

Analysis on the Effect of Unit Non-Response Adjustment using the Survey of Household Finances

Jeeseon Baek^{a,1} · Kyuho Shim^a

^aMethodology Division, Statistical Research Institute, Statistics Korea

(Received November 13, 2012; Revised March 20, 2013; Accepted May 22, 2013)

Abstract

Unit non-response of surveys reduces the efficiency of the estimates and also causes non-response bias especially when there is large difference between respondents and non-respondents. Non-response weighting adjustments have usually been used to compensate for non-response. It is not easy to examine the non-response bias as well as to obtain information on the non-respondents in sample surveys. A household panel survey, called *The Survey of Household Finances*, was conducted in both 2010 and 2011. In this paper, we assume that non-response households in Wave 2 have strong non-response (non-cooperative) tendency. We classify those households into non-response households in Wave 1. Under this assumption, the characteristics of non-response households, the non-response bias and the effect of non-response adjustments are investigated.

Keywords: Non-response, bias, non-response adjustment, The Survey of Household Finances.

1. 연구배경

통계 조사에서의 오차는 크게 표본오차(sampling error)와 비표본오차(non-sampling error)로 구분할 수 있다. 표본오차는 전체 모집단의 일부인 표본을 조사하여 추정함으로써 인해 발생하는 오차이며, 비표본오차는 조사의 계획부터 수집, 집계, 처리 등 표본오차를 제외한 조사의 전 과정에서 발생하는 오차이다. 통계조사 수행 중 부재, 응답거부, 조사 불능 등으로 인해 발생하는 무응답(non-response)은 대표적인 비표본오차의 원인 중 하나로 표본 크기를 축소시켜 추정값의 효율을 떨어뜨리게 된다. 뿐만 아니라 무응답에 포함되는 표본들이 시스템적인 차이를 발생시킨다면 무응답 편향(bias)이 발생하여 추정값의 신뢰도에 문제가 생길 수 있다 (Song과 An, 2009). 무응답은 전체 항목에 대해 응답하지 않은 단위 무응답(unit non-response)과 몇몇 항목에만 응답하지 않은 항목(item non-response)으로 구분할 수 있으며, 무응답 편향을 줄이기 위해 단위 무응답의 경우는 가중값 조정방법을 사용하며, 항목 무응답의 경우는 대체(imputation)방법을 사용하는 것이 일반적이다. Kim 등 (2004)와 Park (2007)은 가계관련 조사의 무응답 편향을 조정하기 위하여 가중값을 이용한 무응답 처리기법에 대한 연구를 수행한바 있다. 본 연구는 무응답 가구 특성과 무응답 편향 정도 그리고 조정시 사용되는 변수에 따른 무응답 조정효과 비교를 통한 무응답 조정효과 파악을 목적으로 동일한 표본가구를 주기적으로 조사하는 패널조사인 1,

¹Corresponding author: Deputy Director, Methodology Division, Statistical Research Institute, Statistics Korea, Statistical Center, 713 Hanbatdaero, Seo-gu, Daejeon 302-280, Korea. E-mail: jsbaek98@korea.kr

2차 가계금융조사 자료를 이용하였다. 1차 조사에서는 응답하였으나, 2차 조사에서 무응답한 가구는 비협조 성향이 강한 가구로 무응답 가구와 성향이 유사할 것으로 판단하고, 2차 가계금융조사의 응답정보를 이용하여 1차 가계금융조사의 응답가구를 응답가구와 무응답가구로 재분류하였다. 재분류된 1차 가계금융조사에 대해 무응답 조정을 실시하고, 전체 1만 가구를 대상으로 한 결과와 비교분석을 실시하였다. 무응답으로 가정한 가구는 2010년 응답을 하고 1년 뒤 조사에서 무응답으로 전환된 가구로 처음부터 응답하지 않은 가구보다는 무응답 성향이 약할 수 있다는 한계점을 가지지만, 무응답에 대한 계량적 분석을 통해 가계금융 관련 조사의 무응답 조정에 유의미한 정보를 제공할 수 있을 것으로 기대한다.

본 연구에서 이용한 통계청의 가계금융조사는 가구의 실물 및 금융자산, 부채 등의 구성과 분포를 조사하여 각종 경제, 사회, 금융정책이나 학문연구의 기초자료로 제공하고자 시작된 조사로 2010년 1차 조사가 실시되었다. 이 조사는 패널조사로 가계부문의 미시적 재무건전성을 주기적으로 파악하여 정부와 금융당국의 부동산 정책, 통화 및 금리정책, 가계신용관리, 금융시스템 발전 등을 위한 기본정보로 활용하게 된다. 조사주기는 1년이며 모집단은 2005년 인구주택총조사 결과 일반조사구와 총조사 이후 신축된 아파트가 대상이 된다. 표본규모는 전국의 10,000가구이며, 층별 조사구를 주거유형, 주택평균 연건평으로 정렬하여 가구 수를 기준으로 확률비례계통추출로 표본조사구를 추출한 후, 계통추출법으로 표본가구를 추출하였다.

본 논문의 구성은 다음과 같다. 2절에서 단위 무응답에 대해 가중값을 이용한 무응답 보정방법에 대해 간략하게 살펴보고 3절에서 자료와 연구설계를 포함한 연구방법을 설명하였다. 4절에서는 가계금융조사의 무응답 가구 특성을 파악하고 각각의 무응답 보정방법에 의한 추정값을 비교 분석하였다. 5절에서는 연구내용을 요약하고 결론을 제시하였다.

