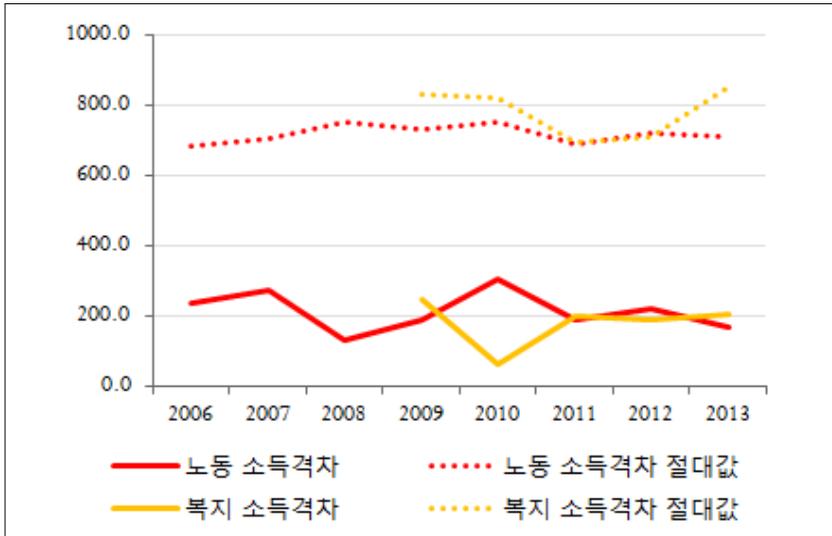
(단위: 만 원/연, %포인트)

	2006	2007	2008	2009	2010	2011	2012	2013
소득격차	236.7	273.9	133.8	187.9	303.0	191.8	223.5	170.5
소득격차 절대값	687.2	706.1	753.0	733.0	750.8	691.7	722.4	710.7
평균대비비율 변화	4.2	2.4	4.9	3.8	5.4	5.1	4.7	0.6
평균대비비율 변화 절대값	32.3	31.1	31.7	30.2	29.1	26.0	26.3	25.5

자료: 한국노동패널 9~17차 자료.

(그림 2-9) 개인소득의 전년 대비 격차 비교

(단위: %)



자료: 한국노동패널 9~17차 자료, 한국복지패널 1~9차 자료.

## 제7절 소 결

이 장은 노동패널 9~17차 자료를 이용하여 가구 및 개인의 소득이동성 추이를 분석하고, 그와 관련된 데이터의 특성을 검토하는 것을 목적으로

로 한다. 다른 데이터에서 나타나는 특성과 비교하기 위해 같은 시기의 소득에 대한 정보를 제공하는 한국복지패널 1~9차 자료의 결과와 비교하였다.

우선 소득이동성 분석을 위해 자료의 특성을 사전적으로 검토한 결과, 가구소득의 경우 매년 약 1~3%에서 결측되는 것으로 나타났으며, 특히 2008년의 소득값이 결측된 비중이 높았다. 소득지위 변동을 파악하기 위해 인접한 두 해의 자료를 이용한 가구 결합데이터 구축 시 최하위분위의 탈락률이 가장 높았고, 다음으로 9, 10분위에서의 평균적인 탈락률이 높았다.

소득이동성을 분위 변동비율, 전년도 소득과의 상관관계, 소득 변화의 절대적 및 상대적 크기 등을 이용하여 분석한 결과, 가구 및 개인의 소득 이동성 추이는 복지패널을 이용한 결과와 크게 다르지 않다는 점을 확인할 수 있었다. 다만 지표별로 변동폭은 복지패널보다 크게 나타나는 경우가 있었다. 2012~13년간 소득지위 변동가구(및 개인)의 비율이 복지패널에서는 상승하는 반면 노동패널에서는 감소하는 것으로 나타났다. 가구 소득의 전년도 소득과 상관관계 변화는 노동패널과 복지패널에서 다소 상이한 양상으로 나타났다. 전년도 소득과의 격차는 노동패널과 복지패널에서 큰 차이가 없었으나, 변동폭은 노동패널의 변동폭이 복지패널보다 항상 높게 나타나고 있었다.

### 제 3 장

## 한국노동패널조사에서 나타난 비정규직 현황과 추세 - 경제활동인구조사와 비교를 중심으로

### 제1절 서론

한국노동패널조사(이하 노동패널)는 10차년도부터 경제활동인구조사(이하 경활)의 근로형태별 부가조사와 동일한 비정규직 판별설문을 조사하였다. 여기서는 이 설문항목 분석을 통해 경활과 비정규직 규모, 추이가 얼마나 유사하게 나타나는지, 차이가 있다면 어떤 차이가 나타나고 있으며, 그 원인은 무엇인지 분석하고자 한다.

노동패널은 2009년을 대상으로 하는 12차년도 조사부터 표본추가를 통해 전국을 대상으로 하는 조사로 변경되었으므로, 12~17차년도 조사자료를 이용하여 비교·분석을 수행하겠다.

노동패널에서는 동일한 패널을 추적함에 따라 경활보다 좀 더 고령화되어 있을 것으로 예상 가능한데, 두 조사 간의 인구분포 차이가 비정규직 규모, 추세 차이의 원인이 될 수 있는지를 분석하였다.

다음으로, 설문구조의 차이나 설문을 조사하는 방식의 차이가 원인이 될 수 있는지 분석하였다. 여기까지의 분석은 2절에 수록하였다.

경활과 노동패널의 비정규직 판별 설문은 비정규직인지를 하나의 문항으로 질문하는 것이 아니라 고용계약이 있는지 등 여러 질문을 조합하는 방식으로 판별하는 복잡한 설문구조를 가지고 있다. 이로 인해 지난해의

응답이 현재의 응답과 어긋나는 비표본오차 발생 가능성을 안고 있다. 이 부분을 좀 더 주의하여 응답받지 않으면 비정규직 추이나 변동에서 오차가 발생할 수 있다. 그러므로 이로 인한 차이가 얼마나 중요한지 분석하였다.

끝으로, 이와 같은 유형의 비표본오차 교정 전후 비정규직의 정규직 또는 실업으로의 이행 확률이 어떻게 달라지는지 분석하였다. 이상 비표본오차 관련 분석을 3절에 담았다.

마지막으로 본 연구의 결론과 이상의 분석으로부터 얻을 수 있는 정책적 함의를 4절에서 정리하였다.

## 제2절 경제활동인구조사와의 비교·분석

### 1. 추 이

노동패널의 비정규직 규모와 추이는 2014년 다소 크게 집계된 것을 제외하면 경황과 비교해 문제 있어 보이지 않는다(표 3-1 참조). 그러나, 세부적으로 구분해 보면 한시근로 비중과 추이가 경황과 반대라는 차이가 나타난다. 특히 과거엔 기간제가 지나치게 적었고, 기간제 추이가 경황과 반대이며, 비기간제 한시근로가 지나치게 많이 잡히는 차이가 있다.

비전형 근로는 전체적으로는 비슷한 규모와 추이를 보인다. 세부적으로 보아도 크게 다른 추이는 나타나지 않으나, 일일근로는 노동패널이 크고, 독립도급은 경황이 큰 정도의 차이가 있다.

시간제는 노동패널의 규모가 다소 작지만 증가하는 추이라는 특징 자체는 유사하게 나타난다. 정리하면, 가장 큰 차이는 한시근로에서 나타남을 알 수 있다.

〈표 3-1〉 한국노동패널조사 통합 표본과 경향 8월 부가조사의 비정규직 추이  
(단위: %)

		2009	2010	2011	2012	2013	2014
비정규직 비중	노동패널	31.4	32.1	32.6	33.8	32.1	35.7
	경향	34.9	33.4	34.2	33.3	32.6	32.4
한시근로 비중	노동패널	20.6	22.3	23.5	26.5	24.6	28.4
	경향	21.3	19.2	19.7	19.2	18.8	18.7
기간제 비중	노동패널	9.1	9.5	9.4	10.2	10.4	12.7
	경향	17.1	14.6	15.2	15.3	15.1	14.6
비전형근로 비중	노동패널	13.1	13.2	12.7	11.7	12.5	12.9
	경향	13.9	13.4	13.9	12.9	12.1	11.3
시간제 비중	노동패널	7.4	7.6	7.9	7.6	8.0	8.8
	경향	8.7	9.5	9.7	10.3	10.3	10.8

주: 임금근로자 중 비중.

〈표 3-2〉 비전형 근로 추이  
(단위: %)

		2009	2010	2011	2012	2013	2014
과건	노동패널	1.9	1.7	1.4	1.4	1.9	1.8
	경향	1.0	1.2	1.1	1.2	1.1	1.0
용역	노동패널	2.6	3.2	3.3	2.8	3.0	3.4
	경향	3.8	3.6	3.8	3.8	3.5	3.2
독립도급	노동패널	2.2	2.0	1.9	2.0	1.8	1.7
	경향	3.9	3.5	3.5	3.1	3.0	2.8
가내	노동패널	0.9	0.7	0.8	0.6	0.6	0.4
	경향	0.6	0.4	0.4	0.4	0.4	0.3
일일	노동패널	7.3	7.0	6.8	6.0	6.3	6.5
	경향	5.4	5.1	5.5	4.9	4.5	4.3

주: 임금근로자 중 비중.

## 2. 인구분포 차이

노동패널 이용자들이 하는 가장 쉬운 의심이 패널 노후화에 따른 고령

화 발생이다. 가중치 부여 전 결과를 보면 이와 같은 의심이 합리적임을 알 수 있다. 하지만, 가중치 부여(통합표본 횡단면 가중치) 후에는 차이가 거의 없다.

그러나 학력별 분포를 보면, 가중치 부여 후에도 저학력자는 노동패널에 다소 적고, 고학력자는 다소 많이 분포하는 것을 알 수 있다.

경황과 인구학적 분포를 정확히 맞추면 비정규직 비중과 추이가 비슷해지는지 보기 위해 경황 8월 부가조사에서 나타나는 성×연령(15~24세/이후 5세 구간/65세 이상)×교육(중졸 이하/고졸/전문대졸/대졸 이상) 분

〈표 3-3〉 연령별 분포 - 2014년 8월 경황과 2014년 한국노동패널조사(17차년도)  
(단위: %)

	경황	노동패널	
		가중치 부여 후	가중치 부여 전
15~24세	14.6	15.3	8.8
25~29세	7.7	8.0	5.0
30~34세	9.2	9.9	8.4
35~39세	8.9	8.1	10.2
40~44세	10.1	9.1	10.6
45~49세	9.8	10.6	9.0
50~54세	10.0	10.0	9.3
55~59세	8.6	8.4	9.2
60~64세	6.0	6.1	7.5
65세 이상	15.1	14.6	22.1

〈표 3-4〉 학력별 분포 - 2014년 8월 경황과 2014년 한국노동패널조사(17차년도)  
(단위: %)

	경황	노동패널
중졸 이하	27.2	24.8
고졸	38.3	34.5
전문대졸	10.6	13.5
대졸 이상	23.9	27.3

포를 정확히 일치시켜 결과를 확인해 보았다.<sup>3)</sup> 여기서는 규모와 추세가 가장 달랐던 한시근로에 대한 결과만 제시하였는데, 인구학적 분포 차이의 교정은 경찰과의 차이를 전혀 좁히지 못해 인구학적 분포 차이로는 경찰과의 차이를 설명할 수 없음을 확인하였다.

다음으로 한시근로 비중이 가장 다른 연령대는 어떤 연령대인지 확인해 보았더니, 고연령대에서 특히 크게 다른 것으로 나타났다. 수치가 다소 튀는 것으로 보이는 2014년은 제외하고 가장 최근 수치인 2013년을 비교해 보면 30대 이하에서는 거의 비슷한 수치를 보이며, 40대부터 차이가 벌어지기 시작해 고연령대는 10%포인트 이상 차이가 나타났다.

한시근로는 기간제와 비기간제 한시근로로 나뉘는데, 기간제 비중 차이보다는 비기간제 한시근로에서 차이가 훨씬 크게 나타나는 것으로 유추된다.

이상의 분석은 젊은 층보다는 고연령층에서 비정규직 비중 응답이 높게 나타나는 것이 경찰과 노동패널의 상이한 규모를 야기한 주요 원인임을 보여준다. 이러한 경향은 특히 비기간제 한시근로에서 두드러지게 나타났다.

〈표 3-5〉 임금근로자 중 한시근로 비중 추이

(단위: %)

	노동패널	경찰	노동패널-교정
2009	20.6	21.3	19.5
2010	22.3	19.2	21.5
2011	23.5	19.7	23.7
2012	26.5	19.2	25.7
2013	24.6	18.8	25.4
2014	28.4	18.7	28.6

3) 분포 일치를 위해 다음 논문의 방법을 사용하였다: DiNardo, John & Fortin, Nicole M & Lemieux, Thomas(1996), "Labor Market Institutions and the Distribution of Wages, 1973~1992: A Semiparametric Approach," *Econometrica, Econometric Society*, Vol. 64(5), pp.1001-1044.

〈표 3-6〉 연령별 임금근로자 중 한시근로 비중

(단위: %)

	한국노동패널조사						경제활동인구조사					
	2009	2010	2011	2012	2013	2014	2009	2010	2011	2012	2013	2014
15~24세	24.3	29.1	27.2	28.8	31.8	33.4	29.4	28.2	25.6	24.4	25.9	23.6
25~29세	15.9	17.0	19.7	20.7	18.0	28.6	20.8	18.9	19.7	17.9	17.5	18.7
30~34세	14.1	13.5	15.7	13.4	13.0	16.5	15.4	14.5	15.6	14.7	13.9	13.5
35~39세	16.7	17.6	15.6	17.4	14.1	16.4	15.3	13.8	15.1	15.9	14.1	14.0
40~44세	21.3	22.2	23.0	26.1	21.1	22.2	17.0	16.0	15.2	15.2	14.8	14.6
45~49세	21.2	23.6	24.0	26.5	24.6	27.2	16.9	14.3	16.2	15.1	14.2	13.7
50~54세	21.5	22.2	26.2	29.9	29.2	29.8	19.2	17.2	17.6	16.9	15.6	15.2
55~59세	29.1	34.0	32.1	39.1	34.1	37.8	27.7	24.0	23.0	22.6	24.1	21.0
60~64세	38.1	38.8	44.9	53.1	54.0	55.7	45.0	34.8	36.5	34.7	30.7	31.9
65+	46.8	56.2	56.0	69.2	64.3	70.3	54.1	51.4	48.0	48.7	50.7	51.8
전 체	20.6	22.3	23.5	26.5	24.6	28.4	21.3	19.2	19.7	19.2	18.8	18.7

〈표 3-7〉 연령별 임금근로자 중 기간제 비중

(단위: %)

	한국노동패널조사						경제활동인구조사					
	2009	2010	2011	2012	2013	2014	2009	2010	2011	2012	2013	2014
15~24세	11.7	14.1	10.9	13.6	12.9	15.4	25.2	23.6	21.4	20.2	21.6	19.0
25~29세	8.0	7.2	10.6	11.4	11.0	16.3	17.4	14.7	15.6	14.5	13.9	14.9
30~34세	7.3	5.6	7.0	4.4	5.7	7.5	12.2	10.7	11.3	11.5	11.4	10.4
35~39세	8.2	8.2	5.6	6.1	5.2	6.5	12.2	9.8	11.4	12.2	11.0	10.1
40~44세	7.0	10.0	9.9	9.6	9.1	9.5	12.5	10.6	10.9	11.2	11.4	10.3
45~49세	8.5	9.8	8.5	8.5	7.7	10.9	11.9	10.0	11.8	11.5	10.5	9.7
50~54세	6.8	6.4	5.4	8.6	9.2	10.1	14.5	12.5	12.1	13.0	11.7	11.0
55~59세	11.9	11.5	11.6	12.0	12.8	13.3	22.7	19.4	18.4	18.8	18.9	16.9
60~64세	19.1	16.4	17.0	23.5	26.1	27.4	38.8	29.2	30.8	28.6	26.5	27.7
65+	28.4	32.3	32.7	37.3	37.9	43.3	47.4	43.2	42.8	42.3	45.3	46.3
전체	9.1	9.4	9.4	10.2	10.4	12.7	17.1	14.6	15.2	15.3	15.1	14.6

### 3. 설문구조, 조사방식의 차이

임금수준별 비정규직 분포를 보면, 2014년을 제외하면 초저임금층에서 다소간의 차이가 있는 것으로 나타난다. 여기서 초저임금층은 중위 시간당 임금의 50% 미만, 저임금층은 중위 시간당 임금의 2/3 미만, 50% 이상, 중하임금층은 중위임금 미만, 중위임금 2/3 이상, 중상임금층은 중위임금 이상, 중위임금 3/2 미만, 고임금층은 중위임금 3/2 이상을 의미한다.

그러나 경황과 차이가 크게 나타난 한시근로는 이 분석에서도 차이가 크다. 특히 노동패널에서 조사된 저임금층은 한시근로 비중이 경황보다 크게 높게 나타났다. 2009년은 노동패널과 경황의 한시근로 비중이 비슷하게 나타나지만 그 뒤로는 노동패널 조사에서 초저임금층 임금근로자의 40.5%가 한시근로이며, 경황은 29.1%만이 한시근로였다.

앞서 분석한 고연령층뿐 아니라 저임금 부문에는 큰 규모의 비공식 고용(informal employment)이 있어 조사를 어떻게 하느냐에 따라 비기간제 한시근로 규모가 크게 변동할 수 있을 것으로 추측된다. 비공식 고용은 특성상 계약관계가 명확치 않아 주관적 판단에 의존하는 문항에서 경황과 동일한 방식으로 조사관행이 수립되어 있지 않다면 차이가 크게 나타날 수 있기 때문이다.

이런 차원에서 특히 문제가 되는 비기간제 한시근로로 분류된 사람들

〈표 3-8〉 임금수준별 비정규직 분포

(단위: %)

	노동패널						경황					
	초저 임금	저 임금	중하 임금	중상 임금	고 임금	전체	초저 임금	저 임금	중하 임금	중상 임금	고 임금	전체
2009	53.3	47.4	36.4	23.5	14.8	31.4	60.6	53.9	41.6	28.5	12.9	34.9
2010	59.6	49.9	35.8	24.4	15.2	32.0	59.5	48.4	40.1	27.5	13.2	33.3
2011	64.8	49.5	36.5	24.0	13.6	32.5	62.5	51.0	39.9	29.5	15.2	34.2
2012	72.5	52.6	40.4	24.7	11.9	33.8	63.2	51.8	38.2	27.6	14.6	33.3
2013	68.8	48.4	35.0	23.5	12.2	32.0	64.0	49.9	36.6	25.6	13.1	32.6
2014	74.7	58.1	40.1	24.5	16.4	35.7	63.1	51.4	37.6	24.2	12.7	32.4

40 노동패널자료 연구(1)

의 사유를 보면 <표 3-10>과 같은 결과가 나온다. 이 표에서 1~7은 다음과 같은 사유로 비기간제 한시근로로 응답한 사람들을 의미한다; 1=이미 정해진 고용계약기간이 만료되기 때문에, 2=묵시적, 관행적으로 계약이 종료될 것이기 때문에, 3=사업주가 그만두려면 언제든지 그만둔다는 조건(임시직 등)으로 채용되었으므로, 4=현재 하는 업무가 끝나기 때문에, 5=현재의 일자리에서 전에 일하던 사람이 복귀하기 때문에, 6=특정 계절

<표 3-9> 임금수준별 한시근로 분포

(단위: %)

	노동패널						경찰					
	초저 임금	저 임금	중하 임금	중상 임금	고 임금	전체	초저 임금	저 임금	중하 임금	중상 임금	고 임금	전체
2009	29.4	32.8	25.1	15.7	10.4	20.6	30.4	36.5	25.6	17.3	8.7	21.3
2010	40.5	34.7	24.5	16.7	11.9	22.4	29.1	29.8	22.5	16.4	9.1	19.2
2011	45.7	35.8	26.8	18.0	9.3	23.5	28.1	28.9	24.0	18.1	9.9	19.7
2012	57.6	43.1	31.2	18.5	9.3	26.5	30.2	27.9	23.2	16.6	10.0	19.2
2013	54.6	36.9	28.2	16.2	9.3	24.6	29.6	30.1	21.3	15.7	9.1	18.8
2014	56.4	48.1	31.9	19.2	13.7	28.5	29.9	30.3	22.1	14.7	8.6	18.7

<표 3-10> 사유별 비기간제 한시근로자의 임금근로자 중 비중

(단위: %)

		0	1	2	3	4	5	6	7	전체
노동 패널	2009	88.5	0.6	0.5	4.0	0.9	0.0	0.2	5.3	100.0
	2010	87.1	0.9	0.8	6.7	1.2	0.1	0.4	2.9	100.0
	2011	85.9	0.5	0.9	8.1	1.1	0.0	0.4	3.2	100.0
	2012	83.7	0.8	1.1	11.0	1.2	0.0	0.3	2.0	100.0
	2013	85.8	0.7	1.3	8.6	1.2	0.0	0.4	2.1	100.0
	2014	84.3	0.4	0.7	8.6	1.2	0.0	0.4	4.5	100.0
경찰	2009	95.8		0.2	2.5	0.3	0.1	0.1	1.0	100.0
	2010	95.4		0.2	2.3	0.2	0.0	0.1	1.8	100.0
	2011	95.6		0.2	1.9	0.3	0.0	0.1	1.9	100.0
	2012	96.1		0.1	1.9	0.2	0.0	0.0	1.6	100.0
	2013	96.3		0.2	1.8	0.2	0.0	0.1	1.4	100.0
	2014	96.0		0.2	1.7	0.2	0.0	0.1	1.9	100.0

주: 0='비기간제 한시근로'가 아닌 임금근로자.

동안만 일할 수 있기 때문에, 7=계약의 반복, 갱신으로 고용이 지속되고 있으므로.

1이나 2, 4, 5, 6은 비교적 응답하기가 쉬운 반면, 3(사업주가 그만두려면 언제든지 그만둔다는 조건으로 채용되었으므로)과 7(계약의 반복, 갱신으로 고용이 지속되고 있으므로)은 계약관계가 명확하지 않을 때 응답하기 까다로운 특성을 가지고 있다. <표 3-10>의 분석결과를 봐도 3과 7의 응답 비중 차이가 노동패널과 경찰 수치 차이의 가장 중요한 원인으로 나타나 이와 같은 추측을 뒷받침한다.

다음으로, 비기간제 한시근로 규모가 왜 크게 차이 나는지 확인하기 위해 비기간제 한시근로 설문구조를 검토해 보았다. 비기간제 한시근로로 응답되려면, ‘회사가 아주 어려워져서 폐업 또는 고용조정을 하거나 귀하가 특별히 잘못을 하지 않는다면, 귀하가 원하는 한 계속 그 직장에 다닐 수 있습니까?’라는 설문을 반드시 거쳐가야 한다.

그런데, 경찰 설문에서는 일일근로로 일한다고 응답한 사람들은 바로 이 설문문항을 뛰어넘게 되어 비기간제 한시근로로 분류될 가능성이 없어진다. 그러나, 노동패널에서는 일일근로 판별 설문은 유형설문지에, 비기간제 한시근로 관련 설문은 취업자 개인 설문지에 있어 일일근로 답변 여부와 관계없이 비기간제 한시근로 설문에 응답하게 되고, 이로 인해 일일근로자의 일부가 비기간제 한시근로자로도 교차 분류될 수 있게 되는 차이가 발생한다.

경찰과 동일하게 일일근로면 비기간제 한시근로로 분류되지 않도록 처리하면 노동패널의 한시근로자 규모는 2009년에는 약 2.6%포인트, 2014년에는 약 4%포인트 감소하는 것으로 나타난다.

