



제1장

혼합조사 추정방법 실무적용방안 검토

- 사교육비조사를 중심으로 -

임정은 · 박라나

제1절 서론

1. 혼합조사 개요

혼합모드조사(Mixed-mode survey, 이하 혼합조사)는 하나의 조사에 2개 이상의 조사 방법을 적용하는 것으로, 악화일로를 견고 있는 조사 환경에서 조사대상자들의 응답 부담과 조사 비용의 증가 없이 고품질의 통계를 생산할 수 있는 방법들 중 하나로 주목 받고 있다.

혼합조사는 저렴한 비용으로 각 조사방법의 약점을 서로 보완할 수 있다는 점에서 각광 받고 있으며, 그 역사 또한 짧지 않다(Leeuw, 2005). 통계청의 2010년 인구주택총조사의 경우, 인터넷조사를 적극 도입함으로써 인건비, 인쇄 및 입력 비용 등 204억원의 예산¹⁾을 절감한 바 있다. 이와 같은 비용 절감 효과뿐 아니라 혼합조사는 응답률을 높이는데 매우 효과적인 것으로도 알려져 있다(Daniel et al., 2012). 여러 가지 조사 방법을 하나의 조사에 이용하면 조사대상자들에 대한 접촉 가능성을 높이고 무응답을 줄임으로써 포함오차와 무응답오차 축소를 통한 총 조사오차(total survey error) 축소 효과를 얻을 수 있으며(de Leeuw, 2005; Dillman et al., 2009), 나아가 고품질의 조사 결과를 얻을 수 있게 된다(Voogt and Saris, 2005).

그러나 혼합조사에 이러한 장점만 있는 것은 아니다. 혼합조사를 이용하여 응답률을

1) 2010년 인구주택총조사에서의 인터넷조사 참여율은 47.9%로 전체 조사의 절반에 가까웠으며, 예산 절감액은 204억원으로 당초 목표액인 164억원(참여율 30%)을 훌쩍 넘긴 바 있다.

높이고 포함오차를 축소한다 하더라도, 조사 방법에 따른 포함률(Coverage rate)의 차이, 다양한 조사방법 중 하나를 선택함으로써 발생하는 선택효과(Selection effects) 및 조사 방법에 따른 측정효과(Measurement effects; 측정오차) 등과 같은 모드효과(Mode effects)는 조사 결과에 또 다른 불안정성으로 작용하게 된다.

단일모드조사(Single-mode survey)에서 보자면, 선택효과는 일종의 무응답오차의 한 형태로 볼 수 있다. 특정 조사방법을 선택하는 경우 선택하지 않은 방법에 대해서는 응답하지 않기 때문이다. 만약 조사에 응답하지 않았거나(non-response), 못했을(non-coverage) 경우의 특성이 응답한 경우(response)와 다르다면, 조사 결과는 편향될 가능성이 매우 높아진다. 물론 조사방법을 선택하는 것 자체가 별다른 문제가 되지는 않는다는 의견도 존재한다. 오히려 단일모드조사라면 응답하지 않았거나, 못했을 대상에 대한 조사를 가능하게 했기 때문에 더 이로운 효과를 가져오기도 한다(Biemer, 2001; Voogt and Saris, 2005; Dillman et al., 2009)는 것이다.

그렇다면 측정효과는 어떠한가? 측정효과란 동일한 조사대상자가 서로 다른 조사 방법으로 조사에 참여할 때 서로 다른 답을 하는데 기인한다(Bowling, 2005; Voogt and Saris, 2005; Weisberg, 2005). 즉, 조사방법으로부터 발생한 측정오차가 측정효과가 되는 것이다. 이러한 차이는 다양한 원인으로부터 발생된다. 조사방법을 선택하는 응답자의 성향, 조사원 효과, 사회적 기대, 불명확한 이전 시점의 기억 등 모두가 측정효과의 원인이 될 수 있다. 2009년 통계청에서는 사회조사를 대상으로 기존의 면접조사와 인터넷조사를 병행하였으며, 자살충동이라는 민감한 질문에 대한 응답이 연령대와 조사방법에 따라 큰 차이가 있는 것으로 확인되었다(임경은, 2010).

이와 같이 선택효과와 측정효과는 조사결과에 직·간접적인 영향을 미치게 되며, 그 영향이 긍정적이냐 부정적이냐에 관계없이 이를 추정 과정에 반영하지 않는다면, 편향된 추정 결과를 얻을 가능성이 매우 높아지게 된다. 따라서 선택에 따른 조사대상자들의 특성과 조사방법에 따른 응답 특성의 차이를 검증하고, 이들이 조사 결과에 미치는 영향을 파악함으로써 그 효과를 추정 과정에 반영하는 것은 매우 중요하다고 할 수 있다.

그러나 혼합조사에 의한 조사 결과를 이용하여 모집단의 특성을 추정 하는 과정에 혼합 조사에 의한 모드효과가 반영되는 경우는 많지 않다. 혼합조사가 다양한 분야에 널리 사용되고 있는데 반해, 이에 대한 방법론적 연구 및 연구 결과의 활용은 매우 미미한 실정이다. 표본조사와 관련된 통계 서적에도 혼합조사와 관련된 내용이 거의 포함되어 있지 않으며, 혼합조사에 대한 연구 및 관련 기사들에서도 조사방법 간 비교나 의미, 방법론적 내용에 대해서는 논의하지 않고, 혼합조사의 적용 형태 및 결과에 대한 내용만 다루지고 있다(Leeuw, 2005). 통계 조사 이후 어떤 추정 방법을 적용하였느냐에 따라 그 결과가 달라진다는 사실에 비춰볼 때, 혼합조사의 특성을 반영한 적절한 추정 방법을 적용하여 통계를 생산하는 일은 통계의 정확성과 신뢰성에 있어 무엇보다 중요한 과정이라 할 수 있을 것이다.



본 연구에서는 혼합조사 중 하나인 통계청의 사교육비조사를 대상으로 선택효과와 측정효과를 고려한 다양한 형태의 추정 방법을 적용함으로써, 이들의 실무 적용 가능성을 확인하는데 그 목적이 있다. 2절에서는 조사방법별 응답자 특성의 차이를 비교하여 선택효과를 확인하고, 3절에서는 주요 관심 변수에 대한 측정효과를 확인하고자 한다. 4절에서는 다양한 추정방법들 중 선택효과와 측정효과 등 모드효과를 가장 잘 설명할 수 있는 추정방법을 탐색하고, 실무적용 가능성에 대해 논의한다.

2. 사교육비조사 개요

사교육비조사는 교육인적자원부 주관으로 2003년에 교육개발원에서 실시한 이후, 통계의 일관성을 유지하고 조사과정에서 발생하는 비표본오차를 줄이기 위해 2007년부터 통계청에서 조사를 대행하고 있다. 사교육비조사는 사교육비 경감대책 및 공교육 내실화 등 교육정책 수립에 활용할 수 있는 통계를 정기적으로 작성·제공하는데 그 목적이 있으며, 사교육 원인, 사교육비 경감대책의 효과 등을 병행하여 분석할 수 있는 사교육 관련 의식항목을 포함하여 수행되고 있다.

조사 대상은 대한민국에 거주하는 초·중·고교 재학생과 그 학부모이며, 2012년 조사에서는 전국 초·중·고교 1,064학교의 1,398학급, 약 43,000여 명의 학부모가 상·하반기 두 차례(6월, 10월)에 걸쳐 조사에 참여하였다. 1차 조사는 3, 4, 5월의 사교육비 지출 항목이 조사되며, 2차 조사에서는 7, 8, 9월의 사교육비 지출 항목이 조사된다. 두 조사에서의 조사 항목(조사 범위)은 교재비를 포함한 주요 과목별 학원비, 개인 및 그룹과외비, 학습지, 인터넷 및 통신 강의 과외비(EBS 제외) 등 사교육비와 방과후 학교 활동비, EBS 관련 교육비 및 어학연수비 등이다.

사교육비조사는 2가지 조사방법을 적용하는 혼합조사로, 2011년도에는 종이조사와 인터넷조사가 별도 추출된 표본(전체 학급 중 각 50%)을 대상으로 수행되었으나, 2012년도에는 인터넷조사 선행 후 종이조사를 순차적으로 진행하였다.

제2절 조사방법별 선택효과

1. 기초자료분석

2011년 사교육비조사는 인터넷조사(전체 조사대상 학급의 50%)와 종이조사(전체 조사대상 학급의 50%)에 참여할 표본이 별도로 추출되어 조사가 진행되었으므로, 조사

방법에 대한 선택이 불가능하였다. 그러나 2012년 조사에서는 인터넷조사를 먼저 권고한 이후, 인터넷으로 조사가 진행되지 않은 경우에 한하여 종이조사가 추가 진행되었다. 따라서 2012년 조사에서는 인터넷 조사에 참여하기 위해 해당 조사방법을 먼저 선택해야했으며, 이에 따른 선택효과가 조사 결과에 반영되었을 것으로 예상된다.

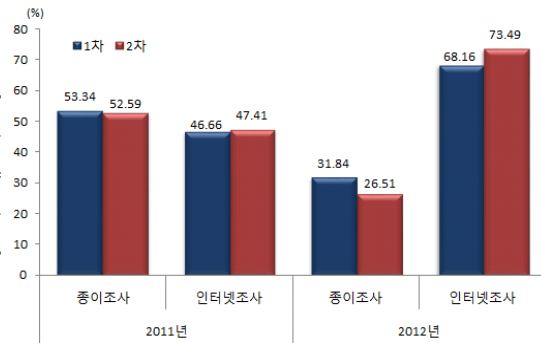
선택효과에 대한 분석에 앞서 2012년도 사교육비조사 자료의 분포를 2011년도 자료와 비교하여 살펴보았다. 다음은 2011년 1차 사교육비조사와 2012년 1차 사교육비조사에서의 조사방법에 따른 일반현황이다.

<표 1-1> 조사방법별 분포

	종이조사		인터넷조사	
	1차	2차	1차	2차
2011	53.34	52.59	46.66	47.41
2012	31.84	26.51	68.16	73.49

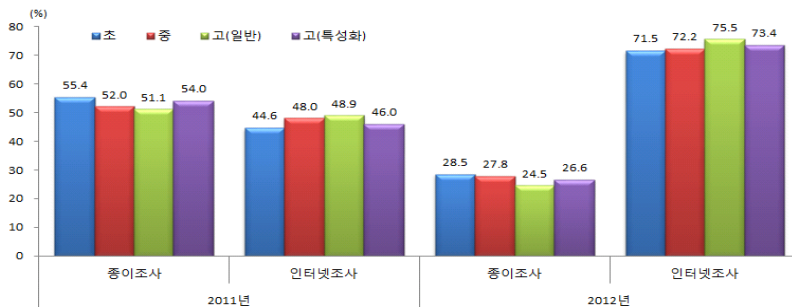
* 조사방법 기준 백분율임

[그림 1-1] 조사방법별 분포



2011년도 사교육비조사 자료는 조사방법에 따라 대략 50%의 응답이 이루어졌다. 이에 반해 2012년도의 경우에는 1차 조사의 경우 68.16%, 2차 조사의 경우 73.49%가 인터넷으로 응답하여, 조사방법에 대한 자발적 선택으로 인터넷 조사에 참여한 응답자가 크게 증가하였음을 알 수 있다.

다음은 학교급별 조사방법의 비율이다.



* 학교급 기준 백분율임

[그림 1-2] 학교급/조사방법별 분포(2차 조사 기준)



2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 학교급별로 살펴보면, 2011년에는 학교급별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다.

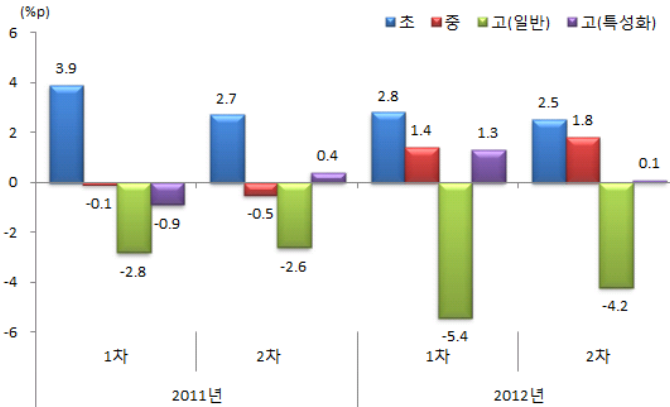
이제 조사방법별로 학교급의 분포를 살펴보자.

<표 1-2> 조사방법/학교급별 분포

(단위: %)

학교급별 분포			초등학교	중학교	일반고	특성화고
2011	1차	종이조사	25.8	25.4	40.8	8.0
		인터넷조사	21.9	25.5	43.6	8.9
	2차	종이조사	25.4	25.3	40.9	8.4
		인터넷조사	22.7	25.8	43.5	8.0
2012	1차	종이조사	25.7	26.7	37.7	10.0
		인터넷조사	22.9	25.3	43.1	8.7
	2차	종이조사	25.8	27.2	38.4	8.7
		인터넷조사	23.3	25.4	42.6	8.6

* 조사방법 기준 백분율임



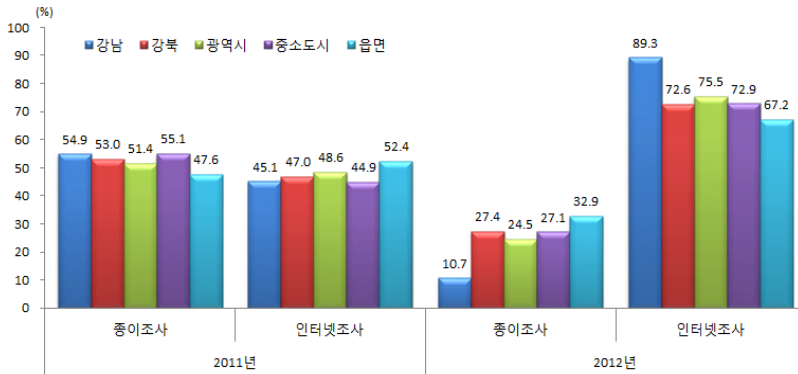
* 조사방법 기준 백분율의 차이임

[그림 1-3] 조사방법/학교급별 비율 차이(종이-인터넷)

2012년 조사에서 조사방법별로 학교급별 비율을 비교한 결과, 1차 조사의 경우 초등학교(2.8%p), 중학교(1.4%p), 특성화고(1.3%p)는 종이조사 비율이 인터넷조사 비율에 비해 높고, 일반고는 인터넷조사 비율이 종이조사 비율에 비해 5.4%p 높은 것으로 나타났다. 이와 같은 경향은 2차 조사에서도 계속되며, 종이조사 비율이 높았던 초·중·특성화고의

인터넷조사 비율은 1차 조사에 비해 다소 상승하였음을 알 수 있다.

이에 반해 2011년 조사에서는 초등학교의 경우 종이조사 비율이 인터넷조사에 비해 다소 높고, 일반고는 인터넷조사 비율이 다소 높았으며, 중학교와 특성화고는 큰 차이가 없었다. 다음은 지역별 조사방법의 비율이다.



* 지역 기준 백분율임

[그림 1-4] 지역/조사방법별 분포(2차 조사 기준)

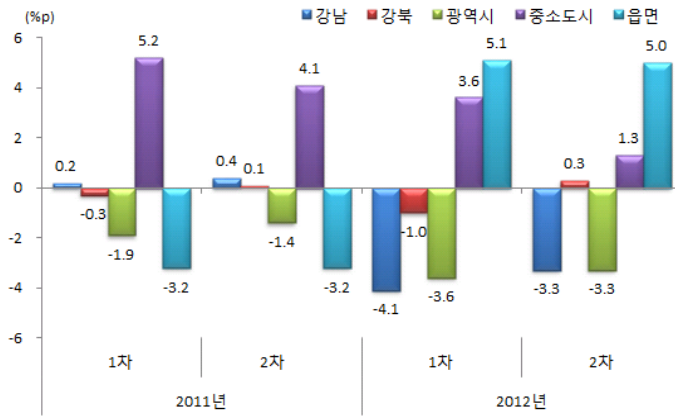
2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 지역별로 살펴 보면, 2011년에는 지역별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다. 특히 강남지역의 인터넷조사 참여 비율(89.3%)이 다른 지역에 비해 높은 것으로 나타났다.

이제 조사방법별로 지역의 분포를 살펴보자.

<표 1-3> 조사방법/지역별 분포

지역별 분포			강남	강북	광역시	중소도시	읍면
2011	1차	종이조사	4.3	7.8	30.0	43.4	14.5
		인터넷조사	4.1	8.1	31.9	38.2	17.7
	2차	종이조사	4.4	8.0	30.2	43.0	14.4
		인터넷조사	4.0	7.9	31.6	38.9	17.6
2012	1차	종이조사	1.4	6.9	29.2	43.6	18.9
		인터넷조사	5.5	7.9	32.8	40.0	13.8
	2차	종이조사	1.7	7.8	29.3	42.1	19.1
		인터넷조사	5.0	7.5	32.6	40.8	14.1

* 조사방법 기준 백분율임



* 조사방법 기준 백분율의 차이임

[그림 1-5] 조사방법/지역별 비율 차이(종이-인터넷)

2012년도 조사에서 조사방법별로 지역별 분포를 살펴보면, 강남 및 강북 등 서울과 광역시의 경우 인터넷조사 비율이 종이조사에 비해 높고, 중소도시와 읍면지역은 종이조사 비율이 다소 높은 것으로 나타났다.

이에 반해 2011년의 경우에는 중소도시에서는 종이조사, 읍면지역에서는 인터넷조사 비율이 높고, 서울 및 광역시의 경우에는 두 조사 비율이 유사한 것으로 나타났다.