2. 단위 무응답 조정

무응답의 발생 메커니즘은 완전임의결측(Missing Completely At Random; MCAR), 임의결측(Missing At Random; MAR) 그리고 비임의결측(Not Missing At Random; NMAR)으로 구분할 수 있다. 대부분의 통계조사에서 부재, 불응, 거부 등 여러 가지 원인으로 무응답이 일어나며, 응답표본과 무응답 표본의 차이가 큰 경우 응답표본만을 이용하게 되면 편향이 발생하게 되므로 편향을 제거하기 위하여 무응답 조정을 실시하게 된다. 무응답 조정은 응답이 확률적으로 일어난다는 가정 하에 주어진 자료를 이용하여 추정된 응답확률의 역수(무응답 보정인자)를 표본의 설계가중값(추출확률의 역수)에 곱하는 것으로 다음과 같이 무응답 보정 가중값을 산출하게 된다.

$$\text{무응답 보정 가중값} = \text{설계가중값} \times \text{무응답 보정인자.}$$

즉, 무응답 조정은 무응답 표본들과 유사한 속성을 갖는 응답표본들의 가중값을 무응답 보정인자를 이용하여 확대함으로써 손실을 보상하는 방법이라고 할 수 있다. 무응답 보정인자를 산출하기 위해서는 무응답 보정층 내 표본가구들이 동일한 응답확률을 갖는다는 가정 하에 유사한 응답성향을 갖는 표본들로 구성된 보정층을 형성하게 되는데, 대표적인 방법으로 CHAID 알고리즘과 로지스틱 또는 프로빗 모형을 이용한 방법이 있다. 무응답 조정에 대한 무응답 보정층을 형성하는 방법에 대한 내용은 Song과 An (2009), Kim 등 (2004)를 참조하기 바란다. CHAID 알고리즘과 로지스틱 모형에 대해서 Dufour 등 (2001)는 실증연구를 통해 무응답편향 축소에 있어서 CHAID 방법이 더 효과적이라고 하였다.

2011 가계금융조사 보고서 (Kostat 등, 2011)에 따르면, 가계금융조사의 무응답 조정은 로지스틱 회귀모형으로 응답성향과 관련된 변수를 검토하여 최종 7개의 변수(가구주 혼인상태, 지역(수도권/광역시/도시/농촌), 가구주 연령, 가구주 산업, 입주형태, 자산(10분위), 전용면적)를 선택하고 CHAID 방

Table 3.1. Design and analyzed households

Household status on the wave 2		Household classification of the wave 1	Number of hh
Surveyed households		Analyzed households	9,944
Response	→	Response hh	8,918
Non-response (Missing, Refusal, Absence)	→	Non-response hh	1,026
Exclusion (Household mergers, Ineligible, Others)		Exclusion	
		Household mergers	56
		Ineligible, Others	
		Total	10,000

법을 이용하여 33개의 무응답 조정층을 형성하고 조정인자와 무응답 보정 가중값을 산출하였다. 미국의 PSID와 영국의 BHPS는 로짓모형을 이용하며, 캐나다의 SLID는 CHAID 알고리즘을 이용하여 무응답 조정을 실시하고 있다.

3. 연구방법

3.1. 자료 및 설계

1차, 2차 가계금융조사 자료에 대해 1차 가계금융조사와 동일한 2차 가계금융조사 가구의 응답정보를 이용하여 1차 가계금융조사 가구를 응답가구와 무응답 가구로 분류하였으며 설계와 분석대상은 Table 3.1에 수록하였다. 1차 웨이브에서 응답한 1만 가구 중 2차 웨이브 조사에서 제외된 56가구는 분석대상에서 제외하고 조사대상인 9,944가구를 분석대상으로 하였다. 2차 웨이브에서 응답한 8,918가구는 협조성향을 가지는 가구로, 행방불명되거나 불응하거나 부재인 1,026가구는 비협조성향 가구로 분류하여, 협조성향의 8,918가구를 2010년에서 응답한 가구로, 비협조성향의 1,026가구를 무응답 가구로 가정하였다.

3.2. 연구내용

본 연구에서는 2010년 조사된 1만 가구에 대해 모든 정보를 가지고 있으므로 준거(reference)조사로 하고, 실험설계에 따라 응답가구 8,918가구, 무응답 가구 1,026가구를 대상으로 무응답 분석 및 조정을 실시하고 준거조사와 비교 분석을 실시하였다. 연구내용은 크게 3가지 단계로 구분된다. 1단계에서 무응답 가구의 특성과 가구의 응답여부와 관련된 가구 특성 및 금융변수를 파악하였다. 먼저 무응답 가구의 특성을 파악하기 위해 응답가구와 무응답가구에 대해 가구주 성별, 연령, 동거여부, 학력 등 가구주 특성변수와 금융변수인 자산, 부채, 소득 분포를 비교하였다. 무응답 조정을 위해 응답여부와 관련된 가구 특성변수를 파악하기 위하여 가구특성과 응답유형의 분포에 대한 카이제곱 독립성 검정을 실시하고, 가구특성별 금융변수 평균을 비교하여 가구 특성변수와 금융변수의 관계를 살펴보았다. 2단계는 무응답 발생과 관련이 있을 것으로 판단되는 변수들에 대해 CHAID 방법을 이용하여 무응답 조정 분석을 실시하였다. 일반적인 조사의 무응답 조정시 무응답 보정층 구성에 사용가능한 변수는 표본추출과 관련된 변수이다. 본 연구에서는 보정층 구성에 따른 무응답 조정효과를 비교하기 위하여 표본추출정보(지역과 주택유형), 가구주정보, 가구금융정보로 구분하고, 편이상 가구주정보와 표본추출정보를 가구정보로 하였다. CHAID 알고리즘에 의해 최종적으로 5가지의 무응답 보정층 “지역변수(수도권, 비수도권)와 주택유형”, “지역변수(16개 시도)와 주택유형”, “가구정보”, “가구정보와 금융정보”, “금융정보”으로 무응답 조정을 실시하였다. 3단계로 무응답 조정 방법에 따른 가중값 분포와 금융변수 추정결과를 비교하