<표 3-11> 한시근로의 규모 변화 - 일일근로 통제 여부별 규모 추이 변화(임금근로자 중 비중)

(단위: %)

		2009	2010	2011	2012	2013	2014
한시 근로	경찰	21.3	19.2	19.7	19.2	18.8	18.7
	노동패널	20.6	22.3	23.5	26.5	24.6	28.4
	일일근로 제외시	18.0	19.0	20.2	23.1	20.5	24.4

### 제3절 정규직-비정규직 변화에 나타나는 비표본오차의 교정

노동패널조사는 한 사람을 지속적으로 추적하는 패널조사의 특성상 정규직-비정규직을 오가는 사람들을 포착할 수 있다. 그런데, 분석을 해보면, 이들 중에는 현실에서 있음직하거나 일관된 변화를 보여주는 경우도 있지만, 한 직장에 있음에도 불구하고 특별한 이유 없이 정규직-비정규직을 오가는 케이스가 섞여 있다. <표 3-12>에서 가~타까지는 한 직장에서 있을 수 있는 변화의 패턴이다. 이를테면, 정규직으로 있다가 은퇴

<표 3-12> 노동패널에서 발견되는 정규직(=0), 비정규직(=1) 변화의 패턴 - 한 직장 내의 경우

	2009	2010	2011	2012	2013	2014
가	0	1	1	1	1	1
나	0	0	1	1	1	1
다	0	0	0	1	1	1
라	0	0	0	0	1	1
마	0	0	0	0	0	1
바	1	0	0	0	0	0
사	1	1	0	0	0	0
아	1	1	1	0	0	0
자	1	1	1	1	0	0
차	1	1	1	1	1	0
카	0	0	0	0	0	0
타	1	1	1	1	1	1
A	0	1	0	1	1	1
B	1	0	1	1	0	0
C	0	1	1	1	1	0
D	0	1	0	1	0	1

후 재고용 같은 이유로 비정규직이 될 수 있을 것이다. 그런데, A~D의 경우 한 직장에서 정규직-비정규직을 왔다갔다 한다는 점에서 현실에서 있음직하지 않은 변화의 패턴이다.

이 장에서는 이러한 사람들의 고용형태를 교정했을 때 비정규직의 규모와 비정규직에서 정규직으로의 전환이 어떻게 변화되는지 분석해 보겠다. 분석을 위해 주된 일자리만 남긴 상태에서 동일 기업 내 비정규 ↔ 정규를 왔다갔다 하는 경우만을 교정하겠다. 다만, 교정을 할 때 어떤 것이 맞는지 특정할 수 없으므로, 관찰기간 6년간 3회를 초과해 비정규직이었으면 비정규직으로 분류하겠다. 즉 A, C, D는 모두 1로, B는 0으로 재입력하였다.

교정한 결과를 보면, 비정규직은 대략 1%포인트 정도 감소하는 것으로

〈표 3-13〉 비정상적으로 판단되는 정규 ↔ 비정규 이동 교정 시 고용형태별 임금근로자 중 비중 추이

(단위: %)

		2009	2010	2011	2012	2013	2014
비정규직 비중	교정	30.5	30.8	31.2	32.8	31.4	34.8
	원계열	31.4	32.1	32.6	33.8	32.1	35.7
	경활	34.9	33.4	34.2	33.3	32.6	32.4
한시근로 비중	교정	20.1	21.4	22.3	25.7	24.0	27.5
	일일근로 제외+비표본오차 교정	17.5	18.1	19.0	22.3	20.0	23.5
	원계열	20.6	22.3	23.5	26.5	24.6	28.4
	경활	21.3	19.2	19.7	19.2	18.8	18.7
기간제	교정	8.7	8.9	9.1	9.9	10.1	12.2
	원계열	9.1	9.5	9.4	10.2	10.4	12.7
	경활	17.1	14.6	15.2	15.3	15.1	14.6
비전형근로 비중	교정	12.7	13.0	12.6	11.6	12.4	12.7
	원계열	13.1	13.2	12.7	11.7	12.5	12.9
	경활	13.9	13.4	13.9	12.9	12.1	11.3
시간제 비중	교정	7.4	7.6	7.9	7.6	8.0	8.8
	원계열	7.4	7.6	7.9	7.6	8.0	8.8
	경활	8.7	9.5	9.7	10.3	10.3	10.8

나타났다. 시간제나 비전형근로는 거의 차이가 없으며, 비정규직 1%포인트 감소는 거의 전적으로 한시근로 감소 때문인 것으로 나타났다.

다만, 이와 같은 교정은 맞는 고용형태가 무엇인지 재조사하여 교정한 것이 아니기 때문에, 실제 조사과정에서 교정이 이루어진다면 이보다 더 크게 비정규직 규모가 영향받을 가능성이 크다.

이상의 분석은 조사과정에서 고용형태 설문에 전년도 값과 대조해 재질문하는 루틴이 포함될 필요가 있음을 보여준다. 노트북을 이용해 조사하기 때문에 전년도 값을 내장해 대조하는 것이 기술적으로 가능할 것이다.

## 제4절 결론 및 정책 함의

비정규직 전체 규모가 2014년 다소 크게 튼 것을 제외하면 경황과 비교해 문제 있어 보이진 않는다. 그러나, 세부로 구분해 보면 한시근로 비중과 추이가 경황과 반대라는 차이, 과거엔 기간제가 지나치게 적었고, 기간제 추이가 경황과 반대이며, 비기간제 한시근로가 지나치게 많이 잡히는 차이가 나타난다.

연령, 성, 교육수준 등 인구분포의 경황과의 차이는 이와 같은 다른 한시근로 규모와 추세를 설명하는 데 전혀 도움이 되지 않는 것으로 나타났다. 가중치를 부여했을 때 인구학적 비중(연령, 성, 교육)은 비교적 잘 조정되어 있는 것으로 판단되는데, 교육분포는 경황과 약간 차이가 있었다. 아마도 인구학적 분포의 차이가 가중치를 부여하면 적기 때문에 이와 같은 결과가 나왔을 것이다.

기간제 추이가 반대이긴 하지만, 그로 인해 역설적으로 최근에는 노동패널과 경황 간 기간제 규모 차이가 2%포인트 내외로 좁혀졌다(임금근로자 중 비중 기준). 과거 수치로 보면 기간제 규모 차이는 주로 1년 미만 계약직 규모 차이에서 유래한다. 이 장에서는 이 차이의 원인에 대해서는 더 다루지 못했다.

한편, 고연령대에서 한시근로, 특히 비기간제 한시근로 비중이 매우 높

은 것으로 나타나는 한편, 저소득층에서 비기간제 한시근로 비중이 월등히 높은 것으로 분석되었다. 고연령대나 저소득층의 일자리는 비공식 고용 특성을 갖는 경우가 많은데, 이로 인해 주관적 판단이 많이 개입되는 비기간제 한시근로 판별설문의 특성상, 경황과 노동패널 간 조사관행 차이가 한시근로의 규모와 추세 차이를 낳은 요인일 수 있다. 또한, 설문구조상 일일근로가 한시근로에 응답하게 되어 있는 노동패널 설문구조가 한시근로 규모를 증가시키는 하나의 원인으로 작용했다.

동일인을 추적 조사하는 노동패널 조사구조를 이용해 동일 기업 내 정규-비정규 전환을 분석해 본 결과 특별한 이유 없이 두 상태를 왔다갔다 하는 사람들이 포착되었다. 이들의 상태변화를 교정해주면 비정규직 규모가 감소하였다. 비정규직 설문이 하나의 설문이 아니고 여러 설문을 조합해 정의하는 특성상 비표본오차가 개입될 소지가 크다. 이상의 분석결과는 우선, 조사 현장에서 과거 정보를 연동시켜 응답이 바뀐 사람들을 대상으로 정말 바뀐 것인지를 확인하는 루틴을 조사 시스템에 탑재하는 것의 필요성을 보여준다. 또한, 경황의 조사관행을 파악하여 노동패널과 무엇이 달라 한시근로 응답이 달라지게 되는지 분석하고, 필요하다면 노동패널 조사 관행에 반영할 필요성을 시사한다.

이상의 분석은 사내하청, 비공식 고용(informal employment) 같은 경황상의 비정규직 못지 않게 사회적으로 문제가 되는 일자리에 대한 분석은 제외된 것이므로 한계가 있다. 또한, 노동패널과 경황의 차이가 꼭 문제인 것이냐도 의문의 여지가 있을 수 있다. 문제가 있으니 일치시켜야 한다는 판단 이전에 조사관행의 차이가 얼마나 원인일지를 파악하는 후속연구가 더 바람직한 접근법일 것 같다. 우리나라에는 계약형태가 불명확한 비공식 고용이 여전히 넓게 포진해 있어 조사방식에 따라 다른 결과가 도출될 소지가 크고, 실제 분석결과로도 주관적 판단에 의존해야 하는 판별분항에서 주로 문제가 발생했기 때문이다. 이와 같은 후속연구는 다양한 질문방식에 어떻게 응답자가 반응하는지에 대한 연구를 통해 좀 더 노동시장의 실태를 깊이 파악하는 데 큰 도움이 될 수 있다는 점에서 생산적일 것으로 기대한다.

## 제 4 장

## 노동패널 고용지표의 신뢰성 검토

## 제1절 머리말

KLIPS는 여러 차례의 패널 품질개선 연구를 통해 CAPI 실험연구, 표본이탈, 임금 및 소득의 히핑, 기간 변수의 신뢰성 등 다양한 패널방법론상의 주제를 다루어왔다(최효미 외, 2012; 윤윤규 외, 2012; 홍민기 외, 2014a; 2014b 등). 그러나 지금까지 주요 고용지표와 관련된 신뢰성 검토는 이루어진 바 없다.

가장 큰 이유는 패널조사의 목적 및 특징이 노동력 상태와 관련된 대표통계를 산출하는 데 있지 않기 때문이다. 예컨대 통계청의 경제활동인구조사는 고용률이나 실업률 같은 지표 산출을 목적으로 설계되었기 때문에 3만 가구 이상 대규모의 순환표본 추출과 1주일 내 단기간 조사 등의 조사체계를 갖는다. 반면 KLIPS와 같은 다목적 패널조사의 경우 조사내용이 경찰조사의 수십 배 이상으로 방대할 뿐만 아니라 조사기간 역시 6개월 이상 소요된다. 그러나 이러한 조사체계상의 차이가 고용지표의 신뢰성을 검증하지 않아도 된다는 근거가 될 수는 없다. KLIPS는 고용상태의 미시적-동태적 변화 파악에 주요한 조사 목적이 있기 때문에 고용지표의 신뢰성이 무엇보다 중요하기 때문이다. 국가 대표통계와 비교했을 때 KLIPS의 고용률과 실업률 차이는 어느 정도인지, 생산가능인구, 취업

자, 실업자수 등은 모집단 특성을 적절하게 반영하는지, 차이의 발생 원인은 무엇인지, 만일 고용지표의 신뢰성 개선이 필요하다면 이는 통계적 처리로 가능한지, 아니면 조사체계상의 조치가 필요한지 등에 대한 검토와 대응이 필요하다.

이러한 배경하에 이 장에서는 KLIPS 고용지표(주로 고용률 및 실업률)와 경황의 비교를 통해 자료의 신뢰성을 검증하고 어떤 점에서 유사하고 어떤 점이 다른지를 파악하고자 한다. 고용지표의 차이를 가져오는 노동력 상태/고용형태의 구성이 어떻게 다르며, 그 원인이 무엇인지를 파악함으로써 자료의 신뢰성을 제고하기 위한 정책적 함의를 도출하고자 한다.

분석의 시간적 범위는 주로 2009년 12차 조사부터 2014년 17차 조사시기에 초점을 맞춘다. 1차 조사인 1998년 표본은 제주도를 제외한 도시가구를 표집했기 때문에 모집단이 상이하여 경제활동인구조사와 지표 비교가 어렵다. 또한 패널조사의 특성상 10년 이상의 시간이 경과하면서 농촌과 도시 간의 이동이 상당한 정도로 발생하여 도시 표본에 대한 대표성 뿐만 아니라 전국 대표성 측면에서도 모호한 상황이다. 2009년의 경우 별도의 표본추가를 실시하여 전국 대표성을 확보하도록 표본과 가중치를 재설정하였다. 특히 전국 대표성 입장에서 볼 때, 동부 표본에서 단독주택과 1~2인 가구가 과소 표집되었고, 읍면부에서는 기존 표본의 비중이 적은 문제가 발생하여 이를 보완하였다(김재광, 2009). 따라서 조사체계상의 차이에도 불구하고 2009년(12차 조사) 자료부터 고용률이나 실업률 등 고용지표의 신뢰성을 비교·검토할 수 있는 토대가 마련되었다고 할 수 있다.

본문의 구성은 다음과 같다. 2절에서는 KLIPS와 경제활동인구조사 간에 고용률과 실업률의 격차 발생 여부 및 정도를 비교·분석한다. 3절에서는 두 조사 간에 고용지표 격차가 왜 발생하는지를 파악한다. 이를 위해 패널조사의 표집 및 비표집 오차를 발생시키는 다양한 원인들을 사전적으로 가설화한 후 각 가설에 어느 정도 부합하는지 여부를 검증한다. 4절에서는 고용지표의 동태적 변화에 대한 분석을 통해 KLIPS의 미시-동태적 분석으로의 활용 가능성을 확인한다. 마지막으로 5절에서는 이상의 분석결과를 종합하여 정책적 함의를 살펴본다.

## 제2절 고용률 및 실업률 추이

2009년 통합표본 기준 KLIPS의 고용률은 54.1%로 경제활동인구조사의 연평균 고용률인 58.6%보다 4.5%포인트 낮은 것으로 파악되었다. 2010년부터 두 조사 간 상대적 고용률 격차가 소폭 완화되기는 했지만 전체적으로 약 3~4%포인트의 격차가 유지되고 있는 추세이다. KLIPS의 고용률 수준 및 추이는 가중치 부여 방식에 따라서도 편차를 보였다.

전체적으로 1998년 표본기준 가중치가 2009년 통합 표본기준 가중치보다 고용률이 낮았다. 또한 횡단면 가중치보다는 종단면 가중치가 더 낮은 고용률을 보인 가운데 양자 간의 격차도 시간이 흐를수록 확대되는 모습을 보였다. 2014년을 기준으로 경찰조사 고용률과의 상대적 격차를 살펴보면 1998년 종단 가중치(-5.7%포인트) < 1998년 횡단 가중치(-4.0%포인트) < 2009년 종단 가중치(-3.8%포인트) < 2009년 횡단 가중치(-2.7%포인트) 순으로 나타났다.

표에는 제시되어 있지 않지만, 1998년 표본의 경우 표본추가 이전(1~11차 조사)까지의 경우를 살펴보면 고용률 격차가 더 큰 폭으로 벌어진다. 1998년 표본은 제주도를 제외한 도시가구 표본이기 때문에 상대적으로 고용률이 낮을 수밖에 없다. 농어촌 가구의 경우 가족농의 형태를 띠기 때문에 고용률이 상대적으로 높기 때문이다.

KLIPS의 실업률은 2009년 표본 최초 표집시점 당시 4.5%로 경찰의 3.6%보다 0.9%포인트 높았으나, 2010년부터는 오히려 경찰보다 낮은 수준을 보였다. 2012년의 경우에는 경찰조사와 KLIPS 간에 실업률 격차가 거의 발생하지 않기도 했다. 그러나 전체적으로는 KLIPS의 표본수가 상대적으로 적기 때문에 실업률 변동폭이 상대적으로 불안정한 추이를 보였다. 또한 경기변화에 대한 응답자의 반응이 경찰조사보다 큰 폭으로 나타나는 특징을 보였다.

고용률 및 실업률 격차가 어디에서 비롯되었는지를 파악하기 위해 2014년(17차) 조사 자료(2009년 횡단 가중치 적용)를 이용하여 성별 및

〈표 4-1〉 전체 고용률 및 실업률

(단위: %)

	경제활동 인구조사 (A)	노동패널(B)				(B)-(A)			
		1998 횡단	1998 종단	2009 횡단	2009 종단	1998 횡단	1998 종단	2009 횡단	2009 종단
고용률									
2009(12차)	58.6	53.1	52.7	54.1	54.1	-5.5	-5.9	-4.5	-4.5
2010(13차)	58.7	53.8	53.5	55.8	55.3	-4.9	-5.2	-2.9	-3.4
2011(14차)	59.1	54.8	54.0	56.8	56.0	-4.3	-5.1	-2.3	-3.1
2012(15차)	59.4	54.5	53.4	56.9	56.1	-4.9	-6.0	-2.5	-3.3
2013(16차)	59.5	55.0	54.0	56.9	56.0	-4.5	-5.5	-2.6	-3.5
2014(17차)	60.2	56.2	54.5	57.5	56.4	-4.0	-5.7	-2.7	-3.8
실업률(지난 1주 기준)									
2009(12차)	3.4	3.8	4.1	4.1	4.1	0.4	0.7	0.7	0.7
2010(13차)	3.4	2.4	2.5	2.4	2.4	-1.0	-0.9	-1.0	-1.0
2011(14차)	3.0	2.2	2.4	2.5	2.6	-0.8	-0.6	-0.5	-0.4
2012(15차)	2.8	2.5	2.6	2.7	2.8	-0.3	-0.2	-0.1	0.0
2013(16차)	2.8	2.4	2.8	2.2	2.4	-0.4	0.0	-0.6	-0.4
2014(17차)	3.2	2.6	2.8	2.3	2.3	-0.6	-0.4	-0.9	-0.9
실업률(지난 4주 기준)									
2009(12차)	3.6	4.3	4.6	4.5	4.5	0.7	1.0	0.9	0.9
2010(13차)	3.7	2.8	3.1	2.7	2.7	-0.9	-0.6	-1.0	-1.0
2011(14차)	3.4	2.7	2.9	3.0	3.1	-0.7	-0.5	-0.4	-0.3
2012(15차)	3.2	3.1	3.2	3.2	3.3	-0.1	0.0	0.0	0.1
2013(16차)	3.1	2.8	3.2	2.6	2.8	-0.3	0.1	-0.5	-0.3
2014(17차)	3.5	3.0	3.1	2.7	2.7	-0.5	-0.4	-0.8	-0.8

연령별 고용률 및 실업률을 같은 연도 통계청 경제활동인구조사 결과와 비교하였다(표 4-2 참조).

분석 결과 두 조사 간의 고용률 및 실업률 격차 발생의 주된 원인은 남녀를 불문하고 청년층의 고용률 및 실업률 격차에 기인한다는 것을 확인할 수 있었다. KLIPS의 20~29세 고용률은 경제활동인구조사에 비해

〈표 4-2〉 2014년(17차) 성별·연령별 실업률 및 고용률 비교

(단위: %)

	고용률			실업률		
	계	남성	여성	계	남성	여성
경제활동인구 조사 (A)						
계	60.2	71.4	49.5	3.5	3.6	3.5
15~19세	7.7	6.8	8.6	9.2	10.2	8.4
20~29세	57.4	55.8	59.0	9.0	10.5	7.6
30~39세	73.9	90.9	56.3	3.1	2.9	3.5
40~49세	79.1	92.7	65.1	2.2	2.1	2.4
50~59세	74.2	87.5	60.9	2.2	2.3	2.2
60세 이상	39.0	51.4	29.3	2.3	2.8	1.6
노동패널-09횡단 가중치 (B)						
계	57.5	69.7	45.4	2.7	2.5	2.8
15~19세	2.8	2.3	3.4	1.3	0.0	2.2
20~29세	46.9	46.7	47.1	8.5	8.6	8.4
30~39세	72.7	92.9	51.4	2.8	2.6	3.3
40~49세	78.0	93.4	62.2	1.7	1.7	1.8
50~59세	75.8	90.6	60.9	0.8	1.0	0.6
60세 이상	36.5	49.4	26.4	1.5	1.5	1.5
두 조사간 격차 (B-A)						
계	-2.7	-1.7	-4.1	-0.8	-1.1	-0.7
15~19세	-4.9	-4.5	-5.2	-7.9	-10.2	-6.2
20~29세	-10.5	-9.1	-11.9	-0.5	-1.9	0.8
30~39세	-1.2	2.0	-4.9	-0.3	-0.3	-0.2
40~49세	-1.1	0.7	-2.9	-0.5	-0.4	-0.6
50~59세	1.6	3.1	0.0	-1.4	-1.3	-1.6
60세 이상	-2.5	-2.0	-2.9	-0.8	-1.3	-0.1

10.5%포인트 낮았는데, 남성(-9.1%포인트)보다는 여성(-11.9%포인트)의 고용률 격차가 더욱 크게 나타났다. 15~19세의 고용률 역시 KLIPS가 4.9%포인트 낮았는데 역시 남성(-4.5%포인트)보다는 여성(-5.2%포인트)의 격차가 더욱 크게 나타났다.

KLIPS의 실업률은 15~19세 연령층에서 크게 낮았다. 경찰조사의 15~19세 실업률 9.2%와 비교할 때 KLIPS의 실업률은 1.3%로 7.9%포인트나 낮은 것으로 확인되었다.

이러한 결과들을 토대로 <표 4-3>과 같이 각 연도별 청년층(15~29세) 고용률 및 실업률 격차를 추가적으로 살펴보았다. 경제활동인구조사의

<표 4-3> 청년층(15~29세) 고용률 및 실업률

(단위: %)

	경제활동 인구조사 (A)	노동패널(B)				(B)-(A)			
		1998 횡단	1998 종단	2009 횡단	2009 종단	1998 횡단	1998 종단	2009 횡단	2009 종단
고용률									
2009(12차)	40.5	32.7	33.2	32.7	32.7	-7.8	-7.3	-7.8	-7.8
2010(13차)	40.3	31.7	32.1	34.0	33.2	-8.6	-8.2	-6.3	-7.1
2011(14차)	40.5	31.4	31.3	35.5	33.9	-9.1	-9.2	-5.0	-6.6
2012(15차)	40.4	30.7	30.8	33.3	31.9	-9.7	-9.6	-7.1	-8.5
2013(16차)	39.7	30.4	31.3	31.5	29.6	-9.3	-8.4	-8.2	-10.1
2014(17차)	40.7	32.0	31.9	32.6	30.1	-8.7	-8.8	-8.1	-10.6
실업률(지난 1주 기준)									
2009(12차)	7.6	9.7	9.4	10.6	10.6	2.1	1.8	3.0	3.0
2010(13차)	7.3	6.9	6.8	6.7	7.0	-0.4	-0.5	-0.6	-0.3
2011(14차)	6.9	5.8	5.7	6.0	6.2	-1.1	-1.2	-0.9	-0.7
2012(15차)	6.7	5.9	5.8	6.5	6.5	-0.8	-0.9	-0.2	-0.2
2013(16차)	7.2	7.7	9.1	6.8	8.1	0.5	1.9	-0.4	0.9
2014(17차)	8.0	9.2	8.4	6.8	7.1	1.2	0.4	-1.2	-0.9
실업률(지난 4주 기준)									
2009(12차)	7.9	10.2	9.9	11.1	11.1	2.3	2.0	3.2	3.2
2010(13차)	7.8	7.6	7.5	7.5	7.7	-0.2	-0.3	-0.3	-0.1
2011(14차)	7.4	7.9	8.2	7.6	7.9	0.5	0.8	0.2	0.5
2012(15차)	7.5	6.9	6.7	7.3	7.4	-0.6	-0.8	-0.2	-0.1
2013(16차)	7.9	8.9	10.2	8.1	9.6	1.0	2.3	0.2	1.7
2014(17차)	9.0	10.5	9.1	8.3	8.4	1.5	0.1	-0.7	-0.6

청년층 고용률은 2014년 40.7%에 이르는 데 비해 KLIPS의 청년층 고용률은 가중치 부여 방식에 따라 30.1%(2009년 종단 가중치)에서 32.6%(2009년 횡단 가중치) 사이에 분포하였다. 1998년 횡단 및 종단 가중치를 부여하더라도 이러한 격차는 완화되지 않았다. 반면 실업률의 경우 일정한 규칙성을 발견하기 어려웠다. 2009년, 2011년, 2013년의 경우 KLIPS의 실업률이 경찰보다 높았지만 다른 연도의 경우 KLIPS의 실업률이 경찰 조사보다 낮게 추정되기도 하였다. KLIPS의 실업률이 상대적으로 편차가 크게 나타나는 것은 표본수가 상대적으로 적기 때문에 청년층 대상의 실업률 추정에 오차가 커질 가능성이 증가하기 때문인 것으로 추정된다.