2. 선택효과

혼합조사에서의 선택효과는 응답자가 선호하는 방법으로 조사에 참여하게 되므로 응답부담 및 무응답률을 낮추는 등 조사 결과에 긍정적 영향을 미칠 수 있다. 그러나 특정 성향을 지닌 응답자들이 특정 조사방법을 선택하는 경향이 강하게 나타났다면, 이는 조사 결과에 어떤 방식으로든 영향을 미치게 되므로 이를 파악하고 대응하는 과정이 필요하다.

먼저 주요 조사 항목 중 방과후 교육비에 대한 무응답률을 확인하여 긍정적 선택효과가 나타났는지 여부를 확인해보았다.

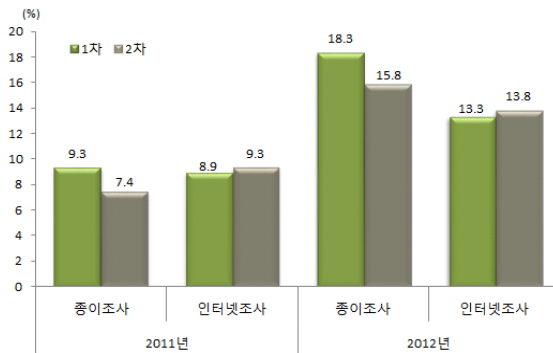
2012년 조사에서는 인터넷조사의 항목무응답 비율이 종이조사에 비해 전반적으로 낮은 경향을 보이는 것으로 확인되었다. 이에 반해 2011년에는 조사방법에 따른 무응답률의 크기가 일괄적이지 않고 비교적 유사한 것으로 확인되었다. 이는 조사방법에 대한 자발적 선택으로 항목무응답이 축소되는 선택효과가 나타난 것으로 볼 수 있다.

〈표 1-4〉 방과후 교육비 무응답 분포

(단위: %)

	2011			2012		
	전체	종이조사	인터넷조사	전체	종이조사	인터넷조사
1차	9.1	9.3	8.9	14.9	18.3	13.3
2차	8.4	7.4	9.3	14.3	15.8	13.8

주) 방과후 교육비 무응답은 방과후 참여시간에는 응답하였으나 교육비에는 응답하지 않은 비율임 (방과후 참여시간이 0인 경우, 방과후 교육비 무응답은 0원으로 처리)

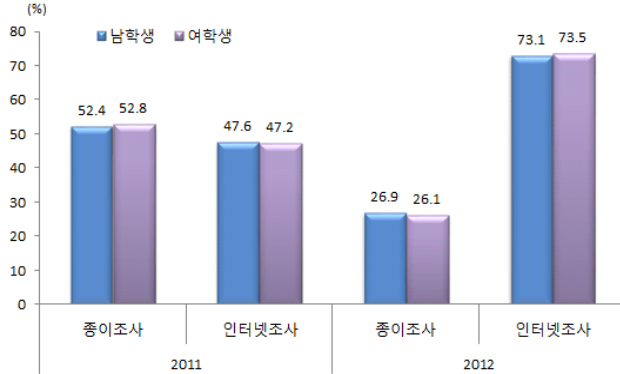


[그림 1-6] 방과후 교육비 무응답 분포

사교육비조사에서 가장 큰 관심 대상인 사교육시간에는 무응답이 존재하지 않았으나, 사교육비에는 상당히 큰 무응답이 존재하였다. 확인 결과, 해당 무응답은 사교육시간이 “0”인 경우로, 사교육을 받지 않아 사교육비에 대한 응답이 불가능했던 것이었다. 따라서 사교육시간에 대한 무응답은 응답해야 하지만 어떤 이유로 인하여 응답하지 않은 일반적인 무응답과는 다른 것이기 때문에 사교육비 관련 무응답 분석은 별도로 진행하지 않았다. 조사방법에 따른 사교육 참여 여부는 조사방법에 따른 측정 효과이므로 다음 절에서 다루도록 한다.

이제 2012년 사교육비조사에 적용된 두 조사방법에 대한 선택에 응답자의 어떤 특성이 영향을 미쳤는지 살펴보자. 이를 위하여 조사방법을 선택할 수 없었던 2011년도의 응답자 분포와 비교하도록 한다.

다음은 학생 성별 조사방법의 비율이다.



* 성별 기준 백분율임

[그림 1-7] 성/조사방법별 분포(2차 조사 기준)

2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 학생의 성별로 살펴보면, 2011년에는 성별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다.

이제 조사방법별로 학생 성별 분포를 살펴보자.

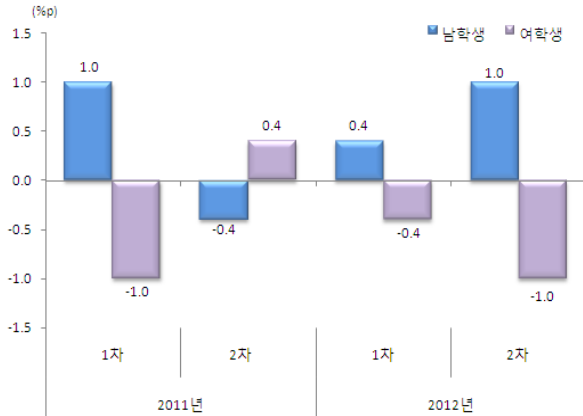
<표 1-5> 조사방법/학생 성별 분포

(단위: %)

		종이조사		인터넷조사	
		1차	2차	1차	2차
2011	남학생	54.1	53.4	53.1	53.8
	여학생	45.9	46.6	46.9	46.2
2012	남학생	52.3	52.7	51.9	51.7
	여학생	47.7	47.3	48.1	48.3

* 조사방법 기준 백분율임

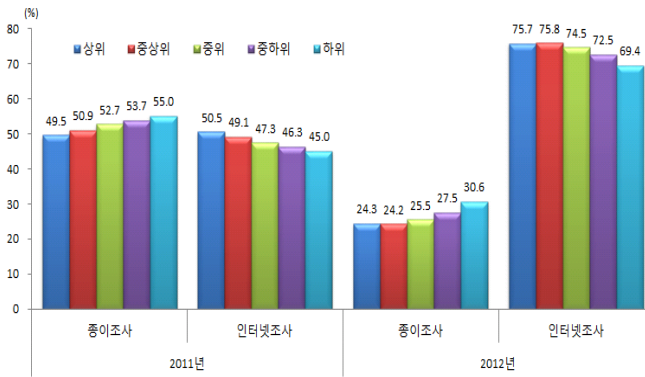
조사방법을 선택할 수 없었던 2011년도와 조사방법을 선택한 2012년도에 학생 성별에 따른 인터넷조사 선택 여부에는 별다른 차이가 없는 것으로 확인되었다.



* 조사방법 기준 백분율의 차이임

[그림 1-8] 성/조사방법별 비율 차이(종이-인터넷)

다음은 성적등급별 조사방법의 비율이다.



* 성적등급 기준 백분율임

[그림 1-9] 성적등급/조사방법별 분포(2차 조사 기준)

2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 학생의 성적 등급별로 살펴보면, 2011년에는 성적등급별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다. 또한 중하위(27.5%) 및 하위(30.6%)의 종이조사 참여 비율이 중위 이상의 참여 비율에 비해 높은 것으로 나타났다.



이제 조사방법별로 학생의 성적등급별 분포를 살펴보자.

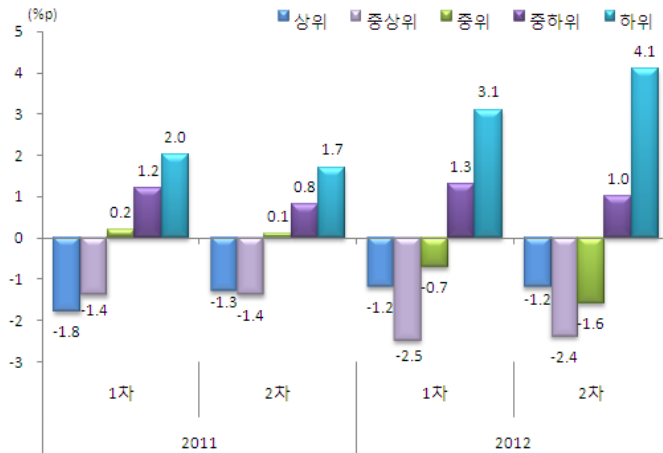
〈표 1-6〉 조사방법/학생 성적등급별 분포

(단위: %)

성적등급별 분포		상위	중상위	중위	중하위	하위	
2011	1차	종이조사	10.0	20.1	30.3	20.2	19.5
		인터넷조사	11.8	21.5	30.1	19.0	17.5
	2차	종이조사	10.3	20.1	30.2	20.0	19.4
		인터넷조사	11.6	21.5	30.1	19.2	17.7
2012	1차	종이조사	9.9	18.7	29.8	21.0	20.7
		인터넷조사	11.1	21.2	30.5	19.7	17.6
	2차	종이조사	9.6	18.3	28.7	20.7	22.6
		인터넷조사	10.8	20.7	30.3	19.7	18.5

* 조사방법 기준 백분률임

학생의 성적등급이 중위이상인 경우 인터넷조사에 참여한 비율이 종이조사에 비해 높고, 중하위 이하에는 종이조사 참여율이 다소 높은 것으로 나타났다.

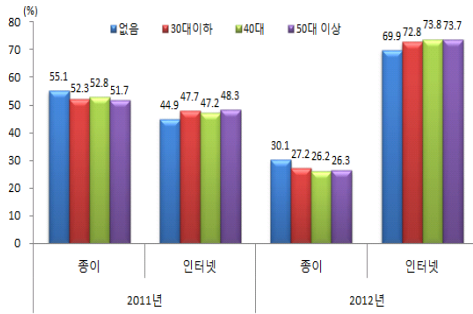


* 조사방법 기준 백분율의 차이임

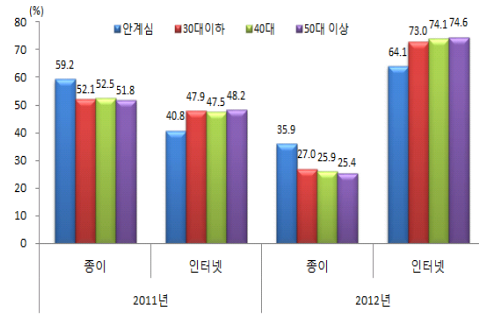
[그림 1-10] 조사방법/성적등급별 비율 차이(종이-인터넷)

다음은 부모 연령별 조사방법의 비율이다.

• 아버지 연령



• 어머니 연령



* 부모 연령 기준 백분율임

[그림 1-11] 부모 연령/조사방법별 분포(2차 조사 기준)

2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 부모의 연령별로 살펴보면, 2011년에는 부모의 연령별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다. 이 때 한부모인 경우(편부 또는 편모인 경우)에 인터넷조사 참여 비율이 다소 낮은 것으로 나타났다.

이제 조사방법별로 부모 연령별 분포를 살펴보자.

<표 1-7> 조사방법/부모 연령별 분포

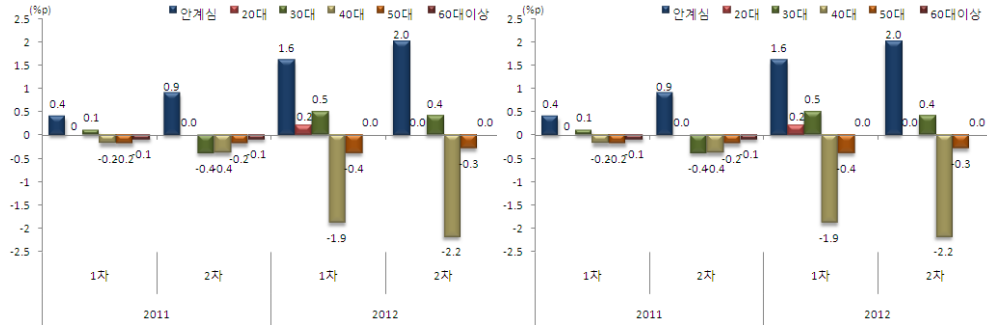
부모 연령별 분포			(단위: %)						
			안계심	20대	30대	40대	50대	60대 이상	
2011	1차	아버지	종이조사	4.2	0.0	8.4	66.8	20.0	0.5
		인터넷조사	4.2	0.1	8.3	66.3	20.6	0.5	
	어머니	종이조사	3.5	0.2	19.5	69.0	7.8	0.1	
		인터넷조사	3.1	0.2	19.4	69.2	8.0	0.2	
	2차	아버지	종이조사	4.6	0.0	8.2	66.6	20.0	0.5
		인터넷조사	4.2	0.0	8.3	66.2	20.8	0.5	
2012	1차	아버지	종이조사	5.7	0.2	11.1	66.3	16.3	0.6
		인터넷조사	4.9	0.1	10.6	68.0	15.9	0.5	
	어머니	종이조사	5.2	0.5	24.5	64.4	5.4	0.1	
		인터넷조사	3.6	0.3	24.0	66.3	5.8	0.1	
	2차	아버지	종이조사	6.1	0.1	10.0	65.9	17.4	0.6
		인터넷조사	5.1	0.1	9.7	67.0	17.6	0.5	
어머니	종이조사	5.6	0.3	22.6	65.4	6.0	0.1		
	인터넷조사	3.6	0.3	22.2	67.6	6.3	0.1		

* 조사방법 기준 백분율임



• 아버지 연령

• 어머니 연령



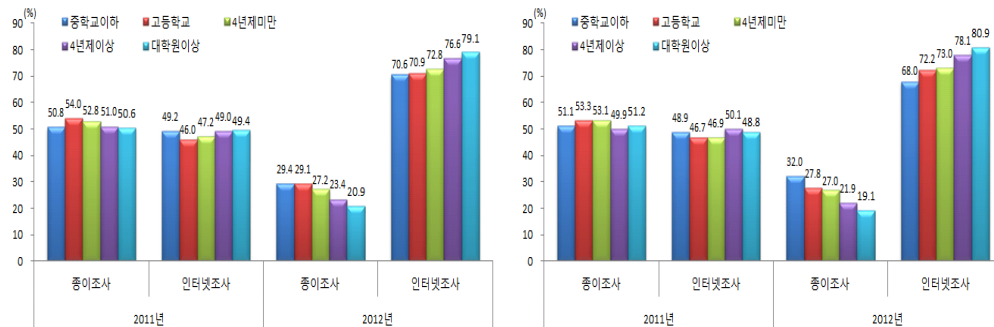
* 조사방법 기준 백분율의 차이임

[그림 1-12] 조사방법/부모 연령별 비율 차이(종이-인터넷)

조사방법을 선택할 수 없었던 2011년의 경우 부모 연령에 따른 조사방법에 별다른 차이가 없었으나, 2012년 조사에서는 아버지와 어머니의 연령이 40대인 경우 인터넷조사 비율이 다소 높고, 한 부모가 없는 경우에는 종이조사 비율이 다소 높은 것으로 나타났다. 다음은 부모의 최종학력별 조사방법의 비율이다.

• 아버지 최종학력

• 어머니 최종학력



* 부모 최종학력 기준 백분율임

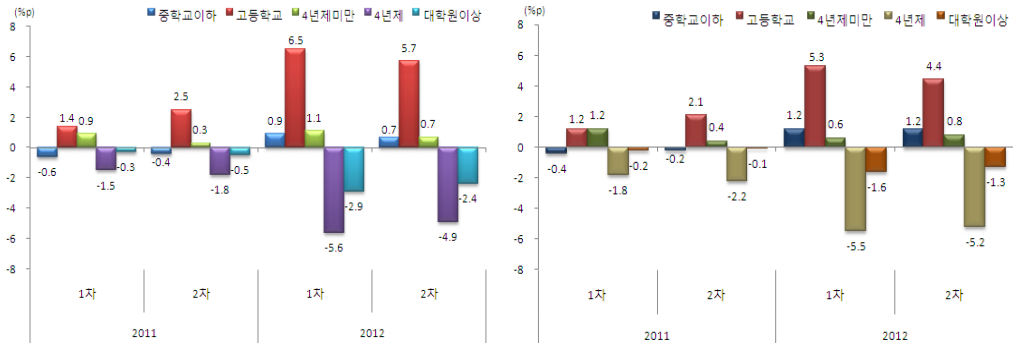
[그림 1-13] 부모 최종학력/조사방법별 분포(2차 조사 기준)

2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 부모의 최종학력별로 살펴보면, 2011년에는 최종학력별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다. 이 때 최종학력이 높을수록 인터넷조사 참여 비율이 높은 것으로 나타났다.

이제 조사방법별로 부모의 최종학력 분포를 살펴보자.

