



제1장 가구패널조사에서 횡단면 가중값 작성방법

백지선 · 심규호

제1절 서론

1. 연구배경 및 목적

통계청은 가구의 자산과 부채 규모, 분포 및 구성 그리고 가계부문의 미시적 재무 건전성을 주기적으로 파악하여 정부의 부동산 정책 및 금리 정책, 가계 신용관리, 금융시스템 발전 등을 위한 기본정보로 활용하는 것을 목적으로 1만 가구를 대상으로 패널조사인 가계금융조사를 2010년, 2011년 두 차례에 걸쳐 실시하고, 전국단위의 통계결과를 발표하였다. 이후 시도단위 통계생산에 대한 필요와 요구에 따라 표본가구수를 2만가구로 확대하고 금융부분의 복지부분에 대한 항목을 추가하여 2012년 제1회 가계금융·복지조사로 확대 변경하여 조사를 실시하였으며, 결과를 공표할 계획이다.

패널조사는 횡단조사와는 다르게 한번 추출된 표본을 계속 추적하여 조사를 실시하는 형태로 조사목적에 따라 고정패널조사(Fixed Panel Survey), 연동패널조사(Rotating Panel Survey), 분할패널조사(Splite Panel Survey)로 구분할 수 있으며, 각각은 장단점을 가지고 있다. 패널조사는 추적조사를 실시함으로써 횡단 및 종단 분석을 가능하게 한다는 장점이 있지만, 그에 따른 표본 관리와 추적원칙, 비용, 조사 후 통계 생산단계에서 복잡한 표본상태 변화를 고려한 횡단 및 종단 가중값 작성 등 다양한 어려운 점들이 발생한다. 가계금융조사는 가구 횡단 가중값만을 작성하여 제공하였으나, 가계금융·복지조사는 횡단과 종단에 대해 가구 및 개인 가중값을 작성하여 제공할 계획이다.

패널조사의 목적에 비추어 보았을 때, 일반적으로 중단면 분석이 중요하지만 횡단면 분석도 그에 못지않게 중요하다. 그러나 조사 차수가 늘어남에 따라 응답 거부, 사망, 이민 등으로 인한 패널이탈 발생에 따른 표본 수 감소는 횡단면 추정의 효율성을 떨어뜨리는 요인이 된다. 이에 대해서 표본을 추가하는 방법과 원표본 개인 외에 추가되는 비표본 개인을 고려하는 방법을 생각해 볼 수 있으며, 후자의 경우는 추적규칙과도 관련되는 문제이기도 하다.

본 과제는 패널마모로 인한 횡단면 추정의 효율성 제고방안으로 비표본 개인을 포함하는 경우에 그에 따른 가중값 작성방법을 살펴보고자 한다. 먼저 웨이브가 진행되면서 패널에 진입하는 비표본 개인과 가구 가중값에 대한 처리 방법에 대한 문헌연구와 사례 연구를 실시하고, 2010년과 2011년 가계금융조사 자료의 적용을 통하여 비표본 개인을 고려한 가중값 처리방법의 영향 및 결과를 분석하였다.

본 보고서의 구성은 다음과 같다. 2절에서 가구패널조사와 추적규칙, 패널조사의 횡단 가중값 부여 방법을 소개하고, 3절에서는 주요 가구패널의 횡단 가중값 작성방법과 가구변화요인별 가중값 처리방법을 검토하였다. 4절에서는 1,2차 가계금융조사 실증분석을 통한 가구 구성변화를 살펴보고 가중값 공유 방법을 적용하여 가구 및 개인 가중값을 산출하였으며, 끝으로 5절에서 연구결과를 정리하였다.

2. 연구내용

본 보고서의 연구내용은 패널조사의 횡단 가중값 방법론을 중심으로 문헌연구와 국내외 사례연구 그리고 가계금융조사에 대한 실증분석으로 구성되어 있다. 패널조사는 동일한 표본을 대상으로 일정한 주기로 조사를 실시하게 되므로 웨이브 진행에 따른 표본 추적규칙, 중단모집단, 원표본, 비원표본 등 일반적인 독립조사와는 차별되는 개념들이 발생한다. 문헌연구를 통하여 이러한 개념들과 시간효과로 발생하는 가구 구성변화반영을 위한 가중값 조정방법인 원표본 가중값 공유방법과 모형방법을 정리하였다. 또한 주요 가구패널조사인 한국노동패널, 재정패널, 미국 PSID, 영국 BHPS, 호주 HILDA의 가중값 작성방법과 가구 변화요인에 따른 가중값 조정방법을 소개하였다. 패널조사에 대한 실증분석으로 가계금융조사를 이용하여 2차 웨이브 가구 및 가구원 변동사항 등을 분석하고, 가구 변화에 따른 가중값 조정을 실시하는 경우 가중값 분포와 가구 및 가구주 특성 분포 그리고 추정값에 미치는 영향을 살펴보았다.



제2절 가구패널조사와 추적규칙

1. 가구패널조사

가. 개요

가구패널조사는 동일한 가구를 주기적으로(보통 연간) 조사하는 것으로, 다양한 주제에 대해 조사를 실시하며, 횡단 및 종단 분석이 가능하다는 장점을 갖는다. 가구패널은 가구원뿐만 아니라 가구에 대한 정보를 얻는 것을 목표로 하며, 가구에 대한 모든 정보를 얻기 위하여 일반적으로 조사대상 가구의 모든 가구원을 조사하게 된다.

패널조사는 일반적인 횡단면 조사와 동일하게 모집단의 대표성을 확보할 수 있도록 표본 설계를 하고, 표본을 추출한 후 조사를 실시하는데, 추출된 표본을 계속해서 조사하게 되므로 1차년도에 표본은 여러 측면에서 매우 중요하다고 할 수 있다. 또한 2차년도 이후 동일한 표본을 조사하면서 응답거부, 부재, 추적불능 등의 사유로 패널이탈(panel attrition)이 발생하게 되며, 그 외에도 출생, 사망, 가구 분가, 새로운 가구원의 진입 등으로 인한 비표본 가구원의 유입 및 패널 구성원 변화가 발생하게 된다. 이러한 패널조사의 특징으로 인해 1차년도에 응답한 가구원을 원표본 가구원(original sample member, OSM)이라고 하며, 이후 결혼으로 인한 배우자, 출생한 자녀, 기타 진입한 구성원 등 2차년도 이후에 원표본 가구에 포함되는 개인을 비표본 가구원(nonsample member)한다.

조사현장에서는 비표본 가구원과 관련하여 발생하게 되는 다양한 경우에 대해 사전에 명확한 처리규칙을 세울 필요가 있으며, 이를 추적규칙(follow-up rule)이라고 한다. 추적규칙에 따라 패널에서 나가거나 들어오는 경우가 발생하게 되며, 이는 비용과도 밀접한 관련이 있다. 원표본 가구원과 추적 조사할 비표본 가구원을 포함한 확장된 원표본 가구원을 영구표본 가구원(permanent sample member, PSM)이라고 한다. 영구표본 가구원의 정의는 각 패널조사에 따라 차이가 있다(<표 1-1> 참조).

패널조사에서 횡단면 추정을 고려할 때, 비표본 가구원은 패널이탈, 사망 등으로 인한 표본 감소에 대해 새로운 추가표본을 추출하는 것보다 접촉 및 조사가 용이하며 비용면에서도 효율적인 측면이 있기 때문에 대부분의 패널조사에서는 비표본 가구원을 고려하고 있다. 각 패널조사는 정의된 영구표본 가구원과 추적규칙을 적용하고 있으며, 이렇게 수집된 비표본 가구원과 탈락된 가구원의 정보는 추정단계에서 고려되어야 한다.

<표 1-1> 패널조사별 영구표본 가구원 정의

Table 1: Following rules in household survey panels

	SLID	PSID	LISS	BHPS	HILDA	SHARE	SHP	SOEP
Wave 1 respondents	PSM	PSM	PSM	PSM	PSM	PSM	PSM	PSM
Births / adoptions to PSM	NA ⁽²⁾	PSM	PSM	PSM	PSM	PSM ⁽⁵⁾	PSM	PSM
Recent Immigrants ⁽¹⁾					PSM ⁽⁴⁾	PSM	PSM ⁽⁶⁾	PSM ⁽⁷⁾
Partner with child from PSM				PSM ⁽³⁾	PSM	PSM	PSM ⁽⁶⁾	PSM
Partner w/o child from PSM						PSM	PSM ⁽⁶⁾	PSM ⁽⁷⁾
Other household entrants							PSM ⁽⁶⁾	PSM ⁽⁷⁾

⁽¹⁾ Recent Immigrants are immigrants who entered the target population after wave 1.

⁽²⁾ Does not apply – respondents stay only for 6 years in panel and respondents younger than 16 are not interviewed.

⁽³⁾ While these sample members are followed, they receive zero weight when they leave the household of other PSMs.

⁽⁴⁾ Since wave 9 (2009).

⁽⁵⁾ Respondents aged 50+; Rare births from younger partners are not followed.

⁽⁶⁾ Since wave 9 (2007).

⁽⁷⁾ Since wave 7 (1990).

출처 : Schonlau, M. et al. (2011), Household survey panels: how much do following rules affect sample size?, Survey Research Methods, 5,2, 53-61.

나. 추적규칙

패널조사에서 추적규칙은 종단 모집단 정의와 관련되는데, <표 1-1>에서 살펴보았듯이 패널조사마다 다르게 적용되고 있다.

1) 종단 모집단과 추적규칙

$U^0(i, i' \in U^0)$ 를 패널 시작 시점의 목표 모집단(target population), $U^{0,t}(k, k' \in U^{0,t})$ 를 t 차 웨이브 시점의 초기 모집단이라 하자. 추적규칙에 의해 $U^{0,t}$ 에 포함될 개인이 결정되어지며, 사망 또는 다른 이유로 모집단에서 빠져나간 경우는 모집단에서 제외된다. 일반적으로 패널조사의 목적은 종단 모집단 $U^{0,t}$ 를 기술하고자 하는데 있으며, $U^{0,t}$ 와 t 차 웨이브 시점의 현재 모집단 U^t 는 구별되어야 한다.

일반적인 추적규칙과 관련하여 종단모집단을 소개하면 다음과 같다(<표 1-1>과 [그림 1-1] 참조).

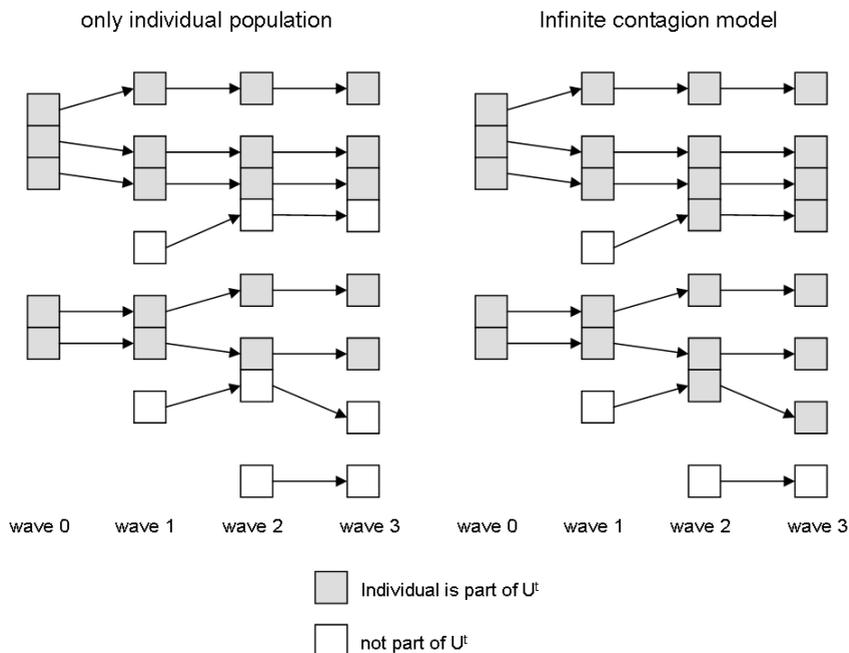
- ① 초기 모집단(only initial population)
- ② 초기 모집단 + 현재 동거인(only initial population plus current cohabiters)
- ③ 초기 모집단의 성인(only adults of the initial population) : PSELL
- ④ 초기 모집단 + 그들의 자녀들(initial population plus their offsprings): PSID
- ⑤ 접촉모형(contagion models) : SOEP



초기 모집단만을 조사하는 것은 가장 단순하지만, 이탈로 인한 표본 수 감소 등으로 인해 대표성 확보에 문제가 생길 수 있다. 초기 모집단에 현재 동거인 또는 자녀를 포함시키는 경우도 있으며, 미국의 PSID는 자녀를 포함시킨다. 접촉모형은 웨이브가 진행되면서 접촉되는 모든 개인들을 포함시키는 모형으로 독일의 SOEP, 스위스의 SHP 등에서 적용하고 있으며 횡단모집단 측면에서 대표성 확보에는 용이하지만 패널관리와 추정단계에서는 많은 어려움이 따른다.