이상의 결과를 종합하면 KLIPS는 고용률로 대표되는 노동시장 참여 규모를 상당한 정도로 ‘과소 추정할 가능성’이 높다는 것을 의미한다. 물론 경제활동인구조사 역시 모집단이 아닌 표본조사에 기인하여 고용률을 추정하므로 오히려 경제활동인구조사의 고용률이 과대 추정된 것일 수 있다. 그럼에도 불구하고 경찰조사 자료는 고용률과 실업률 지표 산출을 목적으로 실시되는 국가 대표통계인 반면, KLIPS는 고용률 추정을 목적으로 설계된 조사는 아니므로 이하에서는 KLIPS의 고용률이 과소 추정되었다는 것을 전제로 논의를 전개해나갈 것이다.

### 제3절 청년층의 노동시장 참여가 낮은 원인

지금부터는 KLIPS에서 청년층의 노동시장 참여가 낮게 추정되는 원인을 파악한다. 앞서 언급하였듯이 실업률의 경우, 표본 규모가 매우 작기 때문에 여기서 의미 있는 원인 분석을 하기 어렵다.<sup>4)</sup> 따라서 KLIPS와 경찰조사 간 고용률 차이를 발생시키는 원인을 파악하는 데 집중하고자 한다.

패널조사 방법론 문헌에 따르면, 패널조사의 오차는 크게 표집 오차(sampling errors)와 비표집 오차(non-sampling errors)로 구분할 수 있다.

4) 예컨대 17차 조사의 경우 KLIPS 15~29세 청년층 실업자수는 52명에 불과하다.

표집 오차는 잘못된 표본 설계에 기인하여 발생하며, 비표집 오차는 표집 오차를 제외한 나머지 모든 오차를 포함한다. 비표집 오차는 크게 패널의 표본이탈(panel attrition), 설문지 설계와 관련된 도구 오차(instrument errors), 개념의 조작적 정의 및 측정방식에 따라 발생하는 측정 오차(measurement errors), 부정확한 기억에 의한 회고 오차(recall errors), 전화, 우편, 면접 등 다양한 조사방식의 선택에 따라 발생하는 모드 효과(mode effects), 응답자와 면접원의 설문방식 및 응답방식에 의해 발생하는 응답자 오차(respondent errors)와 면접원 오차(interviewer errors), 반복조사에 따른 조사 피로감과 부담스러운 문항의 반복적 조사에 따라 응답방식에 편이가 발생하는 패널 조건화(panel conditioning) 등이 있다(Lynn et al., 2009; 이상호, 2012 등 참조).

이러한 접근법을 이용하여 <표 4-4>와 같이 청년층 고용률의 과소 추정 잠재적 요인을 크게 네 가지-표집 오차와 가중치 설계, 패널 이탈, 조사체계상의 문제점, 패널 조건화-로 구분할 수 있다.

첫째, 표집 오차는 가구 및 개인 표본을 추출하는 과정에서 구조적 오류가 발생하여 청년층 인구 대표성에 문제점이 발생했을 가능성이 있다. 만일 표본 추출과정에서 문제가 없었다 하더라도 가중치 설계과정에서 모집단을 충실하게 복원하는 데 오류가 발생했을 가능성도 배제할 수 없다. 문제의 원인이 여기에 있다면 가능한 경우 가중치 조정을 통해 해결할 수 있

<표 4-4> 고용률 과소 추정의 잠재적 요인들

		문제점	해결방안
1. 표집 오차와 가중치 설계		가구표집과정의 오류로 인한 가구원 자체의 인구 대표성 문제	1차 표집시점 가구 가중치 재조정 필요
비 표 집 오 차	2. 패널 이탈	응답대상 가구원의 개인조사 응답률 격차	응답률 반영 가중치 재조정 및 표본이탈 관리 필요
	3. 조사체계	취업자 판별과 관련된 설문문항, 질문방식	잘못된 정의, 스킵 오류 등 기술적인 오류가 아니라 시스템적인 문제일 경우, 전략적 선택 및 조사관리를 통해 해결
		취업자 정의 및 지침상의 문제	
		조사시기, 조사기간 등	
4. 패널 조건화	응답자 및 조사원의 조사 태도상의 문제	조사 관리	

다. 다만 구조적으로 오류가 심각하여 가중치로 조정이 불가능한 수준이라면 조사 재설계가 필요하다.

둘째, 패널 이탈은 패널조사의 비표본 오차를 발생시키는 가장 중요한 요인이다. KLIPS는 가구조사이므로 패널 이탈 역시 가구 자체의 이탈과 가구원 조사 단계에서의 이탈로 구분할 수 있다. 청년층 부부 가구의 체계적인 표본이탈 혹은 가구조사에는 성공했지만 청년층의 취업으로 인해 조사 추적에 실패할 경우 청년층 취업자 조사 성공률이 떨어질 수 있다. 표본이탈 역시 가중치 부여과정에서 2차년도 이후 동태적 응답확률에 대한 보정을 통해 조정 가능하다. 다만 장기적으로 청년층의 이탈이 누적될 경우 역시나 조사의 지속가능성을 훼손하기 때문에 표본이탈을 완화/방지하기 위한 조사체계상의 대응책이 필요하다.

셋째, 조사체계상의 문제점들은 표본이탈을 제외한 대부분의 비표집 오차를 포함하는 영역이다. 취업자 판별과 관련된 문항의 설계 혹은 설문방식, 취업자의 정의와 관련된 지침이 경찰조사와 다를 가능성도 존재한다. 패널 조사의 특성상 청년층 취업률이 낮은 시기에 조사가 집중되거나 조사기간이 길어지면서 발생하는 문제일 수도 있다. 이러한 문제점들은 대부분 가중치를 통해 해결이 불가능하며 조사체계상의 대응책이 필요한 영역이다. 세부 원인에 대한 파악이 이루어진다면 설문지나 지침 변경, 조사시기 조정, 조사기간 단축 등의 맞춤형 대응이 이루어질 수 있다. 그러나 조사체계상의 문제점은 서로 연관된 영역이 복잡하게 얽혀 있을 가능성이 많기 때문에 그만큼 명확한 원인 파악도 어려우며 해법 마련이 힘들다.

마지막으로 패널 조건화는 응답자 및 면접원의 패널조사에서 나타나는 특수한 형태의 비표집 오차이다. KLIPS와 같이 방대한 문항의 패널조사는 응답자와 조사원이 이미 이를 인지하기 때문에 양적으로나 질적으로 응답 부담이 높은 설문에 대한 응답을 회피하려는 경향이 발생할 수 있다. 만일 청년층이 상대적으로 조사에 대한 충성도가 떨어지거나 패널조사에 대한 응답 피로도가 더욱 크다면 취업자 식별과 관련된 문항 응답에서 이러한 경향이 커질 수 있다. 만일 여기에 원인이 있다면 이에 대한 해법 역시 가장 어렵다. 조사 시스템의 개선이 특정 문항의 변경 등과 같이 분리 가능한 대응으로 이루어지는 것이 아니라, 조사 현장의 대변 주

체들인 응답자와 조사원의 태도를 구조적으로 변화시키는 방식으로 이루어져야 하기 때문이다.

## 1. 표집 오차와 가중치

KLIPS의 표본설계는 두 차례의 분기점을 구분하여 살펴보아야 한다. 1차 조사인 1998년 당시 표본 설계는 제주도를 제외한 도시지역 5,000가구를 표집하였다. 이때의 모집단은 1995년 인구 센서스가 기준이다. KLIPS는 조사 차수가 지남에 따라 표본 마모를 보완하고 전국 대표성을 확보하기 위해 2009년 표본추가를 결정하고 1,500가구를 추가적으로 표집하였다. 이때의 모집단은 2005년 인구 센서스 기준 전체 내국인 가구가 된다. 즉 2009년의 표본 재설계를 통해 모집단이 변경되었기 때문에 추가표본뿐만 아니라, 기존 표본 전체를 대상으로 전국 대표성을 확보하도록 원표본 기준 및 가중치 부여방식도 전면적인 재조정이 이루어졌다(표본 설계와 관련된 자세한 사항은 김재광, 2009 참조). 그리고 2009년에 통합적으로 재설계된 표본에 대해 ‘통합표본’이라는 명칭을 부여하였다. 통합표본의 2차년도, 즉 1998년 기준 13차 조사부터는 원표본/신규 표본 가구원의 진입 및 응답률 조정이 이루어진 가중치가 부여되었다. 또한 인구추계조사를 이용하여 15세 이상 개인에 대한 연령집단별(10세별) 스케일 조정을 실시하였다(가중치 부여과정은 김영원 외, 2013 참조).

그렇다면 이러한 표본설계 및 가중치 부여방식은 청년층 고용률 추정과 어떤 관계를 갖는가?

고용률은 15세 이상 생산가능인구와 취업자수의 상대 비율에 의해 결정된다. 그런데 <표 4-5>의 좌측 두 번째 열에 제시된 것과 같이 KLIPS는 2009년 통합표본 이후부터 가중치에 대한 스케일 조정을 통해 청년층(15~29세) 생산가능인구에 대한 추정치가 인구추계조사와 동일한 값을 갖게 되었다. 오히려 경제활동인구조사의 경우 인구추계상의 생산가능인구보다 다소 낮은 것으로 파악되었다.<sup>5)</sup> 즉 생산가능인구의 규모는 가중

5) 경제활동인구조사에서는 섬, 기숙시설, 특수사회시설, 관광호텔 및 외국인거주지역 조사구, 집단시설 등이 제외된다.

〈표 4-5〉 경찰 수준의 청년층 고용률 달성을 위해 KLIPS에서 추가적으로 조사되어야 할 취업자수

(단위: 천 명, 명)

	생산가능인구수 (천 명)			취업자수 (천 명)			KLIPS가 경찰 수준 고용률을 달성하는 데 필요한 취업자수 (E)	차이 (C-E)	청년층 1인당 평균 가중치 (천 명)	추가 파악되어야 할 취업자수 (명)	KLIPS 청년층수 (명)
	KLIPS (A)	경찰조사 (B)	차이 (A-B)	KLIPS (C)	경찰조사 (D)	차이 (C-D)					
12차	10,560	9,780	780	3,452	3,957	-505	4,277	-825	3.5	233	2,977
13차	10,028	9,706	322	3,406	3,914	-508	4,041	-635	3.8	166	2,716
14차	10,382	9,589	793	3,680	3,879	-199	4,205	-525	4.2	125	2,546
15차	10,240	9,517	723	3,413	3,843	-430	4,137	-724	4.4	166	2,448
16차	10,119	9,548	571	3,184	3,793	-609	4,017	-833	4.5	185	2,342
17차	10,038	9,502	536	3,268	3,869	-601	4,085	-817	5.7	143	1,823

주: 13차 조사(2010년)의 KLIPS 생산가능인구는 인구추계자료가 아니라, 2010년 인구 센서스 실측치를 부여함.

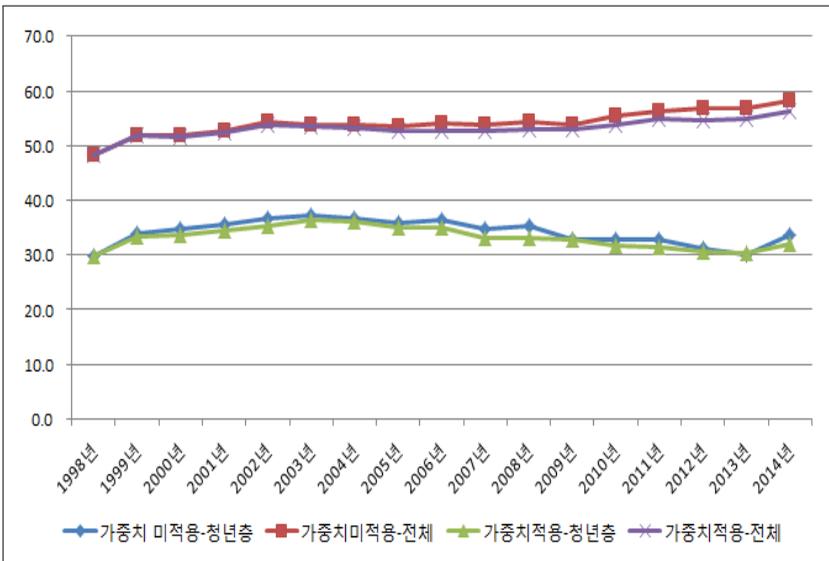
치 부여상의 기술적인 문제로 통계청 가구추계자료와 미미한 차이가 발생하기는 하나 전체적으로 통계청 가구추계와 큰 차이가 발생한다고 볼 수 없다. 만일 생산가능인구 추정에서 문제가 없다고 한다면 결과적으로 고용률의 과소 추정은 취업자의 과소 추정에서 비롯되었다고 볼 수 있다.

그렇다면 취업자의 과소 추정 정도는 어느 정도인가? 경제활동인구조사와 비슷한 고용률이 도출되기 위해서는 생산가능인구 규모에 비례하도록 취업자 규모가 도출되어야 한다. <표 4-5>와 같이 경찰 수준의 고용률이 달성되기 위해서는 추가적으로 825천 명(12차 기준)의 취업자가 증가해야 한다. 12차 조사에서 KLIPS에서 청년층 1인당 평균 가중치가 35백명 규모임을 감안할 때, 가중치를 부여하지 않은 실제 조사인원 기준으로 기존 청년층 취업자 2,977명 외에 233명의 취업자가 추가적으로 파악되어야 한다는 것을 의미한다.

두 번째로 고려할 사항은 가중치 부여과정에서의 왜곡 가능성이다. 가중치 부여과정에서 왜곡이 발생하여 청년층 취업자 과소 추정 문제가 발생했는지 여부를 검증하기 위해서는 가중치를 부여했을 때와 가중치를

(그림 4-1) 가중치 적용 여부에 따른 청년층 고용률 변화

(단위: %)



부여하지 않았을 때의 고용률 변화를 비교하는 방법을 사용할 수 있다. 만일 청년층 가중치를 부여했을 때의 청년층 고용률 격차가 전체 표본에 대해 가중치를 부여했을 때의 고용률 격차보다 현격하게 크다면 이러한 왜곡을 의심할 수 있을 것이다.

[그림 4-1]과 같이 노동패널의 경우 가중치를 부여한 경우와 부여하지 않은 경우 고용률 격차가 전체 표본은 물론이고 청년층 표본의 경우에도 격차가 1% 미만으로 발생하였다. 즉 가중치에 의한 고용률 왜곡 혹은 취업자 과소 추정의 가능성은 미미하다고 볼 수 있다.

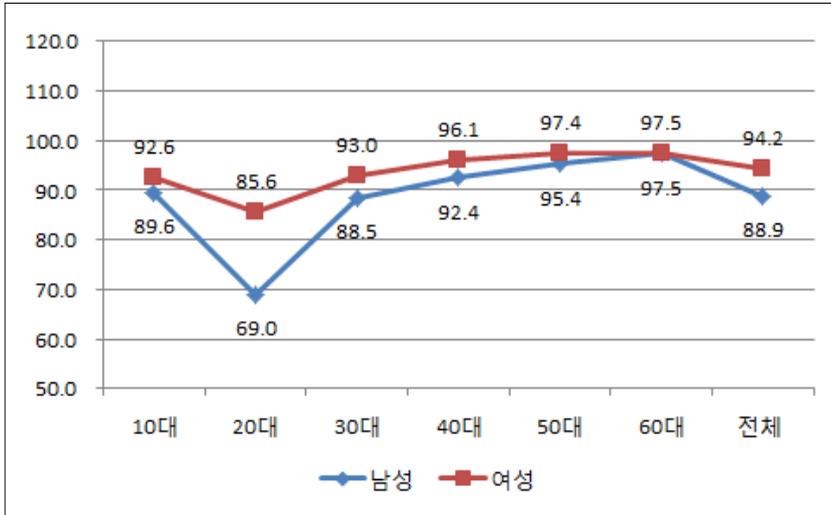
## 2. 개인 응답률

두 번째로 고려할 수 있는 요인은 청년층 취업자에서 상대적으로 더욱 큰 규모의 체계적인 단위무응답(unit non-response)이 발생했을 가능성이다. KLIPS의 조사과정은 우선 가구의 대표 응답자(주로 가구주의 배우자)를 통해 가구조사를 실시한 후, 가구원 중 15세 이상인 자를 대상으로 별도의 개인조사 설문을 실시하는 2단계 구조를 갖는다. 만일 무응답이 취업자와 미취업자에 대해 비례적으로 발생했다면 고용률 추정에 영향을 미치지 않는다. 청년층의 무응답이 전체적으로 다른 연령층에 비해 전체적으로 낮다면 가중치 부여과정에서 연령집단별 스케일 조정을 통해 보정되기 때문이다. 그러나 청년층 중 미취업자에 비해 취업자가 상대적으로 무응답률이 더 높다면 고용률 과소 추정의 원인이 될 것이다.

우선 청년층 취업자가 아닌 청년층 응답률을 살펴보았다. [그림 4-2]는 가구조사에서 15세 이상인 것으로 파악된 가구원의 개인조사 응답률을 분석한 것이다. 그림에서 20대의 응답률은 76.8%로 전체 응답률 91.6%에 비해 14.8%포인트 낮았다. 특히 여성(85.6%)보다는 남성(69.0%) 청년층의 응답률이 더욱 낮았다. 그렇다면 청년층 취업자의 응답률은 왜 더 낮은 것일까? 현장 조사원의 보고에 따르면 청년층의 경우 입시, 통학, 늦은 귀가, 취업, 개인 정보보호 등의 이유로 접촉 및 설득이 상대적으로 어렵다. 이에 비해 설문응답에 대한 보상이 가구단위로 이루어지기 때문에, 경제적 편익조차 크지 않다. 요컨대 청년층의 경우 조사응답에 따른 경제

(그림 4-2) 노동패널 12차 조사의 연령집단별 응답률

(단위: %)



적·비경제적 편익에 비해 비용(시간과 노력)이 클 수 있다.

학교 재학 여부는 청년층의 취업확률과 밀접한 관련성을 갖는다. 재학자보다 비재학자의 고용률이 높기 때문에 비재학자의 무응답률이 상대적으로 높다면 취업자가 더 많이 이탈했을 가능성이 크다고 할 수 있다.

KLIPS 가구자료에는 가구원 각각에 대해 기본적인 인적 정보(성별, 학력, 연령)를 확인하는 문항에서 재학 형태를 식별할 수 있으며, 동거 여부, 비동거 시 비동거 사유도 파악하고 있다.

가구원의 동거형태 역시 청년층 취업률과 밀접한 관계를 가질 수 있다. KLIPS는 문항 분량 및 내용의 민감성 등으로 대부분 대면조사에 의존하며 대리조사 및 전화조사 의존도가 매우 낮다.<sup>6)</sup> 이런 상황에서 비동거 가구원의 경우 원거리에 소재하기 때문에 해당 지역 조사원이 접촉해야 하며, 조사원-응답자 간의 신뢰관계를 통한 설득이 더욱 어려워진다. 비동거 가구원에 대한 연락처 확보에도 어려움이 있기 때문에 추적(locating) 및 접촉(contact)이 더욱 어렵다. 이런 상황에서 청년층이 졸업 후 직장을

6) 한국노동패널 기초분석보고서(이규용 외, 2015: 17)에 따르면 KLIPS의 16차년도 조사 면접 응답 비중은 95.2%에 이른다.

가질 경우, 비동거 형태로 주거지를 달리 할 수 있다. 청년층의 경우 주거 이동을 수반하는 원거리 직장 이동이 빈번하며, 설령 동일 지역에서 일자리를 가지더라도 경제적 독립을 위한 준비단계로 거주지를 변경할 수 있다.

<표 4-6>은 이상의 가설을 검증하기 위해 청년층 가구원의 재학 여부 및 동거형태에 따른 응답률 차이를 분석한 결과이다. 재학생의 응답률은 87.2%로 비재학생(졸업, 수료, 중퇴 등)의 응답률 76.6%보다 10.6%포인트 높았다. 만일 비재학생의 응답률이 재학생 수준이었다면 추가적으로 208명에 대한 조사가 가능하며 비재학생 청년층의 고용률을 60%로만 가정하더라도 125명 정도의 추가적인 취업자 식별이 가능했다고 할 수 있다.<sup>7)</sup>

동거형태에 따른 응답률 역시 편차가 크게 나타났다. 청년층 중 동거 가구원은 개인 응답률이 90.8%지만, 비동거 가구원인 경우 개인 응답률이 29.6%에 그쳤다.

청년층 비동거 무응답자에 대해서는 추가적으로 비동거 사유를 살펴보았다. 비동거-무응답자의 비동거 사유 중 가장 많은 비중을 차지하는 것은 군입대(47.5%)이고, 그다음으로 학업(34.4%)으로 나타났다. 그러나 그 외 취업과 관련된 요인(타 지방 근무)은 11.7%에 그쳤다. 앞서 청년층

<표 4-6> 청년층 가구원의 응답 여부 및 유형

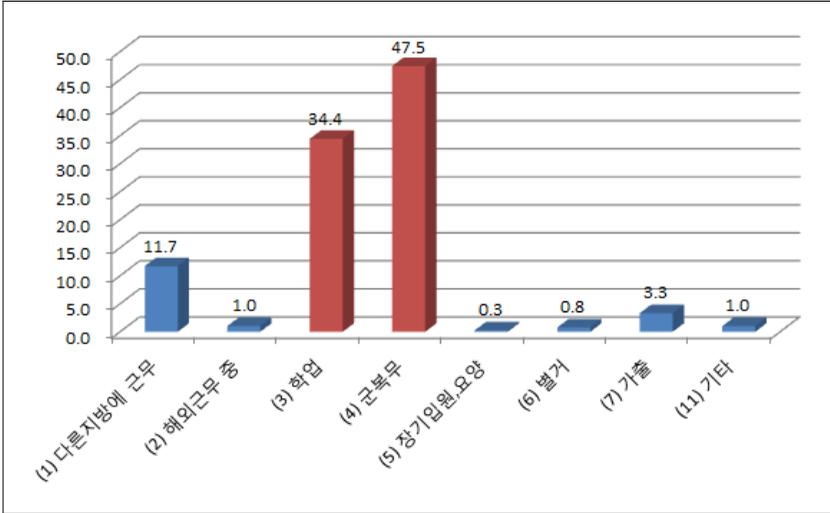
(단위: 명, %)

	합계	재학 형태별		동거 형태별	
		재학	비재학	동거	비동거
합계	3,655 (100.0)	1,689 (100.0)	1,966 (100.0)	3,098 (100.0)	557 (100.0)
응답	2,977 (81.5)	1,472 (87.2)	1,505 (76.6)	2,812 (90.8)	165 (29.6)
무응답	678 (18.5)	217 (12.8)	461 (23.4)	286 (9.2)	392 (70.4)

7) 물론 생산가능인구도 동반 증가하므로 고용률도 비례하여 증가하는 것은 아니다.

