The Weighting Adjustment of Korea Welfare Panel Study

손창균[1], 류제복[2], 홍기학[3], 이기성[4]

요 약

시간의 흐름에 따라 사회 구성원들에 대한 행태 연구나 사회의 변화가 개인의 행동양식에 미치는 영향 등에 대한 조사에서는 어느 한 시점에서의 구성원들에 대한 횡단면 조사와는 다르게 다년간 지속적으로 조사개체를 추적조사 해야 하는 종단면 조사 또는 패널조사를 수행해야 한다.

패널조사는 횡단면 조사와는 달리 최초 표본이 시간이 지남에 따라 조사 대상 표본으로부터 탈락함으로서 발생하는 표본의 마모와 그에 따른 대표성 상실의 문제이다. 그러므로 이러한 표본의 대표성 상실 문제를 적절히 해결하기 위해 적용가능한 방법이 가중치 조정 방법이다. 횡단면 조사에서는 (1)추출가중치의 조정, (2)무응답 가중치 조정, (3)사후층화 가중치 조정과 같이 3단계의 가중치 조정 과정을 수행하지만, 패널 조사의 경우 이와 더불어 원 표본의 대표성을 유지하기 위해 종단면 가중치 (longitudinal weight)를 함께 고려해야 한다.

이러한 관점에서 본 연구에서는 다양한 패널형태에 따른 가중치 조정 방법에 대해 고찰하고, 향후 수행될 한국복지패널(Korea Welfare Panel Study: KWPS)의 가중치 산정에 관한 이론적 근거를 마련함과 동시에 현재 국내에서 수행되고 있는 패널조사의 가중치 조정방법과 비교하고자 한다.

주요용어: 패널조사, 횡단면 가중치, 종단면 가중치, 복합표본조사

1. 서 론

패널조사(Panel survey) 또는 종단면 조사(Longitudinal survey)는 개인 또는 그룹의 행태 연구나 사회의 변화가 개인의 행동양식에 미치는 영향 등에 대한 조사를 다년간 수행하는 조사이다. 당연히 횡단면적인 특성과 시계열적인 특성을 모두 가진 데이터로서 주로 사회, 경제, 교육학 등에서 많이 활용되는 조사 방법이다. 패널자료는 어느 한 시점에 조사된 횡

1) (122-705) 서울시 은평구 불광동 산42-14 한국보건사회연구원 부연구위원
2) (360-764) 충북 청주시 상당구 내덕동 36 청주대학교 바이오정보통계학과 교수
3) (520-715) 전남 나주시 대호동 동신대학교 컴퓨터공학과 교수
4) (565-701) 전북 완주군 삼례읍 후정리 490 우석대학교 e-정보공학과 교수

단면 자료와는 달리 시간의 흐름에 따라 개체들의 동적인 패턴을 연구할 수 있다는 특징을 가진다. 이러한 관점에서 패널 조사의 장점으로는 표본의 크기가 커짐에 따라 자유도가 증가함으로 추정량의 효율이 향상될 수 있으며, 설명변수들 간의 공선성(collinearity) 문제가 감소하며, 추정량의 편향감소 등이 있다. 따라서 개체간의 동적 연관성(dynamic relationship)에 관한 연구가 가능하며, 개체들 간의 이질성(heterogeneity)을 모형화 할 수 있다. 그러나 이러한 장점에도 불구하고 종단면 조사는 최초 표본이 시간이 지남에 따라 조사 대상 표본으로부터 탈락함으로서 발생하는 표본의 마모(attrition)와 그에 따른 대표성 상실의 문제이다. 그러므로 이러한 표본의 대표성 상실 문제를 적절히 해결하기 위해 적용가능한 방법이 가중치 조정 방법이다. 횡단면 조사에서는 (1)추출가중치의 조정, (2)무응답 가중치 조정, (3) 사후층화 가중치 조정과 같이 3단계의 가중치 조정 과정을 수행하지만, 패널 조사의 경우 이와 더불어 원 표본의 대표성을 유지하기 위해 종단면 가중치(longitudinal weight)를 함께 고려해야 한다.

이러한 관점에서 본 연구에서는 외국의 일부 패널 조사에서 적용하고 있는 가중치 조정 방법에 대해 고찰하고, 2006년 11월 현재 진행 중인 한국복지패널(Korea Welfare Panel Study:KWPS)에 대한 향후 분석을 위해 요구되는 가중치 산정에 관한 이론적인 근거를 마련함과 동시에 현재 국내에서 수행되고 있는 다른 여타의 패널조사에 대해 가중치 산정 방법과 비교하고자 한다.

2. 해외 패널의 가중치 조정 방법

2.1 미국의 PSID(Panel Study of Income Dynamics)

PSID는 초기 두개의 독립된 표본으로 구성된다. 하나는 생산가능인구에 대한 횡단면적 표본을 다단계 층화추출에 의해 추출한 표본이며, 다른 하나는 저소득층에 대한 표본이다. 횡단면 표본은 SRC(Survey Research Center)에 의해 추출되었고, 이 표본은 등확률로 추출된 표본으로 1968년에 2,930가구를 성공적으로 면접하였다. 저소득층에 대한 표본은 PSID가 SEO (Survey of Economic Opportunity)표본으로부터 추출한 1,872가구로 구성되며, 불균등확률 표본이다. 전자를 SRC표본이라 하고, 후자를 SEO표본이라 부른다.

초기에 5,000가구로 시작하여 현재 8,700가구에 이르고 있다. 매년 PSID를 수행하며, 조사의 초점은 경제상황과 인구학적인 상황, 특별히 소득원과 소득의 총액, 취업 가구원 구성 변동, 주거위치 등에 대한 사항이다. 그러나 내용이 폭넓기 때문에 사회학적, 심리학적 측도를 포함하고 있다. 1995년 현재 PSID는 약 28년간 개인들이 살아있는 한 50,000명 이상

의 개인들로부터 정보를 수집하였다. 표본은 1968년 이래로 매년 조사된 개인을 포함하며, 1990년에 추가된 남미계통(Hispanic)의 가구주 2,000명의 대표표본을 포함하며, 원 표본 가구원에 의해 형성된 가구원을 포함한다.

1) PSID 표본 추출과 초기 가중치의 계산

가중치 계산을 위해 먼저 PSID의 표본추출과정을 이해할 필요가 있다. 1968년 당시 가족 표본은 다음 두 가지로 구성된다. (1) 공통적으로 미국에 거주하는 횡단면 표본과 (2) OEO(Office of Economic Opportunity)에 대해 통계국(Census Bureau)에 의해 1967년에 조사된 가족들의 부차표본이다.

1969년과 1970년에 표본은 전년도에 계속적으로 조사된 가족에 살고 있는 모든 패널 구성원으로 구성된다. 따라서 전년도 웨이브에 응답하지 않은 구성원에 대해서는 2차년도 조사를 수행하지 않았다. 거처의 횡단면 표본을 SRC(Survey Research Center)의 마스터 프레임으로부터 상주인 전체 추출률로 추출하였다. 마스터 표본은 설계에 유연하기 때문에 1개 이상의 거처를 표본으로 추출하거나 서로 다른 필요성을 가진 조사에 대해 적절한 시점에서 사용가능하다. 이 조사에서 3,000개의 조사가 이루어질 수 있도록 설계되었다. 1968년의 센서스 표본은 재면접과 같은 형태인데, 왜냐하면 이들 가족들이 통계국에서 전년도 조사가 이루어졌기 때문이다. 8가지의 기본적인 추출률을 가진 확률 표본추출이지만, 통계국에 의해 조사된 가족들 중 소득이 $2,000+N($1,000) 이하인 가족에 대해서만 조사되었다. 여기서 N은 가족인원수를 나타낸다. $2,000+N($1,000)값은 1967년에 사용된 연방의 빈곤선(federal poverty line)의 2배와 거의 같다. 이 값 이상인 소득을 갖는 가족은 제외하였으며, 북동부, 북중부, 서부의 3개 지역에 있는 SMSA(Standard Metropolitan Statistical Areas)외부 지역으로서 빈곤가족들에서 제외하였다.

2) 1968년 가중치의 계산

각각의 이들 표본은 1968년 조사에서는 무응답이었다. 왜냐하면 재면접 표본들이 인구센서스 조사에서부터 무응답자들이었기 때문이다. 즉, 이들은 통계국에 의해 조사된 응답자 이름과 주소를 OEO에게 공개하는데 대한 서명을 거부하였다. 또한 OEO로 부터 SRC 에게 일부 표본 주소를 전달하는데 실패하였다.

1968년 가중치를 결정하기위해 다음과 같은 3가지 확률을 계산하였다. (1) SRC 횡단면 표본에서 얻은 확률 (2) 재면접 표본으로부터 얻은 확률 (3) 결합된 표본에서 얻은 확률 이다.(횡단면과 재 면접 표본들을 결합하였을 때, 전체 비추정(overall ratio estimation) 방법은 사용하지 않았는데, 이는 모집단 총합에 관한 정보를 가지고 있지 않았기 때문이다.)
다음은 앞에서 언급한 3가지 추출확률에 대해 살펴보기로 한다.

<표1-1> 횡단면 표본과 재면접 표본의 가중치 계산에 사용된 응답률

지역과 SMSA분류	횡단면 표본		재면접 표본	
	적격응답자 수	응답률	적격 응답자수	응답률
북동부지역				
Self-representing Area	491		444	63%
중심도시	221	61%	330	
도외지역	270	65%	114	
NonSelf-representing Area	394		8	88%
SMSA	235	72%	8	
Non-SMSA	159	84%	*	
북중부지역				
Self-representing Area	308		323	70%
중심도시	134	60%	287	
도외지역	174	80%	36	
NonSelf-representing Area	814		94	67%
SMSA	337	80%	94	
Non-SMSA	477	83%	*	
남부지역				
Self-representing Area	85		291	68%
중심도시	42	83%	260	
도외지역	43	81%	31	
NonSelf-representing Area	1009		927	
SMSA	491	76%	635	79%
Non-SMSA	518	87%	292	85%
서부지역				
Self-representing Area	208		332	64%
중심도시	80	68%	229	
도외지역	128	84%	103	
NonSelf-representing Area	414		127	65%
SMSA	258	79%	127	
Non-SMSA	156	74%	*	

(가) SRC 횡단면 표본에서 얻은 확률

횡단면 표본은 표본추출단시의 미국인 전체에 대해 상수인 비율(0.66/10,080)에 따라 추출되었다. 이때, 응답률은 상수가 아니다. 지리적 위치에 따라, SRC 자체-대표성(self-representing)과 비대표지역에 따라, 자체-대표지역의 중심지역과 도외지 지역에 따라, 그리고 비자체-대표지역(nonself-representing area)에서 SMSA와 non-SMSA에 따라 응답률은 다양하다. 전체적으로 16가지의 응답률을 고려할 수 있다. 횡단면 표본에서 얻은 응답확률은 "**초기 추출률(initial selection rate)× 응답률(response rate)**" 으로서 (0.66/10,080)×(응답률)이다. 예를 들어 뉴욕의 맨하탄에서의 면접확률은 (0.66/10,080)×(61/100), 또는 1/25,037 이 된다.