2) 패널표본

초기 모집단 U^0 로부터 표본설계에 의해 초기표본(또는 원표본) $s^0 \subset U^{0,t}$ 가 추출된다. 초기표본은 종단모집단 정의와 추적규칙에 따라 전개되어, t 차 웨이브 표본 s^t 가 형성된다. 즉, 표본 s^t 는 모집단 $U^{0,t}$ 로부터 추출된 표본이 아니라 시간이 지나면서 가구 구성이 전개된 결과이며 s^t 에는 초기 원표본이 아닌 추적규칙에 의해 패널로 진입한 표본들(비표본)이 포함되어 있다. 패널표본은 초기 원표본 s^0 에만 추출확률이 주어지며, 비표본 s^t 의 추출확률은 알려져 있지 않다.



[그림 1-1] 조사 시점에 따른 패널 전개 사례 : Lynn(2006)에서 발췌

2. 가구패널조사 가중값

패널조사는 시간효과로 인해 목표 모집단의 변화와 가구 구성의 변화가 발생하게 된다. 모집단의 변화는 출생, 이민(immigration)으로 인해 발생하며, 가구 구성의 변화는 결혼, 분가, 출생, 입양, 사망 등으로 인해 발생하게 된다. 앞에서 언급하였듯이 추적규칙에 의해 종단모집단이 결정되며 일반적으로 표본으로 추출된 개인들이 사망(death)이나 이민(emigration)으로 인해 목표모집단을 나갈 때 까지 추적하는 것이 원칙이다.

일반적으로 가중값은 설계가중값, 무응답 조정, 사후 조정의 3단계를 거쳐 작성하며, 1차 웨이브 패널조사의 가중값도 이와 같은 단계를 통해 작성된다. 그러나 2차 웨이브 이후부터는 가구 및 가구 구성원의 변화를 고려한 가중값 작성이 필요하며 본 절에서는 Kalton et al.(1994), 강석훈(2003), Lynn(2006)을 참고하여 비표본 개인을 반영하는 방법에 대한 기본적인 내용을 정리하였다.

초기 패널조사는 주어진 표본설계에 의해 모집단 U^0 에 속한 표본들이 추출되는데, 이 때 표본 i 가 추출될 확률을 π_i 라 하자. 일반적으로 관심 대상인 모집단 총합 $T_{y^0} = \sum_{i \in U^0} y_i^0$ 은 불편추정량인 Horvitz-Thompson(HT) 추정량을 이용하여 다음과 같이 추정할 수 있다.

$$\hat{T}_{y^0} = \sum_{i \in s^0} d_i y_i^0, \quad d_i = 1/\pi_i. \quad (1)$$

t 차 웨이브에서 표본 $k \in U^{0,t}$ 의 특성변수를 y_k^t 라 하면, 모집단 총합을 다음과 같이 정의할 수 있다.

$$T_{y^t} = \sum_{k \in U^t} y_k^t. \quad (2)$$

이 경우 t 차 웨이브의 모집단 총합에 대한 HT 추정량은 다음과 같이 표현될 수 있다.

$$\hat{T}_{y^t} = \sum_{k \in s^t} d_k y_k^t, \quad (3)$$

여기서 d_k 는 s^t 에 진입한 표본 $k \in U^{0,t}$ 에 대한 추출확률의 역수이다. 그러나 이 경우에 s^t 는 $U^{0,t}$ 로부터 추출된 표본이 아니므로 추출확률 d_k 가 주어지지 않는다.



모집단 총합에 대해 불편추정량을 만족하는 가중값으로 추출확률의 역수로 부여하는 방법을 이용하기 위해서는 현재 가구 내 모든 가구원의 원가구가 1차 웨이브에서 추출될 확률이 필요하다. HILDA(Watson, 2004)와 SOEP(Galler, 1987)는 모형을 이용하여 원가구 추출확률을 추정하고 가중값 산출에 이용하는 방법을 적용하였다.

Kalton and Brick(1995)는 주어진 원표본 가중값에 근거하여 비표본의 정보를 활용하면서 동시에 불편추정량이 되는 가중값 부여방법으로 다음과 같은 t 차 웨이브 표본 k 의 가중값 w_k 를 제안하였다.

$$w_k = \sum_{i \in s^0} l_{ki} d_i, \quad \sum_{i \in U^0} l_{ik} = 1. \quad (4)$$

이 방법은 원표본의 가중값을 연결함수(link function) l_{ki} 를 이용하여 비표본에게 재분배(redistribution)하는 것으로 원표본의 가중값을 비표본과 공유(weight share)하게 된다. 이러한 방법을 통해 1차 웨이브 이후 가구 및 개인 가중값을 구할 수 있으며, 식(4)에서 $\sum_{i \in U^0} l_{ki} = 1$ 를 만족하는 연결함수(link function) l_{ki} 는 불편추정량을 제공한다. 불편추정량이 되는 연결함수는 여러 가지 형태로 정의할 수 있으며, 균등 가구 가중값 방법(Equal Household Weighting Scheme)과 균등 개인 가중값 방법(Equal Person Weighting Scheme) 등이 있다. 원표본의 가중값을 공유하는 방법은 BHPS, SIPP 등에서 사용하고 있다. 모형을 이용하여 추출확률을 추정하는 방법(이하 모형방법)과 원표본의 가중값을 분배한 후 공유하는 방법(이하 가중값 공유방법)은 1차 웨이브 가구 추출확률이 작으면 거의 일치하며, 모형방법으로 추정된 새로 진입한 가구원의 추출확률은 원가구내 가구원수에 반비례하는 것으로 알려져 있다(Schonlau, M. et al., 2010). 본 절에서는 일반적으로 사용되는 균등 가중값 공유방법과 모형방법에 대해 소개한다.

가. 모형 방법

추출확률의 역수를 가중값 부여하기 위해서는 현재 표본가구의 모든 구성원의 원가구가 1차 웨이브에 추출될 확률과 결합추출확률을 알아야 한다. 그러나 원표본 가구에 대해서는 추출확률이 알려져 있으나, 그 외 가구에 대해서는 확률이 알려져 있지 않기 때문에 현실적으로 사용하기가 어렵다. Rental(1992)은 결합확률을 무시하고 로짓 모형을 이용하여 추출확률을 계산하는 방법을 제안하였다. HILDA(Watson, 2004)에서는 새롭게 가구로 진입한 가구원의 가구 h_i 에 대한 1차 웨이브 추출확률 p_i 을 회귀모

형을 통해서 추정하고 아래의 식을 이용하여 t 차 웨이브 가구 추출확률 $P(H_i)$ 를 계산하였다.

$$P(H_i) = p(h_1 \cup h_2 \cup \dots \cup h_k) = 1 - (1-p_1)(1-p_2) \dots (1-p_k) .$$

SOEP(Galler, 1987)도 새로 진입한 가구원의 가구에 대한 추출확률 p_l 을 모형으로 추정하였으며, HILDA와는 달리 가구 H_i 추출확률에 대해 결합확률을 무시하고 다음과 같이 $P(H_i)$ 를 계산하여 가중값을 산출하였다.

$$P(H_i) = p_1 + p_2 + \dots + p_k .$$

또한 새로 진입한 그룹을 각각 분리된 단위로 취급하였는데 예를 들어, 기존 원가구에 엄마와 자녀가 진입한 경우, $P(H_i) = p_1 + p_2 + p_3$ 가 된다.

나. 균등 가중값 공유방법

가중값 공유방법을 적용하는 방법은 가구를 기준으로 하는 방법과 개인을 기준으로 하는 방법이 있으며, 가구 기준 방법보다는 개인 기준 방법이 비교적 간단하여 더 많이 사용되고 있다. 두 방법 모두 1차 웨이브 원표본의 가중값을 이용하게 되므로, 가구 내 가구원 중 적어도 한명의 원표본 가구원이 존재해야한다는 한계점을 가지고 있다. 횡단면 가구 가중값과 개인 가중값을 구하는 방법이 있으며, 한 가구 내 모든 개인에게 동일한 가중값을 부여하는 원칙하에 개인 가중값을 작성하는 방법으로, 결과적으로 두 방법은 동일하다.

1) 횡단면 가구 가중값

횡단면 분석에 대표성 및 효율성 측면에서 1차년도 이후 원표본과 동거하는 모든 가구들을 포함시키고자 하는 것으로 비표본 개인으로만 구성된 가구는 제외한다.

① 균등 가구 가중값 방법

균등 가구 가중값 방법(Equal Household Weighting Scheme, EHWS)은 원표본 가구를 기준으로 재분배하는 방식으로 현재 가구내 원표본가구 가중값을 원가구 수로 나눈 평균 가중값



을 현재의 가구 가중값으로 부여하는 방법이다. t 차 웨이브 가구 H_i 의 구성원이 1차 웨이브 가구들 h_j 의 집합으로 구성되어 있다면, 가구 H_i 에 대한 가중값은 다음과 같이 표현된다.

$$w_i = \sum_j l_{ij} w'_{ij},$$

여기서, $w'_{ij} = \begin{cases} 1/p_j, & \text{가구 } h_j \text{ 가 원표본 가구인 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$ 이며, p_j 는 가구 h_j 의 추출 확률이다.

이 경우 불편추정량을 만족하는 가구 가중값은 다음과 같다.

$$w_i = \sum_j w'_{ij} / C_i, \quad C_i \text{는 가구 } H_i \text{에 속하는 원가구수.}$$

이 방법은 원가구 h_j 에 대해 균등한 연결함수 $1/C_i$ 를 부여하여 t 차 웨이브 가구 가중값을 재산출한다.

② 균등 개인 가중값 방법

균등 개인 가중값 방법(Equal Person Weighting Scheme, EPWS)은 원표본 개인을 기준으로 재분배하는 방식으로, 현재 가구내 원표본가구원 가중값을 원가구원수로 나눈 평균 가중값을 현재의 가구 가중값으로 부여하는 방법이다. 이 방법은 미국의 SIPP와 캐나다 SLID에서 사용되고 있다.

원가구 h_j 에서 온 가구 H_i 의 구성원 k 를 I_{ijk} 라 하면, 가구 H_i 에 대한 가중값은 다음과 같이 표현된다.

$$w_i = \sum_j \sum_k l_{ijk} w'_{ijk}$$

여기서, $w'_{ijk} = \begin{cases} 1/p_j, & \text{가구 } h_j \text{ 내 가구원 } k \text{가 원표본 가구원인 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$ 이며, p_j 는 가구 h_j 의 추출확률이다.

이 경우 불편추정량을 만족하는 가구 H_i 가중값 w_i 는 다음과 같다.

$$w_i = \frac{1}{M_i} \sum_j M_{ij} w'_{ij}$$

여기서 M_{ij} 는 원가구 h_j 에서 온 가구 H_i 의 가구원 수이며, $M_i = \sum_j M_{ij}$ 로 H_i 가구내 원가구원 수이다.

2) 횡단면 개인 가중값

t 차년도에 횡단면 분석에 대표성 및 효율성 측면에서 원표본과 연관된 개인들 (associated persons)에게 가중값을 부여하는 것으로 동일한 가구 내 모든 가구원에 대해서는 동일한 가중값이 부여된다.

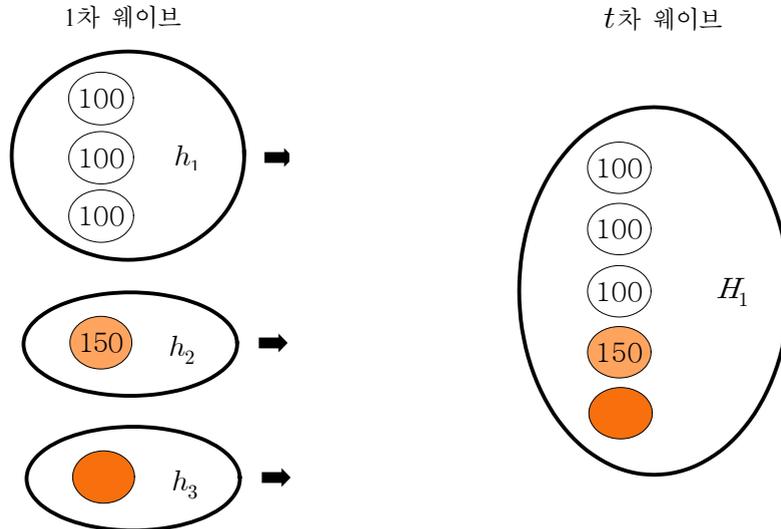
가구 가중값과 동일하며, 자세한 내용은 Kalton and Brick(1995)를 참조하기 바란다. 원가구 h_j 에서 온 가구 H_i 의 구성원 k 를 I_{ijk} 라 하면, 가구 H_i 의 개인 가중값은 다음과 같이 표현된다.

$$w_i = \sum_j \sum_k l_{ijk} w'_{ijk}$$

여기서 w'_{ijk} 는 가구 h_j 가 원표본 가구이면 $1/p_j$ 이고, 그렇지 않은 경우는 0이다.