(그림 4-3) 노동패널 12차 조사의 무응답 사유

(단위: %)



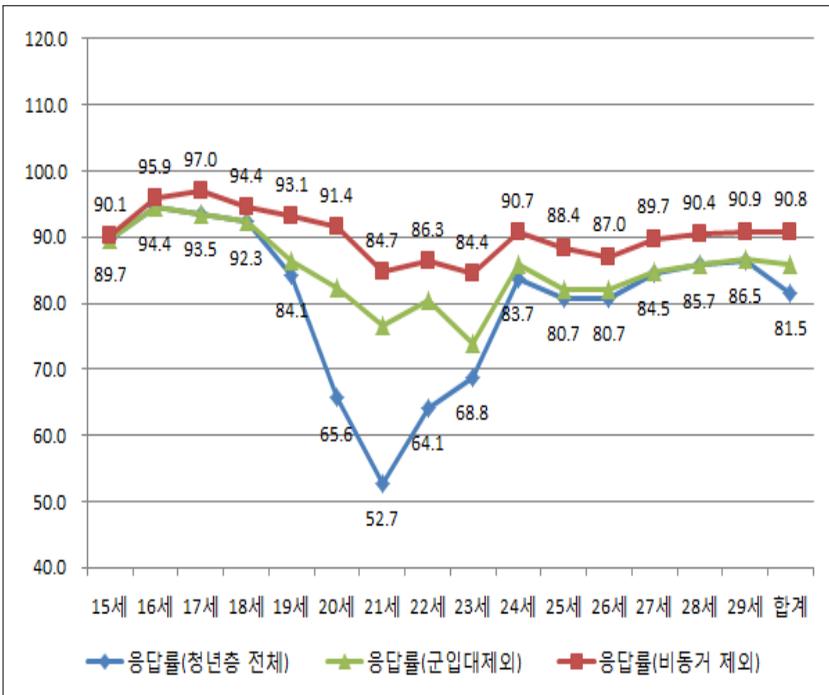
남성의 응답률이 낮은 이유가 군입대에 기인함을 알 수 있다. 군입대는 생산가능인구에서 제외되며, 학업 때문에 비동거-무응답하는 청년층과 다른 지역에 근무하는 청년층의 응답률이 모두 증가하더라도 분모와 분자가 비례적으로 증가하므로 고용률 증가에는 거의 영향을 미치지 못할 것이다. 따라서 앞서 <표 4-2>에서 KLIPS와 경찰조사 간 여성 청년층 고용률 격차가 남성보다 더욱 컸다는 점을 상기하더라도 군입대 요인은 직접적인 취업자 과소 추정의 요인으로 보기 어렵다. 요컨대 학업과 군입대가 대부분을 차지하는 가구원 비동거 요인은 통합표본 첫 해의 청년층 고용률 과소 추정과는 관계없는 것으로 나타난다. 그러나 패널조사의 경우 청년층 비동거 가구원의 무응답은 그 효과가 누적되기 때문에 동태적으로서는 문제가 될 수 있다. 군입대와 학업은 향후 노동시장에 진출할 잠재적 취업자 풀이다. 따라서 가중치를 통해 응답확률을 보정하더라도 한계가 따른다. 앞서 <표 4-5>에서 청년층 1인당 평균 가중치가 12차 조사에서는 35백 명에서 17차 조사에 이르면 57백 명으로 1.6배가 증가하였다. 개인 전체 평균 가중치가 12차 조사 28백 명에서 34백 명으로 1.2배 증가한 것과 비교할 때 훨씬 높은 수준이다. 개인표본 이탈이 많기 때문에 응

답률 조정을 통해 청년층의 1인당 평균 가중치가 더욱 가파르게 증가하는 것이다. 요컨대 비재학생과 비동거 청년층의 낮은 응답률은 취업자 과소 식별의 중요한 요인 중 하나임을 확인할 수 있다.

청년층의 낮은 응답률은 노동패널 표본관리 정책과도 연관된 부분이다. 기존의 KLIPS 표본이탈 억제 정책은 가구, 특히 원표본가구 중심의 표본 유지율을 중심으로 추진되고 있으며, 조사 성공률을 나타내는 제1지표로 ‘원표본가구 유지율’을 사용하고 있다. 그런데 청년층의 경우 혼인, 출산, 취업, 양육 문제 등으로 분가 확률이 높고, 조사 추적 문제로 표본이탈 가능성이 높음에도 불구하고 동등한 수준 혹은 그 이하의 노력을 투여하고 있는 것이다. [그림 4-4]와 같이 현재에도 군입대를 제외할 경우 청년층 응답률은 85.8%까지 증가하기 때문에 비동거 응답률이 심각한 정도로 낮은 수준이라고 볼 수 없다. 그러나 비동거 응답률을 제외할 경우

(그림 4-4) 비동거 제외 시 개인 응답률(12차 조사)

(단위: %)



청년층 응답률은 90.8%까지 증가한다는 점에서 여전히 약 5%포인트가량의 추가적인 응답률 제고 노력이 필요하다. 또한 비동거를 제외하더라도 여전히 다른 연령대보다는 2~3%포인트 정도 응답률이 낮으므로, 청년층 응답률을 제고할 필요가 있다.

### 3. 조사체계의 차이

조사체계상의 문제점들은 취업자 판별과 관련된 문항의 설계 혹은 설문방식, 취업자의 정의와 관련된 지침상의 차이점, 조사 시기와 기간 등 조사 전 단계에서 발생하는 비표집 오차이다.

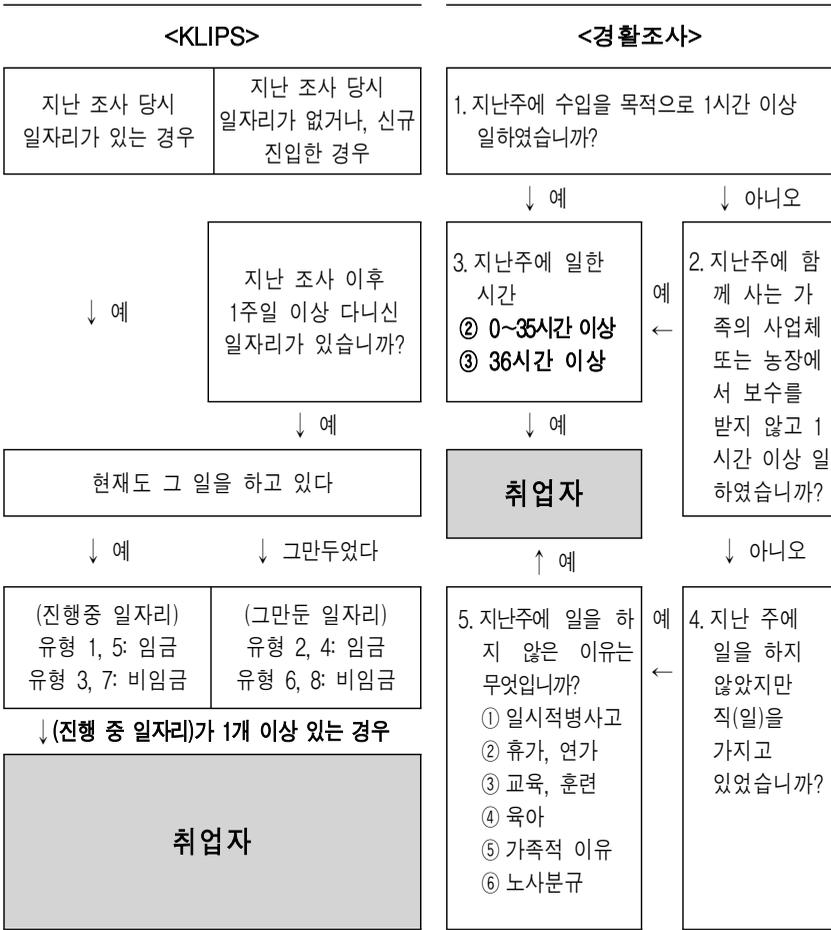
첫째, KLIPS의 취업자 판별구조가 경제활동인구조사와 다를 수 있다. 황수경(2010)에 따르면 설문조사에서 취업여부 및 구직활동과 관련된 질문에서 비논리적인 응답이 빈번하게 발생할 뿐만 아니라, 문항 구성의 순서에 따라서도 다른 응답이 발생한다. 또한 동일한 설문이라 하더라도 개념/용어에 대한 정의의 차이, 면접원과 응답자들에게 실제로 전달되는 조사 지침상의 차이 등으로 인해 동일한 내용의 문항에서도 다른 응답이 가능하다(이상호, 2002).

[그림 4-5]는 경제활동인구조사와 KLIPS의 취업자 판별과정을 도식화한 것이다. 경제활동인구조사가 ‘지난 1주간의 경제활동’을 토대로 취업 여부를 판별하는 데 비해, KLIPS의 경우 취업자 판별 기준에서 조사시점 ‘현재’를 기준으로 사용한다. 즉 현재 일자리 보유 여부에 대한 예비조사와 ‘유형 설문’ 식별과정을 통해 취업자를 선(先) 판별한 후, 경제활동인구조사와 동일한 과정을 거쳐 실업자를 판별하는 방식이다.

다만 KLIPS도 취업자 개인용 설문, 미취업자 개인용 설문 등을 통해 지난 1주간의 활동상황에 대해 경제활동인구조사와 동일한 구조의 설문을 포함하고 있다. 즉 두 가지 방식의 취업자 판별이 가능한 것이다. 그러나 일자리 ‘유형’ 판별 결과와 경활방식의 판별 결과 차이는 연간 0.5% 이내(약 20명)에서 발생하는 것으로 나타났다.<sup>8)</sup> 두 방식 간의 결과 차이가 나타나는 주된 이유는 ‘발령대기 및 일시 휴직’에 대해 일자리 유형 판별

8) 다만 전체 표본을 대상으로 할 경우, 이러한 분류 차이는 3% 정도로 다소 증가하였다.

[그림 4-5] KLIPS와 경활조사의 취업자 판별 설문구조 차이



에서는 취업자로 판별되기 때문인 것으로 확인되었다.

다음으로 취업자 판별과 관련하여 경계성 일자리를 가진 경우 과소 보 고가 나타날 가능성을 분석하였다. 그런데 <표 4-7>은 경제활동인구조 사와 KLIPS 간 비경제활동인구의 분포를 비교한 것이다. 경제활동인구 조사에 비해 KLIPS의 경우 비경제활동인구 중 통학자의 비중이 현저하 게 높은 것을 알 수 있다. 즉 학교 재학생들의 일자리 보유가 낮은 것이 다. 경활조사의 청년층 중 재학생 비중(2009. 8 기준)은 46.0%, KLIPS(12 차 조사)의 재학생 비중은 50.3%로 다소 높은 것이 사실이다. 그러나 재

〈표 4-7〉 2009년 KLIPS와 경제활동인구조사의 비경제활동인구 분포

(단위: 명, %)

비경제활동인구	경활		KLIPS	
	명	%	명	%
계	15,698	(100.0)	17,735	(100.0)
육아	1,599	( 10.2)	698	( 3.9)
가사	5,552	( 35.4)	6,403	( 36.1)
통학	4,292	( 27.3)	5,364	( 30.2)
정규교육기관통학	3,974	( 25.3)	5,115	( 28.8)
입시학원통학	78	( 0.5)	107	( 0.6)
취업목적학원·기관통학	240	( 1.5)	143	( 0.8)
연로	1,606	( 10.2)	1,483	( 8.4)
심신장애	448	( 2.9)	749	( 4.2)
그 외	2,202	( 14.0)	3,037	( 17.1)
취업준비	351	( 2.2)	250	( 1.4)
진학준비	144	( 0.9)	156	( 0.9)
군입대대기	44	( 0.3)	93	( 0.5)
쉬었음	1,475	( 9.4)	2,334	( 13.2)
기타 등	188	( 1.2)	206	( 1.2)

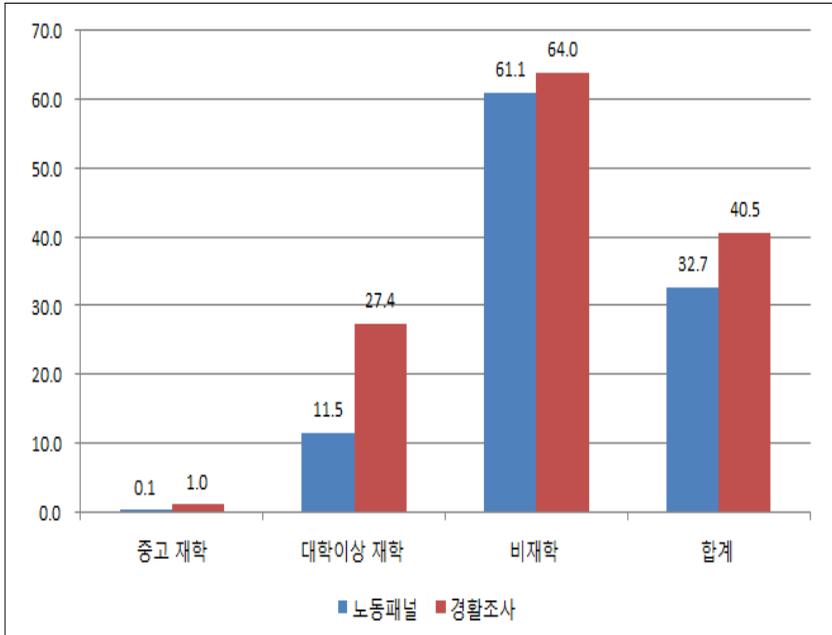
학생의 구성 차이에 기인한 효과는 약 25천 명 정도로 <표 4-4>에서 제시한 취업자수 차이의 20% 내외만을 설명한다.

오히려 더욱 중요한 문제는 재학생 중 취업자수에서 두 조사 간에 현격한 차이가 발생한다는 점이다. [그림 4-6]과 같이 대학 재학 여부에 따른 고용률을 비교하였다. 비재학생의 고용률 격차는 3%포인트 이내로 다른 연령층에서 나타나는 두 조사 간 고용률 격차와 비슷한 수준이었다. 그런데 대학 이상 재학자(두 조사 모두 청년층 인구의 약 20% 정도를 차지)의 고용률 격차는 경제활동인구조사(27.4%)와 KLIPS(11.5%) 간에 16%포인트나 되는 것으로 확인되었다.

그렇다면 재학생의 낮은 고용률은 왜 발생하는가? 우선 재학생의 낮은 고용률은 KLIPS의 조사시점과 관련이 있을 수 있다. 방학 중에는 아르바이트가 증가하는 등 고용률이 높아질 수 있기 때문이다. 만일 청년층의

[그림 4-6] 학교 재학 여부별 고용률 격차(12차 조사)

(단위: %)



응답이 특정 시점에 집중되었거나 각 시점별로 고용률에 차이가 발생한다면 이를 의심해볼 수 있다. 그러나 <표 4-8>과 같이 특정 월에 체계적으로 청년층 고용률이 높거나 낮다고 보기는 어려우며, 경찰조사에서도 월별 고용률 변동은 2%포인트 내외 수준인 것으로 확인되었다.

그렇다면 재학생의 낮은 고용률은 아르바이트를 포함할 경우 증가할까? KLIPS는 미취업자로 판별되었다고 하더라도 고교 및 대학 재학생에 대해서 6차 조사부터 지난 조사 이후 아르바이트 경험 여부, 기간, 근로조건, 현재 진행 여부에 대해서 질문한다. 미취업자에 대해 조사시점 당시 아르바이트가 진행 중일 경우 취업자로 재분류하여 고용률을 산출한다. <표 4-9>와 같이 아르바이트를 포함하더라도 고용률 증가 효과는 1% 이하 수준이며, 전체 고용률에 미치는 효과는 더욱 미미하였다.

〈표 4-8〉 KLIPS의 월별 청년층 조사자수 및 고용률

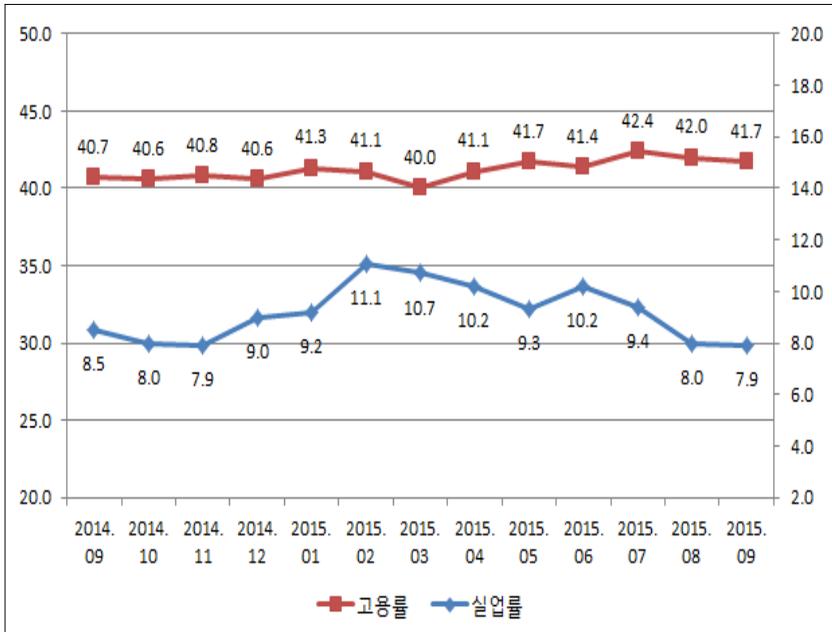
(단위: %)

	12차	13차	14차	15차	16차	17차
	응답수(고용률)	응답수(고용률)	응답수(고용률)	응답수(고용률)	응답수(고용률)	응답수(고용률)
3월	3(61.8)					177(29.8)
4월	656(32.8)					468(35.2)
5월	1,001(33.2)		3(36.4)			538(33.6)
6월	497(31.7)		5(32.2)	7(38.6)	86(33.6)	420(31.8)
7월	388(31.9)	149(32.9)	33(16.4)	565(31.6)	539(33.5)	77(28.2)
8월	238(32.0)	806(32.8)	858(34.7)	744(35.0)	479(31.3)	46(23.7)
9월	150(31.8)	505(31.5)	505(33.5)	504(37.0)	442(26.9)	24(20.8)
10월	42(39.5)	888(34.9)	571(37.6)	310(31.7)	440(30.1)	1
11월	1	231(39.7)	467(36.9)	164(25.6)	173(37.7)	
12월		35(36.7)	22(47.4)	50(25.0)	85(38.2)	

주: 고용률 산출 시 09월단면 가중치 부여.

(그림 4-7) 경제활동인구조사의 월별 청년층(15~29세) 고용률 및 실업률 추이

(단위: %)



〈표 4-9〉 고교 및 대학 재학생 아르바이트 취업 보정 시 고용률 증가 효과

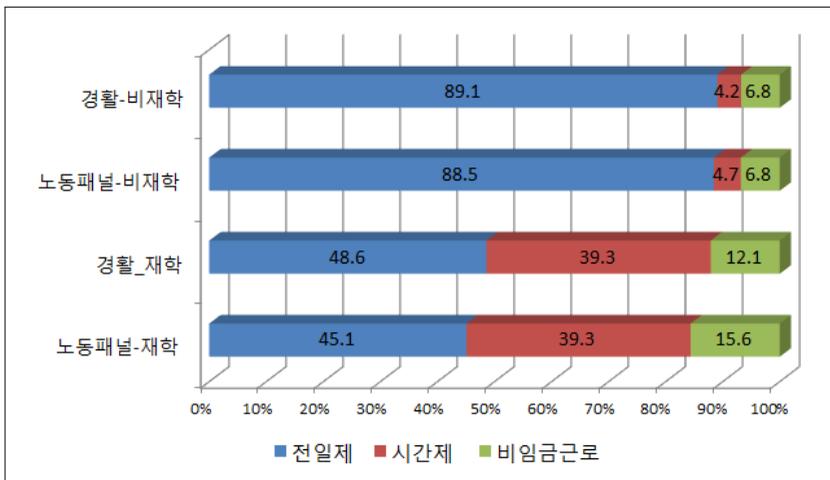
(단위: %)

	아르바이트 보정 청년층 고용률				아르바이트 보정 전체 고용률			
	09횡단면 가중치		09종단면 가중치		09횡단면 가중치		09종단면 가중치	
	고용률	증가분	고용률	증가분	고용률	증가분	고용률	증가분
12차	33.2	(0.5)	33.2	(0.5)	54.3	(0.1)	54.3	(0.1)
13차	34.3	(0.3)	33.6	(0.4)	55.9	(0.1)	55.4	(0.1)
14차	36.0	(0.6)	34.5	(0.7)	57.0	(0.1)	56.2	(0.2)
15차	34.2	(0.8)	32.8	(0.9)	57.1	(0.2)	56.3	(0.2)
16차	32.3	(0.8)	30.6	(1.0)	57.1	(0.2)	56.3	(0.2)
17차	33.2	(0.6)	30.8	(0.7)	57.6	(0.1)	56.5	(0.2)

추가적으로 재학생 취업자의 과소 추정이 청년층 고용형태의 구성에 미치는 영향을 분석하였다. 재학생 취업자의 과소 추정이 상대적으로 시간제 고용 비중의 하락으로 연관될 것으로 예측할 수 있지만, KLIPS와 경찰조사에서 취업자 중 시간제 비중은 39.3%로 동일하였다. 비재학생 취업자의 시간제 비중 역시 경찰은 4.2%, 노동패널은 4.7%로 큰 차이를 보이지

(그림 4-8) KLIPS과 경찰조사 청년층 취업자의 고용형태 구성 비교

(단위: %)



주: 노동패널 09횡단면 가중치 적용.

않았다. 즉 재학생 취업자 판별의 과소 추정 문제는 단순히 시간제에 집중되었을 것이라는 기대와는 달리 끌고루 발생했음을 유추할 수 있다. 따라서 KLIPS의 재학생 취업자 과소 추정 문제는 시간제-아르바이트와 같은 일부 고용형태보다는 청년층 응답자의 전반적인 패널조사에 대한 태도와 관련되었을 가능성이 높다.

#### 4. 소 결

결론적으로 KLIPS의 청년층 고용률 과소 추정 문제의 주된 요인은 비재학생의 상대적으로 낮은 응답률과 재학생의 취업자 식별 때문인 것으로 보인다. 그러나 대학 재학생 취업자수가 과소 식별되는 구체적인 원인과 관련해서는 조사시점 및 기간 효과로 보기도 어려우며, 기존에 파악된 아르바이트가 취업자 식별에 반영되지 않아서 발생한 것도 아니었다.

KLIPS의 재학생 취업자수 과소 식별의 원인과 관련하여 이를 패널조사에 특수한 응답자 태도, 즉 ‘패널 조건화’로 설명할 수 있다. 패널 조건화란, ‘응답자가 이전에 조사를 한 번 이상 해보았기 때문에 발생하는 응답상의 변화’를 의미한다(최효미 외, 2009). KLIPS 설문구조상 대학 재학생의 경우, 경험적으로 취업자의 설문 부담이 상대적으로 많다는 것을 사전에 인지하고 있다. 이때 청년층은 미취업자로 식별되더라도 다른 문항에 응답하는 과정에서 발각될 가능성이 적기 때문에 보유한 일자리에 대해서 생략하는 전략을 취할 때의 기대편익이 더 높을 수 있다. 만일 비재학 취업자가 자신의 취업상황을 속이는 전략을 사용할 경우 가계경제와 관련한 소득 및 소비 문항들, 일상적인 경제 및 노동시장 활동에 대한 질문에서 발각될 확률이 높다. 반면 재학생 청년층들은 주된 활동이 통학 및 학업이므로 일자리를 갖고 있다 하더라도 비경제활동 상태와 관련된 응답 부담이 취업자와 실업자에 비해 현저히 낮은 현재의 문항구조상 이를 파악하기 어렵다.

다만 이 연구에서 패널 조건화 가능성을 언급하는 것은 다른 가설들이 기각되었거나 설명력이 떨어지기 때문이지, 패널 조건화를 직접적으로 입증한 것은 아니다. 예컨대 부업, 교육훈련과 같이 일단 한번 응답하면,

하위 응답문항이 연결되면서 응답 부담의 증가로 귀결되는 지표 분석을 통해 패널 컨디셔닝에 대한 검증이 가능할 것이다.<sup>9)</sup>

결론적으로 KLIPS의 고용률 과소 추정 문제는 패널조사에 특수한 비표집 오차로 이해하는 것이 바람직하다. 그러나 패널조사의 목적이 대표성 있는 고용률 및 실업률 통계를 산출하는 것은 아니므로 이에 대한 해법 역시 신중하게 접근할 필요가 있다.

우선 경제활동인구조사 수준의 고용률로 조정하기 위해 가중치를 부여하는 방식은 권장하는 바가 아니다. 예컨대 국내에서는 청년 패널조사에서 경제활동인구조사에 비례하도록 고용률을 조정하는 가중치를 사용하고 있다.<sup>10)</sup> 그러나 고용률 지표 조정이라는 1개 변수의 교정을 위해 무리하게 가중치를 조정할 경우 다른 주요 변수들에 예상치 못한 영향을 미칠 수 있으며 오히려 전체적인 조사의 대표성 문제를 훼손할 우려가 존재한다. 오히려 청년층의 응답률을 제고하기 위한 인센티브 부여 방안, 미취업자 설문체계 개선, 현장조사 관리 등 비표집 오차 개선방안을 제시하는 것이 바람직하다.