(나) 재면접 표본으로부터 얻은 확률

통계국에 의해 원 표본을 추출하기 위해 사용된 8종류의 기본추출률이 있다. 357개의 PSU가 2가지 서로 다른 추출률을 사용하였다. 초기추출에 따라 통계국에 의해 추출된 어떤 집락에 있는 가구들은 집락내에서는 동일한 추출률을 유지하지만 PSU내의 서로 다른 추출률의 수는 증가한다.

표본으로 선택된 PSU내에서 이름과 주소를 접수받은 표본가족에 대해 재면접이 실시되었다. 접수율의 차이가 매우 다양하기 때문에 PSU에 따른 표본가구의 주소의 접수율에 대한 조정이 필요하며 이러한 조정작업은 백인과 유색인종 가족에 대해 수행되었다.

재면접에 대한 무응답률은 자체-대표지역과 그 외 지역에 따라 4개 지역에 대해 수행되었다.

재면접 표본으로부터 얻은 확률은 다음의 (2.1)과 같이 계산된다.

센서스 표본에 대한 초기추출률×센서스 부차추출률×SRC부차추출률×접수율×응답률 (2.1)

예를 들어 뉴욕, 맨하탄에서 층1에 있는 백인 가족의 재 면접확률은 다음과 같이 계산된다.

(1/3,158)×(1/1)×(1/1)×(20/100)×(63/100)=1/25,063

(다) 결합된 표본에서 얻은 확률

결합된 표본은 다음과 같은 세 부분으로 고려할 수 있다.

㉠ 통계국으로부터 얻은 재 면접 표본

㉡ 남부의 SMSA와 non-SMSA로 부터 얻은 횡단면 표본에 있는 빈곤 가족

㉢ 횡단면 표본의 나머지 부분

세부분 중 처음 두 부분은 동일한 모집단으로부터 2개의 독립된 표본을 추출한 것이므로 어떤 가족은 표본1 또는 표본2 또는 두 부분에서 모두 추출될 수 있다. 따라서 결합된 부분에서 면접확률은 다음과 같이 계산할 수 있다.

재 표본에서 면접확률 + 횡단면표본에서 면접확률 - 두 확률의 곱 (2.2)

예를 들어 맨하탄에서의 추출확률은 (1/25037)+(1/25063)-(1/25037×1/25063)=(1/12525)가 된다. 따라서 맨하탄의 가중치는 이 확률의 역수로서 12,525가 된다.

2.2 미국의 SIPP(Survey of Income and Program Participation)

1970년대 후반에 Department of Health, Education, and Welfare(HEW)가 처음으로

Income Survey Development Program(ISDP)을 시작하였다. ISDP의 내용과 과정의 발전으로 HEW는 조사표의 길이, 조사기간의 길이, 조사 자료의 연계프로그램 등에 관심을 가지게 되었다. 1979년 ISDP 패널은 종단면조사였으며, 응답자는 그들의 소득과 노동력 참여, 다른 특징에 관해 응답하도록 하였다. 이때 응답자들은 그들과 그 외 친인척의 정보를 제공하기위해 매 3개월마다 접촉했다. 여기서 3개월 간격은 면접에 대한 조사주기(reference period) 이다.

SIPP의 주된 목적은 미국의 개인 및 가정의 소득과 분배 정책에 관한 정확하고 포괄적인 소득과 분배정책의 기본적인 결정에 관한 정보를 제공하는데 있다. SIPP는 세금, 자산, 부채, 정부의 지원 프로그램에 참여 등에 관한 데이터를 수집한다. SIPP 데이터는 연방정부, 주정부, 지방정부의 프로그램의 효율성을 평가하는 기초 자료가 된다. SIPP는 1996년 재설계되어 중요한 변화를 주었다. 우선, 1996년 패널은 4년의 간격과 12개 웨이브로 구성된다. 이러한 재설계는 초기 SIPP의 중복(overlapping)패널 구조를 버리고, 표본규모는 증가시켰다. 1996년 초기 표본규모는 40,188개 가구였다. 각각의 SIPP패널에 있는 성인들은 민간인인 미국 모집단에서 가구들의 대표표본으로부터 뽑힌 것이다. SIPP 표본으로 뽑힌 사람들은 패널 생애동안 매4개월에 한번씩 조사에 응하게 된다. 만일 15세 이상인 원표본구성원이 원래의 주소로부터 다른 주소로 이사를 갔다면, 새로운 주소에서 그들을 조사한다. 조사대상 표본은 원 표본 구성원과 함께 거주하는 어린이들 포함한다. 만일 초기 조사 후 이전 조사에 없었던 사람들이 응답자의 가구의 일부가 되었다면 새로운 사람들도 최초 조사한 응답자와 계속 살고 있는 동안 조사한다.

1) SIPP 표본 추출과정

SIPP는 복합표본설계를 사용하고 있으며, 통계국에서 2단계 표본설계에 의해 SIPP 표본을 추출한다. 1단계에서는 PSU를 선택하는 것이고, 2단계에서는 표본 PSU에서 주소를 추출한다. 통계국 면접자는 선택된 주소에서 표본구성원으로 식별하는 과정을 밟는다.

표본 PSU의 추출을 위한 프레임은 미국의 군(county)지역과 독립 시(city)의 리스트로 구성된다. 이와 함께 모집단 수(counts)와 가장최근의 모집단에 대한 센서스 결과로부터 얻은 기타자료를 함께 사용한다. 군(county)는 PSU의 형태를 갖도록 인근 군들로 그룹을 만들어 PSU로 하거나, 하나의 군 자체가 PSU가 된다.

PSU의 구성과정은 적은 PSU는 인근의 비슷한 군들로 그룹을 이루어 하나의 PSU를 이루게 되는 데 이를 nonself-representing(NSR) PSU 이라 하며 층을 구성한다 (남부, 북동부, 중서부, 서부). 이 과정에서 인구학적인 변수나 사회경제학적인 변수들이 최적의 그룹화를 위해 사용된다. NSR PSU들 중 하나의 표본을 층에서 모든 PSU를 대표하도록 각 층에서 추출한다. 특정 기준 이상 큰 규모의 모든 PSU들은 표본에 포함되며, 이를

self-representing(SR) PSU라 한다. 통계국에 의해 유지되는 5개의 분리된 중복되지 않는 추출틀로부터 주소를 추출한다. 이들은 하나의 단위를 구성하며, 주소 조사구 틀(Address enumeration districts)라 한다. 즉, ①지역프레임, ②그룹쿼터프레임(special places 프레임), ③가구단위-포함 프레임, ④ 포함률-개선 프레임, ⑤ 신축건물(new-construction)프레임 등이다. 이때, ①, ②, ③번 프레임은 센서스에 기초한 프레임으로서 가장 최근의 센서스로부터 나온 데이터들이며, 적어도 96%이상의 주소가 완전하게 수록된 자료이다. 그룹쿼터 프레임은 가구단위에 포함되지 않는 하숙집, 호텔방, 시설 등의 나머지 거처들을 포함한다. 3개의 프레임이 SIPP의 표본 주소의 거의 90%를 제공한다. 포함률-개선 프레임은 센서스 당시에는 발견되지 않았지만, 사후 조사로부터 발견된 주소(address)를 포함한 프레임을 말한다. 이 프레임으로부터 추출되는 주소는 매우 적은 수이다. 신축건물 프레임은 새롭게 신축된 건물들의 프레임을 제공한다. 각 표본PSU에서 추출틀에 있는 주소들은 집락으로 그룹화 되어 있으며, 이 집락들을 표본으로 추출한 다음, 추출된 주소 집락들이 조사에 사용된다.

2) 가중치 계산

SIPP의 가중치는 기본적으로 다음과 같은 항목으로 구성된다.

① 표본단위의 추출확률을 나타내는 기본가중치(Base Weight)
② 집락내에서 부차추출에 대한 조정 가중치
③ 이사자를 위한 가중치(2차 웨이브 이후)
④ 무응답자들에 대한 무응답 조정 가중치
⑤ 기지의 모집단 총계로부터 수정을 위한 사후층화 가중치

1차 웨이브에서는 월 당 각 표본 개체(person)에 대해 다음과 같이 4개의 요소로 가중치가 구성된다.

(가) 2차 웨이브 이후의 기본 가중치

참조인 이거나 현재 웨이브에서 그룹쿼터에 있는 각각의 원 표본 개체에 대해 2단계보정에 실시되기 전 이전 웨이브로부터의 가중치이다.

(나) 2차 웨이브 이후의 이주자 조정 가중치

원 표본에는 포함도지 않았지만, SIPP의 1차 웨이브에는 포함되고 1차 웨이브 이후 표본가구에 속한 사람에 대한 보상을 위한 조정 가중치이다. 원 표본에는 포함되지 않았지만, 1차 웨이브에 포함된 성인을 포함한 가구단위에 포함된 개인에 대해서는 가중치가 감소한다. 즉, 2명이 살고 있는 원 표본가구에 세 번째로 1명이 이주하였다면 3명의 성인은 원 표본개인들의 초기 가중치에 2/3를 곱한 가중치를 부여받게 된다.