모집단 총합에 대해 불편추정량을 제공하는 조건 $\sum_j \sum_k l_{ijk} = 1$ 을 만족하는 연결함수 l_{ijk} 는 모든 원가구원에게 $l_{ijk} = 1/M_i$ 를 적용하는 것으로, 이 경우에 가구 H_i 의 신규 진입자를 포함한 모든 가구원에게 $w_i = \frac{1}{M_i} \sum_j \sum_k w'_{ijk}$ 가중값을 부여한다. 이 방법은 가구 가중값의 균등개인 가중값 방법(EPWS)과 동일하다. 다른 방법은 $l_{ijk} = 1/C_i M_{ij}$ 를 적용하는 것으로, 이 경우에는 $w_i = \sum_j \sum_k w'_{ijk} / C_i M_{ij}$ 를 모든 가구원에게 동일하게 가중값으로 부여한다. 이 방법은 가구 가중값의 균등가구 가중값 방법(EHWS)과 동일하다.

균등 가중값 공유방법의 작성 예를 [그림 1-2]에 나타내었다. 횡단면 가구 가중값과 횡단면 개인 가중값이 동일함을 알 수 있다.



- $C_1 = 3$, $M_{h_1} = 3$, $M_{h_2} = 1$, $M_{h_3} = 1$, $M_i = 5$
- 횡단면 가구 가중값
 - ① EHWS: $w_1 = \frac{1}{3}(100 + 150) = 83.33$
 - ② EPWS: $w_1 = \frac{1}{5}(3 \times 100 + 1 \times 150 + 1 \times 0) = 90$
- 횡단면 개인 가중값
 - ① $w_1 = \frac{1}{5}(3 \times 100 + 1 \times 150 + 1 \times 0) = 90$
 - ② $w_1 = \frac{1}{3 \cdot 3}(3 \times 100) + \frac{1}{3}(1 \times 150) + \frac{1}{3}(1 \times 0) = 83.33$

[그림 1-2] 균등 가중값 공유방법의 작성 예

제3절 가구패널조사 사례연구

1. 가구패널조사 횡단 가중값 작성사례

패널조사는 종단면 및 횡단면에 대해 각각 가구 수준 가중값과 개인 수준 가중값을 작성한다. 각 패널조사별로 작성방법에 차이가 존재하는데, 본 절에서는 주요 가구패널의 횡단가중값 작성을 중심으로 정리하였다.

가. 한국노동패널조사(Korea Labor and Income Panel Study, KLIPS)

한국노동패널(KLIPS)은 한국노동연구원에서 1998년부터 시작하여 비농촌지역에 거주하는 패널표본구성원(5000가구에 거주하는 가구원)을 대상으로 매년 1회 경제, 노동, 소득, 소비 등에 대하여 조사하는 패널조사이다. KILPS의 가중값은 Duncan(1995)의 방법을 적용하였으며, 따라서 미국의 PSID에서 사용한 방법과 유사하다. KILPS에서는 Duncan 방법을 사용하여 종단면 개인 및 가구 가중값을 산출한 후 가구 가중값을 비표본 가구원을 포함한 모든 가구원에게 할당하는 방식으로 횡단면 개인 가중값을 작성한다(<표 1-2> 참조). 자세한 내용은 강석훈(2003)과 한국노동패널 유저가이드를 참조하기 바란다.

<표 1-2> KLIPS 가중값 작성방법

(1차 웨이브)

- 가구 가중값
 - 추출확률과 응답확률(최종 조사가가구수/총접촉가구수)을 고려함
- 가구원 가중값
 - 추출된 가구에서는 모든 가구원이 응답하므로 동일 가구 내 가구원에 대해 가구 가중값 부여

(2차 웨이브 이후)

- 가구원 가중값
 - 응답률을 이용한 가구원 무응답 조정 가중값 산출
 - 무응답 조정 가중값 = 1차 웨이브 가구원 가중값 × (1/응답확률)
 - * 무응답조정은 가구와 가구원의 정보에 대해 로짓모형을 이용하여 가구원의 응답확률 추정
 - * 비표본 가구원이나 1차 이후 출생한 가구원은 무응답조정에 미포함
- 가구 가중값
 - 가구원 가중값의 평균을 이용하여 가구 가중값 산출
 - 가구 가중값 = 원표본 가구원 가중값 합/전체 가구원수*
 - * 가구 가중값에 사용하는 전체 가구원에 출생 가구원은 포함되지 않으며, 원표본 가구원과 비표본 가구원만 이용
 - 비표본가구원(결혼, 동거 등)에게는 0의 가중값을 부여
 - 출생 가구원에게는 태어난 조사년도의 가구 가중값을 개인 가중값으로 부여

(횡단면 개인 가중값)

- 가구원 가중값의 평균을 이용하여 산출된 가구 가중값을 모든 가구원에게 부여
 - 가구가 응답하는 경우 가구원이 누락되는 경우가 거의 없으며, 가구의 응답이 조사 참여에 더욱 결정적이라는 판단에 근거



나. 재정패널조사(National Survey of Tax and Benefit, NaSTaB)

재정패널조사는 한국조세연구원에서 주관하며 가구의 소득, 지출, 부채 등 가구 재정 전반에 관련된 내용을 조사하고 있다.

〈표 1-3〉 NaSTaB 가중값 작성방법

(1차 웨이브)

- 가구 가중값과 가구원 가중값
 - 추출확률과 무응답 조정을 고려하여 가구 가중값을 산출
 - 산출된 가구 가중값을 가구내 가구원 가중값으로 부여
 - 가구 가중값에 대한 레이킹(raking) 보정
 - 모집단 정보로 2008년 가구 및 인구 추계통계 이용
 - 레이킹 보정에 사용된 가구변수는 각 지역별 가구주의 성, 연령, 가구원 수
- * 연령은 (39세 미만, 40~49세, 50~59세, 60세 이상), 가구원 수는 (4인 미만, 4인 이상)으로 구분

(2차 웨이브)

- 가구 가중값 산출
 - 1차년도 가구 가중값을 기초로 하여 다음과 같이 가구 형태별로 가중값을 재부여
 - 원 패널가구: 1차년도 최종가중값 그대로 부여
 - 대체가구: 대응되는 탈락가구의 1차년도 최종 가중값 사용
 - 분가가구는 분가 사유에 따라 구분하여 가중값을 부여
 - 분가사유가 결혼인 경우는 분가이전 원가구의 1/2
 - 기타 사유의 분가가구는 원가구 가중값 부여
 - 분가사유가 결혼인 가구에 대한 무응답보정* 실시
- * 연간 총소득 4천만 원을 기준으로 두 그룹으로 나누어 무응답 보정 적용
- 2009년 가구추계자료를 이용하여 레이킹 보정 실시
- 조사대상 가구원 가중값
 - 횡단면 가구 가중값을 가구내 조사대상 가구원(소득이 있는 15세이상 가구원)에 모두 부여

(3차 웨이브)

- 가구 가중값 산출: 가구형태에 따라 가중값을 산출
 - 원패널가구, 대체가구, 1차 웨이브 분가가구에 대해 탈락으로 인한 무응답보정 실시
 - 2차 웨이브 분가가구*에 대해 2차 웨이브와 동일한 방법으로 부여
- *결혼으로 분가한 가구에 대해 2차 웨이브 원가구 연간총소득 4,500만원을 기준으로 두 그룹으로 나누고 무응답보정 실시
- 2010년 가구추계자료를 이용하여 레이킹 보정 실시
- 조사대상 가구원 가중값
 - 횡단면 가구 가중값을 가구 내 조사대상 가구원(소득이 있는 15세 이상 가구원)에 모두 부여

제주도를 제외한 전국 15개 시도에 거주하는 일반가구와 가구원을 모집단으로 하여, 5,014가구의 소득이 있는 가구원(약 7,000명)을 대상으로 2008년에 1차 조사를 실시하였다. 조사설계 등에 관련해서 재정패널조사 기초분석보고서(2008~2010)를 참조할 수 있다. 2010년 3차 조사에서 원표본 유지율이 83.2%(4,172가구)이며, 가구 및 가구원의 횡단면 가중값과 종단면 가중값을 작성하며, 매년 조사에 응답한 가구를 기준으로 횡단면 가중값을 작성한다(<표 1-3> 참조).

다. PDIS (The Panel Study of Income Dynamics)

PSID는 미국 미시간대학에서 주관하며, 복지, 소득, 고용 등 광범위한 정보를 포함하는 1968년부터 시작된 가장 역사가 긴 패널조사이다. Duncan(1995)이 제안한 방법을 사용하여 종단 가중값을 작성하며, 횡단면 분석 시 종단 가중값을 이용한다(<표 1-4> 참조).

<표 1-4> PSID 가중값 작성방법

(1차 웨이브)

- 가구 가중값과 가구원 가중값
 - 추출확률, 무응답 조정, 벤치마킹 조정으로 가구 가중값 산출
 - 가구 가중값을 모든 가구원 가중값으로 부여

(2차 웨이브 이후)

- 가구원 가중값
 - 1차 웨이브 가구원 가중값에 대해 무응답 조정하여 가구원 가중값 산출
 - 무응답조정*은 로짓모형으로 가구와 가구원의 정보 이용률 이용
 - * 비표본 가구원이나 1차 이후 출생한 가구원은 무응답 조정에 미포함
- 가족(family) 가중값
 - 조사에 참여한 모든 가족 구성원의 개인 가중값 평균으로 가족 가중값 산출
 - 가족 가중값 = 원표본 가구원 가중값 합/전체 가구원수*
 - * 전체 가구원수는 원표본가구원과 비표본가구원의 합이며 새롭게 태어난 가구원은 제외
- 비표본 가구원과 1차 웨이브 이후 출생한 가구원 가중값
 - 새롭게 진입한 비표본 가구원 가중값은 0 또는 가족 가중값을 부여
 - 출생 가구원은 응답대상이 되면 가구 가중값을 부여

(횡단 개인 가중값: 1997-2009)*

- PSID는 1997-2009년에 대해 가중값 공유방법을 이용하여 횡단 개인가중값을 구축
- 횡단면 가중값 시작시점으로 종단면 가족 가중값을 사용하였으며, 2단계의 조정을 적용
 - (1단계) 조정된 가족 가중값을 작성하여 표본 및 비표본 가구원의 기본 가중값으로 부여
 - (2단계) 기존 개인 가중값에 대한 사후층화 실시

* Heeringa, S.G. et al. (2011), PSDI Cross-sectional Individual Weights, 1997-2009 참조



라. BHPS (The British Household Survey Panel)

BHPS는 영국의 Essex 대학의 ISER(Institute for Social and Economic Research)에서 주관하는 패널조사로 1991년 시작되어 주거상태, 소비지출, 고용, 재무 등을 조사한다. BHPS는 개인 수준 가중값으로 응답 개인 가중값과 기록 개인(enumerated individuals) 가중값을 작성하며, 대리(proxy) 응답자와 전화 응답자에 대해 응답 개인 가중값은 “0”, 기록 개인 가중값은 양의 값을 부여한다. 응답개인과 기록개인에 대해서는 횡단 및 종단 가중값을 제공하나, 가구에 대해서는 횡단 가중값만 제공한다(Taylor et al., 2010 참조).