• 아버지 최종학력

• 어머니 최종학력



* 조사방법 기준 백분율의 차이임

[그림 1-14] 조사방법/ 부모 최종학력별 비율 차이(종이-인터넷)

<표 1-8> 조사방법/부모 최종학력별 분포

(단위: %)

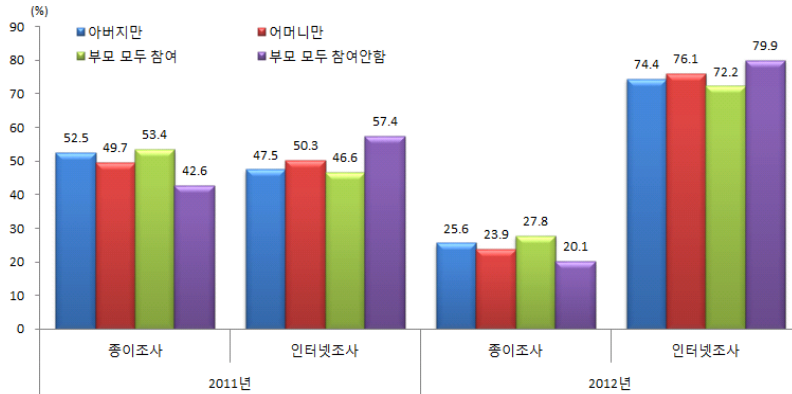
부모 최종학력별 분포			중학교이하	고등학교	4년제 미만	4년제	대학원이상	
2011	1차	아버지	종이조사	4.6	42.1	15.9	29.8	7.6
			인터넷조사	5.2	40.7	15.0	31.3	7.9
		어머니	종이조사	4.9	54.0	16.3	21.5	3.3
	인터넷조사		5.3	52.8	15.1	23.3	3.5	
	2차	아버지	종이조사	4.6	42.6	15.6	29.7	7.6
			인터넷조사	5.0	40.1	15.3	31.5	8.1
어머니		종이조사	4.9	54.5	15.9	21.3	3.4	
		인터넷조사	5.1	52.4	15.5	23.5	3.5	
2012	1차	아버지	종이조사	4.9	44.3	15.9	28.5	6.4
			인터넷조사	4.0	37.8	14.8	34.1	9.3
		어머니	종이조사	5.0	55.0	17.2	20.2	2.6
	인터넷조사		3.8	49.7	16.6	25.7	4.2	
	2차	아버지	종이조사	4.7	44.0	15.8	28.8	6.6
			인터넷조사	4.0	38.3	15.1	33.7	9.0
어머니		종이조사	5.0	54.6	17.5	20.2	2.7	
인터넷조사	3.8	50.2	16.7	25.4	4.0			

* 조사방법 기준 백분율임

조사방법을 선택할 수 있었던 2012년 조사의 경우, 아버지와 어머니의 최종학력이 중학교 이하, 고등학교, 4년제 미만인 경우 종이조사 비율이 다소 높고, 4년제 및 석사 이상 대학원의 경우에는 인터넷조사 비율이 다소 높다.



다음은 부모의 경제활동 참여형태별 조사방법의 비율이다.



* 부모의 경제활동 참여형태 기준 백분율임

[그림 1-15] 부모의 경제활동 참여형태/조사방법별 분포(2차 조사 기준)

2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 부모의 경제활동 참여유형별로 살펴보면, 2011년에는 경제활동 참여유형별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다.

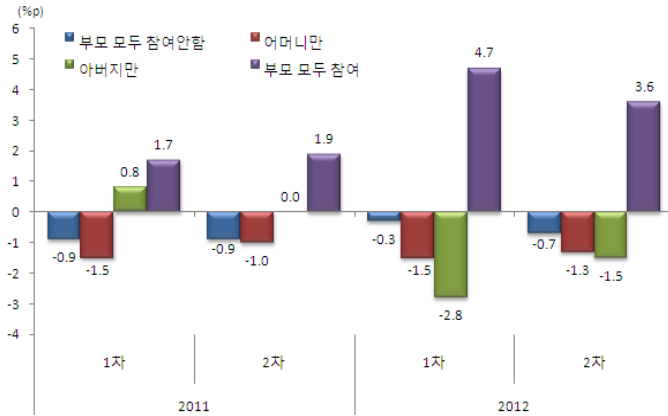
이제 조사방법별로 부모의 경제활동 참여유형 분포를 살펴보자.

<표 1-9> 조사방법/부모 경제활동별 분포

(단위: %)

경제활동별 분포		부모 모두 참여안함	어머니만	아버지만	부모 모두 참여	
2011	1차	종이조사	1.9	8.0	37.8	52.3
		인터넷조사	2.8	9.5	37.0	50.6
	2차	종이조사	1.8	8.1	37.0	53.1
		인터넷조사	2.7	9.1	37.0	51.2
2012	1차	종이조사	2.2	9.0	34.2	54.6
		인터넷조사	2.5	10.5	37.0	49.9
	2차	종이조사	1.8	9.0	34.8	54.5
		인터넷조사	2.5	10.3	36.3	50.9

* 조사방법 기준 백분율임

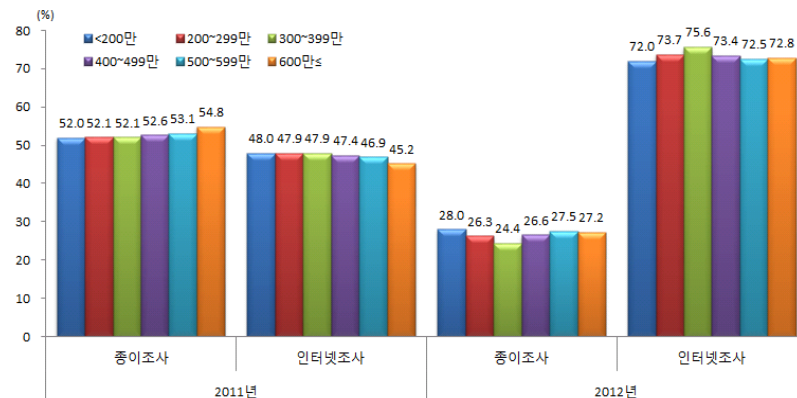


* 조사방법 기준 백분율의 차이임

[그림 1-16] 조사방법/부모의 경제활동 참여형태별 비율 차이(종이-인터넷)

아버지와 어머니가 모두 경제활동에 참여하지 않거나, 어머니만 참여하는 경우에는 조사방법 선택에 관계 없이 인터넷조사 비율이 다소 높다. 반면, 아버지만 참여하는 경우에는 조사방법을 선택한 2012년에 인터넷조사 참여 비율이 더 높아진 것으로 나타났다. 부모 모두가 경제활동에 참여하는 경우에는 종이조사 비율이 다소 높다.

마지막으로 가구의 월평균 소득별 조사방법의 비율이다.



* 가구 월평균 소득 기준 백분율임

[그림 1-17] 가구의 월평균 소득/조사방법별 분포(2차 조사 기준)



2011년과 2012년 2차 조사를 중심으로 인터넷조사와 종이조사 비율을 가구의 월평균 소득별로 살펴보면, 2011년에는 월평균 소득별로 두 조사방법의 비율이 비슷했던 반면 2012년에는 전반적으로 인터넷조사 비율이 높은 것으로 확인되었다.

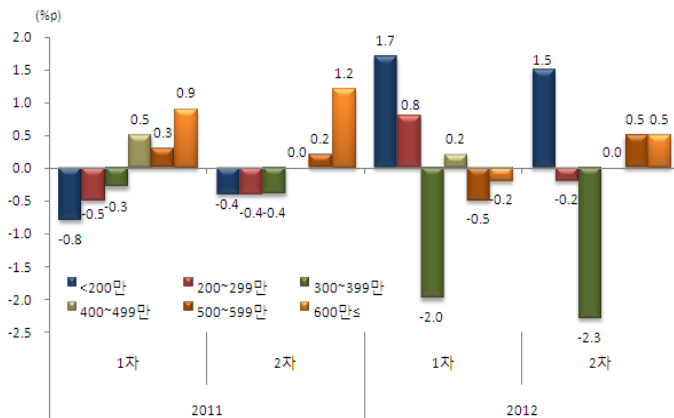
이제 조사방법별로 가구의 월평균 소득 분포를 살펴보자.

〈표 1-10〉 조사방법/가구 월평균 소득별 분포

(단위: %)

월평균 소득별 분포			<200만	200~299만	300~399만	400~499만	500~599만	600만≤
2011	1차	종이조사	20.6	20.4	21.0	14.6	9.3	14.1
		인터넷조사	21.4	20.9	21.3	14.1	9.0	13.2
	2차	종이조사	19.5	20.7	21.5	15.0	9.6	13.9
		인터넷조사	19.9	21.1	21.9	15.0	9.4	12.7
2012	1차	종이조사	21.3	20.8	19.1	15.1	9.4	14.4
		인터넷조사	19.6	20.0	21.1	14.9	9.9	14.6
	2차	종이조사	20.4	20.6	18.9	15.8	10.2	14.1
		인터넷조사	18.9	20.8	21.2	15.8	9.7	13.6

* 조사방법 기준 백분율임



* 조사방법 기준 백분율의 차이임

[그림 1-18] 조사방법/가구의 월평균 소득별 비율 차이(종이-인터넷)

가구의 월평균 소득에 따른 인터넷조사 여부를 확인한 결과, 200만원 미만 가구는 종이조사 참여 비율이 다소 높고, 300만원~400만원 미만인 경우에는 인터넷조사 비율이 다소 높으며, 다른 경우에는 별다른 차이가 나타나지 않는 것으로 확인되었다.

제3절 조사방법별 측정효과

1. 측정효과

2012년 사교육비조사는 먼저 인터넷조사를 권고하고 이어 종이조사가 진행되었다. 이 때 인터넷조사에 응답한 응답자가 종이조사에 임할 경우 다른 응답을 하게 된다면, 조사모드에 따른 측정오차가 발생된 것으로 볼 수 있다. 일반적으로 수입 및 지출 등 민감한 개인 정보 또는 이혼 및 동거 등 사생활에 대한 항목에 조사방법에 따른 뚜렷한 차이가 있는 것으로 알려져 있다(Martin & Lynn, 2011; Revilla, 2010; Vannieuwenhuyze et al., 2010).

Dillman(2000)에 의하면 조사원이 있는 조사와 자기기입식 조사에서는 조사원에 의한 간섭, 시청각 자료의 활용성, 인센티브 차이 등에 의해 응답 경향에 차이가 발생하게 된다. 그러나 사교육비조사는 다른 혼합조사들과는 달리 각 조사방법에 대한 조사 주체가 응답자(자기기입식)로 동일하다는 특징을 가지고 있다. 즉, 인터넷으로 조사에 참여하든, 종이조사표를 이용한 조사에 참여하든지에 관계없이 응답자는 모두 자기기입식 형태로 조사에 참여하게 된다. 따라서 사교육비조사에서의 혼합모드에 의한 모드효과에는 조사원 효과, 시청각 자료의 활용 및 인센티브 효과가 반영되지 않아, 조사방법에 따른 순수한 모드효과를 측정할 수 있다.

먼저 주요 변수(사교육시간 및 사교육비)에 대한 이상점 존재 여부를 살펴보자.

〈표 1-11〉 주당 사교육시간 응답 분포(0 포함)

(단위: 시간)

주당 사교육시간		Min.	Q1	Median	Q3	Max.	
2011	1차조사	종이	0	0	5	10	49
		인터넷	0	0	4	9	48
	2차 조사	종이	0	0	4	10	62
		인터넷	0	0	4	9	46
2012	1차조사	종이	0	0	4	9	48
		인터넷	0	0	4	9	48
	2차 조사	종이	0	0	3	9	60
		인터넷	0	0	3	9	46



〈표 1-12〉 주당 사교육시간 응답 분포(0 제외)

(단위: 시간)

주당 사교육시간			Min.	Q1	Median	Q3	Max.
2011	1차조사	종이	1	5	8	12	49
		인터넷	1	4	8	12	48
	2차 조사	종이	1	5	8	12	62
		인터넷	1	5	8	12	46
2012	1차조사	종이	1	4	7	11	48
		인터넷	1	4	7	11	48
	2차 조사	종이	1	4	7	11	60
		인터넷	1	4	7	11	46

〈표 1-13〉 월평균 사교육비 응답 분포(사교육시간=0 제외)

(단위: 만원)

월평균 사교육비			Min.	Q1	Median	Q3	Max.
2011	1차조사	종이	0.30	18.50	30.00	45.00	300.00
		인터넷	0.07	20.00	30.00	47.45	390.00
	2차 조사	종이	0.33	18.00	30.00	45.00	360.00
		인터넷	0.10	19.67	30.00	48.00	651.67
2012	1차조사	종이	0.40	18.00	30.00	44.00	245.00
		인터넷	0.20	20.00	32.00	50.00	483.30
	2차 조사	종이	0.10	17.48	30.00	44.00	330.00
		인터넷	0.10	20.00	30.00	48.33	665.00

주당 사교육시간은 최대값이 46시간~62시간으로 매일(7일 기준) 6시간 34분 ~ 8시간 51분 정도이며, 60시간 및 62시간은 모두 미술, 46시간 ~ 50시간은 국·영·수에 대한 사교육이었던 것으로 확인되었다. 과목의 특성 및 과외, 학원 등 동일 과목에 대한 다양한 사교육 방법의 적용이 있는 것으로 확인되었으므로, 기입상의 실수가 있거나 극단값이 포함된 것으로 보이지 않는다. 반면 월평균 사교육비로 지출한 금액은 최소값이 7백원~4천원이고, 최대값이 245만원~665만원으로 나타났다. 이 때 사교육비는 사교육 종류에 깊은 연관 관계를 가지므로 각 조사에서의 사교육비 최소값과 최대값에 대한 사교육 종류를 확인하였다.

확인 결과, 2012년의 경우 1차 종이조사에서는 3개월간 사교육비가 1.2만원인 사례가 발견되었으며, 인터넷조사에서는 6천원, 2차 종이조사와 인터넷조사에서는 각 3천원인 조사 결과가 포함되어 있었다. 해당 사교육비가 인터넷 강좌 등에 의한 것이라면 이상값으로 보기 어려우나, 대부분 그룹과의, 개인과의 및 학원수강 등의 방법으로 사교육이 참여하고 있는 것으로 보아 응답자가 축소 응답했을 가능성이 있다.

반면 2012년 1차 인터넷조사의 사교육비 중 최대값인 483.3만원과 2차 인터넷조사의 사교육비 중 최대값인 665만원은 국어, 영어, 수학 및 사회, 과학 등 일반교과에 대한 개인과의 및 학원수강 비용으로 확인되었으며, 발생 가능한 값으로 판단된다.

사교육비는 사교육 종류와 더불어 사교육시간에 깊은 연관 관계를 가지므로 시간당 사교육비에 대한 확인이 필요하다.

〈표 1-14〉 월평균 시간당 사교육비 응답 분포

(단위: 만원)

월평균 시간당 사교육비			Min.	Q1	Median	Q3	Max.
2011	1차조사	종이	0.01	0.60	0.95	1.63	15.00
		인터넷	0.01	0.61	1.04	1.73	30.00
	2차 조사	종이	0.01	0.58	0.93	1.62	20.00
		인터넷	0.01	0.60	1.00	1.67	18.83
2012	1차조사	종이	0.02	0.63	1.00	1.67	22.50
		인터넷	0.01	0.68	1.13	1.90	39.00
	2차 조사	종이	0.00	0.61	0.97	1.67	12.50
		인터넷	0.01	0.66	1.07	1.79	24.50

조사방법을 선택할 수 없었던 2011년 조사의 경우, 1차 조사에서는 인터넷조사에서의 시간당 사교육비가 더 큰 반면, 2차 조사에서는 더 작은 것으로 나타났다. 그러나 조사방법을 선택한 2012년의 경우에는 전반적으로 종이조사에 비해 인터넷조사에서의 사교육비가 더 큰 값을 갖는 것으로 보인다.

다음은 조사방법에 따른 차이가 통계적으로 유의미한 것인지 확인하기 위한 t-test 결과이다.



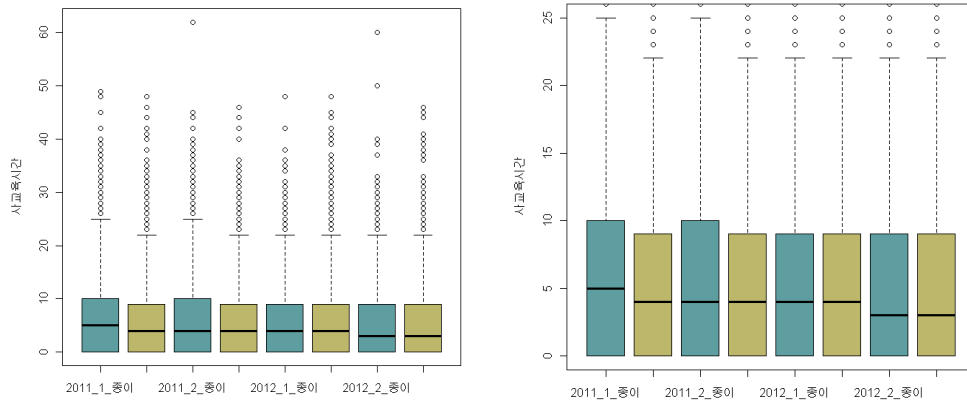
〈표 1-15〉 조사방법에 따른 사교육시간의 차이 분석

(단위: 시간, mean±sd)

주당 사교육시간		종이	인터넷	t-value	p-value
2011	1차조사	5.96±6.11	5.44±6.21	8.92	<.0001
	2차 조사	5.43±6.13	5.27±6.09	2.90	0.004
2012	1차조사	5.16±5.74	5.41±6.11	-3.98	<.0001
	2차 조사	4.89±5.74	4.99±5.85	-1.57	0.115

2011년 사교육시간은 종이조사가 인터넷조사에 비해 유의미하게 긴 것으로 나타난 반면, 2012년에는 인터넷조사가 종이조사에 비해 사교육시간이 긴 것으로 나타났다. 그러나 조사방법에 따른 차이가 모두 유의미한 것으로 나타나지는 않았다.

이는 조사방법을 선택할 수 있었을 때 조사방법에 따른 사교육시간의 차이가 더 작아진 것으로 볼 수 있으며, 이는 선택효과의 긍정적 발현으로 볼 수 있을 것이다.