Figure 4.1. Comparisons of distributions of household characteristics between response and non-response households (%)

었다. 또한 무응답 조정 전과 5가지 방법에 의한 무응답 조정 후 추정결과를 비교하여 무응답 조정으로 인한 편향 조정과 추정결과에 미치는 효과를 검토하였다.

4. 무응답가구 특성 및 조정 분석

4.1. 무응답가구 특성

4.1.1. 응답가구와 무응답가구의 가구 특성 무응답 가구의 특성을 파악하기 위하여 무응답 가구와 응답가구의 가구주 변수에 대한 분포를 비교하였다. 가계금융조사의 가구주 정보로 사용할 수 있는 특성 변수에는 성별, 연령, 동거여부, 혼인상태, 가구원수, 교육정도, 가구주 직업, 가구주 종사상 지위, 가구주 종사 산업 등의 변수가 있다. 가구주 특성 외에 거주지역과 주택유형을 사용하였는데 지역 같은 경우 지역 간의 효과를 효율적으로 분석하기 위해 수도권(서울, 경기, 인천), 비수도권(기타 지역)으로 구분하였다. 응답가구와 무응답 가구를 비교해보았을 때 대부분의 가구주 특성변수에서 차이가 발생하였다. 특히 지역(수도권, 비수도권), 연령, 교육정도, 혼인상태에서 비율의 차이가 큰 것으로 판단되며, 무응답가구에서 응답가구보다 여성, 수도권, 20~40대, 1인 가구, 미혼, 고졸이상의 비율이 높다는 것을 알 수 있다 (Figure 4.1 참조).

4.1.2. 응답가구와 무응답가구의 금융 특성 가구의 금융수준과 응답여부와의 관계를 알아보기 위하여 응답가구와 무응답가구의 금융정보의 평균과 분포를 분석하였다 (Table 4.1, Figure 4.2 참조). 모든 금융항목에서 무응답가구의 평균값이 높으며, 금융부채를 제외한 자산 및 소득에서 유의미한 차이가 있는 것으로 나타났다. 따라서 자산이나 부채가 많은 가구의 무응답은 과소 추정을 초래하는 요인이 될 것이며, 응답여부에 영향을 미치면서 가구 금융변수와 관련이 있는 변수를 이용한 무응답 보정이 필요할 것으로 판단된다.

Table 4.1. Comparisons of financial mean between response and non-response households (Unit: Ten thousand won)

Category	Financial Category	Response	Non-response	t-value	p-value
Assets	Total asset	29981.26	41004.69	-5.41	0.000
	Non-financial asset	23774.20	32503.28	-4.68	0.000
	Financial asset	6207.06	8501.41	-5.67	0.000
	Other than non-financial assets	10470.07	12477.46	-1.33	0.181
Liability	Total liabilities	4962.82	6553.13	-2.59	0.009
	Holdings of debts	3396.95	4142.20	-1.35	0.176
	Receiving money from a lease	1565.88	2410.94	-4.09	0.000
Income	Regular income	3876.56	4785.52	-7.06	0.000

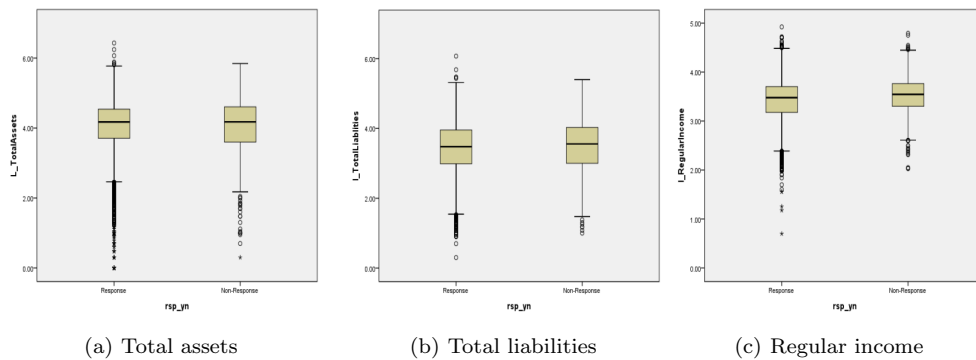


Figure 4.2. Box-plot of household finances data between response and non-response households

4.1.3. 가구특성과 응답여부 앞에서 응답가구와 무응답가구의 가구 특성 분포 비교를 통해 두 가구 집단 간 가구 특성 구성비에 차이가 있음을 살펴보았다. 이제 무응답 조정에 필요한 응답여부에 영향을 주는 정보를 파악하기 위하여 가구 특성(가구주 특성, 지역, 주택유형)과 응답여부의 연관성 분석을 실시하였다.