〈표 1-5〉 BHPS 가중값 작성방법

(1차 웨이브)

- 가구 가중값
 - 설계가중값에 가구 무응답 조정, 사후층화를 적용
 - 가구 무응답 조정은 지역과 거주형태 변수를 이용하여 무응답 조정층을 설정
 - 사후층화는 임차유형, 가구원수, 자동차수를 층화변수로 이용
 - 가중값의 총합이 총 표본수가 되도록 스케일링 조정 실시하여 가구 가중값으로 산출
- 가구원 가중값
 - 기록 개인에 대해 설계가중값에 가구 무응답 조정 적용
 - 응답 개인에 대해 설계가중값에 가구 무응답 조정과 개인 무응답 조정* 적용
 - * 개인무응답조정은 지역, 임차유형, 가구원수, 혼인상태, 취업상태, 성, 연령을 이용하여, 로짓모형 사용
 - 무응답 조정을 적용한 기록개인과 응답개인에 대해 사후층화와 스케일링 조정 실시
 - * 사후층화에 임차유형, 가구원수, 자동차수, 성, 연령을 층화변수로 이용

(2차 웨이브 이후)

- 기록 개인 가중값
 - 사후층화 후 1차 웨이브 기록개인가중값을 초기가중값으로 원표본에 부여, 패널마모조정* 실시
 - *패널마모 조정변수로 가구주, 가구, 개인특성을 이용
 - 균등 가중값 분배방법을 적용하여 모든 기록 개인들(새로 진입한 개인포함)에게 가중값 부여
- 응답개인 가중값
 - 가구내 무응답* 조정
 - * 거부, 대리응답, 인터뷰 불가능 경우를 무응답으로 처리
- 가구 가중값
 - 기록개인 가중값을 총 가구 수를 이용하여 조정

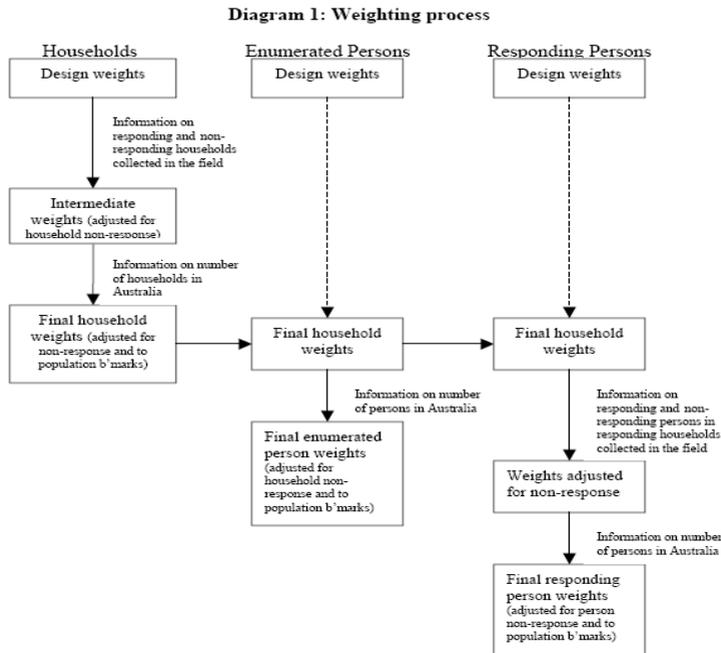
마. HILDA(The Household, Income and Labour Dynamics in Australia Survey)

HILDA는 호주 멜버른 대학에서 주관하는 가구패널조사로 경제, 주관적 웰빙, 노동 등에

대한 내용을 조사하며, 2001년 1차 웨이브 패널에 7,682가구와 19,914가구원이 포함되었다. HILDA도 BHPS와 같이 개인 수준 가중값으로 기록 개인(enumerated person) 가중값과 응답 개인가중값을 작성한다(<표 1-6>, [그림 1-3], [그림 1-4] 참조). 기록 개인은 응답자, 무응답자 그리고 어린이들을 포함한다. 자세한 내용은 Watson (2002,2004)을 참조하기 바란다.

<표 1-6> HILDA 가중값 작성방법

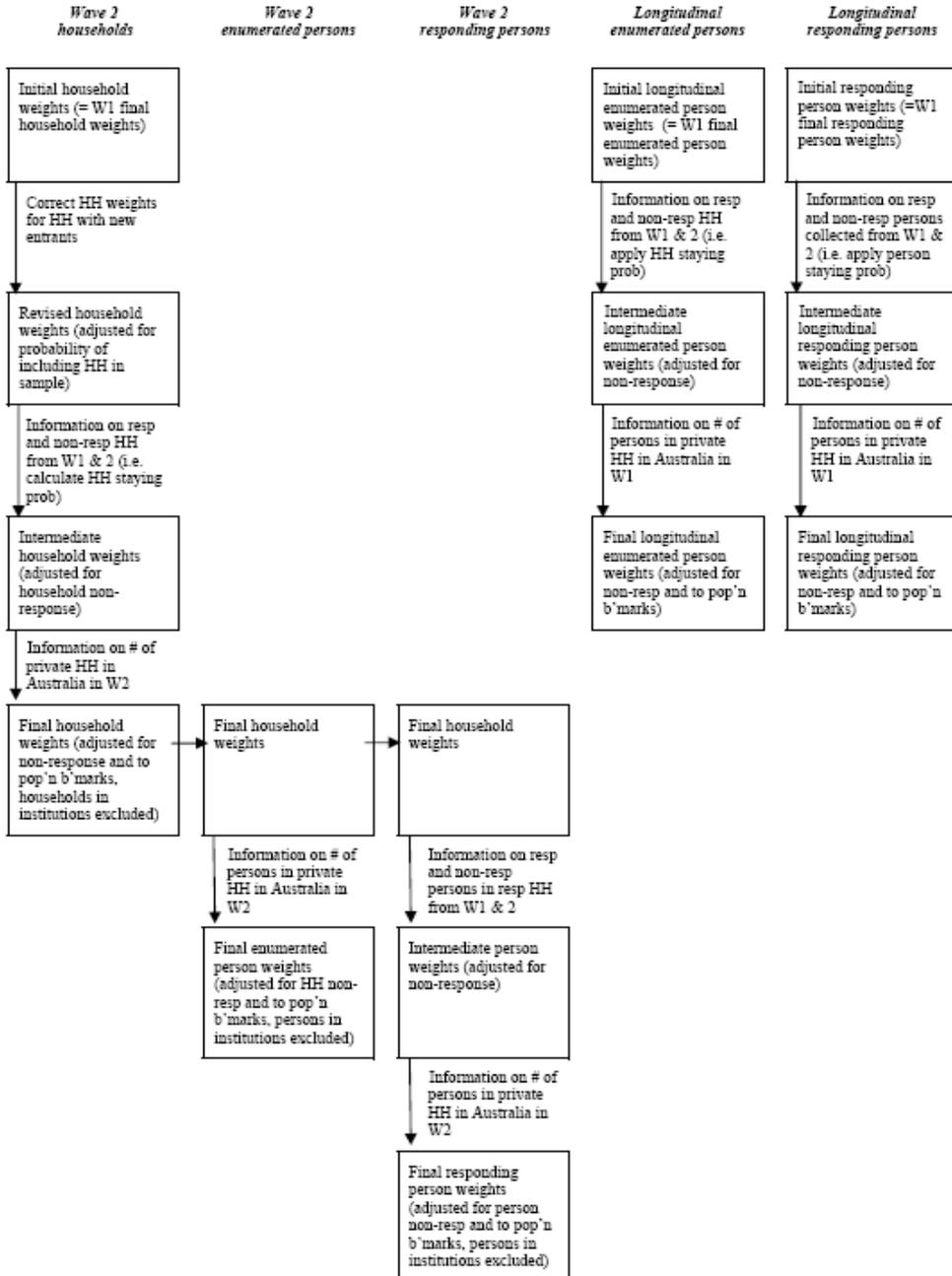
- (1차 웨이브)
- 가구 가중값과 가구원 가중값
 - 가구의 설계가중값에 무응답 조정과 벤치마킹 조정을 실시하여 가구 가중값 산출
 - 가구 가중값을 모든 가구원 가중값으로 부여
- (2차 웨이브 이후)
- 가구 가중값
 - 초기 가구 가중값으로 1차 웨이브 최종 가중값을 부여
 - 새로 진입한 가구원을 반영하기 위해 모형방법으로 초기 가구 가중값을 조정
 - 조정된 가구 가중값에 무응답 조정과 벤치마킹 조정 실시
 - 가구원 가중값
 - 최종 가구 가중값을 모든 가구원 가중값으로 부여
 - 무응답 조정과 벤치마킹 조정을 실시



[그림 1-3] 1차 웨이브 가중값 작성과정



Figure 1: Overview of wave 2 weighting



[그림 1-4] 2차 웨이브 가중값 작성과정

2. 가구 변화요인별 가중값 부여방법

패널조사 사례연구를 통해 가중값 작성방법은 일반적인 독립조사와 유사한 단계를 거쳐서 작성되며, 가구원 증가, 분가 등의 변화에 따른 가구 가중값 또는 가구원 가중값의 조정과정이 일반적인 조사와는 다르게 포함되어 있음을 알 수 있다. <표 1-7>에 주요 가구패널조사의 가중값 조정방법에 대해 간략하게 정리하였다.

<표 1-7> 주요 가구패널조사의 횡단 가중값 조정방법

구분	시작 년도	조사대상	조정방법
PSID(미국)	1968	가구주	비표본 가구원 가중값으로 0 또는 가족 가중값을 부여
BHPS(영국)	1991	가구내 모든 성인들	가중값 공유방법(가구원 가중값 조정 → 가구 가중값)
SHP*(스위스)	1999	가구내 모든 성인들	가중값 공유방법(가구원 가중값 조정 → 가구 가중값)
HILDA(호주)	2001	가구내 모든 성인들	모형 방법(가구 가중값 조정 → 가구원 가중값)
SOEP*(독일)	1984	가구내 모든 성인들	모형 방법(가구 가중값 조정 → 가구원 가중값)

* Swiss Household Panel

* German Socio Economic Panel

또한 가구 및 가구원 가중값에 대한 조정 방법은 가구변화요인과 조사별로 다소 상이한 점이 있으며, 본 절에서는 Schonlau et al. (2010)을 참조하여 가구변화요인에 따른 가중값 조정방법과 영향력에 대해 정리하였다. 각 요인별로 조정방법이 가중값에 미치는 영향은 <표 1-8>과 <표 1-9>를 참조하기 바란다.

가. 가구변화 요인과 가중값 처리방법

- 일반 진입 가구원(regular household entrants)
 - 가구에 새로 진입한 가구원으로 패널 시작 시점에 목표모집단에 존재하였으나 표본으로 추출되지 않은 개인으로 출생/입양, 최근이민자와는 구별된다.
 - 가구 가중값과 가구원 가중값을 감소시키는 영향을 미친다.
- 출생(birth) 및 입양(adoption)
 - 목표모집단의 변화를 나타내므로 일반 진입가구원과는 구별된다.
 - 공유방법은 추가 가중값을 부여하는데, BHPS는 부모 가중값의 평균값을 부여한다.
 - 모형방법에서는 가구 가중값이 그대로 유지된다.



- 최근 이민자(recent immigration)
 - 목표모집단의 변화를 나타내므로 일반 진입과 구별되어야 하나, BHPS와 SOEP는 구별하지 않는다.
 - HILDA와 SHP는 구분하여 취급하는데, SHP는 개인 가중값을 그대로 두고 이민자 가중값으로 가구내 원표본 가구원 가중값의 평균값을 부여하며, HILDA는 가구와 가구원 가중값은 그대로 두고 이민자에 대해 가구 가중값을 부여한다.
- 사망(deaths)과 이민(emigration)
 - 모형방법은 사망자 가중값은 단순하게 제거하며, 가구 가중값은 그대로이다.
 - 공유방법은 사망자 가중값을 단순하게 제거하며, 나머지 가구원 가중값은 그대로 유지, 가구 가중값은 가구원 가중값을 이용하여 재계산한다.
- 합가(household mergers)
 - 합가는 두 가지 타입으로 구분할 수 있으며, 상관없는 표본 가구원들끼리 합치는 경우(unrelated merge)와 패널 초기에는 단독가구로 있다가 다시 합가를 이루는 경우(move-back merge)로 학업을 마친 후 부모 가구로 들어오는 경우이다. 두 경우 모두 드물게 나타난다.
 - 모형방법은 unrelated merge인 경우, 일반적인 진입가구원과 동일한 방법을 적용하면 되며, 다만 추출확률이 알려져 있다는 점에 차이가 있다. move-back merge 경우는 가구 추출확률이 변하지 않으므로 가구 가중값은 그대로 유지된다.
 - 공유방법에서는 기본원칙 대로 개인 가중값을 공유한 후 가구 가중값을 산출한다.
- 가구 분리(household splits)
 - 자녀가 성장하여 독립하거나, 이혼으로 인하여 가구가 분리되는 경우를 말한다.
 - 공유방법에서 분가 가구원의 가중값은 공유되지 않은 가중값을 가져가게 되므로, 원표본가구원과 출생 또는 입양 가구원만 가중값을 가지며, 나머지 가구원은 가중값 “0”을 가지고 가구를 형성하게 된다. 분가가구에 원표본 가구원이 한 명도 없는 경우에는 공유방법을 사용할 수가 없다.
 - 모형방법은 일반적인 진입가구원 방법을 적용하면 된다.
- 고아 응답자들(orphan respondents)
 - 원표본 가구원이 한명도 없는 가구 내 응답자들을 말한다.
 - 공유방법으로는 이러한 응답자들에게는 가중값을 부여하는 것이 불가능하며 “0”으로 부여하면 표본규모가 감소하게 되는 문제가 발생하게 된다.
 - 고아 응답자들의 추출확률을 추정하고, 가중값 공유방법을 사용하는 혼용(hybrid) 방법을 사용할 수도 있다.