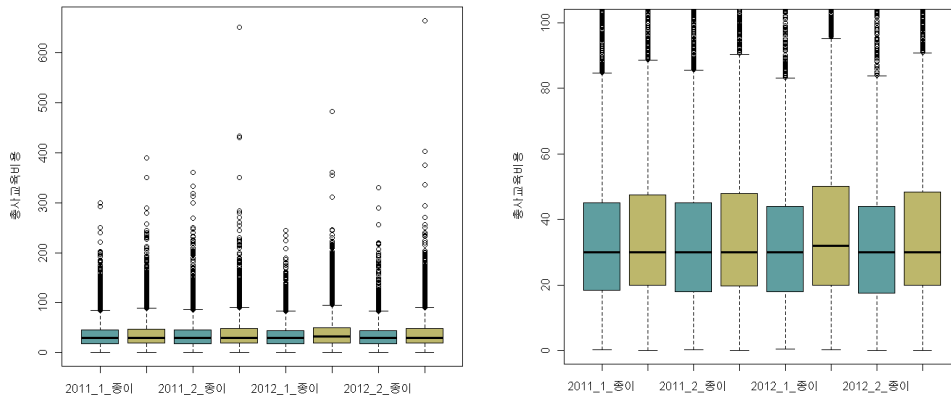
[그림 1-19] 조사방법에 따른 사교육시간 분포

다음은 사교육비에 대한 차이 검정 결과이다.

〈표 1-16〉 조사방법에 따른 사교육비의 차이 분석

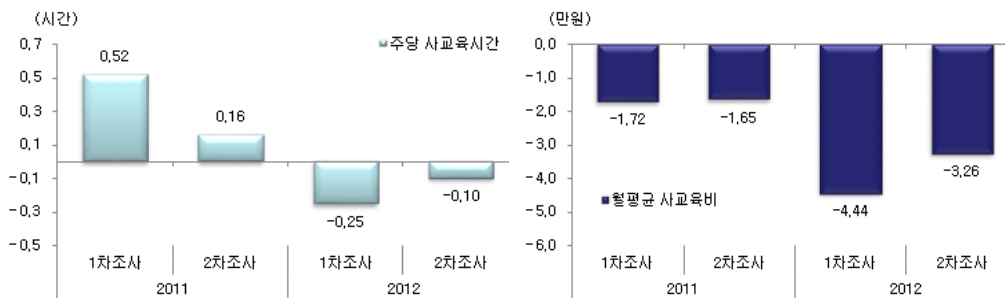
(단위: 만원, mean±sd)

월평균 사교육비		종이	인터넷	t-value	p-value
2011	1차조사	34.88±24.49	36.60±26.61	-5.78	<.0001
	2차 조사	34.79±25.85	36.44±26.87	-5.17	<.0001
2012	1차조사	33.72±22.99	38.16±27.52	-13.11	<.0001
	2차 조사	33.82±24.17	37.08±27.45	-8.61	<.0001



[그림 1-20] 조사방법에 따른 사교육비 분포

2011년과 2012년 모두 인터넷조사에서의 사교육비가 종이조사에 비해 큰 것으로 나타났으며, 그 차이는 사교육시간에 대한 차이 검정 결과와는 달리 2012년 조사에서 더 큰 것으로 확인되었다. 앞서 논의한 바와 같이 사교육비조사에서의 두 조사방법은 모두 자기기입식으로 진행되므로, 이와 같은 차이는 인터넷조사에 의한 측정 효과로 볼 수 있을 것이다.



[그림 1-21] 주당 사교육시간 및 월평균 사교육비 차이(종이-인터넷)

다음은 조사방법에 따른 시간당 사교육비의 차이 검정 결과이다.

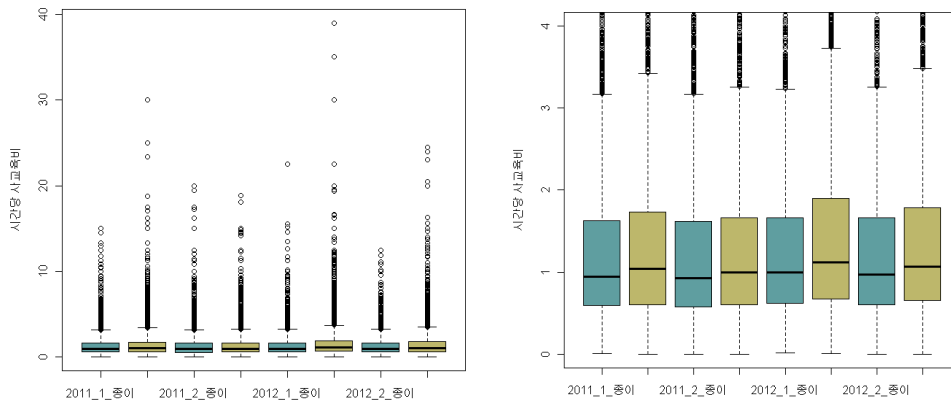


〈표 1-17〉 조사방법에 따른 시간당 사교육비의 차이 분석

(단위: 만원, mean±sd)

시간당 사교육비(월평균)		종이	인터넷	t-value	p-value
2011	1차조사	1.26±1.04	1.37±1.27	-8.17	<.0001
	2차 조사	1.27±1.11	1.35±1.21	-6.02	<.0001
2012	1차조사	1.35±1.17	1.55±1.49	-10.59	<.0001
	2차 조사	1.30±1.06	1.44±1.30	-8.18	<.0001

시간당 사교육비 차이 검정 결과 역시 사교육비와 유사한 형태를 보이는 것으로 확인되었다.



[그림 1-22] 조사방법에 따른 시간당 사교육비 분포

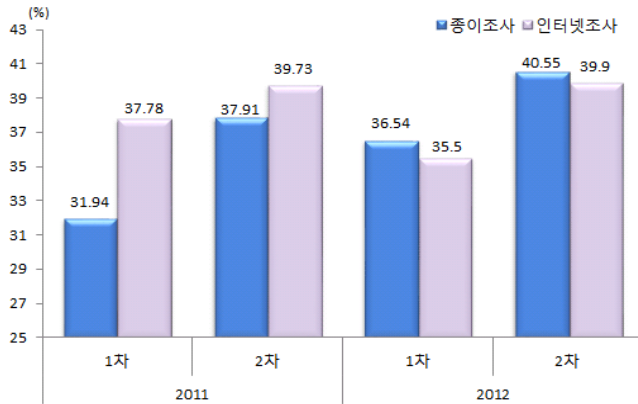
분석 결과들을 종합해보면, 조사방법을 선택할 수 있었던 2012년 사교육비조사에는 선택효과와 측정효과가 포함되어 있는 것으로 볼 수 있다. 이 때 두 조사방법 중 기준이 되는(더 신뢰할 만한) 조사방법은 무엇인가에 대한 논의가 필요하다. 박상언 외(2012)는 종이조사가 인터넷조사에 비해 더 신뢰할 수 있는 것으로 보았으며, Marcel(2012)은 전통적 조사방법인 종이조사, 면접조사 등을 참값으로 보고, 조사 환경(응답자)의 특성에 따라 새롭게 적용된 조사방법이 선택효과 및 측정효과를 가지게 된다고 한다. 그러나 최근 진행되고 있는 혼합조사에서는 전통적인 조사방법에 비해 새로운 조사방법의 적용 비율이 더 높고, 신뢰도 또한 전통적인 조사방법보다 오히려 높다는 전문가 의견도 많다. 이와 같이 참조조사(reference survey)에 대한 전문가 의견이 하나로 모아지지 않고 있으므로, 전통적인 조사방법을 기준으로 삼을 것인지, 다수의 조사대상자가 참여한 새로운 조사방법을 기준으로 삼을지 여부에 대한 고민도 필요하다.

다음은 조사방법에 따른 사교육비 항목 무응답(사교육을 받지 않는 학생) 비율이다.

〈표 1-18〉 조사방법에 따른 사교육비 항목 무응답(사교육을 받지 않는 학생) 비율

(단위: %)

	1차조사		2차 조사		합	
	종이	인터넷	종이	인터넷	종이	인터넷
2011	31.94	37.78	37.91	39.73	34.90	38.76
2012	36.54	35.50	40.55	39.90	38.35	37.77



[그림 1-23] 조사방법에 따른 사교육비 항목 무응답(사교육을 받지 않는 학생) 비율

2011년의 경우 인터넷조사가 종이조사보다 사교육비 항목 무응답 비율이 더 높았으나, 2012년에는 종이조사가 인터넷조사보다 더 높은 것으로 나타났다. 이는 조사방법을 선택할 수 없었던 2011년의 경우 인터넷조사의 신뢰도가 종이조사에 비해 더 낮을 수 있음을 의미함과 동시에, 조사방법에 대한 선택이 가능했던 2012년에는 종이조사의 신뢰도가 상대적으로 더 낮을 가능성이 있음을 의미한다. 또한 2012년 조사에서는 전체 응답자의 다수(1차(68.16%), 2차(73.49%))가 인터넷조사를 선택하였으므로 두 조사방법 중 인터넷조사를 참조조사로 활용하는 방안을 고려할 수 있을 것이다.

그러나 조사방법에 따른 사교육비 관련 주요 지표들의 표준편차가 종이조사에 비해 인터넷조사가 다소 큰 것으로 나타나는 등 참조조사로 인터넷조사를 활용하는 문제는 좀 더 신중하게 결정할 필요가 있다. 본 연구에서는 기존 연구와 같이 종이조사를 참조조사로 이용하며, 추후 인터넷조사를 참조조사로 하는 연구를 지속적으로 진행하고자 한다.



다음은 2012년 사교육비조사에서의 지역별 사교육시간 및 비용에 대한 조사방법에 따른 차이 검정 결과이다.

〈표 1-19〉 지역별 조사방법에 따른 사교육시간의 차이 분석

(단위: 시간, mean±sd)

주당 사교육시간		종이	인터넷	t-value	p-value	
2011	1차조사	강남	8.53±7.06	7.92±6.97	1.90	0.057
		강북	6.84±6.32	6.83±6.77	0.02	0.981
		광역시	6.18±5.96	5.80±6.08	3.75	0.000
		중소도시	5.89±6.10	5.17±6.17	7.92	<.0001
		읍면	4.47±5.63	4.17±5.72	2.19	0.029
	2차 조사	강남	7.95±6.86	7.70±6.66	0.80	0.423
		강북	6.29±6.43	6.14±6.56	0.68	0.496
		광역시	5.79±5.94	5.61±6.02	1.85	0.065
		중소도시	5.30±6.15	5.04±5.99	2.83	0.005
		읍면	3.84±5.63	4.20±5.80	-2.66	0.008
2012	1차조사	강남	7.42±6.38	8.10±6.79	-1.31	0.189
		강북	5.62±5.73	6.18±5.72	-2.55	0.011
		광역시	5.81±5.94	5.62±6.15	1.64	0.101
		중소도시	5.25±5.75	5.26±6.04	-0.05	0.959
		읍면	3.67±4.98	3.84±5.62	-1.28	0.202
	2차 조사	강남	7.42±6.60	7.95±6.91	-1.00	0.319
		강북	5.22±6.11	6.03±6.14	-3.36	0.001
		광역시	5.63±5.79	5.22±5.80	3.55	0.000
		중소도시	4.87±5.73	4.73±5.64	1.47	0.142
		읍면	3.44±5.13	3.61±5.43	-1.21	0.2256

〈표 1-20〉 지역별 조사방법에 따른 사교육비의 차이 분석

(단위: 만원, mean±sd)

월평균 사교육비		종이	인터넷	t-value	p-value	
2011	1차조사	강남	60.76±38.43	61.93±41.77	-0.56	0.572
		강북	43.48±31.06	47.33±32.87	-3.09	0.002
		광역시	32.81±21.52	34.53±24.79	-3.61	0.000
		중소도시	34.11±21.96	36.08±23.07	-4.66	<.0001
		읍면	25.74±17.79	26.89±19.26	-1.90	0.058
	2차 조사	강남	61.88±42.99	63.61±49.73	-0.71	0.476
		강북	44.53±34.32	45.85±33.75	-0.95	0.341
		광역시	32.11±21.72	35.35±24.10	-6.71	<.0001
		중소도시	33.59±22.42	35.26±22.58	-3.88	0.000
		읍면	25.39±17.29	27.10±18.68	-2.77	0.006
2012	1차조사	강남	55.22±37.48	64.48±45.51	-2.30	0.022
		강북	40.66±24.88	46.89±31.93	-4.94	<.0001
		광역시	34.31±24.15	36.03±24.21	-3.14	0.002
		중소도시	33.96±22.13	36.13±23.54	-4.74	<.0001
		읍면	26.04±16.43	28.77±20.77	-4.07	<.0001
	2차 조사	강남	59.94±41.66	63.71±49.57	-8.78	0.380
		강북	42.59±30.12	45.72±31.98	-2.00	0.046
		광역시	33.95±23.91	34.86±22.94	-1.59	0.112
		중소도시	33.36±22.34	35.13±23.45	-3.50	0.000
		읍면	40.66±24.88	46.89±31.93	-4.41	0.000



〈표 1-21〉 지역별 조사방법에 따른 시간당 사교육비의 차이 분석

(단위: 만원, mean±sd)

시간당 사교육비(월평균)		종이	인터넷	t-value	p-value	
2011	1차조사	강남	1.69±1.25	1.95±1.92	-2.99	0.003
		강북	1.51±1.32	1.69±1.80	-2.88	0.004
		광역시	1.21±0.94	1.28±1.11	-3.55	0.000
		중소도시	1.26±1.00	1.41±1.19	-7.27	<.0001
		읍면	1.07±1.03	1.09±0.95	-0.57	0.566
	2차 조사	강남	1.77±1.36	1.84±1.46	-0.86	0.388
		강북	1.65±1.63	1.71±1.74	-0.99	0.322
		광역시	1.19±1.06	1.27±1.07	-3.26	0.001
		중소도시	1.23±0.97	1.34±1.10	-5.60	<.0001
		읍면	1.04±0.93	1.18±1.20	-3.84	0.000
2012	1차조사	강남	1.49±0.98	1.83±1.45	-2.63	0.009
		강북	1.51±1.40	1.68±1.39	-2.56	0.011
		광역시	1.33±1.18	1.55±1.57	-7.45	<.0001
		중소도시	1.39±1.15	1.50±1.45	-4.30	<.0001
		읍면	1.20±1.12	1.42±1.44	-4.80	<.0001
	2차 조사	강남	1.73±1.10	1.79±1.48	-0.57	0.569
		강북	1.54±1.36	1.67±1.47	-1.80	0.073
		광역시	1.25±1.03	1.39±1.25	-5.41	<.0001
		중소도시	1.33±1.05	1.41±1.24	-3.47	0.001
		읍면	1.14±0.95	1.34±1.35	-4.65	<.0001

다음은 2012년 사교육비조사에서의 학교급별 사교육시간 및 비용에 대한 조사방법에 따른 차이 검정 결과이다.

〈표 1-22〉 학교급별 조사방법에 따른 사교육시간의 차이 분석

(단위: 시간, mean±sd)

주당 사교육시간		종이	인터넷	t-value	p-value	
2011	1차조사	초등학교	7.68±5.88	7.50±6.38	1.45	0.148
		중학교	7.71±6.68	6.95±6.68	6.01	<.0001
		일반고	4.64±5.45	4.21±5.46	5.46	<.0001
		특성화고	1.60±3.94	2.06±4.80	-3.21	0.001
	2차 조사	초등학교	7.46±5.98	7.64±5.99	-1.56	0.119
		중학교	7.13±6.70	6.74±6.58	3.14	0.002
		일반고	3.98±5.44	3.77±5.23	2.66	0.008
		특성화고	1.33±3.45	1.90±5.08	-3.91	<.0001
2012	1차조사	초등학교	6.62±5.70	6.87±6.15	-2.05	0.040
		중학교	6.70±6.21	6.80±6.38	-0.78	0.436
		일반고	3.98±4.98	4.56±5.74	-6.82	<.0001
		특성화고	1.73±4.55	1.63±4.26	0.66	0.513
	2차 조사	초등학교	6.50±5.66	6.82±5.63	-2.65	0.008
		중학교	6.26±6.19	6.34±6.25	-0.60	0.550
		일반고	3.62±5.17	3.91±5.46	-3.14	0.002
		특성화고	1.41±3.78	1.39±3.89	0.15	0.879

〈표 1-23〉 학교급별 조사방법에 따른 사교육비의 차이 분석

(단위: 만원, mean±sd)

월평균 사교육비		종이	인터넷	t-value	p-value	
2011	1차조사	초등학교	27.61±18.95	27.14±18.57	1.18	0.237
		중학교	35.47±20.91	38.30±24.45	-5.59	<.0001
		일반고	41.45±28.95	43.49±30.85	-3.57	0.000
		특성화고	24.23±22.14	25.29±22.48	-0.73	0.463
	2차 조사	초등학교	26.88±18.73	27.55±18.70	-1.68	0.094
		중학교	35.20±21.30	38.53±24.49	-6.41	<.0001
		일반고	42.95±32.09	43.15±32.03	-0.31	0.755
		특성화고	23.04±20.20	27.23±24.60	-2.57	0.011
2012	1차조사	초등학교	25.62±16.90	26.65±17.01	-2.63	0.008
		중학교	36.64±22.17	39.83±24.18	-5.90	<.0001
		일반고	39.72±26.08	46.53±32.72	-11.18	<.0001
		특성화고	23.66±21.24	25.48±22.04	-1.18	0.239
	2차 조사	초등학교	25.26±15.90	26.34±17.05	-2.71	0.007
		중학교	36.43±22.01	39.48±23.57	-5.28	<.0001
		일반고	41.09±29.87	45.58±33.88	-5.94	<.0001
		특성화고	23.97±20.87	24.34±21.03	-0.21	0.833



〈표 1-24〉 학교급별 조사방법에 따른 시간당 사교육비의 차이 분석

(단위: 만원, mean±sd)

월평균 시간당 사교육비		중이	인터넷	t-value	p-value	
2011	1차조사	초등학교	0.89±0.65	0.93±0.79	-2.1	0.036
		중학교	1.09±0.82	1.26±1.16	-7.47	<.0001
		일반고	1.72±1.27	1.81±1.47	-3.43	0.001
		특성화고	1.17±0.96	1.20±1.27	-0.42	0.672
	2차 조사	초등학교	0.87±0.62	0.90±0.68	-1.88	0.061
		중학교	1.09±0.84	1.27±1.18	-7.39	<.0001
		일반고	1.80±1.43	1.82±1.40	-0.89	0.372
		특성화고	1.12±1.03	1.17±1.11	-0.68	0.499
2012	1차조사	초등학교	0.95±0.84	1.11±1.06	-7.15	<.0001
		중학교	1.28±1.06	1.44±1.45	-5.69	<.0001
		일반고	1.82±1.38	1.97±1.68	-4.69	<.0001
		특성화고	1.08±1.05	1.29±1.30	-2.63	0.009
	2차 조사	초등학교	0.93±0.74	0.99±0.83	-3.41	0.001
		중학교	1.22±0.93	1.37±1.22	-5.68	<.0001
		일반고	1.77±1.28	1.90±1.51	-3.98	<.0001
		특성화고	1.08±0.94	1.28±1.51	-2.15	0.032

제4절 혼합조사 추정 방안

두 가지 이상의 방법을 하나의 조사에 이용하는 혼합조사에 대한 추정 관련 연구는 크게 최적결합을 이용한 방법과 성향점수를 이용한 방법으로 볼 수 있다. 최적결합을 이용한 추정방법은 서로 다른 방법으로 얻어진 조사 결과들을 각각 추정한 후 가중값을 이용하여 서로 결합하는 방법이며, 성향점수를 이용한 방법은 누군가가 특정 조사에 참여할 것인지 아닌지에 관한 성향을 조정변수로 모형화하여 얻게 된다.