가구 특성에 따른 응답과 무응답 비율에서 대부분의 가구 특성변수에서 차이가 있었으며, 카이제곱 검정을 통한 유의성 검정 결과 가구주 성별을 제외한 대부분의 가구주 특성변수에서 응답과 무응답의 비율이 통계적으로 유의하게 차이가 있는 것으로 분석되었다 (Table 4.2). 유의수준 5%에서 성별을 제외한 가구 특성에서 응답여부에 유의미한 차이가 있으며, 응답과 무응답에 영향을 미치는 것으로 판단되었다.

4.1.4. 가구 특성과 금융정보 가구특성과 가구 금융정보에 대해 응답가구와 무응답가구를 비교한 결과, 대부분에서 차이가 있음을 확인하였으므로, 이 절에서는 가구 특성에 따른 가구 금융정보 수준에 차이가 있는지를 살펴보았다. 가구주 특성변수에 대해 금융정보별 평균값은 지역, 가구원수, 결혼여부, 학력 등에서 차이가 큰 것으로 나타났으며, Figure 4.3에 수록하였다. 지역의 경우 수도권 지역이 대부분의 금융정보에서 큰 평균값을 가지며, 가구원 수는 당연한 결과일수도 있겠지만, 1인 가구에서 대부분의 금융정보의 평균이 낮다는 것을 알 수 있다. 학력의 경우도 고졸이상이 중졸이하보다는 금융정보의 평균이 높은 것으로 나타났다. 혼인상태는 배우자가 있는 가구가 금융정보의 평균이 높고 사별한 경우의 평균이 두 번째로 높은 것을 볼 수 있다.

이러한 차이는 무응답 가구에 높은 금융수준을 가지는 가구들의 비율이 높을 경우, 예를 들어 무응답가구에서 수도권 지역의 비율 또는 고졸이상 가구주 비율이 높으면, 조사 결과는 무응답 편향을 가지게 되며,

Table 4.2. χ^2 -test for households characteristics and response status

Characteristics of Household Heads	χ^2 -value	p-value
Living together**	32.45	0.000
Occupation**	115.86	0.000
Gender	2.54	0.061
Educational attainment**	90.93	0.000
Marital status**	17.06	0.000
Industry**	151.15	0.000
Number of household members*	3.87	0.049
Age**	100.01	0.000
Educational status**	18.64	0.000
Area(metro/non-metro)**	106.00	0.000
Kind of house*	8.63	0.035

(* : p-value < 0.05, ** : p-value < 0.01)

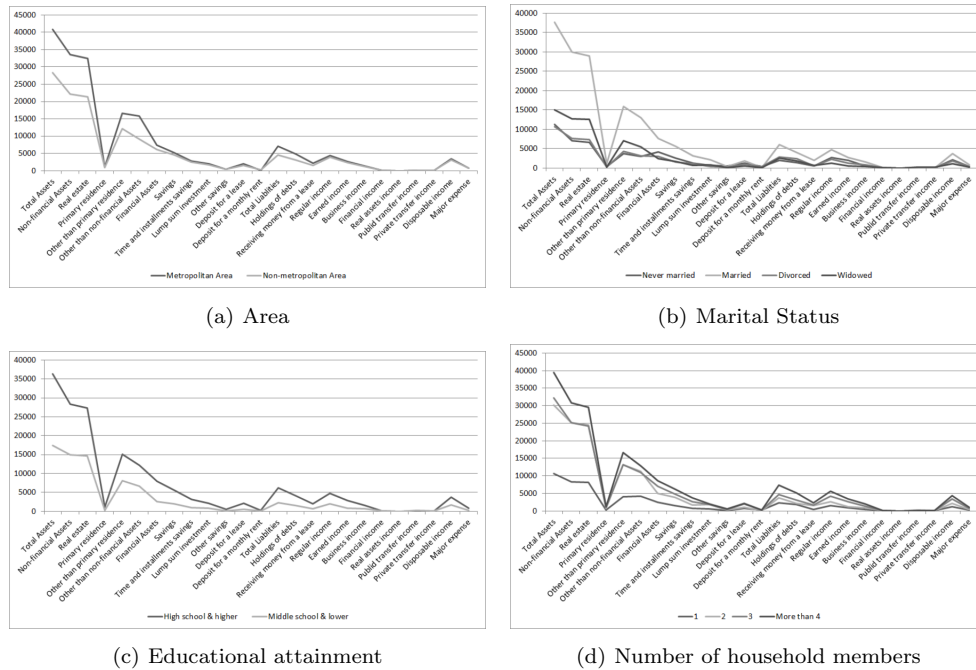


Figure 4.3. Comparisons of financial mean for household characteristic (Unit: Ten thousand won)

과소 추정될 가능성이 높아진다. 따라서 무응답 조정에 주요 추정변수인 금융정보와 상관이 높은 가구 정보나 그 외 활용 가능한 보조정보를 최대한 이용한다면, 무응답으로 인한 편향 조정효과를 얻을 수 있을 것으로 판단된다.

4.2. 무응답 조정 적용 분석

본 절에서는 무응답 조정 효과를 측정하고자 다음과 같이 8가지 경우에 대해 추정결과에 대한 비교분석을 실시하였다.