<표 2-1> 1차 웨이브 가중치

가중치 종류	내 용
1차 웨이브 기본가중치 : BW	표본개체의 추출확률의 역수로 계산된다.
중복-통제 인자 (Duplication-control factor) :DCF	집락의 부차추출에 대한 조정인자이다. 집락들이 종종 기대한 것 보다 크다고 판명될 때 현장에서 부차추출 할 때 이를 조정하기 위한 가중치이다.(1~4까지의 자연수 값)
1차 웨이브 무응답 조정 가중치 : NAF	조정계급 내에서 무응답가구의 서로 다른 비율을 보상하기 위한 값이다. 500개 이상의 무응답조정 계급을 정의하고 있다. $NAF_c = \frac{[(BW의\ 합)*(셀\ c에\ 있는\ 모든\ 표본가구에\ 대한\ DCF)]}{[(BW의\ 합)*(셀\ c에\ 있는\ 모든\ 면접가구에\ 대한\ DCF)]}$
1차웨이브 2단계 보정(calibration) : SSCA	모집단 총계에 대한 월별 추정치와 일치시키기 위해 표본 추정치를 조정하기 위한 과정이다. 다양한 변수를 사용하여 보정을 수행한다. 보다 확실하게 일치성을 보장하기 위해 raking과정을 거친다. 이 조정 작업은 순환그룹에 대해 실시하며, 월별로 모집단 총계의 1/4을 각 그룹에 할당한다. 2단계 전까지의 가중치= BW×DCF×NCF 최종 가중치= BW×DCF×NCF×SSCA

(다) 2차 웨이브 이후의 무응답 조정 가중치

2차 웨이브와 그 이후에 대한 무응답 조정은 초기 면접이후 가구무응답에 대한 보상으로 이용된다. 무응답 조정 층은 가장최근의 웨이브로부터 표본단위의 특성치나 개인의 인구학적 특성치를 이용하여 정의한다. 사용된 정보들은 가구특성치들을 구성한다. 기준개체의 특성치는 다른 가구 특성치를 정의하는데 이용한다.

(라) 2차 웨이브 이후의 2단계 보정

이 조정 작업을 위해 1차 웨이브 방법을 준용한다. 즉, 기준 월(month)별 적절한 모집단 총계를 사용한다.

가구, 가족, 그리고 자녀가족(sub-family)에 대한 기준 월의 최종가중치는 다음과 같이 개인가중치로부터 구한다.

·가구가중치는 가구의 기준이 되는 개체(세입자 또는 집주인)의 개인가중치이다.
·가족가중치는 가족의 기준이 되는 개인의 개인가중치이다.
·친족의 자녀가족에 대한 자녀가족 가중치는 기준이 되는 친족 자녀가족의 개인가중치이다.
·면접 월 최종 가구가중치는 면접 월에 가구의 기준이 되는 개인의 개인가중치이다.

(마) 최종 패널가중치와 역년(calendar year) 가중치

이들 두 가지의 가중치는 적합한 표본에 대해 완전한 가중치를 제공한다. 최종 패널가중

치(final panel weight)는 패널의 1차 웨이브에서 조사범위 내에서 매달 패널을 위해 데이터를 얻게 해주는 표본에 포함된 사람에 대해 계산된다. 그 외의 사람에게는 0의 값이 주어진다. 모든 월 동안 패널에서 자료를 제공한 대부분의 사람은 0이 아닌 가중값을 부여받는다.

또한 최종 역년 가중치(final calendar year weight)는 제한된 날짜에 면접을 수행한 사람에게 부여되는 가중치이며, 조사범위 내에서 역년의 매달 동안 데이터를 제공한 사람에게 부여된다. 그 외의 사람에게는 0의 값이 부여된다. 패널이 시작된 이후 원 표본개체의 가구에 결합한 사람에게는 2차역년 당시 데이터를 제공했다면 2차 역년에서 0이 아닌 역년 가중치를 부여한다. 이들 가중치는 다음의 3가지 가중치로 구성된다.

<표 2-2> 가중치의 종류

가중치 종류	내 용
초기가중치	패널이 시작될 때와 2단계 보정 조정전의 역년가중 기간에 횡단면 가중치로 구성된다.
무응답 조정인자	적합한 표본이 응답을 하지 않았을 때 초기가중치에서 고려된다.
2단계보정인자	이 가중치는 기준월과 면접 월 가중에서 사용한 것과 유사하다.

<표 2-3> SIPP, CPS 그리고 PSID의 비교

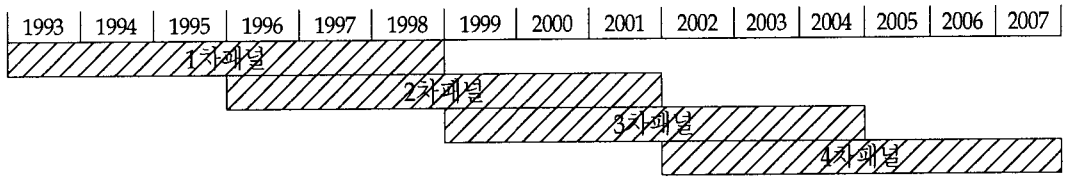
특징	SIPP	CPS (March Income Supplement)	PSID
표본규모와.설계	1996년 패널: 40,188가구 새로운 패널; 주기적; 매4개월마다 조사; 조사 월에 패널에 있는 각각의 원표본 성인	50,000가구;2년 주기에 걸쳐 8개월 동안 표본에 있는 각 가구;순환 그룹설계; 월별조사;1년에 한번 부가조사	9,000가구원(families);저소득가구원의 과대표집; 매년 조사 패널
주내에서 대표성을 가진 표본인가?	아니오	예	아니오
소득 자료 (Income data)	각각의 4개월 웨이브에서 약 70가지의 현금과 현물 원천에 대한 데이터와 대부분의 소득원천에 대한 월별 자료	약 35가지의 현금과 현물원천에 대한 이전 경상년도(calendar year)데이터	특정한 월에 받은 약 25가지의 현금과 현물원천에 대한 이전 경상연도 데이터
세금 자료 (Tax data)	연방, 주, 지역소득세; 근소세, 재산세를 결정하는 정보	없음	연방, 주, 지역소득세; 근소세, 재산세를 결정하는 정보
자산보유 자료 (Asset-holdings data)	1996년도부터 매년에 한번 부동산과 금유자산, 부채 현황; 보조프로그램과 연관관 자산에 대한 측정 빈도를 높임	없음 자가주택소유자 제외	정기적으로 주택의 가치와 모기지 빚에 관한 정보; 가끔 저축행태와 복지에 관한 정보
소비 자료 (Expenditure data)	1996년 이전과 1996년 한해에 한번 또는 전 달에 각 패널 당 적어도 한번 현금 지급한 의료비용, 주거비용(모기지, 임대, 설비), 보육비, 부양비용에 관한 정보.	없음	월세, 모기지 비용; 연간 설비비용; 주당 평균 식료품비용; 보육비

2.3 캐나다의 SLID(Survey of Labor and Income Dynamics)

1993년을 기준년도로 시작한 SLID는 개인들에 대한 종단면패널 조사이다. 이 패널의 목적은 개인들의 경제적 풍요(well-being)의 변화와 이러한 변화의 요인들, 특별히 인구학적인 측면과 가족의 특성, 노동시장 활동 등을 주요인자(key factors)로 측정하고자 하였다.

1) 패널설계

SLID 표본은 2개의 패널로 구성되는데, 각각은 6년간 지속된다. 1차 패널은 1993년 1월에 추출되었고, 1992년 12월 31일 현재 캐나다의 10개 지역을 포괄한다. 2차 패널은 1996년 1월에 추출되었으며, 1995년 12월 31일 현재 10개 지역의 모집단을 포괄한다. 두 표본 모두 인디언 보호지역에 거주하는 자, 군인 복무자, 6개월 이상 기관이나 시설에 거주하는 자는 모집단에서 제외하였다. 이와 같이 매 3년마다 패널의 구성이 변동하게 되는데, 추가적으로 이러한 규칙에 의한 패널 구성 형태를 도식화하면 <그림2-1>과 같이 표현된다. 즉, 1차 패널은 1993년부터 1998년까지이고, 2차 패널은 1996년부터 2001년까지 이다. 따라서 1996년부터 1998년까지는 2개의 패널이 동시에 존재하게 된다. 1999년부터는 1차 패널이 종료되고 2차와 3차 패널이 동시에 존재하게 된다. 2차 패널의 종료시점은 2001년 12월 31일 이며, 2002년 1월에 4차 패널이 시작된다.



<그림 2-1> SLID의 패널 구성 형태

이와 같이 연동 패널(rotation panel)을 사용하게 된 이유는 조사당시의 횡단면 표본의 대표성을 확보하며, 장기간의 패널유지로 인한 패널 마모효과를 감소시키며, 패널의 응답 부담을 경감시키고자 함이다. 각각의 개인은 1년에 총 2회의 면접을 수행한다. 1월에는 노동시장관련 내용을 질문하고, 5월에는 임금과 소득에 관한 설문을 수행한다. 응답 부담을 경감시키기 위해 응답자는 해당 소득관련 설문에 응답하지 않는 대신 5월에 통계청이 개인의 소득자료(Revenue Canada tax file)에 접근을 허가도록 한다.

각각의 패널은 캐나다 노동력조사(Labor Force Survey: LFS)의 1월과 2월에 조사된 표본으로부터 약 15,000가구의 표본을 추출한다. 이때, LFS는 층화다단계 추출방법을 사용하며,

마지막 추출단위는 거처(dwelling) 이다. 따라서 표본으로 선정된 거처안의 모든 가구원을 LFS 표본이 된다. LFS 표본으로 선정된 가구는 6개월 간 표본에 남아있게 되며, 매월 전체 표본의 1/6씩 새로운 표본으로 대체된다. SLID의 1차 패널은 LFS의 6개의 연동 그룹중 2개의 그룹(20,000가구)을 먼저 선정하고, 이들의 약 88%인 17,000가구가 SLID 표본으로 동의하였다. 이들 중 2,000가구를 제외한 15,000가구를 1차 패널로 구성하였다. SLID의 목적에 부합하기 위해 관심단위는 개인이다. 왜냐하면, 가구의 특성상 가구의 변동은 매우 더디게 일어나며, 따라서 종단면 분석을 위한 대상으로 고려하기가 매우 어렵기 때문이다. 횡단면적인 목적으로는 가구와 개인 모두를 분석단위로 고려하였다. SLID의 패널 표본으로 선택되면, 선택된 가구의 모든 구성원은 연령에 무관하여 패널의 종단면 표본의 일부로 남게 된다. 이들은 심지어 이사를 가거나, 사망하거나, 혹은 시설에 입소하거나, 군인으로 복무한다하더라도 패널지속기간인 총 6년간 종단면 표본의 일부로 고려된다.