또한 통계개발원에서는 2012년 정책연구용역으로 “혼합조사에서의 추정 방법 개발 연구”를 진행한 바 있다. 해당 연구에서는 추정모형과 구조모형을 이용한 예측치를 활용하여 혼합조사에서의 추정값을 산출하는 방안을 제시하였다.

본 절에서는 위 3가지 방법을 이용하여 사교육비조사에서의 주요 관심변수에 대한 추정값을 도출하고 이를 기존 추정 결과와 비교함으로써, “사교육비조사”에 대한 혼합조사 추정 방법의 적용 가능성에 대하여 살펴보고자 한다. 이 때 비교 대상이 되는 추정 결과는 표본에 의한 추정값으로 사후가중 적용 결과(공표 결과)와는 차이가 있다.

1. 최적결합방법을 이용한 추정 방안

가. 최적결합방법

동일한 특성을 조사하기 위하여 적용되는 조사방법에 따라 산출되는 추정 결과에는 차이가 발생된다. 이 차이는 적용되는 조사방법별 고유 특성에 따라 자료로부터 확보되는 정보에 차이가 발생하기 때문이다. 따라서 단일 조사방법을 이용하여 응답자의 특정 성질을 조사하기보다는 다양한 조사방법을 적용하고 각 조사방법으로부터 얻어지는 다양한 정보를 함께 이용한다면 모집단과 보다 유사한 형태의 결과를 얻는 일이 가능할 것이다.

서로 다른 방법에 의하여 수집된 자료를 통합하는 방법 중 하나는 최적결합방법(optimal weighting method)이다. Bates and Granger(1969)는 동일한 현상에 두 개 또는 그 이상의 예측(추정)값이 존재하는 경우, 개별 추정값의 분산이 불편성과 안정성을 만족할 때 개별 추정값들을 결합한 결합 추정치의 평균제곱오차가 개별 추정치의 평균제곱오차에 비해 작아짐을 보였다. 또한 North and Nakamoto(1989)와 North et al.(1991)는 서로 다른 강수장에서 얻어진 강우량 자료를 통합하여 최적의 강우량을 추정하는데 최적결합방법을 이용하였다.

사교육비조사는 서로 다른 두 가지 방법으로 이루어졌으므로, 사교육비 관련 주요 특성에 대한 두 가지 추정값이 존재한다. 두 조사(종이조사와 인터넷조사)에 의한 추정값의 분산이 불편성을 만족하고 서로 상관관계가 없다고 가정하면, 사교육비조사의 주요 특성에 대한 결합 추정량과 분산은 다음과 같다.

$$f = wf_1 + (1-w)f_2 \quad (1.1)$$

$$\text{var}(f) = w^2 \text{var}(f_1) + (1-w)^2 \text{var}(f_2) \quad (1.2)$$

$$= w^2 \sigma_1^2 + (1-w)^2 \sigma_2^2$$

여기서 f_1 (f_2)은 종이(인터넷)조사의 추정량이며, σ_1^2 (σ_2^2)은 종이(인터넷)조사의 분산 추정량이다.

이 때 결합 추정값(f)의 오차 분산을 최소화하는 최적 가중값(optimal weight)은 다음과 같다.

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (1.3)$$



만약 가중값 $w = \frac{1}{2}$ (등분산)이면, $var(f) = \frac{var(f_1) + var(f_2)}{4}$ 이 된다.

조사방법의 종류가 M 개인 경우 결합 추정값은 (4)와 같고, 추정값들이 서로 상관관계가 없다고 가정하면 최적 가중값은 (5)와 같다.

$$f = w_1 f_1 + w_2 f_2 + \cdots + w_M f_M \quad (1.4)$$

$$w_m = \frac{\sigma_m^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_M^{-2}} \quad (1.5)$$

여기서 $w_1 + w_2 + \cdots + w_M = 1$ 이다.

나. 최적결합방법의 적용

최적결합방법을 적용하여 2011년 및 2012년 사교육비조사의 주요 관심 지표를 추정한다. 2012년 사교육비조사의 경우 인터넷조사를 먼저 권유한 이후에 인터넷조사에 참여하지 않은 학생 및 학부모를 대상으로 종이조사를 진행하고 있으나, 조사 시점이 연속된 시간 상에 존재하고 조사의 기준 시점이 동일하므로, 조사 시점에 따른 차이는 크지 않을 것으로 예상된다. 따라서 사교육비조사에서의 두 개의 자료는 조사방법에 따른 차이²⁾만이 존재하는 것으로 볼 수 있다.

다음은 최적 가중치를 적용한 추정 결과이다.

2) 두 조사 모두 자기기입식으로 진행되어 조사원에 의한 효과는 포함되어 있지 않으므로, 조사방법에 차이만 존재하는 것으로 볼 수 있음

〈표 1-25〉 주요 특성에 대한 평균 추정 및 비교 결과(최적결합)

			종이	인터넷	결합 추정값(a)	기준 추정값(b)	차이 (a-b)
2011	1차	사교육비참여율	31.940	37.780	34.746	34.670	0.076
		사교육시간	5.959	5.440	5.704	5.717	-0.013
		사교육비	34.875	36.600	35.666	35.642	0.024
		시간당사교육비	1.264	1.374	1.309	1.313	-0.004
	2차	사교육비참여율	37.910	39.730	38.812	38.770	0.042
		사교육시간	5.433	5.265	5.348	5.353	-0.005
		사교육비	34.789	36.436	35.580	35.557	0.023
		시간당사교육비	1.267	1.351	1.305	1.306	-0.001
2012	1차	사교육비참여율	36.690	35.540	36.111	35.900	0.210
		사교육시간	5.159	5.405	5.274	5.327	-0.052
		사교육비	33.721	38.164	35.546	36.766	-1.220
		시간당사교육비	1.354	1.545	1.427	1.485	-0.058
	2차	사교육비참여율	40.550	39.900	40.224	40.080	0.140
		사교육시간	4.891	4.991	4.940	4.964	-0.024
		사교육비	33.817	37.084	35.244	36.225	-0.981
		시간당사교육비	1.299	1.442	1.356	1.404	-0.048

주) 사교육비참여율은 사교육에 참여하지 않는 비율(%)이며, 사교육시간은 주당 사교육시간(시간), 사교육비는 월평균 사교육비(만원), 시간당 사교육비는 월평균 시간당 사교육비(만원)임

분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기준 추정값과 최적결합방법에 의한 추정값 사이에 약간의 차이가 나타난 것으로 확인되었다. 이는 인터넷조사의 분산이 종이조사에 비해 상대적으로 커, 종이조사의 가중값이 다소 늘어난데 기인한 것이다. 최적결합방법을 이용하는 경우 평균의 차이에는 큰 변화가 없으나, 분산은 크게 줄어든 것으로 확인되었다(<표 1-49>).

다음은 학교급별 최적결합 가중치 적용 결과이다.



〈표 1-26〉 학교급별 사교육비참여율 추정 및 비교 결과(최적결합)

(단위: %, %p)

			종이	인터넷	결합 추정값(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	15.63	18.1	16.79	16.68	0.11
		중학교	25.31	30.28	27.66	27.64	0.02
		일반고	37.69	44.7	41.11	41.08	0.03
		특성화고	76.09	73.65	74.91	74.88	0.03
	2차	초등학교	17.58	16.29	16.92	17.00	-0.08
		중학교	30.00	32.2	31.08	31.06	0.02
		일반고	46.73	49.36	48.04	48.02	0.02
		특성화고	79.95	78.17	79.09	79.13	-0.04
2012	1차	초등학교	21.46	17.42	19.28	18.81	0.47
		중학교	28.05	27.56	27.80	27.73	0.07
		일반고	42.39	41.35	41.87	41.65	0.22
		특성화고	77.53	77.93	77.73	77.79	-0.06
	2차	초등학교	22.17	17.7	19.75	18.98	0.77
		중학교	32.58	32.08	32.33	32.22	0.11
		일반고	49.55	48.49	49.02	48.75	0.27
		특성화고	80.45	80.57	80.51	80.54	-0.03

학교급별 사교육에 참여하지 않는 비율에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 최적결합방법에 의한 추정값 사이에 차이는 거의 없고, 분산은 크게 줄어든 것으로 확인되었다(부록 참조).

〈표 1-27〉 학교급별 주당 사교육시간 평균 추정 및 비교 결과(최적결합)

(단위: 시간)

			종이	인터넷	결합 추정값(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	7.676	7.503	7.596	7.602	-0.006
		중학교	7.706	6.954	7.330	7.354	-0.024
		일반고	4.644	4.212	4.429	4.435	-0.007
		특성화고	1.604	2.061	1.788	1.829	-0.042
	2차	초등학교	7.459	7.639	7.549	7.539	0.010
		중학교	7.130	6.740	6.932	6.943	-0.011
		일반고	3.976	3.769	3.869	3.875	-0.006
		특성화고	1.331	1.901	1.511	1.593	-0.082

			종이	인터넷	결합 추정값(a)	기존 추정값(b)	차이 (a-b)
2012	1차	초등학교	6.621	6.870	6.736	6.785	-0.048
		중학교	6.701	6.800	6.749	6.768	-0.018
		일반고	3.978	4.564	4.230	4.394	-0.165
		특성화고	1.726	1.628	1.674	1.662	0.011
	2차	초등학교	6.498	6.824	6.662	6.731	-0.069
		중학교	6.264	6.343	6.303	6.321	-0.018
		일반고	3.623	3.910	3.759	3.839	-0.081
		특성화고	1.410	1.389	1.400	1.394	0.005

학교급별 사교육시간에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 최적결합방법에 의한 추정값 사이에 차이는 거의 없고, 분산은 크게 줄어든 것으로 확인되었다(<부록> 참조).

<표 1-28> 학교급별 월평균 사교육비 평균 추정 및 비교 결과(최적결합)

(단위: 만원)

			종이	인터넷	결합 추정값(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	27.615	27.140	27.373	27.416	-0.043
		중학교	35.472	38.298	36.665	36.746	-0.080
		일반고	41.454	43.489	42.407	42.377	0.030
		특성화고	24.235	25.294	24.756	24.783	-0.027
	2차	초등학교	26.884	27.550	27.218	27.184	0.034
		중학교	35.200	38.530	36.634	36.771	-0.137
		일반고	42.946	43.149	43.048	43.043	0.005
		특성화고	23.043	27.225	24.728	25.055	-0.328
2012	1차	초등학교	25.617	26.653	26.131	26.309	-0.177
		중학교	36.635	39.835	38.096	38.784	-0.688
		일반고	39.719	46.533	42.366	44.583	-2.217
		특성화고	23.662	25.482	24.538	24.838	-0.299
	2차	초등학교	25.265	26.344	25.767	26.048	-0.282
		중학교	36.426	39.477	37.846	38.633	-0.786
		일반고	41.090	45.578	43.053	44.495	-1.442
		특성화고	23.966	24.338	24.150	24.239	-0.088



학교급별 평균사교육비에 대한 분석 결과, 조사방법을 선택했던 2012년도 조사 결과에서의 추정값 간 차이가 선택하지 못했던 2011년도에 비해 다소 발생하였으나, 그 크기가 크지는 않은 것으로 확인되었다. 반면 두 해 모두 분산은 크게 줄어들었다(<부록> 참조).

〈표 1-29〉 학교급별 월평균 시간당 사교육비 평균 추정 및 비교 결과(최적결합)

(단위: 만원)

			종이	인터넷	결합 추정값(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	0.894	0.927	0.908	0.908	-0.001
		중학교	1.087	1.256	1.144	1.163	-0.019
		일반고	1.720	1.810	1.759	1.761	-0.003
		특성화고	1.173	1.204	1.184	1.189	-0.005
	2차	초등학교	0.872	0.898	0.884	0.883	0.000
		중학교	1.094	1.267	1.153	1.176	-0.023
		일반고	1.797	1.822	1.810	1.809	0.001
		특성화고	1.116	1.168	1.140	1.141	-0.001
2012	1차	초등학교	0.954	1.106	1.012	1.056	-0.043
		중학교	1.280	1.443	1.336	1.389	-0.053
		일반고	1.818	1.968	1.878	1.925	-0.047
		특성화고	1.081	1.294	1.165	1.219	-0.053
	2차	초등학교	0.927	0.991	0.956	0.974	-0.018
		중학교	1.225	1.374	1.280	1.332	-0.053
		일반고	1.773	1.903	1.828	1.872	-0.044
		특성화고	1.076	1.279	1.133	1.225	-0.092

학교급별 시간당 사교육비에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 최적결합방법에 의한 추정값 사이에 차이는 거의 없고, 분산은 크게 줄어든 것으로 확인되었다(<부록> 참조).

2. 성향점수를 이용한 추정 방안

성향점수를 이용한 보정 기법은 관측연구 내 비교 집단 간에 존재하는 개체 특성의 차이를 조정하여 각 집단별 개체의 선택 절차(selection mechanism)에 따른 영향을 완화하는

방안으로 도입된 기법이다(Rosenbaum and Rubin, 1983, 1984; D'Agostino, 1998). 표본 조사에서 무응답과 비포함 편향이 랜덤화 가정의 위반에 기인하는 경우, 관측연구에서의 선택 편향과 유사하게 볼 수 있다(김서영 외, 2010).

자원자 표본이나 조사방법을 선택하는 조사의 경우에는 지역, 성별, 연령대 등의 인구통계학적 요소를 이용한 사후층화방법을 이용한 보정으로는 자발적 조사 참여 또는 조사방법의 선택 등에서 오는 편향을 보정하는데 한계가 있다. 이에 많은 연구자들은 성향점수를 이용한 보정이 이러한 선택 편향을 줄이는데 효과적이라는 것을 보인 바 있다(Taylor et al., 2001; Schillewaert and Meulemeester, 2005; Lee, 2006; Harris Interactive, 2008; 김영원과 배예영, 2010; 김서영 외, 2010). 그러나 성향점수를 통한 편향 축소가 모든 변수에 대한 보편적 현상이 아니며, 오히려 편향이 확대되는 경우도 있다. 또한 추정치의 변동이 증가하게 되므로 통계적 효율성이 저하되는 문제가 발생하기도 한다(Lee and Calliant, 2009). 따라서 성향점수의 적용에 의한 편향의 축소와 변동의 증가 등을 종합적으로 고려하여 조사의 총오차(total survey error) 측면에서 적용 여부를 결정하여야 할 것이다.

가. 성향점수를 이용한 보정

2012년 종이와 인터넷으로 각각 진행된 사교육비조사의 경우에는 먼저 인터넷조사를 권한 이후에 인터넷조사에 참여하지 않은 조사대상에 한하여 종이조사를 진행하게 된다. 따라서 인터넷조사를 선택한 응답자들의 특성이 인터넷조사 결과에 포함되어 있을 가능성이 있으며, 이러한 특성을 성향점수를 이용하여 보정하고자 한다.

성향점수를 이용한 보정은 참조조사(reference survey) 자료가 존재한다는 가정이 필요하다. 참조조사 자료는 관찰조사에서의 대조군 역할을 하게 되며, 선택 편향이 포함된 조사 자료는 실험군 역할을 하게 된다. 참조조사는 비교 대상이 되는 조사와 비슷한 시기에 실행되어야 하며, 신뢰성이 인정된 전통적인 조사방법에 의하여 수행되고, 높은 응답률을 갖는 양질의 조사이어야 한다(김영원 외, 2010). 성향점수를 이용한 보정은 성향점수모형을 이용하여 주요 공변량을 대상으로 참조조사의 성향점수 분포를 벤치마킹하는 것이다.