Table 4.3. Distributions of six weights

Kind of weight	mean	s.e	CV(%)	max	min	median
Design Weight	1693.70	212.66	12.56	2167.45	258.10	1722.47
A. CHAID(Area(metro/non-metro, kind of house)	1837.39	240.71	13.10	2514.24	258.10	1873.93
B. CHAID(Province-All, kind of house)	1857.83	250.60	13.49	2535.92	258.10	1890.84
C. CHAID(Household infor)	1864.71	265.15	14.22	2818.77	258.10	1897.11
D. CHAID(Finance infor)	1864.71	249.31	13.37	2648.62	258.10	1897.24
E. CHAID(Finance & Household Infor)	1864.72	266.10	14.27	2818.77	258.10	1877.57

- 2010년 1만 가구의 표본자료
- 2010년 1만 가구 표본자료와 설계 가중값
- 2010년 8,918 응답가구와 설계 가중값
- 2010년 8,918 응답가구와 5가지 무응답 조정 가중값

무응답 조정 가중값은 5가지 무응답 보정방법 - A. CHAID(지역-수도권과 비수도권, 주택유형), B. CHAID(지역-16개시도, 주택유형), C. CHAID(가구정보), D. CHAID(금융정보), E. CHAID(금융정보, 가구정보)에 대해 산출하였다. 2010년 1만 가구 표본자료에 대해 설계 가중값을 이용하여 추정된 결과를 준거(reference)로 하였다. 각 방법별 무응답 조정층에 대한 결과는 부록에 수록하였다.

4.2.1. 무응답 조정 방법별 가중값 분포 비교 5가지 무응답 조정방법으로 인한 가중값 분포에 대한 기초분석을 실시하고 Table 4.3에 수록하였다. 설계가중값의 평균과 표준편차가 각각 1693.70, 212.66이며 무응답 조정의 영향으로 평균값과 표준편차가 커지는 것을 알 수 있다. 표준편차는 CHAID(금융정보와 가구정보)가 266.10으로 가장 크며 CHAD(수도권과 비수도권, 주택유형)가 240.71로 가장 낮다. 변동계수(CV)의 값은 설계가중값이 12.56으로 가장 작으며 CHAID(금융정보와 가구정보)인 경우가 14.27로 가장 커지는 것으로 나타났다.

4.2.2. 무응답 가중값을 적용한 금융변수 추정값 비교 무응답 조정을 포함하여 8가지 방법에 대한 금융변수 추정결과를 Figure 4.4와 Table 4.4에 수록하였다. 설계가중값을 적용한 준거 추정결과와 비교했을 때, 무응답 조정을 실시한 모든 항목에서 과소 추정되는 것으로 나타났으며, 무응답 조정변수로 금융정보 변수를 이용하였을 때, 과소 추정 정도가 가장 작다는 것을 알 수 있다. 무응답 조정을 실시하지 않고 응답가구에 설계가중값만을 적용하여 추정된 결과는 과소추정 정도가 가장 높은 것으로 나타나, 무응답 조정 효과를 파악할 수 있었다.

4.2.3. 무응답 조정 효과 분석 무응답 가중값 적용으로 인한 조정 효과에 대한 척도로 무응답 편향(bias)을 사용하였다. 표본에서 응답자 평균에 대한 무응답 편향과 상대편향은 응답가구와 무응답 가구의 평균 차이와 무응답률에 영향을 받으며, 다음과 같이 정의된다. 무응답 편향은 응답 유무 가구의 평균 차이가 크더라도 무응답률이 낮으면 줄어들게 된다.

$$\text{무응답 편향} = \text{무응답률} \times (\bar{y}_r - \bar{y}_m), \tag{4.1}$$

$$\text{무응답 상대편향} = \text{무응답률} \times \left(\frac{\bar{y}_r - \bar{y}_m}{\bar{y}_m} \right), \tag{4.2}$$

여기서 \bar{y}_r 는 표본에서 응답자 평균, \bar{y}_m 는 표본에서 무응답자 평균이다.

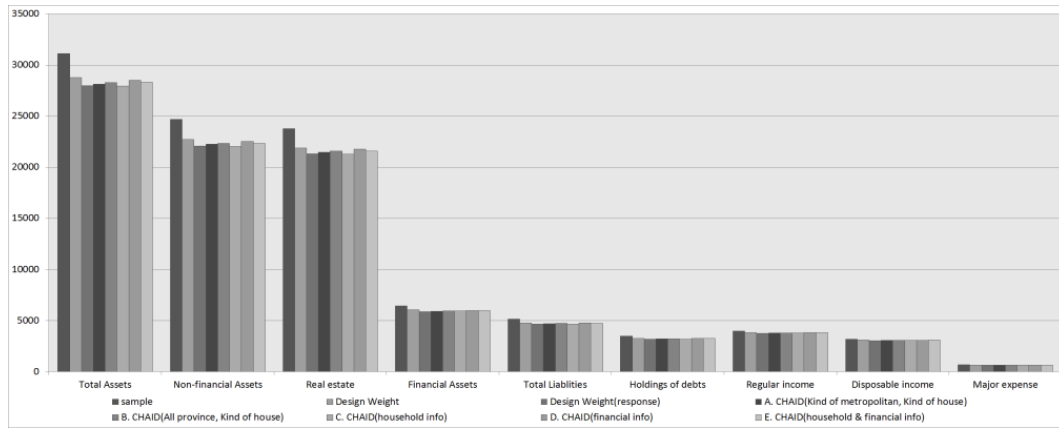


Figure 4.4. Comparisons of estimates of financial mean by eight weights (Unit : Ten thousand won)

Table 4.4. Estimates of financial mean and Relative bias

Finance	Estimates of mean							
	Sample	Reference	Design	A	B	C	D	E
Total Assets	31119.0	28769.0	27966.0	28135.7	28282.4	27950.3	28509.4	28316.7
Total liabilities	5126.9	4782.4	4652.0	4687.2	4722.1	4665.8	4761.9	4729.8
Regular income	3970.3	3828.0	3754.4	3764.5	3771.1	3776.2	3803.3	3811.6
Finance	Relative bias of response household mean with non-response adjustments (%)							
	Sample	Design	A	B	C	D	E	
Total Assets	-26.9	-22.0	-21.6	-21.2	-22.1	-20.5	-21.1	
Total liabilities	-24.3	-21.6	-21.0	-20.4	-21.4	-19.8	-20.3	
Regular income	-19.0	-16.2	-15.9	-15.8	-15.7	-15.1	-14.9	