SLID는 종단면 개인의 특성뿐만 아니라 가구의 특성과 연관되어 있다. 적어도 한 종단면 개인과 같이 살고 있는 모든 사람은 면접대상이 된다. 그러므로 주어진 해당 년도의 횡단면 표본은 참조년도의 12월 31일 기준으로 모든 종단면 개인과 그 당시 함께 살았던 모든 사람으로 구성된다. 만일 기준년도 당시 인디언 보호구역, 민간시설, 군대에 6개월 이상 거주한 것을 제외하고, 10개의 지역(province)중 한 지역에 살았던 사람이라면, 주어진 년도의 12월 31일 기준으로 조사대상에 포함된다. 종단면의 범위에서 보면 모든 횡단면 개인은 조사대상에 포함된다. 종단면 표본이 아닌 피면접자를 동거인(cohabitants)라 한다. 이러한 내용으로부터 각 년도의 횡단면과 종단면 가중치를 구성해야 한다. 종단면가중치(longitudinal weight)는 종단면 표본이 추출되었을 당시의 10개 provinces 모집단에 대한 대표성을 확보하기 위해 고려한 가중치이며, 횡단면 가중치(cross-sectional weight)는 주어진 기준년도의 12월 31일의 10개 province 모집단에 대한 대표성을 확보하기 위한 가중치이다. 1차 패널의 종단면 표본으로부터 구한 추정치는 1992년 12월 31일의 10개 province의 모집단에 대한 값이며, 2차 종단면 표본으로부터 구한 추정치는 1995년 12월 31일의 10개 province의 모집단에 대한 값이다. 1997년 기준년도의 횡단면 표본으로부터 구한 추정치는 1997년12월 31일 province 모집단에 대한 값이다. 1993년 1월에 39,745명의 개인이 1차 패널의 종단면 표본이었고, 1996년 1월에 43,547명이 2차패널의 종단면 표본이었다. 1997년 기준의 횡단면 표본은 81,090명의 개인이었고, 이중 70,372명이 종단면 표본이었다. 다음의 <표 2-4>는 시간의 흐름에 따른 종단면 표본의 변화를 나타낸 것이며, 이와 더불어 <표 2-5>는 1993년부터 1997년까지 기준년도의 횡단면 표본의 변동을 나타낸 것이다.

<표 2-4> 1차 패널과 2차 패널의 종단면 표본의 구성

(단위: 명)

패 널	내 용	1993	1994	1995	1996	1997
Panel 1	면접에서 접촉한 개인					
	기준년도의 12월31일 내	39,456	36,241	34,336	33,159	31,802
	10개 province 외부 거주	28	37	45	150	280
	시설거주	81	119	278	256	280
	사망	180	408	657	908	1,134
	중복, 오류	0	0	1	1	1
	면접에서 접촉 못한 개인					
	강력거절, 추적불가	0	2,940	4,428	5,271	6,248
	총 계	39,745	39,745	39,745	39,745	39,745
Panel 2	면접에서 접촉한 개인					
	기준년도의 12월31일 내	-	-	-	41,767	38,366
	10개 province 외부 거주	-	-	-	12	270
	시설거주	-	-	-	40	120
	사망	-	-	-	234	466
	중복, 오류	-	-	-	0	2
	면접에서 접촉 못한 개인					
	강력거절, 추적불가	-	-	-	1,380	4,323
	총 계	-	-	-	43,547	43,547

<표 2-5> 1차 패널과 2차 패널의 횡단면 표본의 구성

(단위: 명)

	1993	1994	1995	1996	1997
Panel 1					
종단면 개인	39,456	36,241	34,336	33,159	31,802
동거인	2,062	3,640	4,620	5,768	6,655
소 계	41,518	39,881	38,956	38,927	38,457
Panel 2					
종단면 개인	-	-	-	41,767	38,366
동거인	-	-	-	2,351	4,011
소 계	-	-	-	44,118	42,377
총 계	41,518	39,881	38,956	83,045	80,834

2) 패널 가중치 계산

SLID의 가중치는 우선 종단면 가중치와 횡단면 가중치로 구분할 수 있으며, 각각의 가중치를 보다 세분하여 살펴보는 것이 바람직 하지만, 개략적인 가중치 산정 방법만을 소개하기로 한다.

(가) 종단면 가중치

2차 패널표본에 대해 1993년 1월과 2월에 LFS에서 나온 2개의 연동그룹으로부터 응답가구는 총 20,486가구였으며, 이들을 표본으로 선택하였다. 이들 중 SLID 초기 면접에 응답한 가구가 17,659개이며 나머지 2,827개는 무응답 가구이다. 예산상의 이유로 15,000가구만을 SLID 초기표본으로 고려하였다. 결과적으로 15,000/17,659 =0.84인 확률로 응답가구를 추출한 것으로 간주할 수 있다. 무응답가구에 대해서는 0.06으로 무응답이 발생한 것으로 볼 수 있다. 이러한 결과로부터 14,832응답가구(39,745명)로 축소되었고, 174가구(410명)가 무응답 가구였다.

초기 종단면 가중치는 가구추출확률의 역수와 같고, 모든 종단면 가구원은 동일한 종단면 가중치를 가진다. 1차패널과 2차패널의 초기종단면 가중치는 다음과 같이 계산된다.

1차 패널의 초기 종단면 가중치 : $w_{int,p_1} = w_{LFS} \times 3 \times 1.19$ (2.3)

2차 패널의 초기 종단면 가중치 : $w_{int,p_2} = w_{LFS} \times 3$ (2.4)

여기서 w_{LFS}는 무응답이 조정된 LFS 가중치이며, 숫자 3은 LFS 연동그룹의 추출확률의 역수로서 (LFS의 6개 연동그룹중 2개를 선택함) 2/6의 역수이다. 또한 1.19는 1차 패널에서 초기면접에서 응답자들의 추출률(=1/0.84)의 역수이다.

이와 같은 초기가중치를 정의하고, 각 웨이브마다 종단면 개인을 그룹화하여 0인 가중치가 생기지 않도록 한다. 즉, 무응답자, 횡단면적으로 응답자의 범위에 속하는자, 횡단면적으로 범위밖의 개인 등으로 구분한다. 또한 소득과 근로설문 2가지 중 적어도 1가지에 응답한 가구는 응답가구로 취급한다. 만일 2가지 설문에 모든 가구원이 무응답한 경우에는 무응답가구로 고려한다. 또한 SLID는 응답가구에서 무응답 가구원은 응답자로 고려하였다. 응답가구 값으로 무응답 가구원 값을 대체(imputation)하였다. SLID에서 무응답 조정인자는 응답자로 고려된 모든 개인들의 가중치에 적용하였으며, 이때, 응답가구의 어린이와 조사범위 밖의 개인들은 제외하였다. 이러한 무응답가중치 조정은 사후층화로 수행하였다.

다음으로 무응답 조정을 위해 SLID는 개인의 무응답 조정승수(adjustment factors)를 응답자 그룹(RHG)의 응답률의 역수로 정의하였다. 적절한 무응답 승수를 계산하기위해 무응답을 정의하고, 이에 필요한 설명변수를 선택한 다음 로지스틱 회귀를 이용하여 적절한 응답자 그룹을 형성하여 무응답 승수를 구성하였다.

이러한 과정으로부터 다음과 같이 무응답이 조정된 가중치를 정의할 수 있다.

$$w_{ADJST}=\begin{cases} 0 & \text{, 무응답가구의 개인} \\ w_{int} & \text{, 어린이 또는 횡단면적인 제외 개인} \\ w_{int}/R_{RHG} & \text{, 응답가구의 개인} \end{cases} \quad (2.5)$$

여기서 w_{ADJST}는 무응답이 조정된 종단면 가중치이고, w_{int}는 초기 종단면 가중치로서 (2.3)과 (2.4)에서 정의되었다. 또한 R_{RHG}는 RHG에서 가중된 응답률이다.

SLID는 횡단면적인 소득의 추정치와 분산을 추정치에 영향을 주는 영향력 관찰치에 의한 극단가중치를 조정하였다. 영향력 관찰치에 대한 조정값은 0과 1사이의 값으로 계산되며, 2차패널의 2차웨이브에서 단지 2개의 개인의 영향력 관찰값이 존재하였다. 영향력 관찰치에 대한 조정가중치는 다음과 같다.

$$w_{infl} = w_{ADJST} \times \beta_{infl} \quad (2.6)$$

여기서 w_{infl}은 영향력 관찰치에 대한 가중치가 조정된 종단면 가중치이고, w_{ADJST}는 무응답이 조정된 종단면 가중치, β_{infl}는 영향력 관찰값에 대한 조정 승수이다.

다음으로 사후층화를 2개의 패널에 대해 독립적으로 수행하였다. 사후층화로부터 계산된 종단면 가중치는 다음과 같은 식에 의해 계산된다.

$$w_{post} = w_{infl} \times t_L / \sum_L w_{infl} \quad (2.7)$$

여기서 t_L은 사후층 L에 대한 총합이며 주로 캐나다 통계청의 인구과에서 작성된 값이다. w_{post}는 사후층화 조정된 종단면 가중치이다.

최종적으로 사후층화 가중치에 대해 가구와 개인의 정보 보호를 위해 일정 수준의 잡음(noise)가중치를 결합하여 외부 자료로 공개하고 있다. 따라서 사후층화 조정 가중치에 잡음가중치가 고려된 최종적인 종단면 가중치는 다음과 같다.

$$w_{noise}=\begin{cases} w_{post} \pm (e \times a) & \text{, 가중치가 같은 동일가구의 개인} \\ w_{post} & \text{, 그 외의 모든 경우} \end{cases} \quad (2.8)$$

여기서 e는 U(0,1)에서 발생시킨 확률잡음(random noise)이고, a는 종단면 잡음(longitudinal noise)이다.

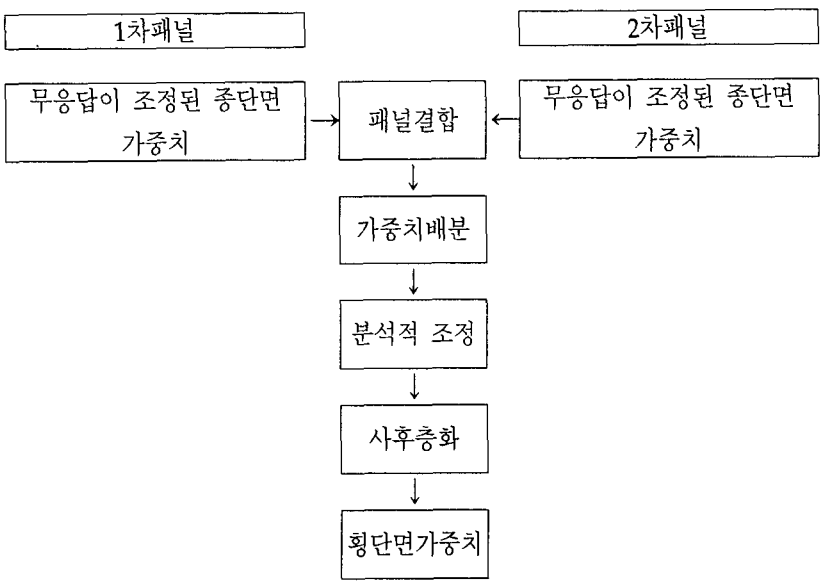
이러한 가중치 조정과정으로부터 10개 provines에 대한 최종적인 종단면 가중치는 다음과 같다.