## 제4절 노동력 상태 이동성 분석

KLIPS와 같은 패널조사는 고용지표에 대한 대표성보다는 노동력 상태의 이동성(mobility) 같은 노동시장 행태의 미시적-동태적 변화 분석을 활성화하는 것이 바람직하다. 외국의 가구패널조사에서는 노동력 상태 변동을 사용자지침서 등에서 기본 분석지표로 제공하고, 더 나아가 다양한 영역으로 분석 대상과 방법을 확장하고 있다. 예컨대 Wilkins et al.(2013, p.38. table 7.3.)은 호주의 가구패널인 HILDA를 이용하여 200

9) 다만 KLIPS는 ‘부업’(응답시점 당시 복수의 일자리 진행 중)에 대한 추정 역시 상대적으로 과소 식별되는 것으로 알려져 있는데, 2012년 KLIPS 취업자의 부업 비율은 1.9%로 2012년 8월 경황조사 부업자 비중인 1.9%와 동일한 수준이다.

10) 청년패널 1차 조사(2007년)의 고용률은 20% 수준이나, 가중치 부여 후 고용률이 40% 수준으로 증가한다.

5~10년의 경제활동상태 간 변동을 네 가지 노동력 상태(전일제, 시간제, 실업, 비경활)로 나누어 분석한 바 있다. Malo and Muñoz-Bullón(2003) 역시 영국 가구패널조사인 BHPS 첫 3개년도 직업력 자료를 이용하여 노동력 상태 간 이동성, 특히 비자발적 이직에 초점을 맞추어 분석하였다.

이 절에서는 Wilkins et al.(2013)과 같은 형태로 경제활동상태와 미취업 상태의 인접 연도 간 이행 및 5년 시점 간의 이행을 분석하였다.

우선 <표 4-10>의 경제활동상태의 인접 연도 간 변화(좌측 열)를 살펴보면, 실업자의 경제활동상태 변화가 매우 역동적으로 이루어지고 있음을 확인할 수 있다. 2013년 → 2014년 사이의 경우 실업자의 31.4%는 다음해 50.4%(전일제 상용직 31.4%, 시간제 혹은 임시일용직 17.4%, 비임금근로 1.6%) 취업으로 이행하며, 37.2%는 비경활인구로 전환되었다. 시간제 혹은 임시일용직은 그다음으로 고용상태의 변화가 큰데, 전일제 상용직으로 전환되는 비중이 5.7%로 비경활인구로 전환되는 비중(10.3%)보다 적게 나타났다.

경제활동상태의 연도 간 누적변화(우측 열)를 살펴보면, 2009년 실업자의 71.2%가 취업으로 이행하였다. 5년 후에도 실업자로 남아 있는 비중은 4.7%에 그치며, 24.1%는 비경활인구로 이행하였다.

시간제 혹은 임시일용직의 경우, 5년 후에는 절반(50.4%)만이 동일한 노동력 상태로 잔존하고 있으며, 나머지는 각각 전일제 상용 혹은 비경활로 양분되었다.

다음으로 <표 4-11>은 앞서와 동일한 방식으로 미취업 상태의 변화를 분석한 결과이다.

실업자와 비경제활동인구로 표본을 제한하여 상태이동 변화를 분석한 결과, 비경제활동인구 내부의 특성에 따라 상이한 상태변화 경로를 가진다는 점을 확인할 수 있었다. 예컨대 2013년 → 2014년 사이 비경활상태의 변화(좌측 열 하단)를 살펴보면, 전년도 실업자의 31.4%가 전일제 상용직으로 이행하고, 17.4%가 임시일용직 혹은 시간제로 이행하였다. 실업자의 2009년 → 2014년까지 누적 상태변화를 살펴보면 실업자로 잔류하는 비율이 4.7%, 비경제활동인구로 이행하는 비율이 24.1%에 그치는 것으로 나타났다.

〈표 4-10〉 경제활동 상태의 동태적 변화

(단위 : %)

	인접 연도 간 변화						기준 시점(2009년) 이후 변화					
	전일제 상용	시간제/ 임시 일용	비임금 근로	실업자	비경활	합계	전일제 상용	시간제/ 임시 일용	비임금 근로	실업자	비경활	합계
	<b>2009년(12차) → 2010년(13차)</b>						<b>2009년(12차) → 2010년(13차)</b>					
전일제상용	89.6	2.6	1.6	1.6	4.7	100.0	89.6	2.6	1.6	1.6	4.7	100.0
시간제/임시일용	7.9	72.2	2.8	2.3	15.0	100.0	7.9	72.2	2.8	2.3	15.0	100.0
비임금근로	2.6	1.5	89.2	0.8	5.9	100.0	2.6	1.5	89.2	0.8	5.9	100.0
실업자	30.1	20.6	8.5	9.2	31.7	100.0	30.1	20.6	8.5	9.2	31.7	100.0
비경활	4.6	4.6	2.4	1.2	87.2	100.0	4.6	4.6	2.4	1.2	87.2	100.0
합 계	29.6	10.5	17.3	1.5	41.1	100.0	29.6	10.5	17.3	1.5	41.1	100.0
	<b>2010년(13차) → 2011년(14차)</b>						<b>2009년(12차) → 2011년(14차)</b>					
전일제상용	89.7	2.1	1.5	1.4	5.3	100.0	84.6	4.1	3.2	1.5	6.7	100.0
시간제/임시일용	8.7	74.1	2.5	2.0	12.7	100.0	13.1	62.9	3.9	2.0	18.2	100.0
비임금근로	2.7	1.7	90.2	0.8	4.6	100.0	4.4	3.1	83.0	1.0	8.5	100.0
실업자	32.7	18.0	3.7	20.5	25.1	100.0	35.7	19.3	9.2	9.2	26.6	100.0
비경활	4.5	5.2	1.4	1.5	87.5	100.0	6.9	7.0	3.3	1.4	81.4	100.0
합 계	30.1	11.0	16.4	1.7	40.8	100.0	30.3	11.3	16.8	1.7	39.9	100.0
	<b>2011년(14차) → 2012년(15차)</b>						<b>2009년(12차) → 2012년(15차)</b>					
전일제상용	89.6	2.4	1.7	1.7	4.7	100.0	80.8	5.0	4.6	1.7	7.9	100.0
시간제/임시일용	9.6	73.4	2.3	2.4	12.3	100.0	17.7	55.9	4.5	2.8	19.1	100.0
비임금근로	2.4	2.0	90.9	0.5	4.3	100.0	6.0	4.8	78.8	0.8	9.6	100.0
실업자	27.3	16.6	8.3	14.3	33.5	100.0	40.4	20.4	6.5	5.4	27.4	100.0
비경활	4.7	5.0	1.3	2.0	86.9	100.0	8.6	8.2	4.3	1.7	77.2	100.0
합 계	30.8	11.2	16.1	1.9	40.0	100.0	31.2	11.7	16.9	2.0	38.2	100.0
	<b>2012년(15차) → 2013년(16차)</b>						<b>2009년(12차) → 2013년(16차)</b>					
전일제상용	90.5	1.6	1.8	1.4	4.7	100.0	77.7	6.1	5.9	1.2	9.1	100.0
시간제/임시일용	8.7	75.1	3.0	1.6	11.5	100.0	21.3	51.5	5.8	0.9	20.4	100.0
비임금근로	3.0	1.9	90.1	0.3	4.7	100.0	7.7	5.0	75.7	0.7	11.0	100.0
실업자	24.8	23.2	3.7	18.3	30.1	100.0	38.5	19.8	8.1	6.0	27.6	100.0
비경활	4.4	4.4	1.4	1.5	88.3	100.0	9.8	9.4	4.7	1.5	74.6	100.0
합 계	31.0	11.2	15.7	1.6	40.5	100.0	32.0	12.0	16.8	1.7	37.5	100.0
	<b>2013년(16차) → 2014년(17차)</b>						<b>2009년(12차) → 2014년(17차)</b>					
전일제상용	91.9	1.4	1.2	1.1	4.4	100.0	75.2	6.3	6.9	1.5	10.2	100.0
시간제/임시일용	5.7	79.4	2.3	2.3	10.3	100.0	21.4	50.4	6.3	0.8	21.2	100.0
비임금근로	1.4	2.4	92.5	0.4	3.4	100.0	8.2	6.5	72.4	0.3	12.7	100.0
실업자	31.4	17.4	1.6	12.4	37.2	100.0	38.0	23.2	10.0	4.7	24.1	100.0
비경활	3.7	3.8	1.2	1.7	89.6	100.0	10.5	9.5	5.1	1.4	73.4	100.0
합 계	30.9	11.6	15.8	1.6	40.2	100.0	32.2	12.3	17.1	1.6	36.9	100.0

주 : 두 기간 사이에 관측치가 모두 존재하는 경우만 수치 산출.  
 자료 : 한국노동패널 2009~14년 풀링 자료, 09표본 개인 중단 가중치 사용.

〈표 4-11〉 미취업 상태의 동태적 변화

(단위: %)

	인접 연도 간 변화						기준 시점(2009년) 이후 변화					
	전일제 상용	시간제 /임시 일용	비임금 근로	실업자	비경황	합계	전일제 상용	시간제 /임시 일용	비임금 근로	실업자	비경황	합계
	<b>2009년(12차) → 2010년(13차)</b>						<b>2009년(12차) → 2010년(13차)</b>					
실업자	30.1	20.6	8.5	9.2	31.7	100.0	30.1	20.6	8.5	9.2	31.7	100.0
구직단념자	6.0	11.7	1.7	6.8	73.9	100.0	6.0	11.7	1.7	6.8	73.9	100.0
구직활동/취업준비	25.8	14.1	1.6	6.2	52.4	100.0	25.8	14.1	1.6	6.2	52.4	100.0
쉬었음	8.8	9.0	4.6	2.3	75.2	100.0	8.8	9.0	4.6	2.3	75.2	100.0
통학/양육 등	3.5	3.5	2.1	0.8	90.1	100.0	3.5	3.5	2.1	0.8	90.1	100.0
합 계	6.0	5.4	2.8	1.7	84.2	100.0	6.0	5.4	2.8	1.7	84.2	100.0
	<b>2010년(13차) → 2011년(14차)</b>						<b>2009년(12차) → 2011년(14차)</b>					
실업자	32.7	18.0	3.7	20.5	25.1	100.0	35.7	19.3	9.2	9.2	26.6	100.0
구직단념자	16.7	14.1	1.9	2.4	65.0	100.0	8.3	17.7	4.8	5.4	63.8	100.0
구직활동/취업준비	23.3	6.9	1.7	8.6	59.5	100.0	35.5	9.6	2.3	5.5	47.2	100.0
쉬었음	7.0	7.5	2.0	2.9	80.6	100.0	10.0	11.0	5.8	1.9	71.3	100.0
통학/양육 등	3.3	4.6	1.3	1.0	89.8	100.0	5.7	6.1	2.9	1.1	84.1	100.0
합 계	5.5	5.6	1.5	2.2	85.2	100.0	5.5	5.8	1.6	2.3	84.9	100.0
	<b>2011년(14차) → 2012년(15차)</b>						<b>2009년(12차) → 2012년(15차)</b>					
실업자	27.3	16.6	8.3	14.3	33.5	100.0	40.4	20.4	6.5	5.4	27.4	100.0
구직단념자	9.1	15.9	4.9	13.1	57.0	100.0	11.6	18.4	7.9	5.1	57.0	100.0
구직활동/취업준비	29.5	6.1	2.0	1.8	60.6	100.0	39.7	12.3	1.6	11.7	34.7	100.0
쉬었음	5.6	7.9	1.7	5.4	79.4	100.0	13.1	11.6	7.1	2.3	66.0	100.0
통학/양육 등	3.7	4.3	1.2	1.3	89.5	100.0	7.1	7.4	3.9	1.3	80.4	100.0
합 계	5.6	5.5	1.6	2.5	84.8	100.0	5.8	5.8	1.7	2.7	84.1	100.0
	<b>2012년(15차) → 2013년(16차)</b>						<b>2009년(12차) → 2013년(16차)</b>					
실업자	24.8	23.2	3.7	18.3	30.1	100.0	38.5	19.8	8.1	6.0	27.6	100.0
구직단념자	10.4	12.2	4.6	21.1	51.7	100.0	11.4	17.5	8.4	3.0	59.8	100.0
구직활동/취업준비	17.0	10.2	4.1	8.3	60.5	100.0	39.0	12.8	2.7	3.8	41.6	100.0
쉬었음	5.8	5.6	2.6	2.6	83.4	100.0	13.4	13.7	7.7	1.9	63.4	100.0
통학/양육 등	3.7	3.9	1.1	0.9	90.5	100.0	8.6	8.5	4.1	1.4	77.4	100.0
합 계	5.3	5.2	1.5	2.3	85.8	100.0	5.6	5.9	1.8	2.5	84.4	100.0
	<b>2013년(16차) → 2014년(17차)</b>						<b>2009년(12차) → 2014년(17차)</b>					
실업자	31.4	17.4	1.6	12.4	37.2	100.0	38.0	23.2	10.0	4.7	24.1	100.0
구직단념자	5.4	21.5	0.0	12.5	60.6	100.0	9.8	18.8	9.0	3.1	59.3	100.0
구직활동/취업준비	13.5	6.9	1.2	6.1	72.5	100.0	42.6	12.4	5.0	4.6	35.4	100.0
쉬었음	4.8	4.7	3.0	2.7	84.8	100.0	12.8	12.5	7.3	1.6	65.9	100.0
통학/양육 등	3.0	3.3	0.9	1.2	91.6	100.0	9.5	8.8	4.6	1.3	75.7	100.0
합 계	4.7	4.3	1.2	2.1	87.7	100.0	4.9	4.7	1.3	2.4	86.7	100.0

주: 두 기간 사이에 관측치가 모두 존재하는 경우만 수치 산출.

자료: 한국노동패널 2009~14년 폴링 자료, 09표본 개인 중단 가중치 사용.

다음으로 구직단념자(취업희망과 취업 가능성이 있으나 적당한 일거리가 부족한 등의 사유로 구직활동을 하지 않은 자)의 경우 5.4%가 다음 해 전일제 상용직으로 이행하며, 21.5%가 시간제 혹은 임시일용직으로 이행하였다. 구직단념자의 경우 5년 시점 동안의 상태변동 역시 거의 차이가 발생하지 않았다.

마지막으로 일시휴직, 구직활동, 취업준비, 발령대기, 취업학원 등의 사유로 비경제활동인구로 식별된 경우 13.5%가 전일제 상용직으로 이행하며, 6.9%가 시간제 혹은 임시일용직으로 이행하였다. 그런데 우측 열과 같이 2009년에서 2014년까지 5년간 누적변화를 살펴본 결과, 일시휴직, 구직활동, 취업준비, 발령대기, 취업학원 등의 전일제 상용직 이행비용은 42.6%로 오히려 실업자(38.0%)보다 높은 것으로 나타났다.

이상의 결과를 종합하면 미취업자의 상태를 세분화했을 때 각각의 이행경로가 서로 상이함을 알 수 있다. 또한 단기적 변화와 중기적 변화가 상이하기 때문에 이를 구분하여 살펴볼 필요가 있다. 비경제활동인구라 하더라도 취업준비 목적의 사유인 경우 상대적으로 양질의 일자리에 취업하기 위한 경우가 많기 때문에 중기적으로는 오히려 상용직으로의 이행률이 더 높다는 점도 흥미롭다. 따라서 비경제활동인구의 특성에 따라 노동시장 진입을 활성화하고자 하는 정책적 개입을 차별화할 필요가 있다.

## 제5절 정책적 함의 및 토론

### 1. 고용지표 신뢰성 향상을 위한 방향

본문에서는 KLIPS와 경황조사의 비교·분석을 통해 KLIPS의 고용률이 경황보다 다소 낮으며, 그 원인은 주로 청년층 고용률 과소 추정에 있음을 확인하였다. KLIPS의 고용률 과소 추정은 표집과정이나 가중치 부여와 같이 표집 오차에 관련된 문제라기보다는 패널조사가 갖는 고유한

비표집 오차에 기인하는 것으로 추정된다. 특히 청년층 비재학생의 패널 이탈과 재학생의 취업 여부에 대한 과소 추정이 가장 큰 영향을 미치는 것으로 분석되었다.

이 문제에 대한 해법 역시 분리해서 세 가지 접근 방법이 필요하다.

첫째, 청년층 응답률이 상대적으로 낮기 때문에 표본유지율 정책의 변화가 필요하다. 현재와 같이 원표본가구원에 대한 표본유지율보다는 분가가구 및 비동거 가구원에 대한 추적 및 접촉 개선 노력과 개인 인센티브의 도입이 필요하다. 또한 지나치게 면접조사에만 의존하기보다는 전화, 이메일, 우편 조사 등 다양한 조사도구를 결합하여 유연한 조사정책이 실시될 필요가 있다.

둘째, 청년층 재학생이 일자리를 갖고 있는 경우 이를 발굴하여 파악할 수 있도록 설문내용의 구조 변화가 필요하다. 특히 현재의 KLIPS 설문구조는 비경제활동인구의 취업 활동 및 학교 생활에 관한 보다 세부적인 파악이 필요하다. 미취업자는 상대적으로 적고 응답 부담 역시 취업자보다 적기 때문에 이에 대한 문항이 추가되더라도 전체적인 응답 부담의 증가는 크지 않을 것으로 예상된다. 오히려 취업자와 미취업자의 응답 부담간 균형을 높임으로써 조사의 질을 제고하는 데 기여할 수 있다.

마지막으로 추가적인 표본 확대가 필요하다. 비록 2009년 표본추가를 통해 전국 대표성을 확보했다고는 하나, 여전히 하위집단의 규모가 작은 경우 신뢰성 확보가 어렵다. 특히 청년층의 경우 절대적 표본수가 여전히 적어서 NEET나 구직자에 대한 세부적인 분석이 어렵다. 청년층의 유입은 적은 반면 유출과 표본이탈은 많기 때문에 향후 이러한 어려움이 더욱 가중될 것으로 예상된다.

## 2. 패널자료 신뢰성 연구의 확장

KLIPS 주요 변수들의 신뢰성 연구는 향후에도 다양한 분야에 걸쳐 확대될 필요가 있다. 2009년 추가표본의 구성으로 인해 KLIPS의 고용관련 변수에 대해 경황조사 등 다른 대표통계들과 보다 직접적인 비교가 가능하게 되었다. 그러므로 고용상태 및 비정규직 외에도 근로시간, 산업 및

직업, 기간 관련 변수들에 대한 다양한 신뢰성 검증 연구로 확대될 필요(홍민기, 2014b, 제5장 신동균의 논의 참조)가 있다.

둘째, 개별 변수의 신뢰성 검증에서 더 나아가, 각 변수에 의해서 발생되는 측정오차가 다른 추정치에는 어떤 영향을 미치는지에 대해 검증할 필요가 있다. 예컨대 청년층 취업자 과소 추정으로 인해서 고용형태 구성, 임금 구성 등에 미치는 효과에 대해서 추가적인 분석이 필요할 수 있다.

셋째, 패널조사의 비표집 오차에 대한 체계적인 연구가 확대되어야 한다. 최근 KLIPS에서는 패널 컨디션닝, 패널 반복응답에 대한 응답자 피로도(panel fatigue) 등이 패널 이탈이나 단위무응답, 히핑, 돌출효과(seam effects) 등 다양한 비표집 오차들을 심화시킬 가능성이 제기되고 있다(최효미 외, 2012; 윤윤규 외, 2012; 홍민기 외, 2014a; 2014b 등 패널자료 품질 연구시리즈 참조). 그럼에도 불구하고 패널 컨디션닝이나 패널 피로도 같은 현상들은 정량적으로 측정하기가 쉽지 않기 때문에 국내에서도 이와 관련된 본격적인 이론 및 실증 작업이 필요하다. Jackle(2008), Lynn et al.(2009) 등은 패널조사 방법론을 체계화하고 서구 주요 가구패널조사의 비표집 오차 문제를 체계적으로 다룬 바 있으나, 국내의 조사환경과 응답자 특성이 외국과 다르기 때문에 이에 대한 별도의 연구가 필요하다.

마지막으로 분석 활용도 제고를 위해서 미시적-동태적 분석 영역 및 방법이 확장되어야 한다. 패널조사의 활용도 제고를 위해 반드시 모든 분석이 엄밀한 모형에 기반할 필요는 없다. 계량모형에 기반한 엄밀한 학술적 연구와는 별개로 일반대중이나 정책담당자들이 쉽게 이해할 수 있도록 고도의 계량적 지식이 필요하지 않은 분석방법을 이용한 동태적 분석도 강화되어야 한다.

## 제 5 장

# 노동패널 중장기 발전방안 - 해외 장기패널과의 비교를 중심으로

### 제1절 문제 제기

한국노동패널은 1990년도에 1차 조사를 시작한 이래 지금까지 18차에 이르는 조사자료를 축적하였다. 초기 시점에 5,000가구를 표본으로 추출하여 여기에 속한 13,321명의 가구원을 면접조사하고 이들을 조사대상 패널로 구축하였다.

패널자료를 장기적으로 축적한다는 것의 의미는 현재의 실태와 변화 과정을 파악할 수 있는 자료가 만들어진다는 뜻이므로 귀중한 자료가 될 것이지만, 그만큼 어려운 과정이기도 하다. 응답자의 수가 점진적으로 줄어드는 것이 필연적이며, 이는 특정 시점에서 횡단면 자료로서 대표성에 문제가 발생할 가능성이 있다. 반복되는 조사에 대한 피로감으로 응답이 부실해질 수 있으며, 과거시점과 최근 시점의 응답 간에 일관성이 없는 경우도 발생하게 된다. 이러한 문제는 한국노동패널이 이미 맞닥뜨린 문제이자, 다른 나라의 장기 패널조사들이 공통적으로 당면하고 있는 문제이다.

이 장에서는 이러한 문제를 극복하면서 장기간의 가구패널자료를 구축해온 사례로 미국의 Panel Study of Income Dynamics(이하 PSID), 영국의 British Household Panel Study(이하 BHPS), 독일의 Socio-Economic

Panel Study(이하 SOEP)를 참고하여 한국노동패널조사의 중장기적인 발전방향을 마련해 보고자 한다.<sup>11)</sup>

여기서는 주로 다음과 같은 하위주제를 중심으로 검토한다. 첫째, 적정 규모의 표본을 확보하기 위하여 추가표집을 해왔는지, 어떤 기준과 방식으로 해왔는지 살펴본다. 대표성 확보를 위한 가중치는 어떻게 관리하며, 특히 다양한 방식의 추가표본이 이루어진 경우 가중치를 어떻게 이용자들에게 제공하는지도 살펴볼 것이다. 둘째, 표본 유지율을 높이면서도 투입예산 대비 더 많은 응답자수를 확보하기 위한 조사방법, 그리고 응답의 신뢰도를 높이기 위한 조사의 노하우를 살펴보고자 한다. 셋째, 연구주체의 다양성과 설문내용을 검토한다. 넷째, 자료 생산과 관리에 투입되는 인력 및 조직을 간략하게 비교·검토한다. 다섯째, 데이터 생산기관 간 협력관계를 형성하기 위한 노력을 소개하고 노동패널이 이러한 네트워크 속에서 어떤 역할을 하고 도움을 받게 될지 살펴본다.