본 논문에서는 무응답 가중값 적용으로 인한 편향조정 효과를 파악하고자 무응답률을 고려하지 않은 무응답 조정된 응답가구 평균에 대한 상대편향을 산출하고 Table 4.4에 항목별 편향 결과를 수록하였다. 설계가중값과 무응답 가중값을 고려한 모든 경우에서 표본에 대한 무응답 편향보다는 감소하는 것을 알 수 있다. 예를 들어, 자산총액에서 표본에 대한 상대편향 -26.9%이며, 금융정보를 이용한 무응답 조정된 경우의 상대편향은 -20.5%이다. 대부분의 항목에서 CHAID(금융정보) 방법을 적용한 경우의 상대편향이 가장 작은 것으로 나타났으며, CHAID(가구정보) 방법의 상대편향이 가장 크게 나타났다. 또한 가구정보를 무응답 조정에 이용한 경우는 지역과 주택유형을 이용한 경우보다 오히려 편향이 커지는 경우도 있는 것으로 나타났다.

다음은 무응답 조정이 추정값에 미치는 영향력을 비교하기 위하여 무응답 조정 방법별 추정값과 준거(reference) 추정값을 비교하여 Table 4.5에 수록하였다. 비교 척도는 무응답 가중값 조정된 응답가구에 대한 준거(reference, 2010년 가계금융조사(1만가구)에 설계가중값 적용한 경우) 추정값의 상대 비율을 사용하였으며, 아래와 같다.

$$\text{상대편향(\%)} = \left(\frac{\bar{y}_{r(NRA)} - \bar{y}_{ref}}{\bar{y}_{ref}} \right) \times 100,$$

여기서 $\bar{y}_{r(NRA)}$ 는 응답가구에 대해 무응답조정을 적용한 경우의 추정값이며, \bar{y}_{ref} 는 준거 추정값이다. Table 4.5에서 자산총액, 부채총액, 경상소득에서 방법별로 정도의 차이는 있으나, 무응답으로 인해 응

Table 4.5. Effect of non-response adjustments (Unit : %)

	Design	A	B	C	D	E
Total Assets	-2.79	-2.20	-1.69	-2.85	-0.90	-1.57
Total liabilities	-2.73	-1.99	-1.26	-2.44	-0.43	-1.10
Regular income	-1.92	-1.66	-1.48	-1.35	-0.64	-0.43

답가구만을 이용한 경우 약 0.5~3.0% 정도로 모두 과소 추정되고 있음을 알 수 있다. 무응답 조정을 실시하지 않고 설계가중값만을 이용한 경우에 과소추정 정도가 가장 높으며, 금융정보를 이용하였을 때 가장 편향이 작다는 것을 알 수 있다. 또한 지역과 주택유형정보만을 이용한 경우보다 가구주정보까지 포함한 경우(C)에 편향이 크게 개선되지 않는 것으로 나타나, 무응답조정에 가구주정보가 효과적이지 않는 것으로 판단되었다.

앞서 언급하였듯이 무응답한 표본에 대해서 정보를 얻는다는 것은 현실적으로 어려운 일이므로 무응답 조정시 일반적으로 표본추출정보를 사용하게 되며, 추가적으로 무응답 가구주에 대한 정보 수집이 가능하다면 가구주 정보를 사용할 수도 있다. 일반적으로 사용 가능한 정보와 가계금융조사의 주요 변수인 금융정보를 무응답 조정에 사용하는 경우의 편향 조정효과에 대한 비교에서 가구주 정보는 효과적이지 않은 것으로 나타나 동일한 가구주 정보내에서도 여전히 무응답 가구의 자산, 부채 소득이 높은 것으로 판단된다. 따라서 가계금융조사의 무응답 조정에는 금융정보와 높은 관련이 있는 정보(예를 들어, 가구별 세금정보 등)를 활용하는 것이 바람직할 것이나, 사용이 불가능하다면 추출정보를 이용하는 것이 무응답 편향 조정에는 효과적인 것으로 사료된다.

5. 결론

통계조사에서 무응답은 추정결과에 편향을 발생시키는 주요 요인으로, 편향 축소를 위하여 추정단계에서 이용 가능한 무응답 표본에 대한 정보를 근거로 무응답 조정을 실시한다. 현실적으로 무응답 표본에 대한 정보를 얻기는 쉽지 않으며, 따라서 무응답편향에 대한 계량적 분석을 실시하는데 어려움이 따른다. 본 논문에서는 가계금융조사의 2차 웨이브 응답여부를 1차 웨이브 자료에 적용하여 응답가구(8,918가구)와 무응답가구(1,026가구)로 분류하고 무응답 분석과 무응답 조정 등 계량적 분석을 실시하였다. 또한 1차 웨이브 응답가구(1만가구)를 준거로 무응답 조정방법에 따른 효과를 비교하였다.