<표 2-6> SLID의 1차 패널과 2차 패널의 최종 종단면 가중치의 중앙값

Provinces	1차 패널		2차패널	
	초기가중치	최종가중치	초기가중치	최종가중치
Newfoundland	207	250	242	285
Prince Edward Island	138	164	95	107
Nova Scotia	237	305	281	322
New Burnswick	279	321	254	296
Quebec	565	699	566	676
Ontario	563	769	558	718
Manitoba	221	309	327	409
Saskachewan	305	370	279	347
Alberta	783	870	693	967
British Columbia	661	848	811	1117

(나) 횡단면 가중치

SLID의 횡단면 가중치는 특정한 기준년도에서 추정치를 계산하기 위해 2개의 패널이 결합되는 형태의 가중치로 나타난다. 횡단면적 가중치의 목표모집단은 기준년도의 12월 31일 현재 10개 provincesp 거주하는 자로서 보호지역, 또는 시설 및 군대시설 거주자는 제외된다. 모든 종단면 개인과 그들과 동거하는 자는 횡단면 표본으로 간주한다. 초기 횡단면 가중치는 무응답이 조정된 종단면 가중치로서 각 시점에서 모집단을 대표한다. SLID의 횡단면 가중치 조정과정은 다음과 같은 절차로 수행된다.



< 그림 2-2> 횡단면 가중치 조정과정의 흐름

횡단면가중치 조정의 첫 번째 단계는 무응답이 조정된 종단면 가중치를 배분 승수(allocation factor)를 이용하여 2개 패널 표본을 결합한다. 이때 패널 분배 승수는 다음과 같이 계산된다.

$$p_1 = \frac{n_1}{n_1 + n_2 \frac{d_1}{d_2}} \quad , \quad p_2 = 1 - p_1 \tag{2.9}$$

여기서 p_1과 p_2는 각각 1차 패널과 2차 패널의 패널 배분승수이며, n_1과 n_2 는 각각 1차 패널과 2차 패널에서 16세 이상의 응답한 종단면 개인의 수이고, d_1과 d_2는 각각 조정된 가중치를 적용한 추정치의 분산과 단순임의 표본의 분산의 비로서 각각 1차 패널과 2차 패널의 설계효과를 나타낸다.

결과적으로 1차 패널과 2차 패널을 결합하여 조정된 횡단면 가중치는 다음과 같이 정의된다.

$$w_{com} = \begin{cases} p_1 w_{ADJST} & \text{, 1차 패널의 개인} \\ (1-p_1) w_{ADJST} & \text{, PAF의 적용에서 제외되지 않은 2차 패널의 개인} \\ w_{ADJST} & \text{, PAF의 적용에서 제외된 2차 패널의 개인} \end{cases} \tag{2.10}$$

1997년을 기준년도로 한 경우 패널배분승수(panel allocation factor)값들을 지역별로 구해보면 다음과 같다.

<표2.7> 1997년 기준년도에 대한 패널 배분승수

Provinces	d_1/d_2	n_1	n_2	p_1	$1-p_1$
Newfoundland	1.71	1,693	1,306	0.43120	0.56880
Prince Edward Island	1.92	568	883	0.25095	0.74905
Nova Scotia	4.02	1,788	1,986	0.18298	0.81702
New Burnswick	1.94	1,752	1,819	033176	0.66824
Quebec	2.75	4,908	5,788	0.23568	0.76432
Ontario	2.87	6,893	8,901	0.21249	0.78751
Manitoba	1.36	1,776	2,111	0.38219	0.61781
Saskachewan	2.21	1,915	1,872	0.31642	0.68358
Alberta	2.87	2,378	2,113	0.28168	0.71832
British Columbia	2.76	2,238	2,509	0.24425	0.75575

<그림 2-2>로부터 횡단면 가중치에 대해 패널 결합 후 중요한 작업중의 하나가 패널가중치의 배분(weight share)이다. 패널조사에서 표본 가구내에 이사를 오거나, 가구내에서 어린이가 새로 태어날 수 있다. 즉, 원표본에 포함되지 않았던 개인들이 신규표본(동거인)의 일

부로 들어오는 경우가 발생한다. 따라서 이들은 추출확률이 없기 때문에 새로운 가중치를 적용할 필요가 발생하게 된다. 이러한 작업이 가중치 배분작업이다.

SLID는 2개의 패널을 결합하는 방식이므로 1차 패널에 속한 가구의 개인이 2차 패널의 개인으로 포함될 가능성이 항상 존재하게 된다. 즉, 각기 다른 2개의 횡단면 표본으로 간주할 경우 동일한 개인이 2개의 횡단면 표본에 동시에 속하게 됨으로 독립적인 패널가중치를 산정하게 되면, 하나의 개인에 서로 다른 2개의 가중치를 고려해야 하는 문제가 발생하기 때문이다. 가중치 배분과정에서 2가지 서로 다른 가중치가 개개인에 대해 계산되는데, 하나는 표본가구의 개인들에게 부여되는 개인가중치(individual weight)와 표본가구의 모든 가구원들에게 동일하게 부여되는 통합가중치(integrated weight)이다.

개인가중치에 대한 분배과정은 우선 종단면 표본으로 선택된 이후 동거인이 이주해온 가구를 구별해야 한다. 해당 가구가 종단면 개인 전체로 구성된 경우에는 가중치 배분작업이 필요하지 않다. 해당 가구의 모든 구성원들에 대해 패널분배승수가 적용된 수정된 횡단면 가중치로 동일하게 횡단면가중치를 부과하면 된다.

다음으로 새로운 동거인이 생긴 가구에 대해서는 모든 종단면 가구구성원들의 무응답이 조정된 횡단면 가중치 w_{com}를 모두 더해야 한다. 그 다음으로 종단면 개인들과 초기에 함께 살았던 동거인의 수를 구한다. 초기에 적어도 1인 이상의 동거인이 가구에 대해 모든 구성원의 분배된 횡단면 가중치는 모든 종단면 개인들의 가중치인 w_{com}를 더하여 가구내의 종단면 개인들과 동거인들의 수를 더한 값으로 나눈 값이 된다. 이러한 절차로부터 개인횡단면가중치는 다음과 같다.

$$w_{share}=\begin{cases} w_{com} & \text{, 가구의 모든 개인이 종단면 이거나, 초기 동거인이 없는 가구의 종단면 개인.} \\ \dfrac{\sum_{h} w_{com}}{n_{L,h}+n_{IPh}} & \text{, 적어도 1인 이상의 동거인이 있고, 초기에 없었던 동거인이 있는 가구의 동거인.} \end{cases} \quad (2.11)$$

여기서 $n_{L,h}$는 가구 h에 있는 종단면 개인의 수이고, $n_{IP,h}$는 가구 h에 있는 초기 동거인의 수이다.

이와 함께 통합 배분된 횡단면가중치는 다음과 같다.

$$w_{share}=\frac{\sum_{h} w_{com}}{n_{L,h}+n_{IPh}} \quad (2.12)$$

또한 지역내에서의 이주에 따른 가중치 조정을 수행해야 한다. 이는 SLID의 종단면적 성

질과 직접적으로 관련이 있으며, 시간지 지남에 따라 표본에 속한 사람들이 현재의 지역에서 다른 지역으로 이주하게 된다. 이러한 효과를 가중치에 반영할 필요가 있다. 즉, 추출확률이 낮은 지역에서 추출확률이 높은 지역으로의 이주로 인한 가중치의 변동을 고려해야 분석과정에서의 왜곡을 피할 수 있기 때문이다. 이주효과를 방영한 조정된 가중치는 다음과 같다.

$$w_{mig} = w_{share} \times a_{mig} \tag{2.13}$$

여기서 a_{mig}는 이주효과에 대한 조정 승수로서 가중치의 평균이나, 중위 가중값, 가중치의 4분위 등을 사용할 수 있으며, SLID에서는 95분위수를 사용했다.

종단면 가중치와 마찬가지로 영향력 관찰치에 대한 가중치는 다음과 같이 이주효과를 조정한 가중치에 영향력 승수를 고려하여 조정할 수 있다.

$$w_{infl} = w_{mig} \times \beta_{infl} \tag{2.14}$$

또한 종단면가중치의 사후층화 조정과 같은 방법으로 식(2.14)의 가중치를 조정하여 사후층화가중치를 구할 수 있다. SLID에서는 지역×성별×연령 그룹으로 분할하여 횡단면 가중치에 대한 사후층화 조정을 수행하였다. 마지막으로 사후층화 조정된 가중치 w_{post}에 동일한 가중치를 갖는 동일한 가구의 개인들의 가중치로 w_{noise}를 다음과 같이 정의한다.

$$w_{noise} = \begin{cases} w_{post} \pm (e \times a) & \text{, 가중치가 같은 동일가구의 개인} \\ w_{post} & \text{, 고 외의 모든 경우} \end{cases} \tag{2.15}$$

3. 국내 패널조사의 가중치 조정 방법

3.1 국내 패널조사의 현황

외국 특히 미국, 캐나다 영국 등의 패널조사가 1960년대부터 수행되기 시작하였으며, 국내에서는 1993년부터 5년간 수행되어 오다가 중단된 대우재단의 대우패널이 국내 패널조사의 효시이다. 이를 기반으로 1998년 한국노동연구원의 노동패널이 시행된 이후 청년패널(2001), 저소득층 자활패널(2002), 청소년패널(2003), 복지패널(2004), 연금패널(2005), 고령자패널(2006), 교육고용패널(2005), 장애인패널(2007예정), 인구패널(예정), 소비자패널(2004), 여성패널(2007, 예정) 등으로 다양하고 많은 패널조사들이 수행되고 있거나 혹은 계획되고 있다.

3절에서는 국내에서 수행되고 있는 각종패널들 중 노동패널과 청소년패널의 가중치 조정 방법에 대해 살펴보고자 한다.