## 제2절 한국노동패널의 발전경로와 현재 상황

### 1. 표본의 규모와 대표성

한국노동패널은 2015년에 18차 조사를 마쳤으나, 현재까지 2014년 17차 조사까지만 그 결과가 공식적으로 발표되어 있다(표 5-1 참조). 가구원의 사망 등으로 소멸한 가구를 제외하면 2014년 현재 남아 있는 원표본 가구, 즉 유효표본가구는 4,741가구이다. 이 중에서 2014년에 실제로 조사된 가구는 3,451가구인데 이는 1차년도 원표본가구 5,000가구의 69%에 해당하며, 유효표본가구수 대비로는 72.8%이다. 한편 원표본가구에서 분가하여 생성되는 분가가구는 매년 늘어나서 2014년에는 2,101가구까지 증가하였다. 유효표본가구와 분가가구를 통틀어 2014년에 조사에 성공한

11) 이 장은 위의 세 패널자료를 생산하는 기관을 직접 방문하여 담당자를 면담한 결과와 이 데이터의 홈페이지가 제공하는 자료에 근거하여 작성되었다.

가구는 5,552가구이다. 조사성공가구에 속한 15세 이상 가구원 중에서 개인설문에 응답한 개인은 10,757명이다.

노동패널조사는 2014년 현재 원표본 유지율 69.0%, 유효표본 유지율 72.8%

〈표 5-1〉 1~17차년도 조사 결과

	조사대상 원가구- 소멸가구	조사성공 가구수 <sup>1)</sup>	원표본 가구수 <sup>2)</sup>	원표본 유지율 <sup>3)</sup>	유효표본 유지율 <sup>4)</sup>	분가 가구수	가구원 응답자수 <sup>5)</sup>	실사 기간
1차(1998)	5,000	5,000	5,000	100.0	100.0	-	13,321	6~10월
2차(1999)	5,000	4,507	4,378	87.6	87.6	129	12,037	7~12월
3차(2000)	5,000	4,266	4,044	80.9	80.9	222	11,205	5~10월
4차(2001)	5,000	4,248	3,866	77.3	77.3	382	11,051	5~10월
5차(2002)	5,000	4,298	3,798	76.0	76.0	500	10,966	4~ 9월
6차(2003)	4,993	4,592	3,862	77.2	77.3	730	11,541	4~ 9월
7차(2004)	4,949	4,761	3,862	77.2	78.0	899	11,661	4~ 9월
8차(2005)	4,935	4,849	3,822	76.4	77.4	1,027	11,580	4~10월
9차(2006)	4,914	5,001	3,820	76.4	77.7	1,181	11,756	4~10월
10차(2007)	4,899	5,069	3,775	75.5	77.1	1,294	11,855	4~10월
11차(2008)	4,862	5,116	3,709	74.2	76.3	1,407	11,734	4~10월
12차(2009)	4,832	5,306	3,657	73.1	75.7	1,649	11,739	4~10월
13차(2010)	4,802	5,379	3,606	72.1	75.1	1,773	11,582	7~12월
14차(2011)	4,786	5,404	3,528	70.6	73.7	1,876	11,377	7~12월
15차(2012)	4,770	5,469	3,517	70.3	73.7	1,952	11,444	6~12월
16차(2013)	4,741	5,501	3,472	69.4	73.2	2,029	11,331	6~12월
17차(2014)	4,741	5,552	3,451	69.0	72.8	2,101	10,757	3~10월

주 : 1) '조사성공가구'는 원표본가구 중에서 실사가 성공적으로 이루어진 유효표본가구와 조사 성공한 분가가구를 포함한 총 조사성공가구임.

2) '원표본가구'는 1차년도 조사 당시 원표본이었던 5,000가구 중 해당 연도에 실사가 성공적으로 이루어진 가구를 의미함.

3) '원표본 유지율'은 1차년도 조사에 성공한 원표본 5,000가구 중에서 해당 연도에 조사 성공한 원표본가구의 비율임.

4) '유효표본 유지율'은 원표본 5,000가구 중에서 가구원의 사망 등으로 소멸한 원가구를 제외한 가구 대비 조사 성공한 원표본가구의 비율임.

5) 15세 이상 가구원으로 면접에 성공한 가구원수.

자료 : 한국노동연구원(2015), 『제17차(2014)년도 한국 가구와 개인의 경제활동 : 한국노동패널 기초분석보고서』, p.12.

로 비교적 높은 패널 유지율을 보이고 있으나, 패널 이탈에 따른 대표성 문제를 해소하기 위하여 2009년에 표본추가를 실시하였다. 또한 이때에는 당초 노동패널이 도시지역인구를 모집단으로 했던 한계를 극복하기 위한 목적도 동시에 추구하였다. 추가표본은 2005년 인구주택총조사를 바탕으로 표집하였으며, 최종적으로 1,415가구에 대한 조사가 이루어졌다. 이에 노동패널은 1998년에 시작한 '1998표본'과 2009년 추가한 표본을 포함하는 '통합표본' 두 가지 버전으로 관리되고 있다.<sup>12)</sup>

노동패널 '통합표본'은 2009년 시점에서 6,721가구 14,489명으로 출발하게 된다. 이후 통합 6차에 해당하는 2014년에는 6,493가구가 조사되어 2009년 대비 86.9%의 표본유지율을 나타낸다. 응답자 개인으로 보면 12,595명이 패널로 유지되고 있다.

2009년도 표본추가로 구축된 '통합표본'은 전국적인 대표성의 문제를 해소하였으나, 전체적으로 표본의 수가 부족하다는 문제는 여전히 해결하지 못하였다. 현재의 표본수로도 전반적인 소득분포나 종사상 지위 분포를 살펴보는 데는 무리가 없지만, 좀 더 세밀한 하위개념으로 분석이 이루어지게 되면 표본수가 적기 때문에 분석에 한계가 발생하는 경우가

〈표 5-2〉 통합표본 1~6차년도 조사 결과

	조사대상 원가구- 소멸가구	조사성공 가구수 <sup>1)</sup>	원표본 가구수 <sup>2)</sup>	원표본 유지율 <sup>3)</sup>	유효표본 유지율 <sup>4)</sup>	분가 가구수	가구원 응답자수 <sup>5)</sup>	실사 기간
1차(2009)	6,721	6,721	6,721	100.0	100.0	0	14,489	4~10월
2차(2010)	6,694	6,398	6,232	92.7	93.1	166	13,641	7~11월
3차(2011)	6,676	6,388	6,082	90.5	91.1	306	13,410	7~11월
4차(2012)	6,641	6,434	6,016	89.5	90.6	418	13,427	6~12월
5차(2013)	6,597	6,457	5,904	87.8	89.5	553	13,303	6~12월
6차(2014)	6,589	6,493	5,840	86.9	88.6	653	12,595	3~10월

주: 1)~5) <표 5-1>과 동일.

자료: 한국노동연구원(2015), 『제17차(2014)년도 한국 가구와 개인의 경제활동: 한국노동패널 기초분석보고서』, p.12.

12) 김재광(미국 아이오와 주립대학교 통계학과)의 『한국노동패널조사 표본추가 연구용역 최종보고서』(2009. 3)를 참조.

있다. 이 점은 다른 나라의 오래된 패널조사들도 모두 겪은 문제로 이 문제를 해소하기 위한 노력을 이어왔다. 이후 살펴보게 될 해외 패널자료들 수준의 표본규모를 확보하기 위한 노력은 노동패널의 주요 과제가 될 것이다.

## 2. 조사방법

노동패널의 조사방법은 거의 대부분 면접원에 의한 직접면접 방식(face-to-face interview)으로 이루어지고 있다. 가구에 대한 설문은 반드시 직접면접을 통하여 조사하도록 하고 있으며, 가구원에 대한 개인설문은 경우에 따라서 전화면접도 원칙적으로 허용하고 있으나, 실제로 그 비중은 매우 낮다. 2014년 7차년도 조사의 경우 전체 개인 응답자의 96.6%가 직접면접 방식으로 조사되었다. 2007년 조사부터는 면접조사에 노트북을 이용하고 있다(Computer Assisted Personal Interviewing, CAPI).

직접면접 방식의 조사는 가장 엄밀한 조사방식이지만, 동시에 가장 비용이 많이 드는 조사방식이기도 하다. 면접원의 인건비가 선진국에 비해서 낮은 점이나, 조사대상의 주거지가 비교적 밀집해 있는 우리나라의 특성이 이러한 조사방법을 유지하는 데 일조하고 있다. 이후 살펴볼 해외 패널조사에서 전화설문은 물론 웹 기반의 온라인 조사를 병행하는 것에 비하여 노동패널은 비교적 전통적인 조사방법을 고수하고 있다고 볼 수 있다. 향후 분석기법에 따른 표본유지율의 변화나 응답의 신뢰성 차이에 대한 면밀한 검토를 거친 후 다른 조사방법을 병행하는 방안에 대한 검토가 필요한 시점에 이른 것으로 판단된다.

## 3. 연구범위

한국노동패널 자료의 조사내용을 간략하게 요약하면 가구 자료와 개인 자료, 그리고 부가조사 자료로 구분할 수 있다(표 5-3 참조). 가구설문은 그 가구에 속한 모든 가구원의 인구사회학적 기본정보와 소득과 소비, 자산 등 가구의 재정적 특징을 조사한다. 개인설문은 조사대상 가구에 속한 15

〈표 5-3〉 1~17차년도 한국노동패널 자료의 구성

		내 용	
가구 자료	KLIPS_H	가구원의 기본정보 가구 특성 관련 설문(소득, 소비, 주거 등)	
개인 자료	KLIPS_P	주된 일자리의 특성	임금근로자 비임금근로자
		응답자 공통	취업자 미취업자
		신규 조사자 문항	
부가조사 자료	KLIPS_A	청년층 조사(3차) 건강과 은퇴(4차) 중고령자 조사(6차) 근로시간과 여가(7차) 노동조합과 노사관계(8차)	청년층 조사(9차) 고용형태 조사(10차) 교육(11차) 시간사용과 삶의 질(17차) 일과 행복(18차)
직업력 자료	KLIPS_W	개인이 가졌던 모든 일자리에 대한 정보	

자료: 한국노동연구원(2015), 『제17차(2014)년도 한국 가구와 개인의 경제활동: 한국노동패널 기초분석보고서』, p.13.

세 이상의 모든 가구원에 대하여 실시하는데, 주로 경제활동상태와 일자리의 특성을 상세히 질문한다. 이렇게 조사한 내용은 직업력 자료의 형식으로도 변환하여 제공한다. 또한 매 조사연도에는 특별한 주제에 집중한 부가조사가 실시된다. 그동안 청년층의 학업과 노동시장 진입, 고령자의 건강과 은퇴, 비정규직의 특성, 근로시간과 여가 등 다양한 주제들이 다루어진 바 있다.

#### 4. 인력과 조직

노동패널은 현재 한국노동연구원에서 생산하고 있으며, 실사(field work)는 한국리서치가 담당하고 있다. 1명의 연구책임자 이외에 4명의 전담인력이 투입되어 있다. 해외의 주요 패널데이터 생산에 관여하는 인력 규모와는 비교할 수 없을 정도로 작다. 이러한 인력 규모로 설문 개발과 실사 관리, 데이터를 활용할 수 있는 형태로 가공하여 배포하는 일, 데이터를 홍보하고 활용도를 높이는 일, 데이터를 이용한 다양한 분석 결과를 출판

하는 일 등을 모두 하고 있다.

### 제3절 미국 PSID의 경험

#### 1. 표본의 규모와 대표성

PSID는 미국의 대표적인 가구패널조사로서 1968년부터 시작하여 지금까지 진행되어 왔다. 지금까지 3,200개의 저널 논문이 이 PSID 자료 분석을 토대로 작성되었고 9개의 미국 정부기관이 이 자료 분석결과를 정책 결정에 사용한다. 그래서 미국 연구재단(National Science Foundation)에서는 PSID를 후원한 것을 그 재단이 실시한 60개의 주요 혁신 중 하나로 선정할 만큼 그 중요성을 인정받고 있다 할 것이다.

PSID의 연혁을 살펴보면 1968년부터 1997년까지는 매년 조사되었다가 이후부터는 격년으로 조사되어 공표되고 있다. 처음 표본 설계 당시에는 Survey of Economic Opportunity(SEO) 표본에서 얻어진 저소득층을 과대 표집하는 형태로 초기 표본을 얻어냈고, 이후로 분가나 생식, 소멸 등에 따라 표본가구가 더욱 분화되었다. 1990년도에는 히스패닉 계열의 표본가구들이 대거 보충되었고 1997년도에는 실사상의 어려움으로 SEO 초기 표본의 일부를 제거하고 대신에 1968년 이후에 미국으로 이민 온 이민자 모집단에서 표본가구를 뽑아 PSID 표본가구에 보충시켰다. 1990년대 후반부터는 설문문항에 보건과 재산, 지출, 자녀 양육, 성인으로 진출하는 과정, 그리고 시간 사용 등에 대한 항목들을 별도의 조사로 추가시켰다.

PSID의 규모는 2009년 자료의 경우, 8,690개의 가족과 23,102명의 개인으로 구성되어 있다. 연령대별로는 0~18세에 7,511명, 19~39세에 7,396명, 40~59세에 5,678명, 60세 이상은 2,517명의 응답자가 표본에 포함되어 있다. 이처럼 전 연령대에 고르게 분포하고 있으면 이 중 반 이상은 5 번 이상의 차수(waves)에 응답한 경험이 있다.

PSID 표본에서는 저소득층이 더 높은 표본탈락률(attrition rate)을 보

여주고 있지만 가중치를 사용하여 보정하고 있으며 다른 벤치마킹 서베 이와 비교했을 때 비슷한 결과를 보여주고 있어서 표본의 대표성이 유지되고 있는 것으로 판단하고 있다. 가중치는 종단면 개인 가중치, 종단면 가구 가중치, 횡단면 개인 가중치를 제공하고 있으며, 이러한 가중치 작성과 관련된 보고서는 PSID 홈페이지에서 제공되고 있다.

PSID와 다른 공식 통계자료와의 일치성을 높이기 위해 가중치 단계에서 일종의 벤치마킹 작업을 실시한다. 가중치 조정을 통한 벤치마킹에 사용되는 변수로는 단순히 인구학적 변수(성별, 연령대, 인종)뿐만 아니라 사회·경제학적 변수(소득, 실업률 등)도 사용된다고 한다. 그런데 빈곤도 등의 지표로 사용될 수 있는 식료품지원제도(Food Stamp) 등의 사용자 비율 등은 PSID에서 지속적으로 과소 대표되는 것으로 잘 알려져 있는데 이를 가중치에서 반영하지는 않는다. 사망으로 인하여 조사가 중단된 경우에는 다른 표본이탈과는 달리 처리한다(mortality correction).

## 2. 조사 및 데이터의 배포

PSID의 조사는 1973년 이후로 거의 대부분 전화조사로 이루어지며, 1993년부터는 CATI(Computer-assisted telephone collection technology)가 사용되었다. PSID 표본은 다른 조사들보다 더 높은 응답률(표본 유지율)을 보이고 있는데 그 이유로는 1분당 1달러에 육박하는 응답 인센티브, 응답 완료 시 소정의 추가 인센티브, 응답가구에 가급적 예전과 동일한 조사원을 배치하기 등과 같은 세심한 배려, 정교한 응답 거절자 설득 기술(refusal conversion techniques),<sup>13)</sup> 뉴스레터 등의 홍보자료, 조사 차수 사이의 연락처 업데이트 시 소정의 추가 인센티브 제공 등을 들 수 있다. 조사는 3월부터 12월까지 진행되며 많은 표본이 조사 초반에 얻어지지만 뒤로 갈수록 재접촉을 통해서 응답 비협조를 응답으로 바꾸는 데 노력을 기울이게 되는데, 이러한 추가적인 노력의 성공 확률은 30~40% 수준이다.

자료가 수집 완료되고 나서 바로 그 다음해 초에 에디팅(자료정제)이

13) 설문참여를 독려하는 편지와 함께 5달러 정도의 사례비를 우편으로 선지급하기도 함.

거의 없는 상태로 일부 연구자들에게 제공된다. 이후에 자료정제 과정을 충분히 거친 후 데이터 센터(Data Center)에서 공식적으로 공개하기까지 약 18개월이 걸린다.

자료정제 과정에서 가장 어려운 부분은 가족 구성원을 과거 자료와 링크를 통해서 일관성 있게 파악하는 일로 판단되고 있다(family composition editing). 이 과정에 활용하기 위하여 Family Identification Mapping System (FIMS)이라는 어플리케이션을 개발하였는데, 이 덕분에 응답자의 가족 관계를 보다 쉽게 파악할 수 있게 되었다.

자료 제공은 온라인 맞춤형으로 제공되고 있는데 거의 7만 명의 개인 표본에서 얻어지는 약 7만 개의 변수들을 ‘data cart’에 담아 다운로드 하는 형태로 편리하게 제공된다. 자료 제공 범위는 응답자의 사생활 보호와 상충되지 않는 범위에서 등록된 연구자에게 공짜로 제공되는데 응답자의 거주지역 정보 등 보다 자세한 정보를 원하는 경우에는 ‘data security plan’을 작성한 후 심사를 거쳐 제공하게 된다. 예를 들어, 응답자의 지리 정보, 자동차 종류와 연식, medicare 관련 정보, 주거비용 정보, 보조금 정보 등은 일반적으로 제공되지 않고 심사를 거쳐 계약을 통해 얻게 된다.

### 3. 연구범위

PSID는 오랜 세월을 거치면서 설문지 내용이 조금씩 바뀌었다. 질문하는 방식이 바뀌기도 하고 새로운 항목이 추가되거나 없어질 수도 있는데 이에 대해 시계열적 에디팅을 다 하지는 못하고 주요 항목에 대한 기본적인 에디팅이 실시된다. PSID는 초기부터 소득과 가족의 인적 구성을 조사함으로써 경제적 안녕의 dynamics를 연구하는 것을 주안점으로 삼았다. 그 후 많은 설문 항목이 추가되었는데 예를 들어 흡연이나 음주 관련 내용은 1968년부터 조사되기 시작했지만 일반적인 건강 관련 설문은 1984년부터 조사되기 시작했고 키와 체중은 1986년, 건강 관련 지출내역은 1999년에 조사되기 시작했다.

#### 4. 인력과 조직

PSID 생산 조직은 다음과 같이 구성된다.

- a. Administration(2명)
- b. Data processing team(8명)
- c. Field operation team(4명)
- d. Application development team(2명)

Administration은 총괄 책임자(PI)와 이전 책임자 그리고 각 세부 업무 책임자 두 명 정도이고, Data processing team은 6~8명의 전담인원이 있어서 수집된 자료 처리와 관련된 일(에디팅 작업, 가중치 작업 등)을 맡는다. Field operation team은 현재 4명이 있으며 Data collection team과 협력하여 자료 수집과 관련된 여러 가지 일(설문지 관련, 응답 인센티브 관련 등)을 맡는다. Application development team은 자료 제공 및 editing 과 관련된 IT 관련 support를 맡는다. 대부분의 인력들은 PSID와 관련하여 전담하게 된다.

자료 수집 자체는 ISR에서 하는 것이 아니라 Survey Research Center 라는 미시건 대학 산하 부설 연구소에서 실시한다. 대략 100명의 interviewer를 고용하고 있다.

PSID는 연구재단(NSF)의 후원으로 시작되었고 아직까지 그곳이 주된 후원자이지만 그 외에도 National Institute on Aging, Eunice Kennedy Shriver Institute on Child Health and Human Development, Economic Research Service of the US Department of Agriculture, the Center on Philanthropy at Indiana University, the Assistant Secretary for Planning and Evaluation of the department of Health and Human Service, the Department of Housing and Urban Development 등에서 지원을 받고 있다.

## 제4절 영국 BHPS의 경험

### 1. 표본의 규모와 대표성

영국의 BHPS는 가구를 단위로 표집하여 그에 속한 16세 이상의 가구원을 개인 응답자로 추적 조사한 데이터이다. 1991년에 잉글랜드 지역에서 5,500여 가구, 1만 명의 가구원을 조사하는 것으로 시작하였다. 1991년에는 스코틀랜드와 웨일즈 지역, 2001년에는 노던 아일랜드 지역으로 모집단의 범위를 넓히는 추가 표집을 시행하였다. BHPS 조사를 수행하고 있던 에섹스대학교 사회경제연구소에 속한 연구진은 표본수를 대폭 확대하고 연구의 범위, 즉 설문내용을 개편하여 BHPS를 영국의 가구와 개인을 대표하는 패널자료로 만드는 새로운 조사 설계를 영국연구재단에 제출하여 2006년에 승인받았다. 이후 2년의 준비기간을 거쳐 2008년부터 ‘Understanding Society(US 또는 UKHLS; UK Household Longitudinal Study)’ 조사를 수행하게 되었다.

US는 기존 BHPS의 표본을 모두 포함하되 표본을 추가하여 4만 가구, 10만 명의 표본으로 새롭게 출발한다는 계획으로 시작되었다. 1차 조사는 2008년에서 2009년에 걸친 기간 동안 이루어졌으며, BHPS의 패널은 이 조사에 2010년부터 포함되었다.<sup>14)</sup> 이 시점에서 BHPS의 표본은 14,500명의 성인과 1,222명의 11~16세 청소년이 남아 있었다.

US의 표본은 다음과 같은 네 가지 요소로 구성되었다; ① 일반표본 26,000가구, ② 소수인종표본 4,000가구와 이들의 비교집단 표본, ③ 혁신패널 1,500가구, ④ 구BHPS 표본 8,400가구.

일반표본 26,000가구는 전체 영국 지역에서 주소지를 기준으로 ‘stratified, clustered, equal probability sample’로 설계되었다. 소수인종 표본은 소수인종이 모여 사는 지역에 가중치를 두어 주요 소수인종이 충

14) US의 신규 조사자에게 질문할 내용(직업력 등)은 기존 BHPS 응답자들에게는 다시 질문할 필요가 없는 것이었으므로 2차 조사부터 조사에 포함시키게 된 것임.

분히 표집되도록 한 후, 이들과 비슷한 특성이 있는 다른 인종을 매칭하여 설문을 설계하였다. 혁신패널은 대표성을 고려하여 확률 표집하지는 않았으며, 새로운 설문을 실험하는 용도로 사용되고 있는 패널이다. 구 BHPS 표본은 2010년 현재 표본에 남아 있던 BHPS 응답자들을 그대로 US에 포함시킨 것이다(Buck and McFall, 2011). 2차 조사에는 1차년도 조사대상의 85%가 유지되었다고 알려지고 있다.

여러 가지 종류의 데이터셋이 있고 이 중에 어떤 것을 결합하여 사용하느냐에 따라 가중치를 달리 주어야 하기 때문에 가중치 사용이 쉽지는 않다. 연구진은 여러 가지 가중치를 만들어놓고 사용자가 골라서 쓰게 하는데, 어떤 가중치를 사용할지 판단을 돕기 위하여 매뉴얼을 제공하고 있다.

## 2. 조사방법

US의 조사방법은 CAPI, CATI뿐 아니라 웹 기반의 온라인 조사인 CAWI까지 가능한 모든 도구를 이용하는 것이다. 응답자에 대한 사례비를 온라인 응답자에게는 30파운드, 면접 응답자에게는 10파운드로 차등적으로 지급한다.

## 3. 연구범위

US는 추가적인 연구주제를 설문에 포함시키면서 더욱 다양한 연구 분야에서 활용하는 계획으로 짜여졌다. 매년 반복적으로 질문하는 내용은 가구원의 기본적인 특성, 주거 특성, 소비지출, 건강상태와 장애, 교육수준과 기대, 경제활동상태와 일자리의 특성, 소득, 자녀양육, 생활 만족도, 정치적 성향 등이다. 면접시점과 그 다음 차수시점 사이에 일어났던 모든 히스토리를 조사하는 주제는 일자리, 출산, 혼인, 주거이동, 교육과 훈련, 건강진단이다.

가족배경과 어린 시절 생활수준, 학력, 결혼, 출산력은 최초 면접 시에 한 번 질문한다. 몇 년에 한 번씩 돌아가면서 질문하는 모듈(module)은

다음과 같은 내용을 포함하고 있다; 여가와 문화생활, 웰빙, 심리상태와 개인선호(신뢰, 위험회피성향 등), 건강 관련 행동(음주, 흡연, 다이어트, 운동), 친지와 사회적 네트워크, 가족관계, 이웃관계, 사회적 지지, 환경에 대한 태도와 행동, 정치행동, 근로환경과 일자리의 질, 시간사용, 자산 등.