성향점수는 조사 대상이 특정 조사에 참여할 것인지 아닌지를 나타내는 변수를 모형화하여 얻게 된다. 보통 태도변수(attitudinal variable)이며 이항변수를 종속변수로 갖는 로지스틱 회귀모형이 사용되며, 조사 대상이 특정 조사에 참여할 확률(성향점수)을 추정하여 이용한다. 사교육비조사의 경우, 전통적 조사방법인 종이조사가 참조조사(대조군)가 되며, 조사 대상의 선택에 의한 인터넷조사는 선택 편향이 포함된 조사(실험군)가 된다.



사교육비조사에 대한 성향점수를 이용한 보정 과정은 다음과 같다. 우선 종이조사와 인터넷조사로 각각 얻어진 사교육비조사의 자료를 조사방법에 관계없이 통합한다. 즉, 사교육비조사의 통합표본은 $s = s^P \cup s^I$ 이고, 표본의 크기는 $n = n^P + n^I$ 가 된다. 이 때 s^P 는 종이조사 표본, s^I 는 인터넷조사 표본이며, n^P 는 종이조사의 표본 크기, n^I 는 인터넷조사의 표본 크기를 나타낸다. 이렇게 통합된 전체 자료 즉, 통합표본에 다음과 같은 로지스틱 모형을 적용하면 성향점수를 추정할 수 있다.

$$\ln \left[\frac{e(x)}{1-e(x)} \right] = \alpha + \beta^T f(x) \quad (1.6)$$

여기서 x 는 공변량 벡터, $f(x)$ 는 공변량 벡터의 함수를 나타낸다.

이 때 i 번째 응답자의 인터넷조사 참여에 대한 성향점수는 $e(x_i) = P(i \in s^I | x_i)$ 이며, 주어진 조건(공변량 x_i)에서 i 번째 응답자가 인터넷조사 참여할 확률을 의미한다.

위와 같은 방법을 통하여 얻은 성향점수를 이용한 보정방법은 매칭, 층화, 공분산 조정 및 가중치방법 등을 들 수 있다(Steinmetz and Tijdens, 2010). 본 연구에서는 이 중 성향점수 보정방법으로 가장 널리 이용되고 있는 층화방법을 이용하도록 한다.

우선 식 (1.6)을 이용하여 모든 응답자들을 대상으로 성향점수를 계산한 후, 추정된 성향점수를 크기 순으로 정렬한 후에 C개의 층으로 분할³⁾한다. 구분된 각 층 내에 응답자들의 성향점수에는 큰 차이가 없도록 구성한다. 구성된 C개의 층을 $s_c = s_c^P \cup s_c^I$ 로 표기하면, c번째 층은 $n_c = n_c^P + n_c^I$ 개의 응답자들로 구성된다. 여기서 s_c^P (s_c^I)는 c번째 층에서 종이(인터넷)조사 표본과 연계된 표본을 의미한다.

다음으로 각 층에서 다음과 같은 조정인자(adjustment factor)를 계산한다.

$$f_c = \frac{\sum_{k \in s_c^P} d_k^P / \sum_{k \in s^P} d_k^P}{\sum_{j \in s_c^I} d_j^I / \sum_{j \in s^I} d_j^I} \quad (1.7)$$

식 (1.7)에서 기본가중치가 선택확률의 역수라면 다음과 같이 표현될 수 있다.

$$f_c = \frac{\hat{N}_c^P / \hat{N}^P}{\hat{N}_c^I / \hat{N}^I} \quad (1.8)$$

c번째 층 내에서 인터넷 표본의 j번째 응답자에 대한 성향점수보정(PSA)인자는

3) Cochran(1968)은 5분위수를 기준으로 하여 다섯 개의 층(c=5)으로 구분하는 것을 권장

다음과 같다.

$$d_j^{I.PSA} = f_c d_j^I = \frac{\hat{N}_c^P / \hat{N}^P}{\hat{N}_c^I / \hat{N}^I} d_j^I \quad (1.9)$$

만약 단순확률추출과 같이 모든 응답자들이 갖는 기본가중치가 같거나, 이를 알 수 없는 경우에는 다음 조정인자를 사용할 수 있다.

$$f_c = \frac{n_c^P / n^P}{n_c^I / n^I} \quad (1.10)$$

식 (1.9)에 의한 성향점수보정을 진행하면 성향점수로 구성된 층에 대한 인터넷조사 표본의 분포와 참조조사인 종이조사 표본의 분포가 같아진다.

이 때 인터넷조사에서 c번째 층에 해당하는 모집단 크기에 대한 추정값은 식 (1.11)과 같으며, s^I 로부터 얻어지는 관심 변수(y)에 대한 평균은 식 (1.12)와 같다.

$$\hat{N}_c^{I.PSA} = \sum_{j \in s_c^I} d_j^{I.PSA} = \hat{N}^I \frac{\hat{N}_c^P}{\hat{N}^P} \quad (1.11)$$

$$\frac{\hat{y}^{I.PSA}}{y} = \frac{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA} y_j}{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA}} \quad (1.12)$$

여기서 추정값 $\hat{y}^{I.PSA}$ 를 구할 때, 참조조사인 종이조사 표본은 사용되지 않는다.

나. 성향점수 보정방법의 적용

성향점수를 이용한 보정방법을 적용하여 2011년 및 2012년 사교육비조사의 주요 관심 지표를 추정한다. 성향점수를 이용한 보정방법은 성향변수의 선택이 매우 중요하며, 성향변수에 무응답이 포함된 경우 성향점수를 계산할 수 없으므로, 선택된 성향변수에는 무응답이 포함되지 않아야 한다. 따라서 성향점수를 이용한 보정을 위해서는 조사방법의 선택에 유의미한 영향을 미칠 것으로 보이는 성향변수를 선택하고, 선택된 성향변수에 무응답이 포함된 경우, 다양한 대체 방법을 활용하여 항목 무응답을 제거하거나, 분석에서 제외하는 등의 조치를 취해야 한다. 그러나 항목 무응답을 대체하면 추정 과정이 한 단계 늘어나고, 분석에서 제외하면 제외된 케이스의 특성이 보정 과정에 반영되지 않는다는 단점이 있으므로, 방법을 적용하는데 있어 신중을 기해야 할 것이다.



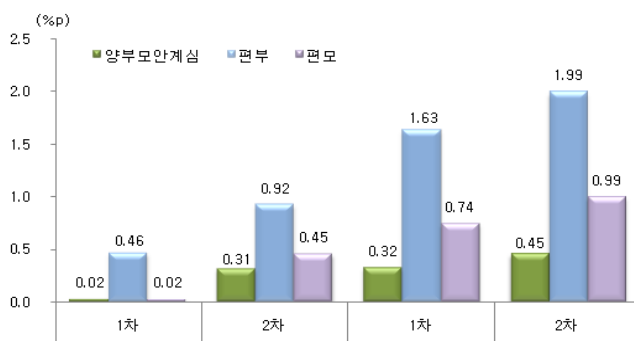
사교육비조사의 경우 활용 가능한 성향변수가 한정되어 있으며, 활용해야 하는 성향 변수 내에 다수의 무응답이 포함되어 있다. 예를 들어 부모의 학력과 경제활동상황은 사교육과 밀접한 관계가 있을뿐더러 조사방법의 선택에도 유의미한 영향을 미칠 것으로 보이나, 양부모가 모두 없는 경우와 편부모인 경우의 학생들은 부모들의 경제활동 및 학력 등에 응답할 수가 없다. 즉, 발생한 항목 무응답은 불가피하게 발생한 것이며, 양부모가 없거나 편부모인 경우 또한 사교육에 영향을 미칠 수 있는 요인 중 하나이므로 이를 대체하여 이용하는 것에는 무리가 있다.

다음은 2011년과 2012년 사교육비조사에 참여한 학생 중 양부모가 모두 안계시거나 편부모인 경우의 비율이다.

<표 1-30> 조사방법별 양부모 또는 한 부모가 없는 비율

(단위: %)

	부모	2011			2012		
		전체	종이조사	인터넷조사	전체	종이조사	인터넷조사
1차	양부모안계심	0.50	0.51	0.49	0.48	0.70	0.38
	편부	3.31	3.52	3.06	4.06	5.18	3.55
	편모	4.22	4.23	4.21	5.18	5.68	4.94
2차	양부모안계심	0.52	0.67	0.36	0.50	0.83	0.38
	편부	3.44	3.88	2.96	4.12	5.58	3.59
	편모	4.43	4.64	4.19	5.37	6.10	5.11



[그림 1-24] 조사방법별 양부모 또는 한 부모가 없는 비율 차이(종이-인터넷)

<표 1-30>과 같이 전체 조사 대상 학생 중 양부모가 없는 학생 비율은 약 0.5%에 달하며, 편부 또는 편모인 학생 비율은 각각 3~5% 가량 되는 것으로 파악되었다. 또한 [그림 1-24]에 의하면 조사방법을 선택할 수 없었던 2011년에는 조사방법에 따른 비율의

차이가 크지 않았으나, 조사방법의 선택이 가능했던 2012년도에는 조사방법 간 차이가 다소 커진 것으로 확인되었다. 따라서 부모의 존재 여부는 조사방법의 선택뿐 아니라 조사 결과에도 상당한 영향을 미쳤을 것으로 볼 수 있다.

이에 필요한 성향변수를 최대한 활용할 수 있도록 변수 간의 관계를 이용하여 성향변수를 재구성하였다. 먼저 활용 가능한 성향변수와 재구성된 변수의 특성은 <표 1-31>과 같다.

<표 1-31> 성향변수와 범주

성향변수	범주	성향변수	범주
지역	5	지역	5
학교급	4	학교급	4
성적등급	5	성적등급	5
아버지연령	6	부모연령 (평균연령)	4
어머니연령	6		(-)
아버지 학력	20(24)	부모학력	4
어머니 학력	20(24)		
경제활동	4	경제활동	4
소득	8	소득	8

주) 2011년도 부모의 학력 중 대학원으로 조사되었던 항목이 2012년도에는 석사와 박사과정으로 구분되어 조사됨(범주의 수 변경, 20 → 24)

부모의 연령은 우선 양부모가 모두 있는 경우에 두 부모의 평균 연령을 계산하고, 편부모인 경우에는 해당 부모의 연령을 이용하여 평균 연령을 구한 후에, 양부모가 모두 없는 경우에 “0”, 부모의 평균 연령이 2~30대인 경우 “1”, 40대인 경우 “2”, 50대 이상인 경우 “3”으로 분류하였다. 부모의 평균 연령에는 양부모가 계시지 않은 경우가 결측치로 포함되어 있다. 부모의 학력은 부모 모두 고졸 이하인 경우 “1”, 부모 중 1명이 대재 이상인 경우 “2”, 부모 모두 대재 이상인 경우 “3”으로 분류하였으며, 편부모인 경우에는 함께 살고 있는 부모의 학력이 고졸 이하인 경우 “1”, 대재 이상인 경우는 “3”으로 각각 분류하였다. 양부모가 모두 없어 부모의 학력을 알 수 없는 경우에는 “0”으로 별도 분류하였다. 부모의 경제활동에서는 부모가 모두 없는 경우의 항목 무응답을 “모두 참여하지 않음”으로 분류하였다.

다음은 로지스틱 회귀모형에 대한 추정 결과이다. 여기서 종속변수는 조사방법(종이조사=0, 인터넷조사=1)이며, 범주형 성향변수 중 세부 범주 내 하나라도 유의미한 영향을 미칠 경우, 모형에 포함하였다. 양부모가 없는 경우, 부모의 연령과 학력에서 중복(모두 “0”)이 발생하게 되어 하나 이상의 변수에 대한 자유도가 감소하는 문제가 발생된다. 이와 같은 문제를 방지하게 위하여 부모의 연령은 평균연령을 이용하도록 한다.



〈표 1-32〉 로지스틱 회귀모형 추정 결과

성향변수		2011년				2012년			
		1차		2차		1차		2차	
		Odds	Pr(> z)	Odds	Pr(> z)	Odds	Pr(> z)	Odds	Pr(> z)
지역 (ref.강남)	강북	1.073	.221	1.068	.254	.298	.000	.329	.000
	광역시	1.127	.017	1.163	.002	.296	.000	.386	.000
	중소도시	.916	.075	.999	.989	.242	.000	.335	.000
	읍면	1.295	.000	1.386	.000	.196	.000	.257	.000
학교급 (ref.초등학교)	중학교	1.239	.000	1.200	.000	1.147	.000	1.110	.001
	일반고	1.376	.000	1.288	.000	1.433	.000	1.345	.000
	특성화고	1.354	.000	1.097	.027	1.160	.001	1.282	.000
부모학력 (ref.부모없음)	고졸이하(모두)	1.941	.000	3.197	.000	2.620	.000	3.666	.000
	대졸이상(한명)	2.184	.000	3.682	.000	3.105	.000	4.340	.000
	대졸이상(모두)	2.163	.000	3.816	.000	3.608	.000	4.984	.000
경제활동 (ref.아버지만)	어머니만	1.195	.000	1.110	.005	1.143	.001	1.131	.003
	모두참여O	.973	.189	.958	.038	.854	.000	.910	.000
	모두참여X	1.560	.000	1.541	.000	1.232	.004	1.525	.000
성직		.934	.000	.946	.000	.929	.000	.918	.000
부모연령		.915	.000	.942	.007	.903	.000	.927	.003
소득		.985	.008	.974	.000	.979	.001	.966	.000

주) 오즈비(Odds)는 인터넷조사를 선택할 확률이 선택하지 않을 확률에 비해 몇 배인가를 나타냄

먼저 로지스틱 회귀모형 추정 결과를 이용하여 i 번째 응답자가 인터넷조사에 참여할 확률($e(x_i) = P(i \in s^I | x_i)$)인 성향점수를 계산한 후, 계산된 성향점수를 크기 순서대로 나열한 후 4분위수와 5분위수를 이용하여 각각 4개($C = 4$)와 5개($C = 5$) 집단으로 분할한다. 이 후 분할된 각 집단별로 종이조사와 인터넷조사의 빈도를 구하고, 식 (8)을 이용하여 성향가중값을 계산하였다.

계산된 성향가중값은 다음과 같다.

〈표 1-33〉 집단별 성향가중값

		C=4				C=5				
		집단1	집단2	집단3	집단4	집단1	집단2	집단3	집단4	집단5
2011	1차	1.306	1.049	0.950	0.772	1.356	1.059	0.983	0.865	0.768
	2차	1.240	1.018	0.915	0.771	1.219	0.998	0.875	0.831	0.694
2012	1차	1.476	1.109	0.898	0.633	1.467	1.147	0.953	0.805	0.689
	2차	1.101	0.842	0.707	0.503	1.112	0.847	0.719	0.653	0.455

집단의 수를 4개와 5개로 각각 구성하여 분석한 결과, 분석 결과에는 별다른 차이가 없는 것으로 확인되었다. 다음은 집단을 4개로 한 성향점수보정에 의한 추정 결과이다. 집단의 수가 5인 경우의 추정 결과는 <부록>을 참고하면 된다.

〈표 1-34〉 주요 특성에 대한 평균 추정 및 비교 결과(C=4; 성향점수)

			종이	인터넷	성향점수 보정(a)	기준 추정값(b)	차이 (a-b)
2011	1차	사교육비참여율	31.940	37.780	33.702	34.670	-0.968
		사교육시간	5.959	5.440	5.842	5.717	0.125
		사교육비	34.875	36.600	35.248	35.642	-0.394
		시간당사교육비	1.264	1.374	1.289	1.313	-0.024
	2차	사교육비참여율	37.910	39.730	38.340	38.770	-0.430
		사교육시간	5.433	5.265	5.423	5.353	0.070
		사교육비	34.789	36.436	35.299	35.557	-0.258
		시간당사교육비	1.267	1.351	1.286	1.306	-0.020
2012	1차	사교육비참여율	36.690	35.540	37.338	35.900	1.438
		사교육시간	5.159	5.405	5.208	5.327	-0.119
		사교육비	33.721	38.164	34.771	36.766	-1.995
		시간당사교육비	1.354	1.545	1.424	1.485	-0.061
	2차	사교육비참여율	40.550	39.900	40.612	40.080	0.532
		사교육시간	4.891	4.991	4.917	4.964	-0.047
		사교육비	33.817	37.084	34.661	36.225	-1.564
		시간당사교육비	1.299	1.442	1.359	1.404	-0.046

주) 사교육비참여율은 사교육에 참여하지 않는 비율(%)이며, 사교육시간은 주당 사교육시간(시간), 사교육비는 월평균 사교육비(만원), 시간당 사교육비는 월평균 시간당 사교육비(만원)임



분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 성향점수 보정방법에 의한 추정값 사이에 소폭의 차이가 나타난 것으로 확인되었다. 조사방법을 선택할 수 없었던 2011년의 경우에는 그다지 큰 차이가 발견되지 않았으나, 조사방법을 선택한 2012년도에는 월평균 사교육비에 다소 차이가 발생한 것으로 나타났다.

다음은 학교급별 성향점수보정에 의한 추정 결과이다. 앞서 살펴본 것과 같이 집단 수에 따른 분석 결과에 별다른 차이가 없으므로, $C=5$ 인 경우의 결과는 생략한다.