3.2 한국가구경제 패널조사(KHPS)

대우경제연구소에서 아시아권에서 첨으로 실시한 패널조사로서 1993년부터 매1년 주기로 수행된 패널조사이다.

1) 표본추출

제주도를 제외한 전국의 일반가구를 모집단으로 고려하고, 이때, 외국인가구, 특수시설(고아원, 양로원, 기도원 등)의 시설은 조사대상에서 제외하였다. 목표 표본수는 4,500가구로 설정하고, 조사 완료된 4,00가구는 1가구당 2,747가구를 대표하게 된다. 1차추출단위(PSU)는 시·군·구로 하고, 2차추출단위(SSU)는 읍·면·동, 3차추출단위(TSU)는 통·반·리로 하는 3단계집락추출법을 사용하였다. 각 단계에서는 계통추출법을 사용하여 집락을 추출하고, 최종적으로 표본집락에서는 임의추출에 의해 표본가구를 선정하였다. 6대 도시에서는 통·반·리 당 8가구를 표본으로 선정하였고, 기타지역의 통·반·리에서는 7가구를 표본가구로 추출하였다. 패널조사에 사용된 추출틀은 다음과 같이 5가지를 이용하였다.

○ 전국의 시·군·구 리스트 (가구수 포함, 1992년 행정구역총감)
○ 전국의 읍·면·동 리스트 (1990년 인구센서스, 분할된 동지역은 1992년행정구역총감)
○ 추출된 읍면동의 통반리 리스트(통반리의 가구수 포함, 1993년 4월-5월 조사)
○ 추출된 통반리별 주소리스트(통반리의 가구수 포함, 1993년 4월-5월 조사)
○ 추출된 주소의 가구주 리스트 (전화번호 포함, 1993년 6월-7월 조사)

통계청 조사구를 기반으로 한 추출틀은 인구이동이 심한 경우 센서스 조사시점과 패널조사시점의 차이로 인구수 또는 가구수의 변동이 반영되지 않기 때문에 본 조사의 추출틀을 조사시작 약1-3개월 전에 작성한 통반리 번지의 가구수 리스트를 추출틀로 이용하였다.

2) 가중치조정 방법

대우패널의 가중치는 1차년도 에만 작성되었고, 가중치의 형태는 각 추출단계별 추출확률의 역수로 계산되었다.

○ 추출확률 :

$$
\begin{aligned}
p_{ijkl} &= (a \times p_i)(b_i \times p_{ij})(c_{ij} \times p_{ijk})\frac{n_{ijk}}{N_{ijk}} \\
&= \left(a \times \frac{N_i}{N}\right)\left(b_i \times \frac{N_{ij}}{N_i}\right)\left(c_{ij} \times \frac{N_{ijk}}{N_{ij}}\right) \\
&= (ab_ic_{ij}) \times \frac{N_{ijk}}{N}
\end{aligned}
\tag{3.1}
$$

여기서 a 는 표본PSU의 수, b_i는 i번째 표본PUS 내의 표본SSU의 수, c_{ij}는 i번째 PSU, j번째 SSU 내의 표본 TSU의 수이다. n_{ijk}는 i번째 PSU, j번째 SSU, k번째 표본TSU내의 가구수이며, N_{ijk}는 i번째 PSU, j번째 SSU, k번째 TSU내의 총가구수이다. p_i는 i번째 PSU가 표본으로 추출될 확률, p_{ij}는 i번째 PSU의 j번째 SSU가 추출될 확률, p_{ijk}는 i번째 PSU, j번째 SSU내의 k번째 TSU가 추출될 확률이다.

○ 가중치 :

$$W_{ijkl} = constant \times 1/p_{ijkl} \tag{3.2}$$

3.3 한국 노동 패널조사(KLIPS)

한국 노동패널은 국내 패널조사 중 대표적인 조사로서 2006년 12월 현재 8차년도 패널조사가 이루어진 상태이다. 노동패널은 노동시장과 관련된 기초 자료를 생산하여 고용정책의 수립과 향후 평가에 이용하고자 1998년에 시작된 조사이다. 비농촌지역에 거주하는 한국의 가구 및 가구원을 대표하는 5,000가구(17,505명)의 개인을 대상으로 매년 개인의 경제활동, 노동시장이동, 소득활동 및 소비, 교육및 직업훈련, 사회생활 등에 관해 추적 조사하는 종단면 조사이다. 1998년 1차년도에 조사된 5,000가구의 가구원들은 비농촌지역의 거주인구를 대표한다. 2차년도 조사에서는 1차년도 원표본가구원들에 대한 매년조사가 수행되고, 이사 및 분가한 경우에는 추정조사를 수행한다. 원표본가구에서 출생한 자녀들은 표본가구원으로 추가되며, 원표본가구원 또는 그 자녀가 결혼 등으로 배우자가 있는 경우 그 배우자도 혼인관계가 지속되는 한 조사대상으로 포함한다.

1) 표본추출

1995년 인구센서스의 10% 표본조사구를 모집단으로 한다. 표본추출과정에서 1995년 인구센서스의 10% 표본조사구중에서 5,000가구를 직접 추출하지 않고, 1997년 고용구조특별조사(:'97고특)의 결과와 상호 비교를 위해 추출된 표본이 '97고특 조사의 표본에 속하도록 하기위해 지역별로 층화한 후 층내에서는 '97고특 조사의 층화 기준을 적용하였다. 조사구의 추출방법은 계통추출방법을 사용하였고, 제주도를 제외한 시부만을 대상으로 1,000개의 조사구를 선정하였고, 각 조사구에서 '97고특 조사의 조사대상가구중 5~6가구를 단순임의 추출하였다. 계통 추출된 조사구가 '97고특 조사구가 아닌 경우에는 가장 가까운 '97고특조사구를 표본조사구로 선정하였다. 각 조사구내에서 특정가구가 추출될 확률은 조사구내의 총가구수, '97고특에서 성공한 가구수, KLIPS에서 추출한 가구수에 따라 결정된다. 약 5개월

간의 조사('98.6.2 ~ '98.10.13)로 부터 5,000가구 약 17505명중 면접에 성공한 가구원은 13,317명이고, 가구수로는 75.3%인 3,773가구가 면접에 성공하였고, 나머지 24.5%인 1,227가구는 대체하였다. 이러한 표본대체의 주요한 사유로는 주소불명이 약 1.7% 이사로 인한 추적불가가 6.1%, 이사후 추적하였으나 응답거절인 경우가 0.6%, 응답을 강력히 거절하는 경우 11.8%(591가구)이고 기타이유가 4.3%로 나타났다.

2) 가중치 조정방법

1차년도에는 2단층화집락계통추출법을 사용하여 표본가구를 선정하였다. 1단계에서는 1995년 인구센서스의 10% 표본조사구중에서 도시지역 조사구 19,025개를 먼저 선정하고, 다음으로 이중에서 1,000개의 조사구를 선정하였다. 이 과정에서 최종 표본으로 선정된 조사구는 951개 조사구이다.

1차년도의 조사에서는 1차패널에 대해서 횡단면 조사이므로 통상적인 횡단면 조사의 가중치를 그대로 적용하게 된다. 일반적인 가중치 부부의 단계는 먼저 추출확률을 계산하고, 다음으로 무응답을 조정하고, 마지막으로 사후층화조정을 수행하게 된다. 앞에서 언급한 외국 패널의 경우를 살펴보면 개별 가중치 산정 과정은 다양하지만, 큰 틀에서는 3가지 과정으로 가중치를 작성한다. KLIPS의 경우 개인단위의 사후층화는 통계청의 인구추계자료가 있으나, 세부적인 정보가 없는 관계로 추출확률과 응답률에 의한 가중치 계산만 수행하였다.

<표 3.1> KLIPS의 1차 패널 개요

지 역	인구센서스 10%조사구		KLIPS조사구	PSU추출확률	표본가구수	응답가구수	응답률
	조사구수	도시조사구수					
서 울	5,186	5,186	227	0.43771693%	1,603	1,362	84.97%
부 산	1,841	1,841	97	0.52688756%	600	485	80.83%
대 구	1,212	1,212	64	0.52805281%	452	320	70.80%
인 천	1,116	1,116	59	0.52867384%	414	295	71.26%
광 주	607	607	32	0.52718287%	197	160	81.22%
대 전	592	592	31	0.52364865%	188	155	82.45%
울 산	456	456	24	0.52631579%	150	120	80.00%
경 기	3,624	3,542	167	0.51551141%	1,223	853	69.75%
강 원	725	504	26	0.51873016%	200	130	65.00%
충 북	694	422	22	0.52132701%	145	110	75.86%
충 남	865	488	25	0.51229508%	152	125	82.24%
전 북	967	757	40	0.52840158%	273	200	73.26%
전 남	996	487	25	0.51334702%	164	125	76.22%
경 북	1,372	1,201	54	0.52889324%	356	270	75.84%
경 남	1,422	1,094	58	0.53015453%	359	290	80.78%
전 국 (제주제외)	21,675	19,025	951	0.49989859%	6,476	5,000	77.21%

KLIPS는 SLID와는 다르게 가구 및 개인을 분석단위로 하기 때문에 개인별 가중치와 가구별 가중치를 각각 구해야 한다.

(가) 1차웨이브의 가중치 산정 방법

1차웨이브에서는 가구와 그 가구에 속한 가구원의 추출확률이 동일하기 때문에 개인별 가중치화 가구가중치는 같은 값을 가진다. 추출확률과 응답확률을 모두 고려한 가구가중치는 다음과 같이 계산된다.

○ 서울 및 6대광역시

0.1×(표본조사구수/도시조사구수)×('97고특조사가구수/ED 내 전체가구수)×(총접촉가구수/'97고특조사가구수)×(최종조사가구수/총 접촉가구수)

○ 도의 동부

0.1×(해당 도의 동부 표본조사구수/해당도의 동부 조사구수)×('97고특조사가구수/ED 내 전체가구수)×(총접촉가구수/'97고특조사가구수)×(최종조사가구수/총 접촉가구수)

○도의 읍면부

0.1×(해당도의 시에 속한 표본읍면부 조사구수/해당도의 시에 속한 읍면부 조사구수)×('97고특조사가구수/ED 내 전체가구수)×(총접촉가구수/'97고특조사가구수)×(최종조사가구수/총 접촉가구수)

이때 표본으로 선정된 모든 가구원이 응답하였으므로 동일가구내의 가구원 가중치는 해당가구가중치와 같게 된다.