〈표 1-8〉 가구변화가 횡단 개인 가중값에 미치는 영향

Table 1: The effect of household changes on cross sectional individual weights for different household panels

	BHPS	SHP	PSID	HILDA	SOEP
Method for computing weights	hh weight = average of individual weights	hh weight = average of individual weights	hh weight = average of individual weights	individual weights = household weight	individual weights = household weight
Method for assigning weight to new Entrants	Weight Share	Weight Share	Zero Weight	Modeling	Modeling
Regular Household Entrants	down-weighted	down-weighted	zero weight	down-weighted	down-weighted
Immigrants	like other household entrants	average of (individual) OSM weights	like other household entrants	unchanged	like other household entrants
Birth / adoptions	receive average weight of parents	does not apply (panel is 11 yrs old, weights are assigned at age 14)	average weight of parents; if only one parent: 1/2 weight of head of household	receive household weight	receive household weight
Household Split	unchanged	zero in households without OSM, otherwise unchanged	unchanged	unchanged (splitting hhs receive the same hh weight)	unchanged (splitting hhs receive the same hh weight)
Merging households	unchanged	unchanged	unchanged	"unrelated merge": like regular household entrants	"move back merge": receive weight from new head of household
Death	unchanged for others	unchanged for others	unchanged for others	unchanged for others	unchanged for others

출처 : Schonlau, M. and Kroh, M. (2010)



〈표 1-9〉 가구변화가 가구 가중값에 미치는 영향

Table 2: The effect of household changes on household weights for each household panel.
 Notation: HH= household, OSM = Original sample member, TSM= Temporary sample member

	BHPS	SHP	PSID	HILDA	SOEP
Regular Household Entrants	down-weighted	down-weighted	unchanged (entrant is not OSM)	down-weighted to account for multiple pathways of being selected	down-weighted to account for multiple pathways of being selected
Births / adoptions	average is recomputed	does not apply (panel is 11 yrs old, weights will be assigned at age 14)	average is recomputed	unchanged	unchanged
Immigrants	treated like other household entrants	unchanged	treated like other household entrants	unchanged	treated like other household entrants
Household Split	average is recomputed	households without OSM: 0. Otherwise weight share	Averages are computed for each household separately	The same HH weight is carried over to both new households	The same HH weight is carried over to both new households
Merging households	Average is computed for merged household	Average is computed for merged household	Average is computed for merged household	"unrelated merge": like regular household entrants	"move-back merge": former household weight of the new head of household is used
Death	OSM death: down-weighted. TSM death: up-weighted	OSM death: down-weighted. TSM death: up-weighted	average is recomputed	unchanged	unchanged

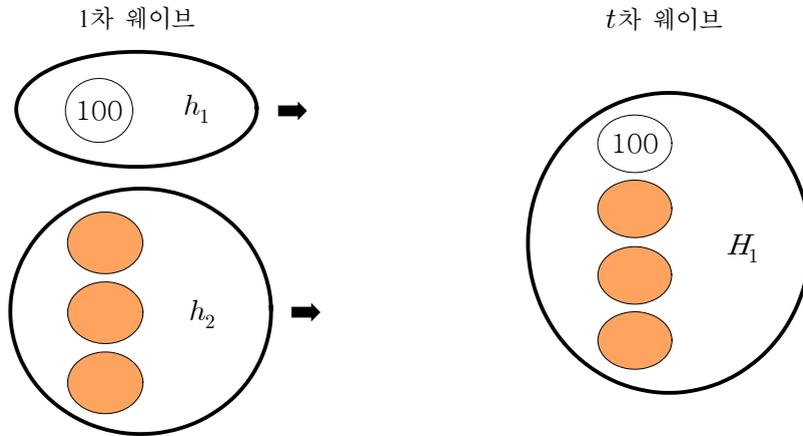
출처 : Schonlau, M. and Kroh, M. (2010)

3. 균등 가중값 공유방법 적용 사례

본 절에서는 몇 가지 경우에 대해 균등 가중값(EHWS와 EPWS)을 적용한 사례를 살펴보았다.

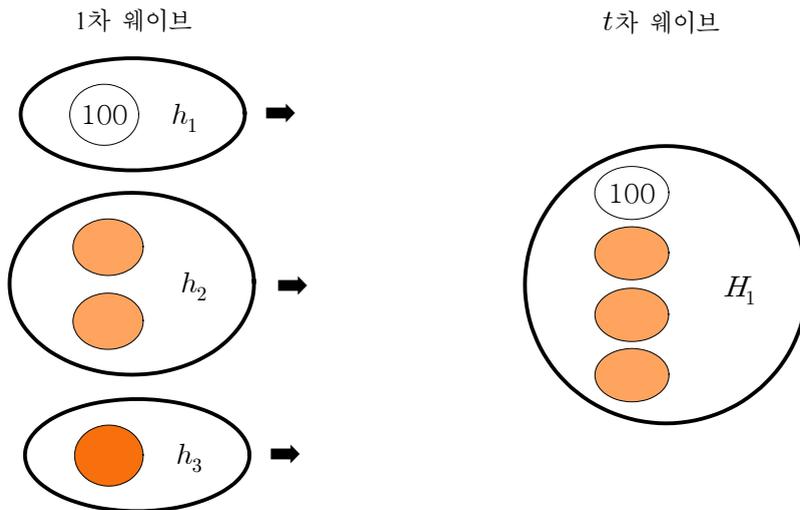
(사례1) 일반적인 가구 진입(regular household entrants)

- 경우1 : 현재 표본 가구가 1차 웨이브의 원표본가구 h_1 의 1명, 원가구 h_2 의 3명으로 구성



$$\text{EHWS: } w_1 = \frac{1}{2}(100 + 0) = 50, \text{ EPWS: } w_1 = \frac{1}{4}(100 + 3 \cdot 0) = 25$$

- 경우2 : 현재 표본가구가 1차 웨이브 원표본가구 h_1 의 1명, 원가구 h_2 의 2명, 원가구 h_3 의 1명으로 구성

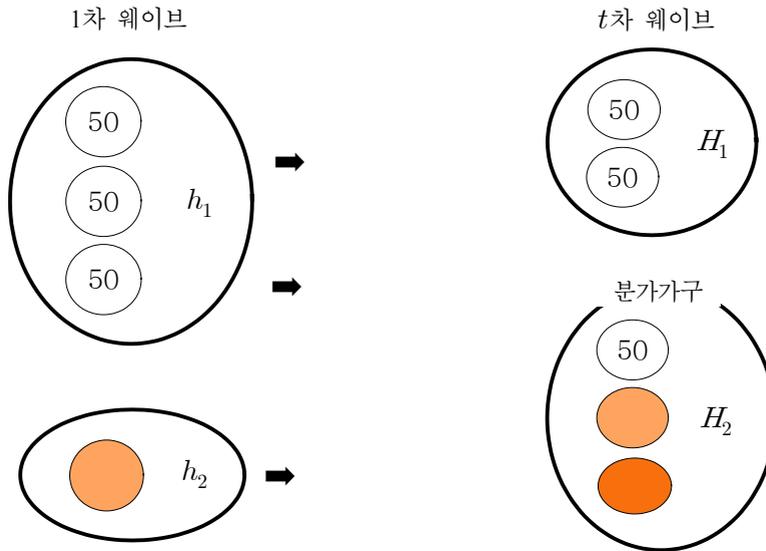


$$\text{EHWS: } w_1 = \frac{1}{3}(100 + 0) = 33.33, \text{ EPWS: } w_1 = \frac{1}{4}(100 + 3 \cdot 0) = 25$$



(사례2) 분가

- 현재 표본가구 H_1 은 원표본 가구 h_1 의 가구원 2명으로 구성, 표본가구 H_2 는 원표본 가구 h_1 의 가구원이 결혼으로 분가가구를 형성하고 아이를 출생한 경우

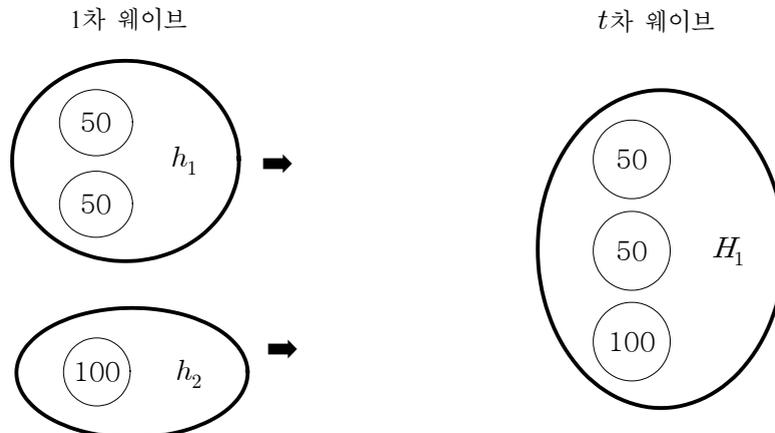


$$w_{H_1} = 50$$

$$\text{EHWS: } w_{H_2} = \frac{1}{2}(50 + 0) = 25, \text{ EPWS: } w_{H_2} = \frac{1}{2}(50 + 0) = 25$$

(사례3) 합가

- 현재 표본가구 H_1 이 원표본가구 h_1 의 2명, 원표본가구 h_2 의 1명으로 구성



$$\text{EHWS: } w_{H_1} = \frac{1}{2}(50 + 100) = 75, \text{ EPWS: } w_{H_1} = \frac{1}{3}(2 * 50 + 100) = 66.67$$

제4절 가계금융조사 분석

본 절에서는 앞에서 소개하였던 가중값 공유방법을 가계금융조사 자료에 적용하여 비표본 개인을 고려한 가중값 조정을 실시하고 추정결과에 미치는 영향력을 살펴 보았다.

1. 가계금융조사 개요

가계금융조사는 가구특성별 자산과 부채 규모, 분포 등 가계의 미시적 재무 건전성을 파악하기 위한 통계청의 패널조사로 조사단위는 가구(household)이나 학업으로 떨어져 살고 있는 미혼자녀와 직장으로 외지에서 살고 있는 배우자를 포함하여 조사하므로 개념적으로는 가족(family) 단위에 가깝다고 할 수 있다.

가. 조사대상 관련 기본개념

- 가구
 - 1인 또는 2인 이상이 모여 주거 또는 생계를 같이하는 사람들의 모임
- 가구원
 - 주민등록 등재여부와 상관없이 사실상 가구를 구성하고 있는 구성원
 - 취업이나, 학업 등의 이유로 동거하지는 않지만, 생계를 같이 하는 가구원은 조사대상
 - 기존 패널가구에서 분가한 가구원과 함께 살게 된 가구원도 조사대상
- 조사대상가구 - 전국 동읍면에 거주하는 1인 이상의 표본가구
- 조사제외가구
 - 학업 때문에 떨어져 사는 미혼자녀나 직장 때문에 외지에 살고 있는 배우자 가구
 - 15세 미만 가구원으로 구성되거나, 비혈연 가구원으로 구성된 가구
 - 사회시설에 거주하거나 외국인으로 구성된 가구
- 조사제외 가구원
 - 같은 집에 살고 있으나 소득 및 지출을 달리하는 친인척
 - 가사도우미, 운전사 등 고용인과 하숙생, 혈연관계가 없는 동거인



나. 분석 자료 및 기초분석

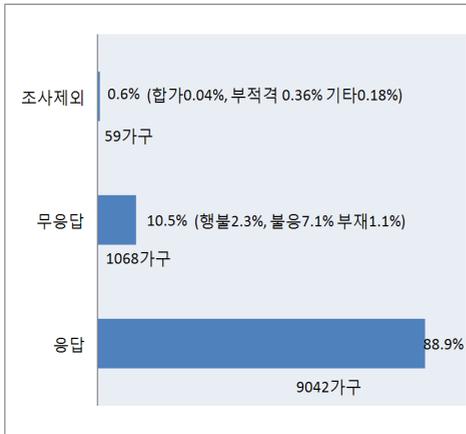
분석자료는 2011년 2차 가계금융조사 자료이며, 패널이탈로 인해 추가된 표본은 제외하였다. 2차 웨이브에서 원표본가구의 변화상태는 <표 1-10>에 수록하였다. 원표본가구의 2차 웨이브 가구상태는 유지, 행방불명, 불응, 부재, 합가, 부적격, 기타로 구분되어 있다. 여기서 부적격은 1인가구에서 조사가 불가능한 경우와 조사대상 제외가구인 경우이며, 기타는 1인가구 사망 및 이민 경우이다. 표본상태는 표본가구의 2차 웨이브 상태를 나타내는 것으로 원가구, 분가가구(원가구의 1차 분리)로 나누어진다.