〈표 1-35〉 학교급별 사교육 비참여율 추정 및 비교 결과(성향점수)

(단위: %, %p)

			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	15.63	18.10	16.787	16.68	0.107
		중학교	25.31	30.28	27.791	27.64	0.151
		일반고	37.69	44.70	41.065	41.08	-0.015
		특성화고	76.09	73.65	74.575	74.88	-0.305
	2차	초등학교	17.58	16.29	17.083	17.00	0.083
		중학교	30.00	32.20	31.294	31.06	0.234
		일반고	46.73	49.36	48.021	48.02	0.001
		특성화고	79.95	78.17	78.758	79.13	-0.372
2012	1차	초등학교	21.46	17.42	19.815	18.81	1.005
		중학교	28.05	27.56	29.500	27.73	1.770
		일반고	42.39	41.35	44.197	41.65	2.547
		특성화고	77.53	77.93	78.430	77.79	0.640
	2차	초등학교	22.17	17.70	19.820	18.98	0.840
		중학교	32.58	32.08	33.895	32.22	1.675
		일반고	49.55	48.49	50.738	48.75	1.988
		특성화고	80.45	80.57	81.123	80.54	0.583

학교급별 사교육 참여율에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 성향점수보정에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다. 이 때 조사방법을 선택할 수 없었던 2011년에 비해 조사방법을 선택했던 2012년도에 추정값 간 차이가 다소 큰 것으로 확인되었다.

〈표 1-36〉 학교급별 주당 사교육시간 평균 추정 및 비교 결과(성향점수)

(단위: 시간)

			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	7.676	7.503	7.600	7.602	-0.002
		중학교	7.706	6.954	7.352	7.354	-0.002
		일반고	4.644	4.212	4.450	4.435	0.015
		특성화고	1.604	2.061	1.871	1.829	0.042
	2차	초등학교	7.459	7.639	7.562	7.539	0.023
		중학교	7.130	6.740	6.919	6.943	-0.024
		일반고	3.976	3.769	3.895	3.875	0.020
		특성화고	1.331	1.901	1.637	1.593	0.044
2012	1차	초등학교	6.621	6.870	6.734	6.785	-0.051
		중학교	6.701	6.800	6.527	6.768	-0.241
		일반고	3.978	4.564	4.146	4.394	-0.248
		특성화고	1.726	1.628	1.604	1.662	-0.058
	2차	초등학교	6.498	6.824	6.669	6.731	-0.062
		중학교	6.264	6.343	6.101	6.321	-0.220
		일반고	3.623	3.910	3.650	3.839	-0.189
		특성화고	1.410	1.389	1.338	1.394	-0.056

학교급별 사교육시간에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 성향점수보정에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다.

〈표 1-37〉 학교급별 월평균 사교육비 평균 추정 및 비교 결과(성향점수)

(단위: 만원)

			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	27.615	27.140	27.617	27.416	0.201
		중학교	35.472	38.298	36.841	36.746	0.095
		일반고	41.454	43.489	42.305	42.377	-0.072
		특성화고	24.235	25.294	24.885	24.783	0.102
	2차	초등학교	26.884	27.550	27.476	27.184	0.292
		중학교	35.200	38.530	36.840	36.771	0.069
		일반고	42.946	43.149	43.151	43.043	0.108
		특성화고	23.043	27.225	25.322	25.055	0.267



			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2012	1차	초등학교	25.617	26.653	25.915	26.309	-0.394
		중학교	36.635	39.835	37.438	38.784	-1.346
		일반고	39.719	46.533	42.681	44.583	-1.902
		특성화고	23.662	25.482	24.456	24.838	-0.382
	2차	초등학교	25.265	26.344	25.751	26.048	-0.297
		중학교	36.426	39.477	37.644	38.633	-0.989
		일반고	41.090	45.578	42.857	44.495	-1.638
		특성화고	23.966	24.338	23.856	24.239	-0.383

학교급별 평균 사교육비에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 성향점수보정에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다. 이 때 조사방법을 선택할 수 없었던 2011년에 비해 조사방법을 선택했던 2012년도에 추정값 간 차이가 약간 더 큰 것으로 확인되었다.

〈표 1-38〉 학교급별 월평균 시간당 사교육비 평균 추정 및 비교 결과(성향점수)

(단위: 만원)

			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2011	1차	초등학교	0.894	0.927	0.914	0.908	0.006
		중학교	1.087	1.256	1.167	1.163	0.003
		일반고	1.720	1.810	1.758	1.761	-0.003
		특성화고	1.173	1.204	1.182	1.189	-0.007
	2차	초등학교	0.872	0.898	0.887	0.883	0.004
		중학교	1.094	1.267	1.180	1.176	0.004
		일반고	1.797	1.822	1.807	1.809	-0.002
		특성화고	1.116	1.168	1.147	1.141	0.007
2012	1차	초등학교	0.954	1.106	1.042	1.056	-0.014
		중학교	1.280	1.443	1.377	1.389	-0.012
		일반고	1.818	1.968	1.890	1.925	-0.035
		특성화고	1.081	1.294	1.218	1.219	0.000
	2차	초등학교	0.927	0.991	0.967	0.974	-0.007
		중학교	1.225	1.374	1.327	1.332	-0.005
		일반고	1.773	1.903	1.849	1.872	-0.023
		특성화고	1.076	1.279	1.219	1.225	-0.006

학교급별 시간당 사교육비에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 성향점수보정에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다.

성향점수보정을 적용하여 구한 추정값은 보정 이전과 전반적으로 큰 차이가 없는 것으로 확인되었다. 그러나 결합추정방법에 비하여 분산이 다소 큰 것으로 확인되었으며, 학교급별 주요 특성에 대해서는 성향점수를 적용하기 이전보다 분산이 커지는 경우도 있는 것으로 나타났다. 이는 한정된 성향변수를 이용해야 하는 한계에서 최적의 성향점수를 도출하지 못한데 기인한 것으로 볼 수 있으며, 조사방법을 선택하는데 영향을 미치는 주요 성향변수를 추가 수집하지 못할 경우, 성향점수를 이용한 보정은 기존의 추정 방법에 비해 큰 잇점이 있다고 보기 어려울 것으로 판단된다.

3. 측정오차모형을 이용한 예측기법 기반 추정 방안

2012년 통계개발원에서는 혼합조사에서의 추정방법 개발을 위한 연구용역(박상언 외 2012)을 진행하였다. 박상언 외(2012)는 참값에 대한 보조변수와의 구조적 관계를 나타내는 구조 모형과 구조 모형 외에 관측값과 참값의 관계를 나타내는 측정 오차에 대한 측정 모형을 설정하고, 설정된 모형의 모수를 추정한 후 예측치를 계산하여 자료를 보정하는 방법을 제안하였다.

연구 보고서에 의하면, 모형 설정을 위한 보조 변수로는 학생과 부모의 인구통계학적 변수들(지역, 교육수준, 성별, 학생 성적, 나이 및 소득)을 이용하였으며, 종속변수(사교육 시간 등)가 0의 값을 갖는 경우 모형의 설명력이 떨어지므로 다음과 같은 구조 모형(Tobit 모형)을 이용하였다. 사교육비조사의 경우 조사방법에 따른 실제 자료가 별도로 존재하므로, 각 조사방법에 따른 관측치들의 모형을 다음과 같이 고려한다.

$$z_{bi} = z_{ai} + u_i, \quad u_i \sim N(0, \sigma_u^2) \quad (1.13)$$

$$z_{ai} = \beta' x_i + e_i, \quad e_i \sim N(0, \sigma_e^2) \quad (1.14)$$

$$y_{bi} = z_{ai} I(z_{bi} > 0) \quad (1.15)$$

식 (1.13)은 측정 모형으로 두 조사방법별 잠재 변인의 차이에 대한 모형이며, 각 관측치들로부터 두 개의 잠재 변인 (z_a , z_b)를 발생 시킨 후 그로부터 모수를 추정하게 된다. 모수 추정은 EM 알고리즘을 이용하였으며, E-step에서의 몬테카를로 방법에서의 fractional weight는 $w_{ij}^* = 1/m$ 을 사용하게 된다.

만약 가상의 (z_{ai} , z_{bi}) 값이 실제 관측되었다면, 모수 추정은 다음과 같이 표현될 수 있다.



$$\sum_{i \in S} (z_{ai} - \hat{\beta}' x_i) x_i = 0 \quad (1.16)$$

$$\hat{\sigma}_e^2 = \frac{1}{n} \left\{ \sum_{i \in S} (z_{ai} - \hat{\beta}' x_i)^2 \right\} \quad (1.17)$$

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{i \in S} (z_{bi} - z_{ai})^2 \quad (1.18)$$

그러나 (z_{ai}, z_{bi}) 값이 관측되는 것이 아니라, 이들을 imputation 한 값을 이용하게 되므로 실제 모수 추정은 다음과 같이 계산된다.

$$\sum_{i \in S} \sum_{j=1}^m w_{ij}^* (z_{ai}^{*(j)} - \hat{\beta}' x_i) x_i = 0 \quad (1.19)$$

$$\hat{\sigma}_e^2 = \frac{1}{n} \left\{ \sum_{i \in S} \sum_{j=1}^m w_{ij}^* (z_{ai}^{*(j)} - \hat{\beta}' x_i)^2 \right\} \quad (1.20)$$

$$\hat{\sigma}_u^2 = \frac{1}{n} \left\{ \sum_{i \in S} \sum_{j=1}^m w_{ij}^* (z_{bi}^{*(j)} - z_{ai}^{*(j)})^2 \right\} \quad (1.21)$$

박상언 외는 연구 보고서에서 $m = 50$ 을 사용하였으며, 편의를 위하여 보조 변수에 결측값이 있는 경우는 자료에서 제외하고 분석을 진행하였다.

이와 같이 측정 오차 모형을 이용한 예측 기법을 바탕으로 제안된 추정량은 다음과 같다.

$$\hat{\theta} = \left\{ \sum_{i \in S} w_i \right\}^{-1} \left\{ \sum_{i \in S_A} w_i y_{ai} + \sum_{i \in S_B} w_i \sum_{j=1}^m w_{ij}^* y_{ai}^{*(j)} \right\} \quad (1.22)$$

측정 오차 모형을 이용한 예측 기법 적용 결과를 실무 적용하기 위하여, 우선 보조 변수에 결측값이 있는 자료를 제외하지 않는 방안을 고려하였다. 보조 변수에 결측값이 있는 경우는 양부모가 없는 경우와 편부모인 경우의 학생으로 전체 조사 대상의 약 5.5%에 해당(양부모가 없는 경우 약 0.5% 내외, 편부모인 경우 약 5% 내외)한다. 이들을 분석 대상에서 제외할 경우 추정 결과에 영향을 미칠 가능성이 있으므로, 성향 점수를 이용한 보정 방안에서와 같이 부모 관련 보조 변수에 대한 재분류를 통하여

조사 대상이 분석에서 제외되지 않도록 하였다. 구체적인 재분류 결과는 성향점수를 이용한 보정 방법을 참고하면 된다.

지역, 학교급, 성적, 부모 연령 및 학력, 경제활동 형태 및 소득 등을 보조 변수로 한 사교육시간에 대한 모형 추정 결과는 다음과 같다. 이 때 참조조사는 종이조사이다.

<표 1-39> 사교육시간에 대한 모형 선택 결과(참조=종이)

보조변수		2011년				2012년			
		1차		2차		1차		2차	
		$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$
지역 (ref.강남)	강북	-1.395	-1.218	-1.599	-1.664	-2.436	-1.845	-2.446	-2.049
	광역시	-2.102	-2.198	-2.402	-2.429	-2.708	-2.362	-2.827	-2.807
	중소도시	-2.211	-2.324	-2.616	-2.705	-3.135	-2.635	-3.337	-3.243
	읍면	-3.788	-4.241	-4.507	-4.473	-5.180	-4.600	-5.317	-4.945
학교급 (ref.초등학교)	중학교	0.309	0.035	-0.192	-0.405	-0.178	-0.218	-0.664	-0.885
	일반고	-3.885	-4.103	-4.916	-5.140	-3.913	-3.639	-4.654	-4.824
	특성화고	-8.965	-8.747	-9.796	-9.730	-8.466	-8.705	-9.418	-9.944
부모학력 (ref.부모없음)	고졸이하(모두)	5.346	4.447	3.935	2.876	-0.431	-0.160	1.428	1.203
	대졸이상(한명)	6.355	5.497	5.033	3.987	0.560	0.990	2.085	2.295
	대졸이상(모두)	6.639	5.825	5.549	4.523	0.745	1.169	2.915	2.708
경제활동 (ref.아버지만)	어머니만	-0.868	-1.194	-0.933	-1.225	-0.638	-0.981	-0.303	-0.702
	모두참여O	0.273	0.248	0.109	-0.027	0.300	0.150	0.570	0.169
	모두참여X	-3.256	-3.738	-3.184	-3.971	-3.816	-4.040	-2.741	-3.462
성적		-0.818	-0.858	-1.008	-0.947	-0.697	-0.701	-0.684	-0.693
부모연령		-0.416	-0.431	-0.516	-0.518	-0.437	-0.439	-0.811	-0.602
소득		0.907	0.905	0.890	0.918	0.894	0.897	0.989	1.004
$\hat{\sigma}_e$		7.338	7.264	7.847	7.712	7.376	6.995	7.704	7.417
$\hat{\sigma}_u$		3.531	4.006	2.013	2.934	2.774	3.773	1.452	2.563

주) $\hat{\beta}_{(0)}$ 은 초기 모수 추정값이며, $\hat{\beta}_{EM}$ 은 EM-algorithm을 통하여 얻은 최종 모수 추정값임

부모 연령, 성적, 소득에 대한 모수 추정 결과($\hat{\beta}$)는 다른 모든 보조 변수들이 고정되어 있을 때 부모의 연령이 한 단위 증가할 때 부모 연령의 $\hat{\beta}$ 만큼 사교육시간이 증가한다는 것을 의미한다. 성적 및 소득의 모수 추정 결과 역시 동일한 의미로 해석할 수 있다. 예를 들어, 2012년 1차 조사에서 가구의 소득이 한 단위 증가할 때 사교육시간은 0.897만큼 증가한다는 것을 의미한다.



지역, 학교급, 부모 학력 및 경제활동 등 범주형 보조 변수의 경우에는 다른 보조 변수들이 고정되어 있을 때 기준이 되는 수준(지역-강남, 학교급-초등학교, 부모 학력-양부모 없음, 경제활동-아버지만 참여)이 아닌 다른 수준을 값으로 갖는 경우, 해당 보조 변수의 $\hat{\beta}$ 만큼 사교육시간이 증가한다는 것을 의미한다. 예를 들어, 2012년 2차 조사에서 강북 소재 학교에 다니는 학생은 다른 보조 변수들이 고정되어 있을 때 강남 소재 학교에 다니는 학생에 비하여 사교육시간이 2.049만큼 감소한다고 해석할 수 있다.

다음은 종이조사를 참조조사로 하는 사교육비에 대한 모형 선택 결과이다. 모수에 대한 해석은 사교육시간에서와 같은 방법으로 진행하면 된다.

<표 1-40> 사교육비에 대한 모형 선택 결과(참조=종이)

보조변수		2011년				2012년			
		1차		2차		1차		2차	
		$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$
지역 (ref.강남)	강북	-0.747	-1.130	-0.089	-0.632	-0.395	-0.571	-2.093	-0.594
	광역시	-3.367	-3.357	-4.099	-3.706	-0.944	-1.777	-3.770	-2.220
	중소도시	-3.135	-3.020	-4.381	-3.812	-0.814	-1.558	-3.047	-1.701
	읍면	-3.536	-3.107	-4.612	-3.702	-2.178	-1.915	-4.816	-2.439
학교급 (ref.초등학교)	중학교	2.157	2.496	2.490	2.857	3.548	3.580	2.921	3.270
	일반고	9.343	9.264	10.422	10.301	9.843	9.630	9.102	9.219
	특성화고	4.283	7.083	4.192	7.577	2.234	6.792	1.813	6.249
부모학력 (ref.부모없음)	고졸이하(모두)	2.491	-2.283	-0.154	-2.460	2.517	0.107	-3.385	-3.719
	대졸이상(한명)	3.444	-1.958	0.624	-2.298	3.902	0.430	-2.585	-3.722
	대졸이상(모두)	4.276	-0.959	1.780	-1.278	3.982	0.994	-2.303	-2.996
경제활동 (ref.아버지만)	어머니만	-0.824	-0.090	-1.064	-0.526	-0.255	-0.055	-0.748	-0.138
	모두참여O	-1.187	-1.141	-0.742	-0.828	-0.478	-0.699	-0.733	-0.837
	모두참여X	0.359	1.997	-0.332	1.775	2.121	2.435	-0.217	0.952
성적		-0.014	0.157	-0.099	0.129	-0.050	0.138	-0.193	0.006
부모연령		0.490	0.538	0.291	0.475	0.397	0.750	1.068	1.044
소득		0.811	0.572	0.832	0.542	0.829	0.541	0.722	0.466
$\hat{\sigma}_e$		11.434	9.801	12.152	10.194	13.262	10.007	11.876	8.783
$\hat{\sigma}_u$		69.045	44.468	82.320	50.335	83.662	45.734	83.779	50.883

주) $\hat{\beta}_{(0)}$ 은 초기 모수 추정값이며, $\hat{\beta}_{EM}$ 은 EM-algorithm을 통하여 얻은 최종 모수 추정값임

다음은 인터넷조사를 참조조사로 하는 사교육시간 및 사교육비에 대한 모형 선택 결과이다.