(나) 2차 웨이브이후의 가중치 산정방법

KLIPS에서는 2차웨이브 이후 Duncan(1995)의 가중치 부여방법을 적용하고 있다. Duncan의 방법은 우선 미국의 대표적인 패널조사인 PSID에서 적용한 방법으로 논리적으로 일관성이 있기 때문이다. 이는 하나의 가구단위 분석이 현실적으로 어렵다는 점을 고려한 방법이다. 이에 대한 절차를 살펴보면 다음과 같다.

○ 1단계 : 초기웨이브에서 가구가중치를 구한다. 이때 추출확률과 조사과정의 응답률을 함께 고려하고, 마지막으로 사후층화조정까지 수행한 가중치를 구한다.

○ 2단계 : 초기웨이브에서 구한 가구가중치를 연령이나 응답여부에 관계없이 모든 가구원의 가구원가중치로 사용한다. 이 가중치는 15세 이상의 가구원 뿐만아니라 15세 이하의 모든 가구원에게도 동일하게 적용한다.

○ 3단계 : 2차웨이브 이후부터는 가구원들의 상이한 응답률을 이용하여 가구원들의 가중치를 조정한다. 이때 2차웨이브에서는 해당가구에 존재하지만, 1차웨이브에서는 응답하지 않았던 비표본가구원이나 1차웨이브이후 새로 태어난 자녀의 경우에는 개인차원의 무응답 조정과정에 포함시키지 않는다.

○ 4단계 : 2차웨이브에서 산출된 개인가중치의 가구내 평균을 이용하여 2차웨이브의 가구가중치를 산출한다. 결혼, 동거등의 사유로 새롭게 진입한 비표본가구원에게는 0의 가중치를 부여한다.

개인차원의 분석에서 개인가중치를 사용하게 되면 비표본가구원을 제외하고 분석하게 되며 비표본가구원을 포함하여 분석할 경우에는 평균치로 계산된 가구가중치를 사용하면 된다. 1차 웨이브에서 4차 웨이브까지 위와 같은 절차로부터 계산된 개인가중치와 가구가중치 및 횡단면과 종단면 가중치는 각각 다음과 같다.

<표 3.2> 각 웨이브별 가중치

웨이브	변수	표본수	평균	표준편차
1 wave	w1_p*	13,321	2,255.039	416.7548
	w1_h*	13,321	2,255.039	416.7548
2 wave	pid	12,039	252,418.2	145127.8
	w1_p	11,237	2,256.868	413.7408
	w1_h	11,237	2,256.868	413.7408
	w2_p	11,708	2,671.798	546.3235
	w2_h	12,032	2,600.247	588.4898
3 wave	pid	11,205	254,092.7	147,331.3
	w1_p	10,141	2,264.563	417.4775
	w1_h	10,141	2,264.563	417.4775
	w2_p	10,080	2,667.529	544.3491
	w2_h	10,324	2,604.68	584.397
	w3_p	10,798	3,072.339	758.9452
	w3_h	11,194	2,963.651	771.8895
4 wave	pid	11,051	259,720.6	149,571.5
	w1_p	8,957	2,264.824	418.4443
	w1_h	8,957	2,264.824	418.4443
	w2_p	9,317	2,667.034	544.2263
	w2_h	9,521	2,608.579	583.2122
	w3_p	9,374	3,062.756	750.1324
	w3_h	9,647	2,968.351	764.0813
	w4_p	10,499	3,450.922	980.4998
	w4_h	11,021	3,287.472	969.8315

(*주: 첨자 p는 개인가중치, h는 가구가중치를 의미한다.)

4. KWPS의 가중치 조정 방법

4.1 KWPS의 개요

한국복지패널 조사는 외환위기 이후 빈곤층(또는 working poor)및 차상위계층의 가구형태, 소득수준, 취업상태가 급격히 변화하고 있어, 이들의 규모와 상태변화를 동적으로 파악하여 정책지원을 위한 기초 자료를 생산하고, 소득계층별 경제활동 상태별, 연령별 등 각 인구집단의 생활실태와 복지욕구 등을 역동적으로 파악하고 정책의 효과를 평가함으로써 정책형성과 피드백에 기여하기위한 조사이다. 따라서 도시의 일반 가구를 대상으로 하는 KLIPS와는 지향하는 정책관점이 서로 다르다고 할 수 있으며, 조사대상가구는 일반가구와 저소득층 가구를 각각 50%씩 추출하여 이들에 관한 다양한 복지실태와 욕구 등에 관한 조사를 하고자 한다. 한국복지패널조사는 가구용 설문지와 가구원용 설문지(15세 이상), 아동용 설문지로 구성되어있으며, 가구원용 설문지는 15세 이상 중고등학생을 제외한 경제활동인구 모두에 대해 응답을 받도록 하고 있다. 또한 아동용 설문지는 1차 패널조사에서만 조사되는 부가용 설문지로서 향후 3년마다 수행할 예정이다.

4.2 표본추출과 가중치 조정

1) 표본 추출

한국 보건사회연구원의 "2006 국민기초생활실태조사(2006.6.30~10.1)"의 표본설계 당시 2005년도 인구센서스 90% 조사구에 대해 이용 가능한 자료로서는 조사구별 가구 수, 조사구 형태, 주택형태 뿐이었기 때문에, 이를 바탕으로 조사구당 평균 60가구인 약 30,000가구를 추출하기 위해 517개 조사구의 기초 자료를 집계하여 지역별, 조사구 유형별, 읍면동별, 주택형태별로 분류하여 분포를 파악하였다.

전체 517개 표본조사구중 문제조사구를 제외한 487개 조사구에 대한 조사 자료를 기초로 7,000가구를 소득기준으로 중위소득 60%이하인 3,500가구와 중위소득 60% 이상인 3,500가구를 각각 표본으로 추출하여 조사를 수행하였다. 이때 저소득층과 일반 가구층을 구분하기 위한 기준은 가구소득을 이용하여 다음과 같은 3가지 대안을 고려하였고, 최종적으로 "공공부조이전경상소득" 의 중위소득 60%를 기준으로 저소득층 가구와 일반가구로 구분하였다.

<표 4.1> 지역별 표본조사구 분포

시 도	계	일반	아파트	계	구	시	군	동	읍	면
서울	110	68	42	25	25	-	-	110	-	-
부산	43	29	14	15	14	-	1	41	2	-
대구	30	23	7	8	7	-	1	27	3	-
인천	30	22	8	8	7	-	1	27	1	2
광주	16	6	10	5	5	-	-	16	-	-
대전	17	8	9	5	5	-	-	17	-	-
울산	14	9	5	5	4	-	1	10	3	1
경기	91	59	32	25	17	16	2	72	4	15
강원	18	9	9	11	-	5	6	10	5	3
충북	16	11	5	8	-	2	6	8	3	5
충남	21	14	7	10	-	3	7	4	6	11
전북	21	16	5	8	2	2	4	13	3	5
전남	22	16	6	11	-	5	6	7	9	6
경북	30	23	7	17	-	7	10	11	4	15
경남	33	19	14	15	-	8	7	15	6	12
제주	5	3	2	3	-	2	1	4	1	-
전체	517	335	182	179	86	59	53	392	50	75

<표 4.2> 저소득가구의 분류기준

중위소득	경상소득		가처분소득		공공부조이전 경상소득	
	가구수	백분율(%)	가구수	백분율(%)	가구수	백분율(%)
< 40%	2,481	10.0	2,489	10.09	3,477	13.96
< 50%	4,016	16.12	3,880	15.62	4,757	19.04
< 60%	5,227	22.56	5,473	22.25	6,128	24.76

국민기초생활실태조사의 487개 조사구중 조사과정에서 향후 재개발이 계획된 지역과 조사에 대한 강력거절 지역을 제외한 나머지 446개 조사구와 각 지역별 가구수 현황은 다음과 같다.

<표 4.3>의 지역별 표본가구수는 소득기준으로 일반가구층과 저소득가구층에 각각 배분된 3,500가구를 지역의 크기와 가구규모별로 배분하여 일반가구와 저소득층 가구 표본을 추출하였다. 전체적인 저소득층 가구의 비율은 약 25% 정도로 추정할 때, 저소득층 표본가구는 패널마모율을 고려하여 과대표본추출(oversampling) 하였다.

<표 4.3> 지역별 조사구수와 가구분포현황

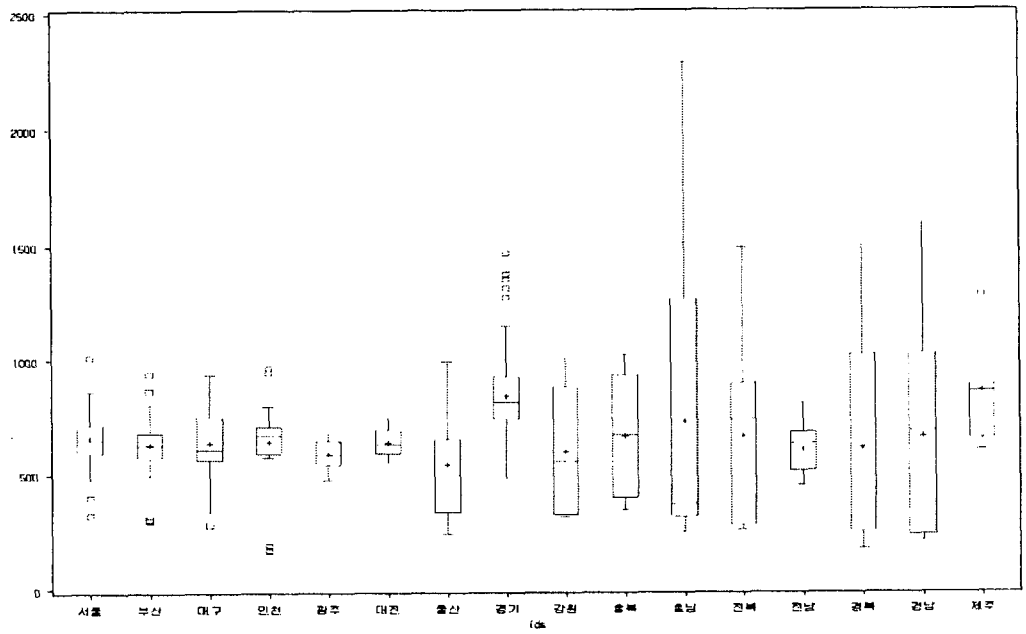
지 역	조사구수	일반가구	저소득가구	합 계
서 울	93	811	506	1,317
부 산	33	254	272	526
대 구	25	187	227	414
인 천	28	228	193	421
광 주	15	114	130	244
대 전	14	118	89	207
울 산	14	120	82	202
경 기	76	644	471	1,115
강 원	14	102	131	233
충 북	14	108	113	221
충 남	20	153	168	321
전 북	20	138	209	347
전 남	19	104	273	377
경 북	26	152	339	491
경 남	30	229	254	483
제 주	5	38	43	81
합 계	**446**	**3,500**	**3,500**	**7,000**