1차 웨이브 10,000가구는 합가로 인해 9,996가구가 되고, 분가로 173가구가 추가되어 2차 웨이브 패널가구는 총 10,169가구이다. 이 중 조사제외 가구는 59가구(0.6%)로 나타났다. 합가 4가구, 부적격 34가구, 기타 18가구로 구성되어 있다. 나머지 조사대상 가구인 10,110가구(99.4%) 중 응답가구는 9,045가구이며, 무응답가구는 1,068가구(행방불명 237가구, 불응 717가구, 부재 111가구)이다.

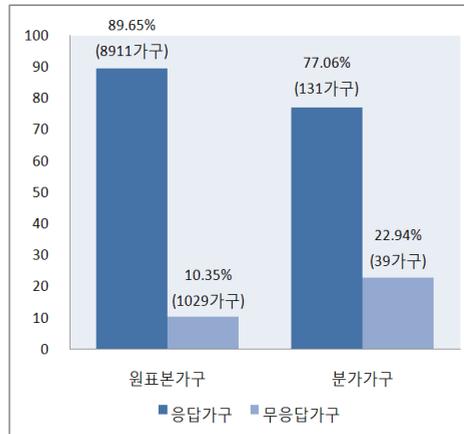
<표 1-10> 2차 웨이브에서 원표본 가구 변화상태

가구상태 표본상태	조사대상(10,110가구)				조사제외(59가구)			총합
	응답	무응답			합가	부적격	기타	
	유지	행방불명	불응	부재				
원표본가구	8,911	235	694	100	(8→) 4	34	18	(10,000→) 9996
분가가구	131	2	26	11	0	3	0	173
소계	9,042 (88.95)	237 (2.33)	720 (7.05)	111 (1.09)	4 (0.04)	37 (0.36)	18 (0.18)	10,169

[그림 1-5]에서 원표본가구와 분가가구의 응답여부 분포를 보면 조사대상 10,110가구 중 무응답률은 원표본가구에서 10.35%이며, 분가가구에서 22.94%로 분가가구의 무응답률이 2배가량 높게 나타났으며, 불응으로 인한 무응답이 많다는 것을 알 수 있다.



(a) 가구상태(응답, 무응답, 조사제외) 분포



(b) 원표본가구와 분가가구의 응답분포

[그림 1-5] 2차 웨이브 가구상태 분포

원표본 가구원의 변동 상태는 <표 1-11>에 수록하였다. 2차 웨이브 가구원 30,872명 중 원표본 가구원은 29,908명이며 원표본 가구로 새로 진입한 가구원은 964명이다. 원표본 가구원 29,908명 중 그대로 유지된 가구원은 29,363명이며, 분가한 가구원은 192명, 사망한 가구원 104명, 기타감소 249명이다. 새로 진입한 가구원 964명은 출생으로 232명, 그 외 결혼 등으로 인한 기타증가 732명이다.

<표 1-11> 원표본 가구원 여부와 가구원 변동사항

원표본가구원 여부와 가구원변동사항	유지*	분가	사망	기타감소	기타증가*	출생	총합
원표본가구원	29,363	192	104	249	0	0	29,908
새로진입	0	0	0	0	732	232	964
총합	29,363	192	104	249	732	232	30,872

* 기타증가에서 누락, 조사착오(143명)는 유지로 재분류

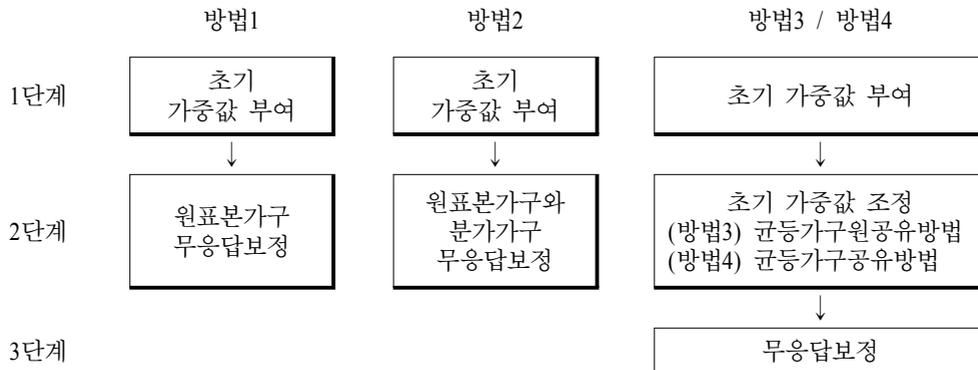
* 단순합가, 결혼, 기타증가를 모두 합하여 기타증가로 재분류



2. 횡단 가구 가중값 작성방법 결과 비교

가. 가중값 작성방법

2차 웨이브 초기 가구 가중값으로 2010년 최종 가중값을 부여하고 4가지 방법으로 가중값을 산출하였다. 분가가구의 초기 가중값으로는 분가이전 원표본 가구 최종 가중값을 부여하였다. 분석 목적이 가중값 분배 적용 효과를 보고자 하는 것이므로 무응답 조정까지 실시하였으며, 사후조정은 실시하지 않았다. 무응답 조정은 2011년 가계금융조사에서 사용한 33개의 무응답층을 그대로 적용하였다. 가중값 공유방법으로 균등 가구 가중값과 균등가구원 가중값 방법이 있으며 적용한 방법은 [그림 1-6]에 나타내었다. 방법1과 방법2는 가중값 공유방법을 적용하지 않은 경우이며, 방법3과 방법4는 각각 균등가구원 공유방법과 균등가구 공유방법을 이용한 경우이다. 방법3은 가구내 원표본 가구원들의 가중합을 가구원수로 나누어 가구 가중값을 조정한 것이다. 방법4는 결혼증가가 없는 가구는 원가구 가중값을 부여하고 결혼으로 증가가 있는 가구는 원가구 가중값의 1/2을 부여하는 방식으로 조정한 것이다. 방법1은 2011년 가계금융조사의 무응답 보정방법으로 원표본가구에 대해서만 무응답보정을 실시한 것이고, 방법2는 분가가구에 대해서도 무응답 보정을 실시한 경우이다.



[그림 1-6] 가중값 적용 방법

나. 가중값 분포 비교

1차년도 최종 가중값을 2차년도 초기 가중값으로 부여하는 경우, 비표본가구원이 발생한 가구의 초기 가중값이 크게 부여되는 가능성이 있으며, 이것은 가구원수 또는 가구주 연령과 같은 가구 특성 분포에 영향을 미칠 수 있다. <표 1-12>의 방법별 가중값 분포를 살펴보면, 1단계 초기 가중값을 그대로 이용하여 무응답조정을 실시한 경우보다

공유방법으로 조정 후 무응답조정을 실시한 경우가 가중합은 감소하고, 표준편차는 조금 커진다는 것을 알 수 있다. 방법별로 1가구의 평균 가중값(평균 대표가구수)은 방법(1)과 방법(2)가 각각 1855가구, 1863가구이며 방법(3)과 방법(4)는 각각 1819가구, 1850가구인 것으로 나타났다. 방법별 가중값 분포는 [그림 1-7]에 나타내었다.

〈표 1-12〉 방법별 가중값 분포

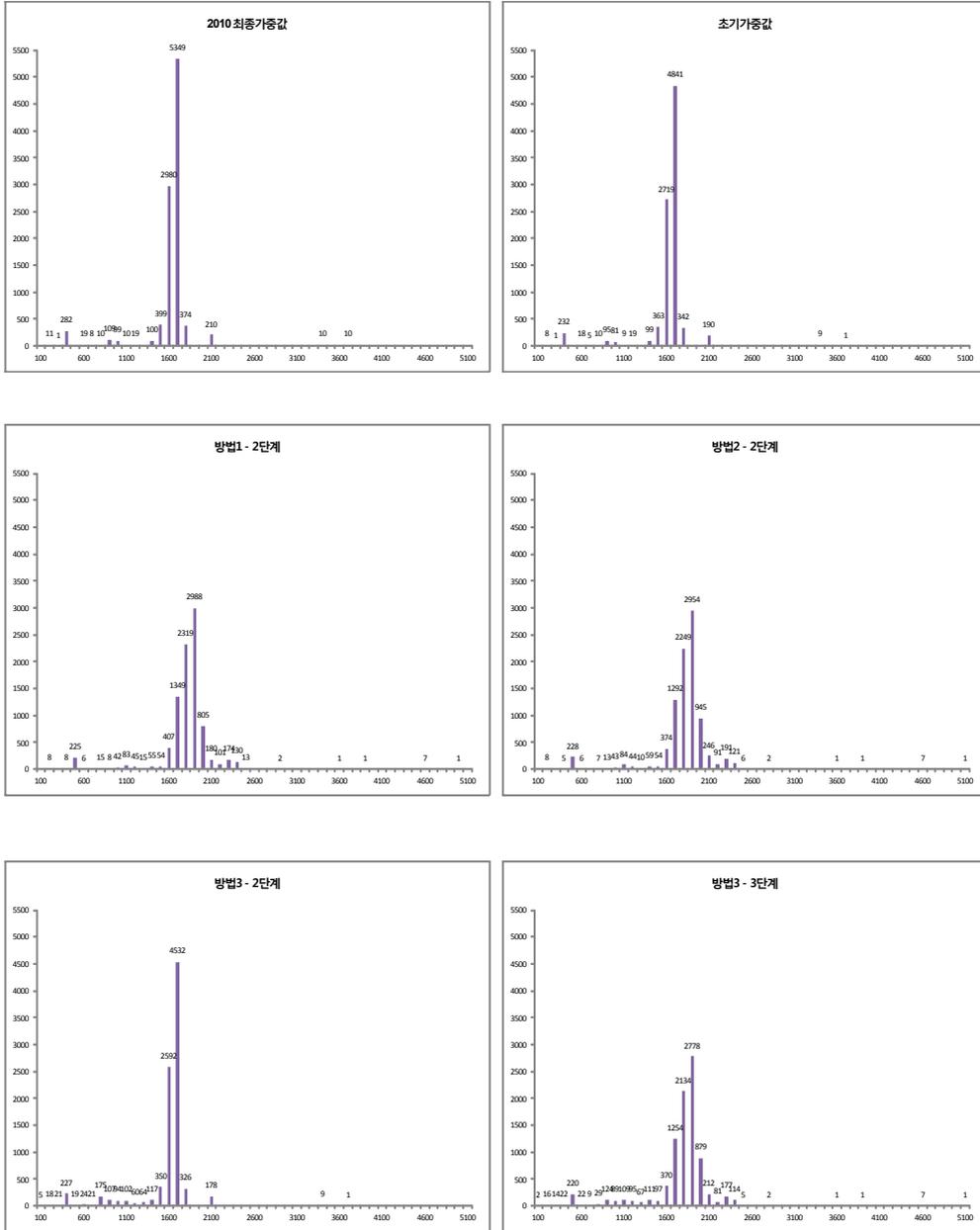
구 분	평균	중앙값	최소값	최대값	범위	표준편차	합
2010년 최종가중값	1664.84	1713.67	253.85	3786.85	3533	272.60	16,648,377
2011년 초기가중값	1667.39	1713.67	253.85	3786.85	3533.00	256.44	15,076,548
방법(1)	1854.98	1895.29	291.93	5074.38	4782.45	305.21	16,772,698
방법(2)	1863.04	1900.07	297.00	5112.25	4815.25	303.71	16,845,594
EPWS 공유	1628.39	1711.76	156.84	3786.85	3630.01	304.33	14,723,877
방법(3)	1818.60	1890.30	181.94	5112.25	4930.31	351.74	16,443,763
EHWS 공유	1656.44	1713.67	233.77	3786.85	3563.08	272.85	14,977,567
방법(4)	1850.45	1895.29	266.29	5112.25	4845.96	319.68	16,731,759

가중값 공유방법이 가구변화에 따라 미치는 영향을 알아보기 위하여 변화요인별 발생 가구수와 가중값 변화를 살펴보았다(〈표 1-13〉 참조). 조사대상가구에서 응답한 9,042가구 중 88%에 해당되는 7,922가구가 가구원 변동없이 그대로 유지되었으며, 6%에 해당되는 531가구에서 비표본가구원이 발생하였다. 원표본 가구원의 사망이 발생한 가구는 95가구로 1%정도였으며, 약 4%에 해당되는 328가구에서 기타감소가 발생하였다. 가구원 증가, 사망, 기타감소에서 공유방법을 적용한 경우 가중값이 감소 조정되었으며, 특히 비표본 가구원이 발생한 증가에서 가중값 감소가 많이 이루어짐을 알 수 있다.