소수인종 샘플과 이들의 비교집단으로 선정된 표본에 대해서는 추가적으로 언어, 주거 이동의 모든 과정, 고용에서의 차별 경험, 정체성, 종교생활 등을 질문한다(Buck & McFall, 2011).

이 외에도 US는 몇 가지 새로운 시도를 하고 있는데, 대표적인 것이 설문조사 결과에 행정데이터를 연결하여 설문에 포함시키지 않은 정보를 정확하게 얻고자 하는 것이다. 이를 위하여 미리 응답자에게 행정데이터를 연결하여 사용해도 좋다는 동의서를 받았으며, 관련 행정기관과의 협의도 마쳤다. 그러나 건강, 세금, 급여, 주택, 에너지 등의 내용을 포함하는 행정 DB를 관리하는 행정조직의 실질적인 협조가 원활하지 않아서 현재까지 교육 관련 DB의 일부만이 연결되어 있는 상태이다.

바이오마커(biomarker)도 그동안 1회 시행하였다. 기존 방법으로 조사하면 약 10년에 1회 정도 시행할 수 있는 예산이 확보되었을 뿐이므로 비용을 절감하여 좀 더 자주 시행할 수 있는 방법을 강구하고 있다.

#### 4. 인력과 조직

에섹스대학교 내 사회경제연구소에서 US 데이터 생산과 배포, 분석에 관련된 연구를 진행하는 인력은 약 45명이다.<sup>15)</sup> 총예산은 조사예산 약 7백만 파운드와 연구예산 약 2백만 파운드이다.

15) 이들이 모두 US 업무만 전담하는 것은 아님.

## 제5절 독일 SOEP의 경험

### 1. 표본의 규모와 대표성

독일의 사회경제패널조사(The German Socio-Economic Panel Study, SOEP)는 베를린에 소재한 DIW 연구소가 수행하는 조사로서, 1984년에 시작하여 2015년에 32차 조사를 완료하였다. 표본수는 2015년 시점에 13,000가구, 성인 25,000명과 아동 5,000명을 포함하고 있다.

그동안 여러 번에 걸쳐서 다양한 방식의 추가표집이 이루어졌지만, 그 내용에 있어서는 상당한 규칙성과 합리성이 있다(Box 1 참조). 1984년 초년도 조사에서는 당시 서독지역의 거주자(A표본)와 그 시점에 외국인으로 분류된 자(B표본)를 합쳐서 약 5,900가구, 응답자 12,000명을 조사하였다(표 5-4 참조). 1990년에는 동서독 통일을 기하여 1989년까지 동독지역이었던 곳에 거주하는 자를 대상으로 표본을 추출하여 기본표본에 포함시켰다(C표본). 1994/5년 조사에서는 그간의 표본이탈을 보충하기 위하여 추가표집을 실시하였다. 1994/5년에는 1984년 이후에 독일지역으로 들어온 이민자를 대상으로 표본을 추출하여 기본표본에 추가시켰다(D표본).

1998년에는 이탈표본을 보충하는 의미의 추가표집을 실시하였는데, 이들 중에서 일부는 다시 혁신표본(SOEP Innovation Sample)으로 옮겨지면서 기본표본에서는 제외된다(E표본). 한편, 2009년에 어떤 응답자 사례 방식이 응답률을 높이는 데 도움이 되는지 실험하기 위한 표집을 실시하였는데(I표본 또는 인센티브 표본), 이 표본은 모두 혁신 샘플로 옮겨지고 기본표본에서는 제외되었다.

2002년에 실시한 ‘고소득 표본(G표본)’은 가장 독특한 방식으로 추출되었다. 고소득자는 서베이에서 항상 과소 추정되는 경향이 있기 때문에 이에 대한 보충이 필요하다고 판단되었다. 이를 위해 기본적인 표집을 폭넓게 실시한 이후에 전화조사를 통해서 가구의 월소득을 확인하고 일정 수준(7,500마르크, 3,835유로) 이상이면 이 표본에 포함시켰다. 다음 연도의

## 〈Box 1〉 SOEP 표본추가 과정

1984	A	“Residents in the FRG”
1984	B	“Foreigners in the FRG”
1990	C	“German Residents in the GDR”
1994/5	D	“Immigrants”
1998	E	“Refreshment” parts of subsample E have been extracted into the SOEP Innovation Sample.
2000	F	“Refreshment”
2002	G	“High Income” independent from all other subsamples.
2006	H	“Refreshment”
2009	I	“Incentive sample” This sample remained in the main data distribution for its first two waves. With the data distribution of 2012, subsample I has been extracted into the SOEP Innovation Sample.
2011	J	“Refreshment sample”
2012	K L	“Refreshment sample”
2013	M	“migration sample”
2014		integration of FiD sample

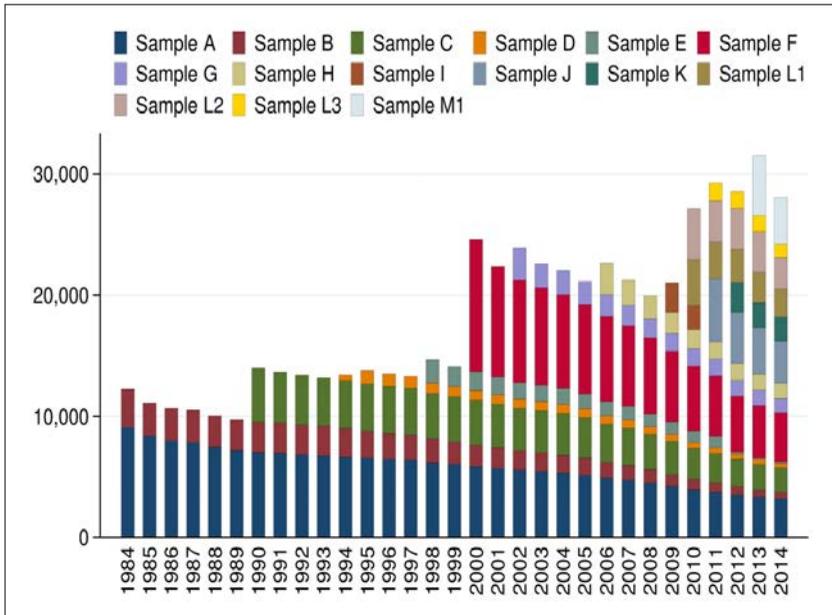
자료 : Gerstorf & Schupp(2014), p.43.

## 〈표 5-4〉 SOEP 표본의 종류별 표집시점의 규모

Sample	Year	Households (net)	Persons (gross)	Respondents (net)
A	1984	4,528	11,422	9,076
B	1984	1,393	4,830	3,169
C	1990	2,179	6,131	4,453
D1	1994	236	733	471
D1/D2	1995	541	1,668	1,078
E	1998	1,057	2,446	1,910
F	2000	6,043	14,510	10,880
G	2002	1,224	3,538	2,671
H	2006	1,506	3,407	2,616
I	2009	1,495	3,428	2,432
J	2011	3,136	6,873	5,161
L1	2010	2,074	7,939	3,770
L2	2010	2,500	9,063	4,227
L3	2011	924	3,645	1,487
K	2012	1,526	3,286	2,473
M1	2013	2,723	8,522	4,964

자료 : Gerstorf & Schupp(2014), p.43.

(그림 5-1) 전체 표본의 크기(개인 응답자수 기준)



자료 : Gerstorff & Schupp(2014), p.43.

조사부터는 월 4,500유로 이상이면 추적하는 기준을 세우고 실행하였다. 그 후의 표집은 전국적으로 이탈표본을 보충하는 표집(2000, 2006, 2011, 2012)과 그 이전의 외국인 조사 이후 독일에 들어온 이민자를 모집단으로 하는 표본을 추출하는 방식으로 운영하였다.

## 2. 조사방법

SOEP는 조사방법(Survey Mode)로 CAPI와 PAPI를 모두 사용하고 있으며 최근에는 CAWI를 도입하였다. CAWI 사용으로 약 30%의 실사비용을 절감한 것으로 연구진은 추산하고 있다. 그러나 조사방법 적용에 있어서 응답자에게 미리 웹서베이와 면접원에 의한 면접 중에서 선택하게 하였는데, CAWI로 응답하겠다고 하고 나서 하지 않아서 결국 면접원이 방문하여 CAPI로 진행한 경우가 상당수 발생하였다. 예를 들어 FiD 표본의 경우 80%가 웹서베이를 선택하였으나 결국은 45%만이 웹으로 조사

되었다. 특히 새롭게 표집된 응답자의 경우 처음부터 웹서베이로 진행하면 이런 문제가 적었으나, 과거부터 조사에 참여해오던 응답자를 웹서베이로 바꾸어 조사하는 것은 성공 가능성이 낮은 것으로 연구진은 판단하고 있다.

이 밖에도 웹서베이의 경우 설문 의 완성률(completion rate)이 낮은 문제가 있고, 소득과 같은 특정 설문에서는 무응답률이 상대적으로 더 높게 나타나는 경향이 있다. 그럼에도 불구하고 젊은 층의 경우 문제가 크지 않고 전체적으로는 비용절감 효과가 있는 것으로 판단되었다.

SOEP 조사는 과거에는 응답자에게 주는 사례로 꽃이나 복권을 선물했다. 그러다가 2009년에 어떤 사례가 응답률을 높이는 데 도움이 되는지 실험하기 위하여 응답자를 4개 집단으로 나누어, 각각 ① 복권, ② 가구당 5유로와 개인 응답자당 5유로, ③ 가구당 5유로와 개인 응답자당 10유로, ④ ①과 ② 중에서 선택을 제공하였다. 그 결과 현금이 더 효과적이라는 실험결과를 얻었다. 최근에는 과거부터 오랫동안 표본에 속해 있던 그룹에는 여전히 복권을 선물하고, 새롭게 표집된 표본의 경우에는 현금을 제공하는 방식으로 운영하고 있다. 가구 설문과 성인 개인 설문, 아동 설문에 모두 다 응답하면 가구당 평균 약 50유로의 사례금을 제공받게 되는 것으로 나타났다.

### 3. 연구범위

기본조사에 해당하는 설문은 인구학적 기본 특성과, 교육 관련 정보, 노동시장 활동과 일자리 변동, 소득과 사회보장, 주거, 건강, 선호와 가치, 삶의 만족도 등이다. 이 외에 주제별 모듈이 몇 년에 한 번씩 돌아가면서 조사되는데 1984년부터 2012년까지 조사된 모듈은 (BOX 2)의 내용과 같다.

SOEP 역시 영국의 US와 마찬가지로 혁신표본을 운영하고 있다. 독일의 경우 영국과 다른 점은 단순히 조사방법이나 새로운 설문을 테스트 하는 대상으로만 활용하는 것이 아니라, 데이터 사용자들이 요구하는 설문 내용(모듈)을 실시해주는 방식으로도 활용하고 있다는 점이다. 이 표본은

<Box 2> SOEP의 연도별 주제 모듈

<b>Topic Modules</b> → Replication every ~ 5-10 years			
1984	Employment biography	<b>2000</b>	<b>Further education or training</b>
1985	Marriage and family biography	2001	Social networks, working cond.
1986	Social origins (Bio), neighborh.	2002	<i>Social Security, assets</i>
1987	<i>Social security, early retirement</i>	2003	Ecology and envir. behavior
1988	<i>Assets</i>	<b>2004</b>	<b>Further education or training</b>
<b>1989</b>	<b>Further education or training</b>	2005	<i>Time use and preferences</i>
1990	<i>Time use and preferences</i>	2006	Family and social networks
1991	Family and social services	2007	<i>Social Security, assets</i>
1992	<i>Social security and poverty</i>	<b>2008</b>	<b>Further education or training</b>
<b>1993</b>	<b>Further education or training</b>	2009	Social networks, working cond.
1994	Neighborhood, values	2010	Consumption, savings, health
1995	<i>Time use and preferences</i>	2011	Family and social networks
1996	Family and social network	2012	<i>Social Security, assets</i>
1997	<i>Social security and poverty</i>		
1998	Ecology and envir. behavior		
1999	Expectations, Use of time		

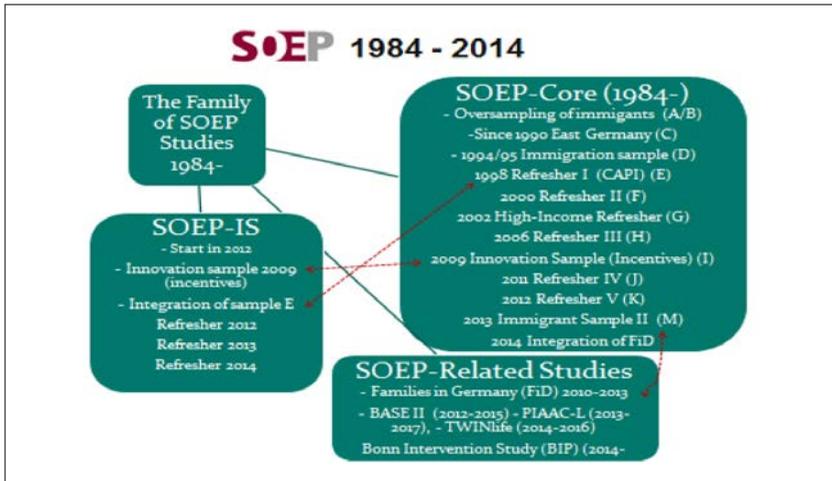
자료 : Gerstorff & Schupp(2014), p.43.

1998년 표본의 일부와 2009년 표본 전체를 합쳐서 3,716가구, 7,322명의 표본으로 시작하였으며, 2012년, 2013년, 2014년 각각 약간씩의 표본을 추가하였다. 2015년 현재는 약 3,500가구, 5,500명의 표본이 남아 있는 상태이다.

이 밖에도 SOEP의 독특한 점은 여러 가지 관련 프로젝트들을 결합하고 있다는 점이다. 라이프니츠협회로부터 제공되는 기본조사(core survey)를 위한 연구비 이외에 SOEP 연구진은 주요 정부부처로부터 추가적인 조사 연구 프로젝트를 수주하여 동시에 진행한다. 이때 어떤 프로젝트는 애초에 SOEP 기본조사 표본에 더하여 필요한 표본을 추가하는 방식으로 진행되는데, IAB와 SOEP 팀이 2013년부터 진행한 인구가동조사가 그러한 예이다. 이 조사과정에서 SOEP는 필요한 표본을 추가하여 기본조사 표본에 포함시켰고, IAB가 가지고 있는 고용보험 데이터베이스의 행정데이터를 결합하여 강력한 데이터를 만들었다.

‘독일 가족 연구(Families in Germany, FiD)’의 경우 2010년부터 2013년까지 SOEP 기본조사와는 분리된 연구로 진행되었으나, 이때 표집된

(그림 5-2) SOEP 표본과 관련 프로젝트의 관계



자료 : Gerstorff & Schupp(2014), p.43.

표본은 연구 프로젝트 종료 후 기본표본에 포함되었다. 이러한 관련 연구 사례는 매우 다양하다. 어떤 프로젝트를 추가로 수행할 것인지, 이때 표본은 어떻게 관리하며 기본조사와의 관계는 어떻게 할 것인지 등에 대한 의사결정은 SOEP의 연구위원회(Science Committee)에서 이루어진다.

#### 4. 인력과 조직

DIW의 SOEP 연구팀의 전담인력은 2014년 현재 45명이며, 조직은 크게 세 가지 업무영역으로 나누어 관리된다; ① 데이터 생산(Survey Methodology and Management) 파트에 11명, ② 데이터 관리(Data Operation) 파트에 13명, ③ 분석(Applied Panel Analysis and Knowledge Transfer) 파트에 11명. 여기에 소장과 파트장, 행정지원인력을 합치면 45명이 된다. 이 외에 RA나 대학원생이 36명 있다. 관련 프로젝트를 위해 다른 기관에 속한 연구자가 파견되어 공동으로 연구하는 경우도 있다.

이 조사의 연구비는 라이프니츠협회(LiebnizGemeinschaft)라고 하는 연구재단(Foundation of Scientific Research)이 제공한다.<sup>16)</sup> 기본조사를 위한 예산은 1년에 약 8.5백만 유로, 관련 프로젝트에 투입되는 예산은

1.5백만 유로 정도이다. 라이프니츠협회는 대학이 아닌 연구소에서 수행하는 인문사회 및 이공계 연구비를 총괄 지원하는 기구이다. 대부분의 연구는 처음에는 ‘프로젝트’ 개념으로 시작하지만, SOEP 조사의 경우처럼 안정성이 담보되어야 한다고 판단을 받은 경우는 7년에 한 번씩만 연구 예산심사를 받으면서 연구비를 지원받는다.

## 제6절 소 결

### 1. 주요 데이터 비교가 주는 시사점

이 절에서 검토한 주요 패널조사 현황을 <표 5-5>에 요약하여 제시하였다. 미국의 PSID는 거의 50년에 육박하는 역사를 자랑하며 최장기 패널의 자리를 지키고 있으며 안정감 있는 조사를 진행하고 있다. 데이터 배포 방식에서 가장 편리하고 세련된 방식으로 데이터 사용자 맞춤형 자료추출도구를 개발한 것이 눈에 띄는 장점이다.

영국의 US(구BHPS)는 여러 조사들 중에서 가장 급진적인 개혁을 한 사례이며, 예산투입 면에서도 압도적인 증액을 한 것이 눈에 띈다. 새롭게 구축한 패널 표본의 규모(4만 가구, 10만 명)가 방대하다는 점에서도 가장 두드러진다. 설문 내용도 매우 포괄적으로 개편하였다. 무엇보다 바이오마커와 행정자료 연계 등 매우 어렵지만 모든 패널조사가 바람직한 조사설계로 추구하는 내용을 담고 있다.

독일의 SOEP는 영국의 US와 맞먹는 정도의 많은 예산을 투입하면서 표본의 규모 자체는 조금 작은 편이다. 그럼에도 불구하고 SOEP의 활용도는 모든 패널조사 자료들 중에서 최고 수준이다. EU와 OECD 수준에서 운영하는 다양한 데이터베이스와 원자료 제공시스템(Luxemburg Income Studies)에서 독일을 대표하는 횡단면 및 종단면 자료로 활용되고 있다. SOEP는 몇 년에 한 번씩 필요에 따라 추가표집을 실시하기 때

16) 이 재단의 재원은 연방정부와 주정부로부터 나오는 것임.

문에 데이터의 관리와 대표성 확보 문제에 어려움을 겪을 수 있는 상황임에도 불구하고 표본추출 시점에서 추출확률을 명확히 함으로써 가장 간단하고 명료한 가중치 정보를 제공하고 있다. 또한 SOEP는 여러 가지 혁신적인 연구와 조사 설계들을 꾸준히 담아내고 있다. 주요 연구 프로젝트들을 SOEP와의 연계성 속에서 소화하고 그 결과로 표집된 표본을 메인 샘플에 추가하면서 기본연구와 관련연구 사이에 높은 시너지 효과를 내고 있다.

이상과 같이 세 국가의 대표적인 가구패널조사 자료의 역사와 현황을 살펴본 결과, 다음과 같은 시사점을 얻게 된다. 첫째, 국가의 사회·경제 정책 수립과 다양한 연구주제를 소화하기 위해서는 표본 규모의 확대가 필수적이다. 노동패널의 경우도 최소한 2만 명 이상의 개인 응답자를 조만

〈표 5-5〉 주요 패널데이터 비교

	BHPS(Understanding Society)	SOEP
시작연도	1991년(BHPS) + 2008 (US)	1984년
최근차수	18차(+7차)	32차(2015)
최초표본	8,000가구 20,000명 (16세 이상 개인 16,844명)	
최근표본	(US) 4만 가구, 10만 명	약 3만 명
표본추가	1999년 스코틀랜드와 웨일즈 지역 2001년 노던 아일랜드 지역 2008년 Understanding Society로 전면 개편	(본문참조)
조사주기	매년	매년
현금보상	온라인 응답자: 30파운드 면접 응답자: 10파운드	복권 또는 5유로
조사도구	CAPI/CATI/CAWI	CAPI/CATI/CAWI
조사기관	에섹스대학교 경제사회연구소	WDI(베를린 경제연구소)
실사기관	NetCen(National Center) 민간비영리	TNS
예산	900만 파운드 UK Economic & Social Research Council (+정부부처)	8.5million유로(core 조사) + 1.5million유로(관련연구) LiebnizGemeinschaft(결국 중앙정부와 베를린 주정부)

## 〈표 5-5〉의 계속

	PSID	KLIPS
시작연도	1968년	1998년
최초표본	4,802가구/18,230명	5,000가구/13,321명(도시지역)
최근표본	9,063가구/24,952명(2013년)	6,838가구/13,169명(2014년)
표본추가	1968년 1,872개의 저소득층 가구 +2,930개의 미국 대표 가구 1990년 라틴 샘플 2,043개 추가 1996년 라틴 샘플 탈락 1997년 이민자 500가구 추가 1997년 저소득층 약 2,200개 탈락	2009년 전국표본으로 확대하면서 1,415가구 추가
원표본 유지율	- /개인 59.9%(16차)	가구 69.4%/개인 57.8(16차)
최근차수	38차	17차
조사주기	격년(1997 이후)	매년
현금보상	\$70 ※약 1분에 \$1 지급을 기준으로 2차에 \$5씩 상향조정 ※평균응답시간: 82분	1가구당 4만 원 ※개인 사례비가 없으나, 2015년부터 신규 조사자에 대해 5,000원 도서상품권 지급
조사도구	CATI(1994년)	CAPI(2008년)
데이터 추출도구	웹 기반 'data center' ※원하는 차수, 원하는 변수만 차별적 으로 데이터 다운로드 가능	홈페이지 공개 ※데이터 병합 프로그램인 smart_ klips 시범 도입. 원데이터 모두 를 다운 받아야 함.
조사기관	Institute for Social Research at the University of Michigan	한국노동연구원
실사기관	대학에서 설문조사원을 고용	한국리서치

간 확보하고, 필요에 따라 지속적으로 확대해나가는 방안을 강구해야 한다.

둘째, 조사의 내용을 확대하고 다른 조사들에서 도입한 혁신적인 설계를 노동패널도 도입할 때가 되었다. 조사내용의 폭을 넓히는 것은 데이터의 활용도를 높이는 데 기여하고, 따라서 데이터의 가치를 높이게 된다. 현실적으로 빠른 시일 내에 추구해 볼 수 있는 혁신은 행정데이터의 결합이다. 해외 패널조사들이 모두 행정데이터와 연계를 추구하지만, 가장 성

공적인 사례이자 우리가 벤치마킹 할 만한 사례는 SOEP를 독일 연방고용청의 고용보험 행정데이터와 연계한 사례이다.

셋째, 노동패널은 조사방법(survey mode)을 CAPI로 전환한 이래 다른 시도는 없는 실정이다. 가장 조사비용이 비싼 직접면접 방식만을 고수할 것이 아니라 다른 방법을 병행하는 안을 검토하여야 한다. 다만, 조사방법의 변화가 응답률이나 응답의 내용에 영향을 주지 않도록 면밀한 사전 검토와 준비가 있어야 할 것이다.

넷째, 데이터의 활용도를 높이기 위하여 데이터의 구조를 사용자가 이해하기에 좀 더 쉬운 방식으로 안내하고, 가중치도 보다 간단하게 제공하는 것이 바람직할 것이다. 데이터 추출방식도 지금보다 편리성을 제고할 필요가 있다.

## 2. 더 많은 표본수 확보를 위한 고려사항

노동패널조사의 장기 플랜에서 가장 핵심적인 부분은 표본추가와 종단면 가중치 작성 등과 관련된 통계학적 부분과 패널 표본 관리 등과 관련된 실사적 부분을 생각해 볼 수 있을 것이다. 통계학적 부분은 표본의 모집단 대표성과 밀접한 관계를 가질 수 있겠지만 초기 표본이 모집단 대표성을 갖는 경우에는 결국 표본 마모로 인한 패널의 대표성 훼손을 보완하는 것이 실제적으로 중요한 문제가 될 것이다. 만약 초기 표본의 모집단 대표성이 의심되는 경우에는 초기 표본의 인구학적 분포와 그 당시의 모집단 인구학적 분포를 비교함으로써 그 대표성 여부를 판단할 수도 있을 것이다. 패널 표본의 모집단 대표성은 노동패널조사 자료의 학술적 신뢰성과 표본자료로부터 일반화(generalization)된 결론을 얻을 수 있으나 여부를 결정할 것이다. 실사적 부분 역시 응답자료의 질과 밀접한 관계를 가지므로 조사결과의 신뢰도에 중요한 영향을 미친다고 볼 수 있다.