2) 가중치 조정

전체 2005년도 인구센서스 자료의 90% 모집단으로부터 각 지역별의 크기에 비례하여 표본조사구를 추출하였으므로 지역별 모집단 조사구와 표본조사구의 비율로서 PSU의 추출확률을 계산할 수 있다. 지역별로 각 조사구당 추출확률(PSU)를 계산하고, 국민기초생활실태조사로부터 조사구당 무응답가구에 대한 가중치를 조정하고, 최종적으로 제외조사구에 대한 SCALE 조정을 마친 후의 지역별 가중치의 분포를 구하면 다음의 <표 4.4>와 같다. 이 가중치는 지역별 조사구의 크기에 따라 확률비례로 부여된 가중치이고, 하나의 조사구에 추출된 가구는 동일한 추출확률을 가지게 됨으로 PSU당 가구가중치로 사용된다. 이와 더불어 1차 웨이브에서는 종단면가중치와 횡단면 가중치가 서로 동일하게 사용됨으로 이 가중치는 횡단면 가구가중치가 된다.

<표 4.4> 지역별 가중치의 분포

지 역	조사가구수	평균	표준편차	최소값	최대값
서 울	4,682	771.076	114.117	380.207	1,186.70
부 산	1,676	772.676	158.780	369.631	1,156.62
대 구	1,277	762.375	189.042	336.772	1,123.74
인 천	1,425	685.409	197.912	178.758	1,031.35
광 주	765	624.025	67.823	506.413	726.73
대 전	714	721.159	68.717	621.880	845.49
울 산	716	538.488	218.627	237.105	991.72
경 기	3,868	977.420	221.032	561.196	1,709.86
강 원	708	712.523	334.362	373.836	1,216.04
충 북	714	756.045	301.117	391.731	1,164.16
충 남	1,020	762.085	606.529	255.649	2,403.57
전 북	1,020	693.340	349.385	267.249	1,566.89
전 남	969	698.307	119.634	519.541	935.94
경 북	1,312	676.340	478.543	191.967	1,670.58
경 남	1,530	699.424	435.659	215.499	1,696.13
제 주	255	856.531	241.114	600.339	1,282.91
계	22,651				



<그림 4.1> 지역별 가중치 분포

<그림 4.1>로 부터 지역별 가중치 분포를 살펴보면, 충남의 조사구에서 매우 큰 가중치가 발견되었다. 향후 추정치의 산정에서 이와 같이 매우 큰 가중치는 smoothing 시키는 방법으로 조정이 이루어져야 할 것으로 판단된다. 종단면 가구가중치와 횡단면 가구가중치는 앞에서 언급한 바와 같이 동일한 값을 가지게 되며, 표본가구의 개인들에 대한 가중치는 가구의 모든 가구원이 응답하므로 가구가중치와 동일하게 부여되며, 이 값은 사후층화조정 이전의 2006 국민기초생활실태조사의 가중치와 같다.

○ 가구가중치와 개인가중치: $w_{L,h} = w_{C,h} = w_{L,p} = w_{C,p}$

(10/9) × (표본조사구/최종조사구수) × (지역별 총조사구수/지역별 조사구수) × (접촉가구수/최종조사완료가구수)

여기서 L은 종단면, C는 횡단면을 나타내며, h는 가구, p는 개인을 의미한다.

이와 함께, 표본으로 선정된 가구가 이주나 조사거절로 인하여 무응답이 발생한 경우에는 해당 조사구 내에서 무응답 가구와 가장 유사한 가구로 대체 응답하도록 하였으며, 가능한 한 현장 대체는 허용하지 않도록 하였다.

4.3 KWPS의 향후 과제

패널조사는 개인 또는 가구의 동적인 현상을 파악하는 중요한 조사 방법으로 현재 국내에서 급속히 확산되어 가고 있다. 패널 조사의 장단점은 이미 많은 연구에서 파악되고 있지만, 다양한 장점에도 불구하고, 시간의 흐름에 따라 표본에서 탈락하는 가구가 증가한다는 사실과 이를 방지하기 위한 패널 유지비용이 일반 횡단면 조사에 비해 상대적으로 월등히 많이 소요된다는 점이다.

이러한 문제점은 모든 패널조사에서 공통적으로 안고 있는 문제이지만, 가능한 한 많은 예산을 투입하여 표본마모율을 축소하고자 노력해야 할 것이다.

1) 1차웨이브 이후의 가중치 조정

한국복지패널 조사에서의 가중치 조정과정은 1차웨이브에서의 가중치를 기본으로 하여 향후 2차웨브 이상의 조사에서의 가중치를 조정하고자 한다. 1차웨이브의 기준년도는 2005년도 12월 31일 기준이며, 이 기준년도는 종단면 가중치를 산정하는 기준이 될 것이다. 2006국민기초생활실태조사에서 재개발, 자연재해, 강력거절 조사구(주로 아파트조사구)에 대

해서는 적절한 조사구의 무응답 가중치를 부여함으로서 무응답 조사구 문제를 해결하였다. 또한 2차웨이브 이상 조사가 진행됨에 따라 가구차원에서는 이주 또는 분가 등으로 인하여 가구의 변동이 발생함으로 이를 위한 추적조사및 가중치 조정이 반드시 필요하다. 또한 개인 차원에서는 1차웨이브 당시의 가구원이 아닌 개인이 2차웨이브에서는 가구에 살고 있는 경우나 새로 태어난 자녀 등에 대해 개인차원의 변동에 대한 가중치의 조정이 필요하다. 이러한 가중치 조정과정은 가능하면 객관적이고, 타당한 방법을 적용하여 추정치의 편향을 가능한 축소시키고자 한다.

2) 가중치의 masking 작업

통상적으로 동일한 조사구에서의 가구 가중치는 같은 가중치를 가지며, 또한 동일한 가구내에서의 개인들의 가중치는 서로 같은 값을 가진다. 따라서 가중치의 특성을 파악하면, 개인정보를 일정 수준까지는 파악할 수 있는 단서를 제공할 수 있다. 이러한 개인 정보 누설을 방지하기위해 SLID의 경우 내부데이터에 대해서는 별도의 masking을 하지 않지만, 외부로 공표되는 자료에 대해서는 반드시 가중치에 masking 작업을 추가하는 것으로 나타났다.

그러므로 KWPS자료의 특성상 개인정보의 누출을 방지하고자 최종가중치에 일정 수준의 잡음을 부가한 가중치를 대외자료로 공표하고자 한다. 이러한 작업은 이미 SLID에서 진행하고 있으며, 이를 기반으로 KWPS의 데이터 환경에 적합한 masking 작업을 진행하고자 한다.

5. 결론 및 토의

한국복지패널조사(KWPS)는 일반 가구와 저소득층 가구에 대해 국민의 복지실태 및 욕구에 대한 동적분석을 위해 수행되는 국가적인 조사사업이다. 특히 외환위기 이후 저소득 계층의 상태가 급격히 변화하고 있으며, 이들에 대한 국가적인 사회안전망 구축, 저소득 계층의 탈 빈곤의 양태 등에 대한 기초자료가 전무한 상태이므로 이러한 요구에 따라 복지패널조사가 수행되고 있다.

향후 기초데이터를 이용한 다양한 분석이 예상되며, 이를 위한 기초자료를 생산하기 위한 노력들이 많은 부문에서 필요한 실정이다. 이를 위해 KWPS의 데이터는 사회복지학적인 측면, 경제학적인 측면, 통계학적인 측면이 모두 고려된 종합적인 데이터로서 국민에 대한 복지정책의 수립과 정책평가의 기초자료로 이용되길 희망한다.

참고문헌

[1] 강석훈(1999), KLIPS의 1차웨이브 가중치 부여방안에 대한 연구, 한국노동패널연구.

[2] _____(2003), KLIP의 가중치 부여방안 연구, 한국노동패널연구 2003-04.

[3] 국민연금연구원 패널팀(2005), 국내외 패널 및 주요 사회조사동향분석, 국민연금연구원 Working paper, 2005.

[3] 금재호(1998), 캐나다노동패널조사, 한국노동연구원.

[4] 신동균(1998), 미국패널데이터의 현황과 시사점-PSID, NLSY, KHPS, KLIPS를 중심으로-, 한국노동패널연구 97-02.

[5] Duncan, G.(1995), "A Simple Method for Weighting in Household Panel Survey", Working paper, Northwestern University.

[6] Earnst, L.(1989), "Weighting Issues for Longitudinal and Family Estimates", In Kasprzyk, D. Duncan, G., and Singh, M. eds, Panel Survey, Weily, pp.139-177.

[7] Institute for Social Research (1972), A Panel Study of Income Dynamics, University of Michigan.

[8] Kalton, G., and Brick, M.(1994), "Weighting Schemes for Household Panel Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association, 789-790.

[9] KILPS 1차년도 - 5차년도 USER's GUIDE , 한국노동연구원

[10] Lecvesque, I., and Franklin, S.(2000), "Longitudinal and Cross-Sectional Weighting of Survey of Labour and Income Dynamics 1997 Reference year", Statistics of Canada Research Paper series.

[11] Naud, J. F.(2004), "Combined-panel longitudinal weighting Survey of Labour and Income Dynamics", Statistics of Canada Research Paper series.

[12] Latouche, M., Dufor, J., and Merkouris, T.(2000), "Corss-sectional Weighting: Combining Two or More Panels", Statistics of Canada Research Paper series.

[13] Westat(2001), Survey of Income and Program Participation User's Guide, 3rd eds.

[14] McGonagle, K. A., and Schoeni, R.F.(2006), The Panel Study of Income Dynamics : Overview and Summary of Scientific Contributions After Nearly 40 years", Institute for Social Research, University of Michigan.

[15] Heeringa, S.G., and Connor, J.H.(1999), "1997 Panel Study of Income Dynamics Analysis Weights for Sample Families and Individuals", Institute for Social Research, University of Michigan.