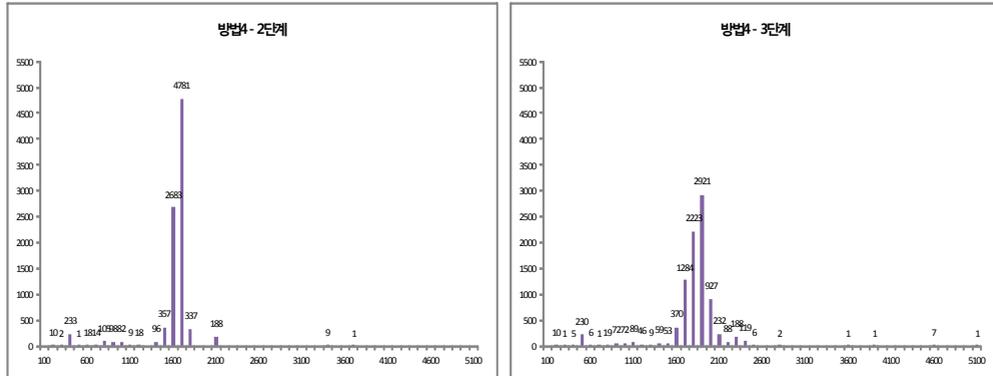
〈표 1-13〉 변화요인별 가구수와 가구 가중값 변화

변화요인과 가구		가구수	초기 가중값	EPWS	EHWS
유지	원가구	7,871	13,130,323	13,130,323	13,130,323
	분가가구	51	85,857	85,857	85,857
	소계	7,922	13,216,180	13,216,180	13,216,180
증가*	원가구	453	746,255	457,536	697,469
	분가가구	78	127,176	63,224	76,981
	소계	531	873,431	520,760	774,450
사망	원가구	95	160,286	153,809	160,286
	분가가구	0	-	-	-
감소	원가구	328	539,545	516,869	532,606
	분가가구	0	-	-	-

* 변동요인 중 증가는 결혼 등의 사유로 비표본가구원이 발생한 경우임



[그림 1-7] 가중값 분포



[그림 1-7] 가중값 분포 - 계속

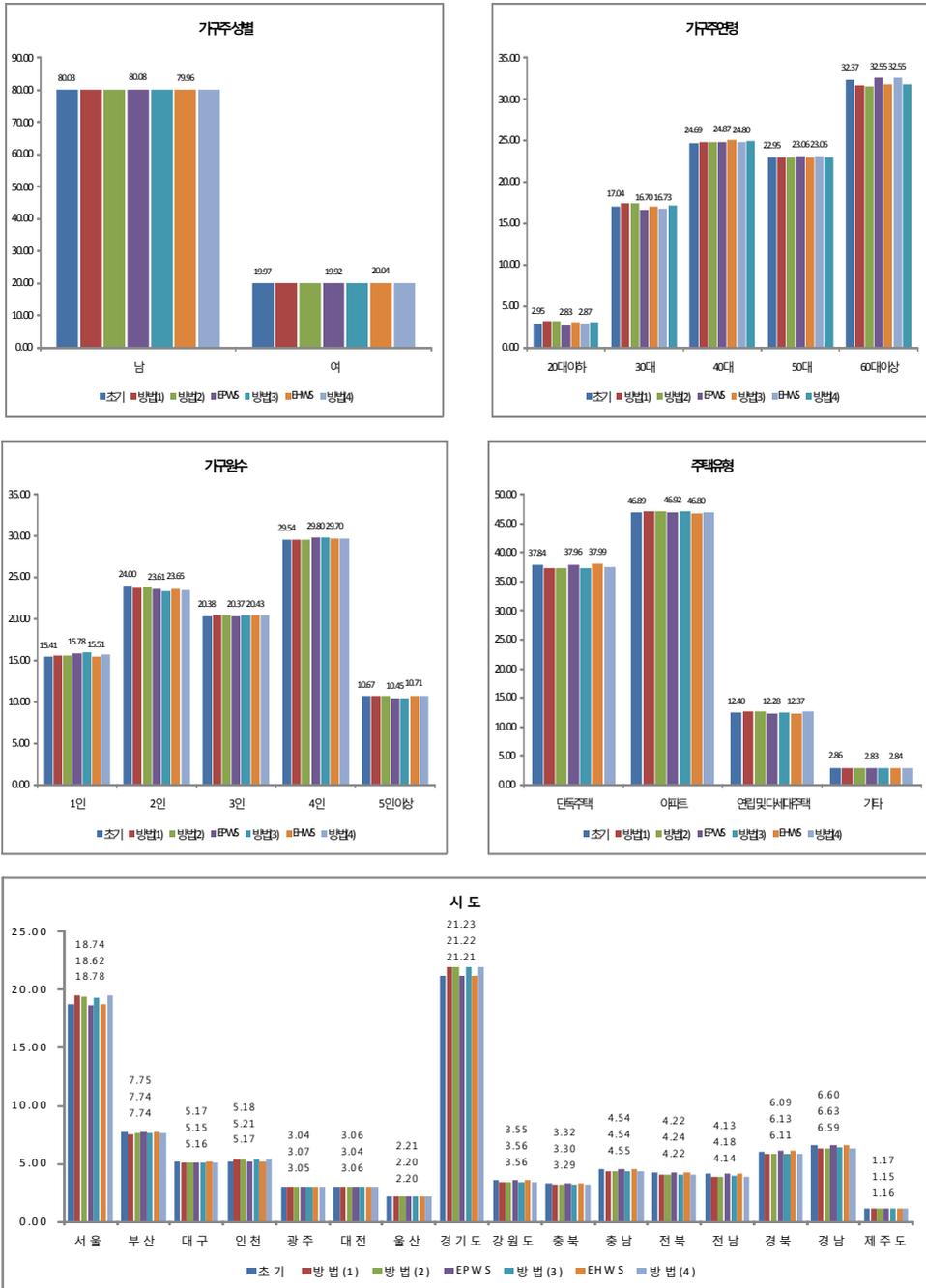
다. 가구 특성 분포 비교

방법별 가구 및 가구주 특성분포를 <표 1-14>와 [그림 1-8]에 수록하였다. 가구원수 분포에 대해 초기 가중값과 조정된 가중값을 비교해보면 조정한 경우 가구주 성별 분포는 남자가 0.05%p 상향되고 여자는 0.05%p 하향되는 정도로 크게 영향을 미치지 않는 것으로 나타났다. 연령분포에서는 20대 이하와 30대에서 초기 가중값 보다 각각 0.12%p, 0.34%p 낮게 조정되었으며, 그 외 연령대에서는 0.1~0.2%p정도 상향 조정되었다. 가구원수 분포에서 조정한 경우 1인 및 4인 가구의 비율이 각각 0.37%p, 0.26%p 높아지고, 2인 및 5인 이상 가구의 비율이 각각 0.39%p, 0.22%p 낮아졌으며, 주택유형 분포는 단독주택 비율이 0.12%p 상향조정되고, 연립 및 다세대주택 비율은 0.12%p 하향 조정되었다. 16개 시도 분포는 조정한 경우에 서울지역 비율이 0.12%p 정도 낮아졌다. 원표본 가중값을 공유하는 방법으로 조정하였을 때, 다른 특성에 비해 가구주 연령과 가구원 수의 분포에 영향을 많이 미치는 것을 알 수 있다. 또한 초기 가중값에 무응답 조정만을 적용한 방법(1)(2)와 가중값 공유 조정을 추가한 방법(3)(4)를 비교하였을 때, 가구 및 가구주 특성 비율 분포에 약간의 차이가 존재함을 알 수 있다.



〈표 1-14〉 가구 및 가구주 특성 분포

특성 구분	조정 방법에 따른 비율 분포(%)							비율 차이(%p)					
	초기	(1)	(2)	EPWS	(3)	EHWS	(4)	EPWS -초기	(3)-(1)	(3)-(2)	EHWS -(1)	(4)-(2)	
성별	남	80.03	80.04	80.03	80.08	80.08	79.96	79.95	0.05	0.04	0.05	-0.08	-0.09
	여	19.97	19.96	19.97	19.92	19.92	20.04	20.05	-0.05	-0.04	-0.05	0.08	0.09
연령	20대 이하	2.95	3.14	3.19	2.83	3.08	2.87	3.12	-0.12	-0.06	-0.11	-0.27	-0.02
	30대	17.04	17.42	17.45	16.70	17.09	16.73	17.13	-0.34	-0.33	-0.36	-0.69	-0.29
	40대	24.69	24.86	24.85	24.87	25.03	24.80	24.96	0.18	0.17	0.18	-0.06	0.1
	50대	22.95	22.94	22.92	23.06	23.03	23.05	23.03	0.11	0.09	0.11	0.11	0.09
	60대 이상	32.37	31.62	31.58	32.55	31.77	32.55	31.77	0.18	0.15	0.19	0.93	0.15
가구 원수	1인	15.41	15.60	15.62	15.78	16.00	15.51	15.72	0.37	0.4	0.38	-0.09	0.12
	2인	24.00	23.80	23.83	23.61	23.42	23.65	23.47	-0.39	-0.38	-0.41	-0.15	-0.33
	3인	20.38	20.42	20.42	20.37	20.40	20.43	20.47	-0.01	-0.02	-0.02	0.01	0.05
	4인	29.54	29.53	29.50	29.80	29.76	29.70	29.67	0.26	0.23	0.26	0.17	0.14
	5인 이상	10.67	10.65	10.64	10.45	10.41	10.71	10.67	-0.22	-0.24	-0.23	0.06	0.02
주택 유형	단독주택	37.84	37.35	37.30	37.96	37.42	37.99	37.45	0.12	0.07	0.12	0.64	0.10
	아파트	46.89	47.05	47.07	46.92	47.11	46.80	46.98	0.03	0.06	0.04	-0.25	-0.07
	연립 및 다세대	12.40	12.68	12.68	12.28	12.56	12.37	12.65	-0.12	-0.12	-0.12	-0.31	-0.03
	기타	2.86	2.92	2.94	2.83	2.91	2.84	2.92	-0.03	-0.01	-0.03	-0.08	0.00
지역	서울	18.74	19.49	19.45	18.62	19.32	18.78	19.49	-0.12	-0.17	-0.13	-0.71	0.00
	부산	7.75	7.59	7.62	7.74	7.61	7.74	7.61	-0.01	0.02	-0.01	0.15	0.02
	대구	5.17	5.10	5.11	5.15	5.09	5.16	5.10	-0.02	-0.01	-0.02	0.06	0.00
	인천	5.18	5.34	5.35	5.21	5.38	5.17	5.34	0.03	0.04	0.03	-0.17	0.00
	광주	3.04	3.00	3.01	3.07	3.04	3.05	3.01	0.03	0.04	0.03	0.05	0.01
	대전	3.06	3.02	3.03	3.04	3.02	3.06	3.03	-0.02	0.00	-0.01	0.04	0.01
	울산	2.21	2.18	2.18	2.20	2.18	2.20	2.17	-0.01	0.00	0.00	0.02	-0.01
	경기도	21.23	21.98	21.96	21.22	21.94	21.21	21.93	-0.01	-0.04	-0.02	-0.77	-0.05
	강원도	3.55	3.44	3.44	3.56	3.45	3.56	3.45	0.01	0.01	0.01	0.12	0.01
	충북	3.32	3.20	3.21	3.30	3.20	3.29	3.18	-0.02	0.00	-0.01	0.09	-0.02
	충남	4.54	4.33	4.33	4.54	4.33	4.55	4.34	0.00	0.00	0.00	0.22	0.01
	전북	4.22	4.08	4.09	4.24	4.11	4.22	4.09	0.02	0.03	0.02	0.14	0.01
	전남	4.13	3.91	3.90	4.18	3.96	4.14	3.92	0.05	0.05	0.06	0.23	0.01
	경북	6.09	5.81	5.81	6.13	5.84	6.11	5.82	0.04	0.03	0.03	0.3	0.01
경남	6.60	6.37	6.37	6.63	6.40	6.59	6.36	0.03	0.03	0.03	0.22	-0.01	
제주도	1.17	1.16	1.16	1.15	1.15	1.16	1.16	-0.02	-0.01	-0.01	0.00	0.00	



주 : 초기, EPWS, EHWS에 해당되는 비율만을 표시하였으며, 시도 부분은 위에서 아래로 초기, EPWS, EHWS 순서임

[그림 1-8] 가구 및 가구주 특성별 비율(%)



1) 방법별 추정값 비교

가중값 적용방법에 따른 추정 결과를 <표 1-15>에 수록하였으며, 방법별 자산, 부채, 소득 추정값에 큰 차이는 없음을 알 수 있다. 자산에서는 공유방법을 적용한 방법(3)과 (4)의 경우가 방법(1)과 (2)에 비해 약간 높게 추정되며, 부채에서는 방법(3)은 조금 낮게 방법(4)는 조금 높게 추정되었다. 또한 소득부분에서도 유사한 경향을 나타내는 것으로 보여진다.