주요 연구결과는 다음과 같다. 무응답가구는 시도, 주택유형, 가구주 교육수준 등 가구특성과 자산, 소득, 부채 등 금융수준에서 응답 가구와 차이를 보였다. 즉, 무응답가구는 응답가구에 비해 자산, 부채, 소득이 높으며, 수도권 거주, 여성, 20~40대, 1인 가구, 미혼, 고졸이상의 비율이 높은 것으로 분석되었다. 응답에 유의미한 영향을 미치는 변수 정보를 이용하여 무응답 조정을 실시한 결과, 응답한 가구만을 이용하여 추정한 결과는 모든 방법에서 과소 추정되며, 응답한 가구에 대해 설계가중값을 이용하여 추정한 결과보다는 무응답 조정을 적용하였을 때, 과소 추정 정도가 축소되는 것으로 나타났다. 무응답 조정효과는 가구정보, 추출정보인 지역과 주택유형, 가구정보와 금융정보, 금융정보 순서로 크게 나타나는 것으로 분석되었다. 특히 자산에서 가구정보를 이용한 무응답 조정 효과는 좋지 않은 것으로 나타났다.

연구결과로부터 모든 조사에 대해 일반적으로 적용할 수는 없지만, 가계의 부채, 자산, 소득과 같은 금융조사에서는 무응답 조정변수로는 가구주 정보보다는 지역과 주택유형 변수가 더 유용하며, 가능하다면 금융정보와 관련이 깊은 보조변수를 활용하는 것이 바람직하다는 결론을 내릴 수 있었다. 또한 모든 무응답 조정방법에서 과소추정되는 것으로 볼 때, 무응답 발생 메커니즘이 비임의결측(NMAR)을 따를 가능성과 함께 무응답 유형별(불응, 부재, 행방불명) 조정방법에 대한 심층적 연구가 필요할 것으로 사료된다.

부록

A. CHAID(Area(metro/non-metro), Kind of house))

Non-response layer with area(metro/non-metro), Kind of house

Layer	Area	Kind of House
1	Non-metropolitan	Other
2	Non-metropolitan	Apartment; Row house & apartment unit in a house
3	Non-metropolitan	Detached dwelling
4	Metropolitan	

Ratio of Non-response adjustment with Area(metro/non-metro), Kind of house

Layer	Size of sample				Ratio of non-response adjustment		
	sum	response	non-response	ratio	sum of response weight	sum of non-response weight	ratio
1	88	75	13	14.77	130825	22270	1.17
2	2671	2459	212	7.94	4118537	345936	1.08
3	2455	2298	157	6.40	4010081	276037	1.07
4	4730	4086	644	13.62	6871944	1065909	1.16

B. CHAID(Province-All, Kind of house)

Non-response layer with Province-All, Kind of house

Layer	Province	Kind of house
1	Chungnam; Gyeongbuk; Busan; Gyeongnam; Gwangju	Other
2	Chungnam; Gyeongbuk; Busan; Gyeongnam; Gwangju	Detached dwelling; apartment; apartment unit in a house
3	Gyeonggi; Daejeon; Ulsan; Daegu; Chungnam	
4	Gangwon; Jeonbuk; Jeju; Jeonnam	
5	Seoul; Incheon	

Ratio of Non-response adjustment with Province-all, Kind of house

Layer	Size of sample				Ratio of non-response adjustment		
	sum	response	non-response	ratio	sum of response weight	sum of non-response weight	ratio
1	45	35	10	22.22	61201	17269	1.28
2	2612	2426	186	7.12	4174774	310693	1.07
3	3565	3163	402	11.28	5313203	660627	1.12
4	1208	1161	47	3.89	1958577	81532	1.04
5	2514	2133	381	15.16	3623630	640030	1.17

C. CHAID(Household info)

Ratio of Non-response adjustment with Household info(Marital status, cohabits, industry, kind of house, age, area(met/non-met))

Layer	Layer	Ratio of non-response adjustment		
		sum of response weight	sum of non-response weight	ratio
1	'divorced', 'cohabits', 'non-cohabits', 'metropolitan'	1743064.40	104478.07	1.06
2	'divorced', 'cohabits', 'non-cohabits', 'non-metropolitan'	1566993.85	241938.13	1.15
3	'divorced', '1person household', '50~59', 'more than 60'	452511.69	57265.56	1.13

4	'divorced', '1person household', '30~39', '40~49'	57509.17	15554.80	1.27
5	'never married', 'cohabits', 'non-cohabits'	4785331.71	418142.30	1.09
6	'never married', '1person household', 'apartment'	2822389.26	322872.10	1.11
7	'never married', '1person household', 'detached dwelling', 'apartment unit in a house', 'other'	890818.03	142287.86	1.16
8	'widowed'	235461.33	17142.78	1.07
9	'married', 'construction', 'Print, Video, communicate, 'IT service', non-edu', 'primary school', 'middle school', 'colleage', 'university', 'graduate'	820755.34	3517.94	1.00
10	'married', 'construction', 'Print, Video, communicate, IT service', 'high school'	534630.81	209587.78	1.39
11	'married', 'agri, forestry & hunting', 'mining'	174516.21	24384.90	1.14
12	'married', 'manufacturing', 'public service & others', 'wholesale & retail trade', 'metropolitan'	133600.11	46921.39	1.35
13	'married', 'manufacturing', 'public service & others', 'wholesale & retail trade', 'non-metropolitan'	260398.20	32787.52	1.13
14	'married', 'electricity', 'transportation', 'hotel and restaunt', 'finance', 'cohabits'	526992.99	45345.05	1.09
15	'married', 'electricity', 'transportation', 'hotel and restaunt', 'finance', 'non-cohabits', '1person'	126415.79	27927.27	1.22

D. CHAID(Financial info)

Ratio of non-response adjustment with financial info(regular income, non-financial assets, major expense)