### 가. 표본추가

종단면 조사에서 표본추가를 하는 이유는 표본이탈과 관련된 보완이

주된 목적이지만 때로는 특정 영역(또는 도메인)에 대한 보다 높은 관심이 생겨서 이러한 특정 도메인에 대한 표본을 과대 표집하여 영역 추정(domain estimation)에 도움이 되도록 하는 목적으로도 실시한다. PSID의 경우에도 1990년에 히스패닉 계열의 가구들을 추가하였는데 이러한 표본추가는 히스패닉 계열 가구에 대한 관심이 높아진 것을 반영한 것이라고 볼 수 있다.

KLIPS는 1998년 제1차 조사에서 5,000가구를 표본가구로 시작하였고 2012년 제15차 조사에서는 원표본 가운데 3,517가구를 조사하여 70.3%의 원표본 유지율을 기록하였다. 2009년 12차 조사에서는 표본의 전국 대표성 확보를 위해 약 1,500개 가구의 표본추가를 실시하였다. 즉 1998년 원표본은 모집단 대표성을 확보하지 못했지만 2009년 추가표본을 통해서 통합된 표본은 모집단 대표성을 확보했다고 볼 수 있다.

그렇지만 패널 마모 등으로 인하여 실제 얻어지는 표본은 대표성을 확보하지 못할 것이므로 고질적인 무응답 가구 등을 버린 후에 다시 대표성을 확보하는 형태로 향후 추가적인 표본 재설계를 고려해 볼 수 있을 것이다. 이러한 표본 재설계 시 고려할 내용은 다음과 같은 것을 생각해 볼 수 있다.

- ① 조사에 매우 비협조적인 표본가구(예를 들어 1회 응답 가구)를 표본에서 제거할 것인가? 그러한 경우 어떻게 보완을 할 것인가?
- ② 만약 표본을 추가한다면 어떤 가구들을 더 추가하는 것으로 할 것인가?
- ③ 표본가구가 모집단 대표성을 가지는지를 어떻게 체크 할 수 있을 것인가?

이러한 질문들에 대하여 하나의 대답을 하기보다는 가능한 대답들을 살펴보고 각각의 경우에서 어떤 결과를 가져올 수 있을지에 대하여 살펴보고자 한다. 먼저 첫 번째 질문의 경우, 조사에 매우 비협조적인 표본가구(예, 강력거절)를 계속 표본으로 남겨두는 경우 매년 방문에 따른 비용 증가 대비 응답률이 현저히 낮을 수 있고, 이는 전체적인 패널 유지율 계산에 나쁜 영향을 미치게 된다. 그러한 경우, 새로운 표본 구축 시에 이러한

강력 비협조 가구를 표본에서 제외하는 것을 고려할 수 있을 것이다. 이렇게 해서 탈락된 가구에 대해 동일 조사구 내에서 다른 가구로 substitution 하는 방법을 고려하거나 아니면 표본추가에서 새롭게 표본을 얻어내는 방법을 고려할 수도 있는데 전자의 경우 편리하긴 하나 substitution 가구를 정하는 방법을 편의대로 해서는 안 되고 이 역시 랜덤 추출 과정을 거쳐서 정해야 한다.

다음으로는 표본추가를 하는 방법을 고려할 수 있다. 표본추가를 할 수 있는 예산이 확보되면 현재 표본가구를 제외한 나머지 모집단에서 표본가구를 추가하는 방향을 고려할 수 있는데 여기서 중요한 것은 어떤 특성의 가구들을 과다 표집하는 것을 고려할 것인가 하는 점이다. 한 가지 방법은 현재 표본을 분석하고 그 현재 표본에서 과소 집계된 인구학적 변수들을 파악하여 그것을 보완하여 표본을 보완하는 방법이다. 예를 들어, 표본의 지역별, 연령대별, 성별 분포를 보고 통계청 자료에서 얻을 수 있는 모집단 지역별, 연령대별, 성별 분포와 비교한 후 어느 지역에서 어느 연령대가 더 부족한지를 계산하여 이를 보완하는 형태로 과대 표집 비율을 계산하는 형태이다. 이러한 아이디어는 실제로 2009년 표본추가 시에 사용되었다.

이와는 달리 어떤 계층에 대한 사회적 관심이 높아져서 그 계층(도메인)에 대한 표본가구를 과대 표집하는 방법을 고려할 수도 있다. 예를 들어 2015년 현재 비정규직 종사자 위주로 표본추가를 하고 싶다면 비정규직 코호트를 추가하는 형식으로 패널을 추가할 수 있다. 이러한 방법의 장점은 물론 관심 계층에 표본수가 많아지므로 원하는 분석을 할 수 있다는 것이겠지만 단점으로는 표본추가 관련 초기비용이 많아진다는 것이다. 이러한 식의 표본추가는 two-phase sampling을 이용하는 것인데 first-phase에서 원하는 표본보다 5배수 정도의 표본을 먼저 뽑아서 그 표본으로부터 정규직/비정규직 여부를 파악한 후에 second-phase에서는 1차에서 비정규직으로 판명된 가구원 가구만을 표본추가 대상 가구 모집단으로 간주하고 그걸로부터 샘플링을 하는 것이다. 따라서 이러한 비정규직 코호트 표본을 추출하기 위한 초기 실사비용이 많아지게 된다. 이렇게 해서 얻어진 코호트 표본은 비정규직 도메인 분석에서는 그냥 사용될 수 있

지만 기존 표본과 통합하여 사용하게 될 때에는 반드시 횡단면 가중치를 이용하여 분석해야 한다.

#### 나. 표본이탈에 대한 가중치 처리

패널 차수가 늘어감에 따라 표본 유지율이 줄어드는 것은 어쩔 수 없는 현상이다. 이러한 패널 마모(또는 탈락)를 보완하기 위해서 가중치를 사용하는데 가중치 작성은 좀 더 정교하게 설계될 필요가 있다. 현재 사용되고 있는 횡단면 가중치 보정 방법은 기본적으로 지역이나 성별, 연령대와 같은 인구학적 변수를 이용한 보정 방법으로 소득이나 교육수준 같은 경제학적 변수를 반영하지 않아서 패널 마모율이 이러한 경제학적 요인에 따라 다르게 나타날 경우 얻어지는 추정량의 편향을 줄여주지 못한다. 또한 1인 가구의 증가와 같은 가구적 특성이 반영되는 가중치를 사용해주는 것이 표본의 대표성을 높여 주는 데 매우 중요한 역할을 할 것으로 보인다. 따라서 이러한 경제학적 변수 및 가구적 변수까지 반영하는 새로운 가중치 방법론 개발이 중요한 다음 과제로 보인다.

가중치 작업은 종단면적 대표성과 횡단면적 대표성 두 가지 측면에서 생각해 볼 수 있는데 종단면적 대표성이라 함은 현재의 패널 표본이 추출 당시의 대표성을 얼마나 반영하는지를 고려하는 것이다. 즉, 제1차년도에는 소득이나 교육 관련 모든 정보에 대해 응답을 했으므로 1차년도를 기준으로 어떠한 가구가 더 많이 탈락했는지를 알아낼 수 있을 것이다. 이러한 정보를 반영하여 가중치를 작성하면 패널 마모로 인한 체계적인 편향을 어느 정도 줄일 수 있을 것으로 판단된다. 이러한 종단면적 대표성이 어느 정도 확보되면 추가적으로 횡단면적 대표성을 높이기 위해 횡단면 가중치를 더 보정하는 것을 생각해 볼 수 있을 것이다. 이러한 작업은 통계청 자료를 이용하여 벤치마킹을 하면 되는데 필자의 개인적인 생각으로는 패널 조사에서 지나치게 횡단면적 대표성을 강조할 필요는 없어 보이지만 종단면적 대표성(즉, 패널 마모로 인한 편향)은 반드시 보완해야 한다.

#### 다. 실사 측면에서의 고려사항

높은 패널 유지율을 위해서는 실사 측면에서 다양한 방법을 강구해야 한다. PSID의 사례를 참조하여 KLIPS의 패널 유지율 제고를 위해 고려할 사항은 다음과 같다.

1. 인센티브 시스템의 정비
2. 조사방법 관련
3. 설문지 관련
4. 테스트 관련

PSID를 주관하는 Institute for Social Research에는 통계학자 외에도 서베이 방법론을 연구하는 survey methodologist들이 있다. 이들이 주로 하는 연구 중 하나는 어떻게 questionnaire design을 할 것인가, 그리고 어떻게 조사하는 것이 응답을 높이는 효과가 있는가인데 우리나라는 이러한 분야에 대한 연구가 매우 부족한 실정이다. 인지 심리학자의 도움을 받아 설문지를 검토하고 질문하는 방식에 대해 더 깊이 있는 검토를 해볼 필요가 있어 보인다.

이러한 연구들은 테스트를 바탕으로 한다. 어떤 방식(예를 들어 문항을 A의 형태로 묻는 것이 더 응답을 높이는지 아니면 B의 형태로 묻는 것이 더 응답을 높이는지)에 대한 질문이 있으면 이를 실제 조사에서 randomized experiment의 형태로 실험을 하는 것이다. 그래서 어떤 방식에 따른 응답률이 유의한 차이가 있는지를 얻어진 데이터로 판단하고 이를 바탕으로 실사 방식을 계속 업그레이드 하는 것이다. 이런 식으로 실사 관련 노하우를 축적해 가되 그것에 대한 과학적 근거를 제시할 수 있어야 할 것이다.

지금까지 KLIPS의 향후 장기 계획에 도움이 될 만한 내용들을 정리해 보았다. 이를 위해 크게 표본추가와 가중치 보정에 대한 개론적인 내용들을 다루었고 아울러 실사 부분에서의 개선 방향 수립에 대한 제안도 정리하였다. 표본추가는 예산이 확보된 경우 기존 표본의 마모된 부분을 보완해주는 형태로 할 것인지 아니면 특정 도메인에 대한 코호트를 확보하는 형태로 할 것인지를 결정해야 한다. 그러한 결정이 나면 표본추가는 가능

하지만 후자의 경우 초기 비용이 더 많이 들게 된다. 가중치 부분은 특히 종단면 가중치에서 표본 마모를 보완할 때 초기연도의 표본가구 정보를 충분히 반영하는 방향으로 가중치를 작성하는 방법을 고려하는 것이 단순하게 인구학적 변수만을 고려하는 가중치 보정 방법보다 훨씬 더 효과적인 것으로 판단된다. 이와 관련하여 향후 무응답 분야의 전문가를 통한 추가 연구를 하는 것이 필요할 것으로 판단된다.

### 3. 데이터 생산기관 간 협력

데이터의 가치는 좋은 연구에 많이 활용될수록 높아지는 것이고, 그중에서도 중요한 것이 국가 간 비교연구에 얼마나 활용 가능한가의 문제이다. 이러한 사실을 모든 데이터 생산에 종사하는 연구자들이 잘 알고 있기 때문에 국가 간 비교 가능한 데이터를 만들기 위한 협력사업들이 꾸준히 진행되어왔다. 한국노동패널은 국제적인 가구패널조사 연구팀들의 네트워크에 참여하여 노하우를 공유하는 것이 바람직하다.

현재 국제적인 가구패널조사 연구팀들의 네트워크는 두 가지 방식으로 진행되고 있다. 하나는 데이터 자체를 비교 가능한 틀로 재구성하여 제공하는 데이터베이스를 구축하는 사업이다. 미국의 코넬대학교를 근거지로 하여 Cross National Equivalent Files(CNEF) 사업이 추진되고 있고, 여기에는 이 보고서에서 벤치마킹 한 세 국가의 데이터 이외에도 호주와 스위스의 데이터가 추가되어 있다. 우리나라의 노동패널도 이 사업에 참여하고 있으며, 최근 데이터를 업데이트 해서 업로드 할 계획이다.

다른 협력 네트워크는 2년에 한 번씩 각 나라의 가구패널조사 연구진들이 모여서 조사방법과 관련된 워크숍을 개최하는 방식으로 활발하게 진행되고 있다(패널조사방법론 워크숍: Panel Survey Method Workshop). 이 워크숍은 각국의 데이터 생산기관이 돌아가면서 개최하는데, 2016년에는 SOEP를 생산하는 독일의 DIW 연구소가 개최한다. 우리 노동패널팀도 한 차례 워크숍에 참석한 적이 있으나 향후 더욱 활발한 참여로 협력사업을 떠나가는 것이 바람직할 것이다.

마지막으로 중국의 China Family Panel Studies(CFPS)에 주목할 필요

가 있다. 이 데이터는 중국정부가 예산을 투입하고 북경대학에서 주관하여 생산하는 가구패널 데이터이다. 2010년에 첫 번째 조사가 시행되었는데 15,000가구, 3만 명의 개인 응답자 표본으로 출발하였다. 미국의 PSID 연구팀의 자문을 받으면서 설계하였기 때문에 국제적인 가구패널자료의 기준에 부합하게 조사되고 있는 것으로 보인다. 이 데이터는 아직 본격적으로 국제적인 협력 네트워크에 진입하고 있지 않으므로 개별적으로라도 협력사업을 추진할 만하다. 한편, 게이오대학에서 조사되고 있다고 알려진 일본 가구패널에 대해서는 국제적으로 알려진 것이 매우 적다. 국제 협력 네트워크에 등장하지 않고 있는 것은 물론이다. 이 데이터와의 협력 사업도 검토 대상이다.

### 〈제5장 부록〉 종단면 가중치 작성 방안

패널조사 자료의 통계학적 분석에서 가장 핵심적인 부분은 패널 마모로 인한 가중치 처리 방법론이라고 할 수 있을 것이다. 이러한 가중치는 종단면 가중치라고도 불리는데 종단면 분석을 하고자 할 때 사용하게 되는 가중치이다. 이러한 종단면 가중치를 작성할 때 인구학적 변수만을 사용하여 보정을 하는 경우가 종종 있는데 이러한 방법은 패널 표본 중 저소득층이 더 빨리 마모하는 경향이 있음을 고려할 때 소득 관련 체계적인 편향이 존재할 위험이 있다.

이러한 문제점을 해결하기 위해 성별, 연령대, 지역 외에도 학력과 소득 관련 변수를 종단면 가중치 작업에 고려하는 방법을 생각해 볼 수 있다. 패널 마모에 대한 가중치 보정에서 인구학적 변수뿐만 아니라 사회·경제적 변수 역시 고려하고자 하는 것이다. 이러한 사회·경제적 변수에 대한 벤치마킹은 1차년도 표본 분포를 기준으로 한다. 즉, 제1차년도 원표본의 해당 변수들에 대한 분포가 현재 연도 응답자 표본에서 유지될 수 있도록 가중치를 보정하는 것이다.

이를 위해서 분가 자료를 제외하고 먼저 가중치 작업을 실시하는데 가구 가중치의 경우 1차년도 원표본에서의 가구주 성별, 가구주 연령대, 지역별 분호 외에도 가구주 학력과 가구 소득을 구간으로 정해서 각각의 주변 분포(marginal distribution)를 계산한다.<sup>17)</sup>

그 후에 현재 연도 응답자 가구의 해당 변수에 대해 1차년도의 주변 분포와 동일한 표본 분포를 갖도록 가중치 보정 승수를 계산하는데 이와 관련한 방법으로는 raking ratio estimation 방법을 고려할 수 있다. Raking ratio 방법은 다차원 범주형 자료에서 주어진 주변 분포에 맞도록 가중치 보정을 해주는 전통적인 방법으로 많이 사용되고 있다.<sup>18)</sup> 또한 연령대 분포를 맞추어 줄 때에는 연도가 많이 경과되었으므로 이를 반영하여 1차년도 당시의 연령대로 변환하여 맞추어야 한다.

이렇게 해서 가중치 보정 승수가 결정되면 이 가중치 승수가 지나치게

17) 그 외에도 자료 분석 결과에 따라 더 많은 변수를 넣을 수 있음.

18) 그 외에도 필요에 따라 다른 calibration weighting 방법을 고려할 수 있음.

커지지 않도록 절삭(trimming) 방법을 적용하고 설계 가중치를 곱하여 최종 가중치를 결정한다. 필요한 경우 셀 병합이 이루어지기도 한다.

이렇게 해서 원자료 가구에 대해 가구 가중치가 결정되면 분가 가구의 가구 가중치를 계산해야 하는데 분가 가구는 모집단으로 유입된 가구를 대표하므로 앞에서 계산된 부모 가구의 가중치를 그대로 계승하여 유입 모집단을 대표하는 표본가구를 구성한 후 그 후 패널 마모되는 분량만큼 비슷한 가중치 조정을 사용하여 보정한다.

그다음으로는 개인 가중치 작업을 실시한다. 개인 가중치는 가구 가중치를 기본 가중치로 하되 그것을 다시 개인 변수에 대해 raking ratio 방법을 적용하여 구한다. 이때 사용되는 변수로는 성별, 연령대, 지역 및 학력과 가구 소득이 될수 있을 것이다. 벤치마킹 변수를 많이 사용하면 가중치의 분포가 더욱 극단적으로 될 수 있으므로 적절한 수준에서 변수를 선택해야 한다. 이와 관련한 보다 자세한 연구는 향후 과제로 남겨둔다.

## 참고문헌

- 김재광(2009), 『한국노동패널조사 표본 추가 연구용역 최종보고서』, 한국노동연구원.
- 박민규 · 김영원 · 변종석 · 송중호 · 임찬수 · 김사라(2013), 『한국노동패널가중치연구』, 한국노동연구원.
- 성명재 · 김현숙(2006), 『분배구조 개선을 위한 조세정책 방향』, 한국조세연구원.
- 윤윤규 · 홍민기 · 이상호 · 배기준 · 최효미 · 김기민 · 장숙량 · 조성일 · 천희란(2012), 『패널자료 품질개선 연구(Ⅱ)』, 한국노동연구원.
- 이규용 · 이지은 · 신선옥 · 이혜정 · 김기홍(2015), 『제16차(2013년도) 한국 가구와 개인의 경제활동』, 한국노동연구원.
- 이병희 · 홍민기 · 이현주 · 강신옥 · 장지연(2013), 『경제적 불평등과 노동시장 연구』, 한국노동연구원.
- 이상호(2012), 『패널조사의 비표본오차와 CAPI에 관한 이론적 검토』, 최효미 외(2012), 제2장, pp.4~33.
- 최효미 · 이상호 · 성재민 · 김기민 · 배기준(2012), 『패널자료의 품질 개선을 위한 조사기법연구-컴퓨터를 이용한 조사(CAPI) 방법-』, 한국노동연구원.
- 통계청(2015), 『경제활동인구조사 지침서』.
- 한국노동연구원(2015), 『제17차(2014)년도 한국 가구와 개인의 경제활동: 한국노동패널 기초분석보고서』.
- 홍민기(2015), 『소득 히핑(Heaping)과 양극화 지수』, 원고.
- \_\_\_\_\_(2015), 『최상위 소득 비중의 장기 추세(1958~2013)』, 경제발전연구.
- 홍민기 · 김재광 · 한치록 · 김기민(2014a), 『패널자료 품질개선 연구(Ⅲ)』, 한국노동연구원.

- 홍민기·최효미(2014), 『노동패널의 표본이탈과 불평등지표』, 조사연구.
- 홍민기·한치록·김재광·신동균·김기민·이고은(2014), 『패널자료 품질개선 연구(IV)』, 한국노동연구원.
- 황수경(2010), 『실업률 측정의 문제점과 보완적 실업지표 연구』, 『노동경제논집』 33(3), pp.89~127.
- 황수경·양정승(2002), 『직업력 자료의 구성과 활용가능성』, 한국노동패널 워킹페이퍼시리즈, 『한국노동패널연구』 2002-02.
- Bonnet, O., P. Bono, G. Chappelle, and E. Wasmer(2014), “Does Housing Capital Contribute to Inequality? A Comment on Thomas Piketty’s Capital in the 21st Century,” SciencesPo. Discussion Paper 2014-07.
- Buck, Nick and Stephanie McFall(2011), “Understanding Society: design overview,” *Longitudinal and Life Course Studies*, Volume 3, Issue 1, pp.5-17.
- Burkhauser, R. V. and T. M. Smeeding(2001), “The Role of Micro-level Panel Data in Policy Research,” *Journal of Applied Social Science Studies* 121(4), pp.469-500.
- Davies, P.(2004), “Is Evidence Based Government Possible?,” Jerry Lee Lecture presented at the 4th Annual Campbell Collaboration Colloquium, 19 February, Washington DC.
- Duncan, G. J., F. T. Juster, and J. N. Morgan(1987), “The role of panel studies in research on economic behaviour,” *Transportation Research* 21A (4/5), pp.249-263.
- Fields, G.(2008), “Income mobility,” Electronic version from Cornell University, ILR School, <http://digitalcommons.ilr.cornell.edu/articles/453>.
- Fields, S. and E. Ok(1996), “The Meaning and Measurement of Income Mobility,” *Journal of Economic Theory* 71, pp.349-377.
- Gerstorff, Sandra & Jurgen Schupp(2014), “SOEP Wave Report 2014,”

DIW SOEP.

- Hobcraft, John & Amenda Sacker(2011), “The Origins of Understanding Society,” *Longitudinal and Life Course Studies*, Volume 3, Issue 1, pp.1-4.
- Homburg, S.(2014), “Critical Remarks on Piketty’s ‘Capital in the Twenty-first Century’,” Discussion Paper 530, Institute of Public Economics, Leibniz University of Hannover.
- Jäckle, Annette(2008), “The Causes of Seam Effects in Panel Surveys,” ISER Working Paper Series.
- Knibbe, M.(2014), “The growth of capital: Piketty, Harrod-Domar, Solow and the long run development of the rate of interest,” *Real-world Economics Review* 69.
- Lynn, Peter(2009), *Methodology of Longitudinal Surveys*, WILEY.
- Malo, M. A. and F. Muñoz-Bullón(2003), “Employment status mobility from a lifecycle perspective: A sequence analysis of work-histories in the BHPS,” *Demographic Research* 9(7), pp.119-162.
- Wijsekere, G.(2009), “The Role of Household Panel Surveys in Evidence-based Policy,” Australasian Evaluation Society International Evaluation Conference, 31 August-4 September, 2009 Canberra Australian Capital Territory.
- Wilkins, R.(Eds.)(2013), *Families, Incomes and Jobs*, Volume 8: A Statistical Report on Wave 1 to 10 of the Household, Income and Labour Dynamics in Australia Survey, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

◆ 執筆陣

- 홍민기(한국노동연구원 연구위원)
- 강신욱(한국보건사회연구원 연구위원)
- 성재민(한국노동연구원 부연구위원)
- 이상호(한국고용정보원 부연구위원)
- 장지연(한국노동연구원 선임연구위원)
- 김재광(미국 아이오와 주립대학교 통계학과 교수)
- 이지은(한국노동연구원 전문위원)

노동패널자료 연구 (I)

- |         |  |
|---------|--|
| ▪ 발행연월일 | 2015년 12월 24일 인쇄<br>2015년 12월 30일 발행   |
| ▪ 발 행 인 | 방 하 남  |
| ▪ 발 행 처 | <b>한국노동연구원</b><br>☎ 30147 세종특별자치시 시청대로 370<br>세종국책연구단지 경제정책동<br>☎ 대표 (044) 287-6080 Fax (044) 287-6089 |
| ▪ 조판·인쇄 | 거목정보산업(주) (044) 863-6566   |
| ▪ 등록 일자 | 1988년 9월 13일   |
| ▪ 등록 번호 | 제13-155호   |

ISBN 979-11-260-0027-2

정가 6,000원