<표 1-15> 가중값 적용방법에 따른 주요 항목 추정 결과

구분	방법(1)		방법(2)		EPWS		방법(3)		EHWS		방법(4)		
	평균	표준 오차	평균	표준 오차	평균	표준 오차	평균	표준 오차	평균	표준 오차	평균	표준 오차	
자산	자산 총액	29,437.00	629.18	29,420.00	627.69	29,169.00	626.45	29,451.00	634.87	29,233.00	623.49	29,510.00	631.60
	금융 자산	6,731.71	123.63	6,732.46	123.51	6,651.64	121.49	6,735.20	125.38	6,651.41	120.40	6,733.83	124.18
	실물 자산	22,705.00	586.86	22,688.00	585.36	22,517.00	585.63	22,715.00	591.94	22,581.00	582.91	22,776.00	588.98
	기타	946.85	27.66	946.96	27.62	941.55	27.90	946.50	28.14	941.00	27.54	945.76	27.78
부채	부채 총액	5,196.29	196.67	5,191.21	196.09	5,096.15	194.20	5,175.05	197.15	5,121.54	194.20	5,200.68	197.07
소득	경상 소득	3,978.64	46.30	3,979.30	46.23	3,943.53	46.20	3,966.52	46.51	3,950.79	46.08	3,973.66	46.38
	근로 소득	2,407.49	30.33	2,409.40	30.31	2,370.51	30.43	2,391.75	30.49	2,378.05	30.29	2,399.48	30.36
	사업 소득	1,245.83	39.74	1,244.58	39.66	1,243.54	39.52	1,247.01	39.90	1,245.12	39.46	1,248.33	39.81
	금융 소득	65.60	6.20	65.57	6.18	65.66	6.21	66.41	6.32	65.13	6.11	65.86	6.22
	실물 자산	15.36	1.68	15.34	1.68	15.79	1.72	15.60	1.72	15.63	1.70	15.44	1.69
	공적 이전	174.06	5.55	173.92	5.54	177.69	5.64	175.67	5.62	176.79	5.59	174.75	5.57
	사적 이전	70.29	4.11	70.49	4.22	70.34	3.90	70.09	3.90	70.06	3.85	69.80	3.85
지출	지출 총액	683.18	9.28	683.23	9.26	676.51	9.25	682.00	9.30	677.79	9.25	683.22	9.31

제5절 결 론

패널조사는 종단분석과 횡단분석이 모두 가능하다는 장점을 가지고 있으며, 각 목적에 따라 가중값 작성이 필요하다. 또한 패널조사는 웨이브가 진행되면서 패널마모로 인해 표본감소가 발생하게 되며, 이것은 횡단면 추정의 상대표준오차 증가 및 대표성에도 영향을 미치게 된다. 표본감소를 보정하기 위하여 원표본 가구원에 추적 조사할 비표본 가구원을 추가한 확장된 원표본 가구원 개념을 도입하고 이를 영구표본 가구원(permanent sample member)이라 한다. 이 때 추적 조사할 비표본 가구원은 종단 모집단과 추적규칙(follow-up rule)에 근거하여 결정하게 된다.

본 연구에서는 문헌연구와 국내외 사례연구를 바탕으로 가구 패널조사에서 횡단추정을 위한 횡단면 가중값 산출방법에 대해 살펴보았다. 패널조사의 횡단 가중값 작성은 일반적인 횡단조사의 가중값 산출단계(추출확률-무응답보정-사후조정)와 기본적으로 동일하다. 다만 패널조사 시작 시점에서 추출된 표본(원표본)을 대상으로 주기적으로 조사를 실시하며, 원표본에 대해서만 추출확률이 존재하므로 2차 웨이브 이후 가구변화(원표본 가구원 및 비표본가구원 변화) 발생에 따라 비표본의 추출확률과 관련하여 가중값 조정 단계가 필요하다.

2차 웨이브 이후 횡단면 추정값에 대한 불편추정량을 제공하면서 가구변화에 따른 추출확률을 고려하여 가중값을 부여하는 방법으로 원표본의 가중값을 공유하는 방법과 모형을 통해 추출확률을 추정하는 방법이 있다(Kalton et al, 1994). 이 두 방법은 1차 웨이브 표본 추출확률이 작으면 거의 일치하며, 모형방법에 의해 추정된 새로 진입한 가구원의 추출확률은 원가구내 가구원수에 반비례한다(Matthias Schonlau, Martin Kroh 2010, SOEP papers)고 알려져 있다. 사례연구를 통해 대부분의 패널조사에서 가구 변화에 대한 가중값 조정단계를 거치며, 조정방법으로 가중값 공유방법 또는 모형방법을 사용하고 있는 것으로 파악되었다. 또한 미국의 PSID와 한국노동연구원의 KLIPS는 Duncan 방법을 사용하고 있음을 알 수 있었다.

2011년 2차 가계금융조사는 패널 이탈가구(약 10%)에 대해 적정 표본 규모를 유지하기 위해 추가표본으로 2,000가구를 추출하여 총 10,500여 가구에 대해 조사를 실시하였다. 그리고 1차 웨이브 원표본과 2차 웨이브 추가표본에 대해 각각 설계가중값, 무응답보정 및 사후층화를 거친 후, 두 가중값을 결합하고 2010년 인구총조사 잠정자료로 벤치마킹하여 시도별로 합보정한 횡단면 가중값을 작성하였다.

본 연구에서는 가중값 조정에 대한 실증분석을 위하여 2차 가계금융조사에서 추가표



본을 제외한 원표본 자료에 대해 가구변화에 따른 가구 가중값을 조정하는 가중값 공유 방법을 적용하여 횡단 추정치에 미치는 영향을 파악하여 보았다.

2차 가계금융조사 자료 분석에서 2차 웨이브의 가구는 총 10,169가구이며 조사대상 가구는 99.4%에 해당되는 10,110가구이다. 조사대상 가구 중 응답가구는 9,042가구이며 나머지 1,068가구는 행방불명, 불응, 부재로 무응답한 가구이다. 분가가구의 무응답 비율 22.94%로 원표본가구의 10.35%보다 높은 것으로 나타났다. 응답한 9,042가구 내 총 가구원 30,872명 중 원표본 가구원은 29,908명이며, 결혼 등으로 새로 진입한 가구원은 964명이다. 원표본 가구원 중 분가한 가구원은 192명, 사망한 가구원은 104명, 기타감소는 249명, 나머지 29,363명(약 98%)은 유지되었다. 따라서 새로 진입한 가구원이 발생한 가구, 가구원 감소 가구, 또는 분가가구가 가중값 조정 대상이 된다.

가중값 조정방법으로 가중값 공유방법인 균등 가구원 공유방법과 균등가구 공유방법을 적용하였으며, 조정을 하지 않은 경우와 가중값 분포, 가구 및 가구주 특성, 추정결과를 비교하였다. 가중값은 조정을 실시한 경우에서 평균은 감소되고 표준편차는 조금 증가하였다. 변화요인별로는 새로 진입한 가구원이 발생한 가구의 가중값 조정(가중합이 약 40%p 정도 감소)이 가장 크게 이루어졌으며, 분가가구에서 더 크게 조정이(가중합이 약 50%p 정도 감소)되었다. 가중값 조정은 가구주 연령과 가구원수와 같은 가구 특성분포에 어느 정도 영향을 미치는 것으로 나타난 반면, 가구주 성별, 주택유형, 16개시도 분포에는 큰 영향을 미치지 않는 것으로 나타났다. 추정결과에서는 조정여부에 따라 큰 차이는 없으며, 조정방법을 적용하였을 때, 자산은 약간 높게 추정되었으며, 부채와 경상소득에서는 약간 낮게 추정되었다.

패널조사는 웨이브가 진행되면서 점점 더 복잡한 경우가 발생하게 되며, 가중값 작성방법은 추정결과에 큰 영향을 미치게 될 것이다. 분석결과가 단지 2차 웨이브 자료를 이용한 것으로 한계가 있지만, 원표본 가중값을 그대로 이용할 때 변화가구에 대해 상대적으로 큰 가중값이 부여될 가능성이 있으며, 조정을 통해 감소시킬 필요가 있다고 판단된다. 또한 추정결과에는 큰 차이가 없었지만, 가구 분포에도 영향을 미친다는 것을 확인할 수 있다.

가구패널조사에서는 일반적으로 가구와 가구원에 대한 가중값을 작성하게 된다. 본 연구에서 시도하지는 않았지만 가중값 조정에서 모형을 이용하여 추출확률을 추정하는 방법은 극단적인 가중값을 발생 시킬 가능성이 크며, 원표본 추출확률모형에 대한 충분한 검토와 이해가 필요할 것이다. 또한 새로운 원표본이 발생할 때마다 추출확률을 추정해야 하는 어려움이 있다. 반면 균등 공유 방법은 자료관리 및 활용측면에서 비교적 간편하게 적용할 수 있는 방법으로 판단되며, 이때 가구 기준 방법보다는 개인 기준 방법

이 실무 적용에는 더 효율적일 것으로 사료된다. 다만 이 방법을 적용하기 위해서는 개인 단위의 가구원 변동사항에 대한 보다 정확한 정보의 수집 및 관리가 필요할 것이다.

2012년 1차 가계금융·복지조사가 실시되었으며, 2013년 2차 조사가 실시되게 된다. 가계금융·복지조사는 가계금융조사와는 다르게 가구뿐만 아니라 가구원 가중값도 작성되어야 한다. 2차 웨이브 이후 횡단 가중값은 다른 패널조사와 마찬가지로 1차 웨이브 가중값에 대해 가구 또는 가구원 변화 조정, 무응답 조정, 사후 조정 과정을 거쳐 작성되게 될 것이며, 자료가 수집된 후 가중값 작성 단계별 과정에서 충분한 검토가 필요할 것으로 판단된다.



참고문헌

- 강석훈 (2003). KLIPS의 가중치 부여방안 연구, 한국노동패널연구 2003-04, 한국노동연구원.
- 한국조세연구원(2008), 1차년도 재정패널조사 기초분석보고서.
- _____ (2009), 2차년도 재정패널조사 기초분석보고서.
- _____ (2010), 3차년도 재정패널조사 기초분석보고서.
- 한국노동패널 1~11차년도 조사자료 User's Guide, 한국노동연구원.
- Duncan, G. (1995) A Simple Method for Weighting in Household Panel Surveys, Working Paper, Northwestern University.
- Geller, HP. (1987). Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In Labenslagen im Wandel: Analysen, edited by H.-J. Krupp and U. Hanefeld. Frankfurt: Campus.
- Heeringa, S.G., Berglund, P.A., Khan, A., Lee, S., Gouskova, E. (2011) PSID Cross-sectional Individual Weights, 1997-2009, Technical Report, Univ. of Michigan.
- Kalton, G. and Brick, J.M. (1995). Weighting schemes for household panel surveys, Survey Methodology, 21,1,33-44, Statistics Canada.
- Lynn, P. (2006). Methodology of Longitudinal Surveys, Chichester: John Wiley & Sons Inc.
- Rental, U. (1992) Weighting Procedures and Sampling Variance in Household Panels, Working paper, German Institute for Economic Research.
- Schonlau, M. and Kroh, M. (2010). On the Equivalence of Common Approaches to Cross Sectional Weights in Household Panel Surveys, SOEP papers on Multidisciplinary Panel Data Research.
- Schonlau, M., Watson, N., Kroh, M. (2011), Household survey panels: how much do following rules affect sample size?, Survey Research Methods, 5,2, 53-61.
- Taylor, M.F., Brice, J., Buck, N. and Prentice-Lane, E. (2010), British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices, Colchester: University of Essex.
- Watson, N. and Tim R.L. Fry (2002). HILDA Survey : Wave1 Weighting, HILDA PROJECT TECHNICAL PAPER SERIES NO.3/02, Melbourne Ins.
- Watson, N. (2004), Wave2 Weighting, In HILDA Project Technical Paper Series NO.4/04, Melbourne Ins.