Layer	Ratio of non-response adjustment		
	sum of response weight	sum of non-response weight	ratio
1 'non-financial assets>60180', 'regular income>7990'	784607.14	60832.52	1.08
2 'non-financial assets>60180', 'regular income<=7990'	1612346.74	74859.53	1.05
3 '60180>='non-financial assets>10045', 'regular income>1232'	1489873.85	140835.56	1.09
4 '60180>='non-financial assets>10045', 'regular income<=1232'	107665.13	23905.33	1.22
5 '10045>='non-financial assets>1100', 'major expense>1400'	791961.50	34695.06	1.04
6 '10045>='non-financial assets>1100', '360<major expense<=1400'	5355586.06	545169.88	1.10
7 '10045>='non-financial assets>1100', 'major expense<=360'	835915.64	111151.06	1.13
8 '1100>='non-financial assets', 'regular income>698', 'major expense>199'	441143.35	108863.34	1.25
9 '1100>='non-financial assets', 'regular income>698', 'major expense<=199'	2278916.68	412154.85	1.18

E. CHAID(Financial info & Household info)

Ratio of Non-response adjustment with financial info(regular income, non-financial assets) and house info(Marital status, cohabits, kind of house, age, area(met/non-met))

Layer	Ratio of non-response adjustment		
	sum of response weight	sum of non-response weight	ratio
1 'divorced', 'cohabits', 'non-cohabits', 'metropolitan'	1847542.47	1743064.40	1.06
2 'divorced', 'cohabits', 'non-cohabits', 'non-metropolitan'	967308.52	906760.76	1.07
3 'divorced', '1person', '50~59', 'more than 60'	509777.25	452511.69	1.13
4 'divorced', '1person', '30~39', '40~49'	916652.18	899158.78	1.02
5 'never married', 'cohabits', 'non-cohabits'	504203.21	473213.35	1.07
6 'never married', '1person', 'non-financial assets>199'	7134707.07	6376022.02	1.12
7 'never married', '1person', 'non-financial assets<=199'	1359766.36	1273227.51	1.07

8	'widowed'	127755.20	117343.01	1.09
9	'married', 'regular income>7990', 'non-financial assets >60180'	135503.64	108763.39	1.25
10	'married', 'regular income>7990', '15400<non-financial assets<=60180'	669336.48	603924.37	1.11
11	'married', 'regular income>7990', '5494<non-financial assets<=15400'	525481.94	420845.51	1.25
12	'married', 'regular income>7990', 'non-financial assets <=5494'	536749.49	372058.60	1.44
13	'married', '1872<regular income<=7990', 'more than 60'	406370.23	337088.43	1.21
14	'married', '1872<regular income<=7990', '20~29', '30~39', '40~49'	180521.50	133600.11	1.35
15	'married', '1232<regular income<=1872'	293185.72	260398.20	1.13
16	'married', 'regular income<=1232', 'apartment', 'apartment unit in a house', 'other'	572338.04	526992.99	1.09
17	'married', 'regular income<=1232', 'detached dwelling'	154343.06	126415.79	1.22

References

- Dufour, J., Gagnon, F., Morin, Y., Renaud, M. and Sarndal, C. E. (2001). A better understanding of weight transformation through a measure of change, *Survey Methodology*, **27**, 97–108.
- Kim, J.-K., Han, G.-S. and Yoon, Y. (2004). Nonresponse weighting adjustment in Korea household income and expenditure survey, *Journal of The Korean Official Statistics*, **9**, 79–102.
- Kostat, Financial Supervisory Service, Bank of Korea (2011). Report on the Survey of Household Finance, 2011.
- Park, S. I. (2007). Nonresponse weighting adjustment in Household Wealth Survey, Hannam University.
- Song, J. W. and An, H. (2009). *Nonresponse Data Analysis*, Statistical Training Institute.

가계금융조사를 활용한 단위무응답 조정효과 분석

백지선^{a,1} · 심규호^a

^a통계개발원 조사연구실

(2012년 11월 13일 접수, 2013년 3월 20일 수정, 2013년 5월 22일 채택)

요약

통계조사에서 단위 무응답은 표본크기를 축소시켜 추정값의 효율을 떨어뜨릴 뿐만 아니라, 주요 관심변수에 대해 응답자와 무응답자간의 차이가 존재하는 경우 무응답 편향(bias)을 발생시키는 요인으로 일반적으로 가중값 조정방법을 이용하여 보정한다. 대부분의 조사에서 무응답 표본에 대한 정보를 얻는 것은 현실적으로 매우 어려우며, 무응답으로 인한 추정값의 편향 정도를 파악하는 것 또한 쉽지 않다. 통계청에서는 2010년, 2011년 가구패널조사인 가계금융조사를 실시하였다. 2010년 1차 웨이브는 1만 가구에 대해 모두 응답한 정보를 가지고 있으며, 2011년 2차 웨이브에는 부재, 거부 등으로 인한 무응답 표본(약 10%)이 존재한다. 본 연구에서는 2차 웨이브의 무응답 표본들을 무응답 성향이 강한 표본으로 가정하고 1차 웨이브에서 무응답가구로 분류한 후 무응답 가구의 특성과 무응답 편향 정도 및 무응답 조정 효과를 분석하였다.

주요용어: 무응답, 편향, 가계금융조사, 무응답 조정.

¹교신저자: (302-280) 대전 서구 한밭대로 713 대전센터, 통계청 통계개발원 조사연구실, 통계사무관.
E-mail: jsbaek98@korea.kr