<표 1-41> 사교육시간에 대한 모형 선택 결과(참조=인터넷)

보조변수		2011년				2012년			
		1차		2차		1차		2차	
		$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$
지역 (ref.강남)	강북	-0.204	0.045	-0.618	-0.420	-0.171	-0.136	-0.943	-0.861
	광역시	-4.194	-3.824	-5.156	-4.853	-3.364	-3.338	-4.650	-4.575
	중소도시	-7.577	-8.212	-8.663	-9.195	-8.476	-8.647	-9.421	-9.461
	읍면	-0.705	-0.967	-1.484	-1.499	-2.028	-2.265	-2.119	-2.427
학교급 (ref.초등학교)	중학교	-2.033	-1.970	-2.152	-2.257	-2.469	-2.663	-2.861	-3.020
	일반고	-2.397	-2.272	-2.662	-2.760	-2.657	-2.944	-3.285	-3.489
	특성화고	-4.118	-3.910	-3.986	-4.278	-4.620	-5.109	-4.864	-5.344
부모학력 (ref.부모없음)	고졸이하(모두)	3.477	4.303	3.264	4.187	1.460	1.193	2.162	2.210
	대졸이상(한명)	4.655	5.476	4.396	5.371	2.712	2.497	3.403	3.408
	대졸이상(모두)	4.944	5.819	4.942	5.991	2.844	2.691	3.530	3.721
경제활동 (ref.아버지만)	어머니만	-1.279	-1.176	-1.481	-1.256	-1.091	-0.945	-0.740	-0.619
	모두참여O	0.277	0.235	-0.176	-0.066	0.057	0.082	-0.015	0.097
	모두참여X	-3.407	-3.279	-4.057	-3.547	-3.704	-3.581	-3.549	-3.273
성적		-0.971	-0.941	-0.822	-0.979	-0.649	-0.700	-0.657	-0.707
부모연령		-0.557	-0.539	-0.578	-0.593	-0.562	-0.549	-0.503	-0.607
소득		0.823	0.903	0.873	0.912	0.850	0.875	0.952	0.972
$\hat{\sigma}_e$		8.143	7.297	8.102	7.521	7.880	7.528	7.840	7.620
$\hat{\sigma}_u$		7.338	4.298	7.847	4.461	7.376	4.354	7.704	4.547

주) $\hat{\beta}_{(0)}$ 은 초기 모수 추정값이며, $\hat{\beta}_{EM}$ 은 EM-algorithm을 통하여 얻은 최종 모수 추정값임

<표 1-42> 사교육비에 대한 모형 선택 결과(참조=인터넷)

보조변수		2011년				2012년			
		1차		2차		1차		2차	
		$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$
지역 (ref.강남)	강북	1.229	1.016	1.480	1.257	1.244	1.272	1.367	1.414
	광역시	3.343	3.192	3.626	3.502	3.271	3.375	3.412	3.491
	중소도시	1.468	2.608	1.421	2.638	1.146	3.114	1.351	2.944
	읍면	-0.790	-0.651	-0.222	-0.227	-0.203	-0.304	0.022	-0.046
학교급 (ref.초등학교)	중학교	-1.929	-1.494	-1.563	-1.312	-0.387	-0.454	-0.794	-0.743
	일반고	-1.699	-1.315	-1.430	-1.174	-0.729	-0.705	-0.850	-0.724
	특성화고	-2.222	-1.521	-1.358	-1.043	-0.539	-0.383	-0.641	-0.464



보조변수		2011년				2012년			
		1차		2차		1차		2차	
		$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{EM}$
부모학력 (ref.부모없음)	고졸이하 (모두)	1.303	-0.407	-0.325	-1.305	-0.231	-0.840	-0.386	-1.001
	대졸이상 (한명)	1.526	-0.331	-0.102	-1.266	-0.062	-0.888	-0.376	-1.152
	대졸이상 (모두)	1.854	0.000	0.206	-0.953	0.171	-0.650	0.072	-0.760
경제활동 (ref.아버지만)	어머니만	0.198	0.282	-0.277	0.051	0.258	0.297	0.090	0.072
	모두참여 O	-0.323	-0.359	-0.327	-0.293	-0.142	-0.194	-0.153	-0.202
	모두참여 X	0.512	0.935	-0.387	0.499	-0.540	0.503	0.434	1.011
부모연령		0.152	0.210	0.151	0.214	0.206	0.271	0.264	0.331
소득		0.259	0.163	0.239	0.157	0.276	0.139	0.276	0.142
성적		-0.053	0.042	-0.069	0.025	-0.021	0.065	-0.003	0.068
$\hat{\sigma}_e$		4.761	3.758	4.507	3.620	5.745	4.999	4.910	4.360
$\hat{\sigma}_u$		28.832	14.250	29.587	14.707	28.876	16.144	28.262	16.533

주) $\hat{\beta}_{(0)}$ 은 초기 모수 추정값이며, $\hat{\beta}_{EM}$ 은 EM-algorithm을 통하여 얻은 최종 모수 추정값임

사교육시간에 대한 모수 추정 결과에 대하여 종이조사를 참조조사로 한 경우와 비교해 보면, 2012년 1차 조사에서 가구의 소득이 한 단위 증가할 때 사교육시간은 0.875만큼 증가하고, 2차 조사에서 강북 소재 학교에 다니는 학생은 다른 보조 변수들이 고정되어 있을 때 강남 소재 학교에 다니는 학생에 비하여 사교육시간이 0.861만큼 감소하는 등, 모수 추정 결과의 부호는 동일하나 그 크기에는 상당한 차이가 있는 것으로 나타났다.

사교육비의 경우에는 참조조사의 선택에 따라 모수 추정 결과의 부호까지 차이가 있는 것으로 나타났다. 이는 사교육비가 사교육시간에 비해 조사방법에 따른 차이가 큰데 기인한다. 즉, 참조조사에 따라 추정 결과에 차이가 발생할 수 있으며, 조사방법에 따라 응답 결과에 차이가 큰 경우에는 그 차이가 더 커지는 것으로 확인되었다. 따라서 참조조사의 선택에 보다 신중을 기해야 할 것으로 보인다.

추정된 모형을 이용한 예측 기법을 2011년과 2012년 사교육비조사에 적용하면 다음과 같은 결과를 얻을 수 있다. 앞서 선택한 모형을 이용하여 사교육비 관련 관심이 있는 주요 특성에 대하여 imputation을 완료한 이후에 식 (22)를 이용하여 얻은 추정값을 기술한 것이다.

〈표 1-43〉 주요 특성에 대한 평균 추정 및 비교 결과(모형적합)

			종이	인터넷	모형적합		기존 추정값 (c)	차이	
					종이 기준(a)	인터넷 기준(b)		(a-c)	(b-c)
2011	1차	사교육비참여율	31.946	37.781	34.298	34.878	34.670	-0.371	0.209
		사교육시간	5.959	5.440	5.399	5.202	5.717	-0.318	-0.515
		사교육비	34.875	36.600	35.137	35.48	35.642	-0.505	-0.162
		시간당사교육비	1.264	1.374	1.285	1.358	1.313	-0.028	0.045
	2차	사교육비참여율	37.912	39.772	38.643	38.935	38.793	-0.15	0.142
		사교육시간	5.433	5.260	5.096	4.873	5.351	-0.255	-0.478
		사교육비	34.789	36.380	35.355	35.343	35.531	-0.176	-0.188
		시간당사교육비	1.267	1.350	1.286	1.347	1.306	-0.02	0.041
2012	1차	사교육비참여율	36.691	35.537	35.811	36.071	35.905	-0.094	0.166
		사교육시간	5.159	5.405	4.845	5.027	5.327	-0.482	-0.3
		사교육비	33.721	38.164	36.074	36.842	36.766	-0.692	0.076
		시간당사교육비	1.354	1.545	1.392	1.535	1.485	-0.093	0.05
	2차	사교육비참여율	40.551	39.904	40.113	40.256	40.076	0.037	0.18
		사교육시간	4.891	4.991	4.624	4.721	4.964	-0.34	-0.243
		사교육비	33.817	37.084	35.904	36.203	36.225	-0.321	-0.022
		시간당사교육비	1.299	1.442	1.324	1.445	1.404	-0.08	0.041

주) 사교육비참여율은 사교육에 참여하지 않는 비율(%)이며, 사교육시간은 주당 사교육시간(시간), 사교육비는 월평균 사교육비(만원), 시간당 사교육비는 월평균 시간당 사교육비(만원)임

사교육시간의 경우 모형을 이용한 예측 기법을 이용한 추정값이 종이 또는 인터넷 조사 값들보다 작게 추정되었음을 확인할 수 있다. 이는 사교육에 참여하지 않는 학생들의 경우 사교육시간을 “0”으로 표기하게 되며, 사교육에 참여하지 않는 학생 비율이 더 많은 조사(2011년-인터넷, 2012년-종이)에서 평균값이 작아지는 현상이 발생하게 된다. 그 이유는 현상 활용하고 있는 토빗 모형의 경우, 0을 기준으로 이보다 작은 값은 “0”으로 절삭하게 되므로 평균보다 작은 관측값들의 보정값들이 더 커지는 효과가 차단된다. 반면 평균보다 큰 관측값들의 보정값들은 평균 쪽으로 가까워지며 더 작아지게 되므로, 전체적으로 과소 추정된 것이다.

사교육비의 경우에는 조사방법을 선택할 수 있었던 2012년 결과에서 인터넷조사를 참조조사로 할 때의 추정값이 종이조사를 기준으로 할 때에 비해 다소 크게 나타나는



것으로 확인되었다. 이는 인터넷조사의 사교육비가 종이조사에 비해 큰 값을 갖는데 기인한 것으로, 조사방법에 따른 응답 차이가 존재하는 경우(측정 오차가 존재하는 경우), 참조조사의 선택에 따라 추정값이 과대 또는 과소 추정될 수 있음을 의미한다.

다음은 학교급별 모형 적합에 의한 예측 기법의 추정 결과이다.

〈표 1-44〉 학교급별 사교육 비참여율 추정 및 비교 결과(모형적합)

(단위: %, %p)

			종이	인터넷	모형적합		기존 추정값 (c)	차이	
					종이 기준(a)	인터넷 기준(b)		(a-c)	(b-c)
2011	1차	초등학교	15.628	18.105	15.947	16.455	16.684	-0.737	-0.229
		중학교	25.311	30.285	26.979	27.408	27.639	-0.66	-0.231
		일반고	37.705	44.696	40.796	41.556	41.084	-0.288	0.472
		특성화고	76.087	73.645	76.338	76.219	74.882	1.456	1.337
	2차	초등학교	17.575	16.336	16.895	16.307	17.024	-0.129	-0.717
		중학교	29.998	32.204	30.442	30.972	31.056	-0.614	-0.084
		일반고	46.731	49.357	47.974	48.533	48.016	-0.042	0.517
		특성화고	79.950	78.171	80.576	80.339	79.132	1.444	1.207
2012	1차	초등학교	21.465	17.422	18.696	18.893	18.811	-0.115	0.082
		중학교	28.053	27.564	26.658	27.59	27.726	-1.068	-0.136
		일반고	42.389	41.353	41.551	41.967	41.653	-0.102	0.314
		특성화고	77.535	77.926	78.328	79.177	77.789	0.539	1.388
	2차	초등학교	22.169	17.702	18.089	18.888	18.976	-0.887	-0.088
		중학교	32.579	32.084	31.216	32.129	32.222	-1.006	-0.093
		일반고	49.554	48.491	48.861	49.087	48.752	0.109	0.335
		특성화고	80.446	80.572	82.447	81.334	80.538	1.909	0.796

학교급별 사교육 비참여율에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 예측 기법에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다. 이 때 조사방법을 선택할 수 없었던 2011년에 비해 조사방법을 선택했던 2012년도에 추정값 간 차이가 다소 큰 것으로 확인되었으며, 특성화고를 제외한 나머지 학교급에서는 인터넷조사를 참조조사로 한 경우의 추정값이 다소 큰 것으로 나타났다.

〈표 1-45〉 학교급별 주당 사교육시간 평균 추정 및 비교 결과(모형적합)

(단위: 시간)

			종이	인터넷	모형적합		기존 추정값 (c)	차이	
					종이 기준(a)	인터넷 기준(b)		(a-c)	(b-c)
2011	1차	초등학교	7.676	7.503	7.266	7.141	7.602	-0.336	-0.461
		중학교	7.706	6.954	7.142	6.594	7.354	-0.212	-0.76
		일반고	4.644	4.212	4.153	4.027	4.435	-0.282	-0.408
		특성화고	1.604	2.061	1.465	1.604	1.829	-0.364	-0.225
	2차	초등학교	7.459	7.630	7.398	7.099	7.535	-0.137	-0.436
		중학교	7.130	6.740	6.622	6.837	6.943	-0.321	-0.106
		일반고	3.976	3.769	3.640	3.489	3.875	-0.235	-0.386
		특성화고	1.331	1.901	1.195	1.400	1.593	-0.399	-0.193
2012	1차	초등학교	6.621	6.870	6.308	6.409	6.785	-0.477	-0.376
		중학교	6.701	6.800	6.192	6.359	6.768	-0.576	-0.409
		일반고	3.978	4.564	3.883	4.191	4.394	-0.511	-0.203
		특성화고	1.726	1.628	1.513	1.416	1.662	-0.149	-0.246
	2차	초등학교	6.498	6.824	6.223	6.600	6.731	-0.508	-0.131
		중학교	6.264	6.343	5.828	5.999	6.321	-0.493	-0.322
		일반고	3.623	3.910	3.445	3.663	3.839	-0.395	-0.176
		특성화고	1.410	1.389	1.102	1.244	1.394	-0.292	-0.150

학교급별 사교육시간에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 예측 기법에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났으며, 인터넷조사를 기준으로 한 경우의 추정값이 종이조사를 기준으로 한 경우에 비해 다소 큰 것으로 확인되었다.



〈표 1-46〉 학교급별 월평균 사교육비 평균 추정 및 비교 결과(모형적합)

(단위: 만원)

			종이	인터넷	모형적합		기존 추정값 (c)	차이	
					종이 기준(a)	인터넷 기준(b)		(a-c)	(b-c)
					2011	1차	초등학교	27.615	27.140
		중학교	35.472	38.298	36.054	36.386	36.746	-0.692	-0.36
		일반고	41.454	43.489	41.449	41.917	42.377	-0.928	-0.46
		특성화고	24.235	25.294	25.029	25.161	24.783	0.246	0.378
	2차	초등학교	26.884	27.311	27.197	26.755	27.075	0.122	-0.32
		중학교	35.200	38.530	35.927	40.409	36.771	-0.844	3.638
		일반고	42.946	43.149	42.677	42.446	43.043	-0.366	-0.597
		특성화고	23.043	27.225	24.731	26.014	25.055	-0.324	0.959
2012	1차	초등학교	25.617	26.653	25.942	26.354	26.309	-0.366	0.045
		중학교	36.635	39.835	37.441	38.679	38.784	-1.343	-0.105
		일반고	39.719	46.533	43.118	44.443	44.583	-1.465	-0.14
		특성화고	23.662	25.482	28.127	25.834	24.838	3.29	0.996
	2차	초등학교	25.265	26.344	25.500	26.143	26.048	-0.548	0.095
		중학교	36.426	39.477	37.284	38.386	38.633	-1.348	-0.247
		일반고	41.090	45.578	43.242	44.346	44.495	-1.253	-0.149
		특성화고	23.966	24.338	24.510	24.575	24.239	0.272	0.336

학교급별 평균 사교육비에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 예측 기법에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다. 이 때 조사방법을 선택할 수 없었던 2011년에 비해 조사방법을 선택했던 2012년도에 추정값 간 차이가 약간 더 큰 것으로 확인되었으며, 인터넷조사를 기준으로 한 경우의 추정값이 종이조사를 기준으로 한 경우에 비해 다소 큰 것으로 나타났다.

〈표 1-47〉 학교급별 월평균 시간당 사교육비 평균 추정 및 비교 결과(모형적합)

(단위: 만원)

			종이	인터넷	모형적합		기존 추정값 (c)	차이	
					종이 기준(a)	인터넷 기준(b)		(a-c)	(b-c)
2011	1차	초등학교	0.894	0.927	0.883	0.909	0.908	-0.025	0.001
		중학교	1.087	1.256	1.092	1.216	1.163	-0.072	0.053
		일반고	1.720	1.810	1.736	1.806	1.761	-0.026	0.045
		특성화고	1.173	1.204	1.237	1.279	1.189	0.048	0.09
	2차	초등학교	0.872	0.892	0.854	0.876	0.881	-0.027	-0.005
		중학교	1.094	1.267	1.109	1.228	1.176	-0.066	0.052
		일반고	1.797	1.822	1.812	1.846	1.809	0.003	0.037
		특성화고	1.116	1.168	1.224	1.247	1.141	0.083	0.106
2012	1차	초등학교	0.954	1.106	0.956	1.085	1.056	-0.099	0.029
		중학교	1.280	1.443	1.280	1.433	1.389	-0.109	0.044
		일반고	1.818	1.968	1.835	1.972	1.925	-0.09	0.047
		특성화고	1.081	1.294	1.231	1.359	1.219	0.012	0.14
	2차	초등학교	0.927	0.991	0.922	0.97	0.974	-0.052	-0.004
		중학교	1.225	1.374	1.231	1.369	1.332	-0.101	0.037
		일반고	1.773	1.903	1.788	1.917	1.872	-0.084	0.045
		특성화고	1.076	1.279	1.166	1.352	1.225	-0.059	0.127

학교급별 시간당 사교육비에 대한 분석 결과, 종이조사와 인터넷조사를 구분하지 않은 상태에서 얻어진 기존 추정값과 예측 기법에 의한 추정값 사이에 차이는 거의 없는 것으로 나타났다. 시간당 사교육비는 사교육시간과 사교육비를 이용하여 계산된 값이므로, 사교육시간과 사교육비와 마찬가지로 인터넷조사를 기준으로 한 경우의 추정값이 종이조사를 기준으로 한 경우에 비해 다소 큰 것으로 확인되었다.

모형 적합을 이용한 예측 기법으로 구한 추정값은 보정 이전과 전반적으로 큰 차이가 없는 것으로 확인되었다.



제5절 결론

혼합조사로 진행된 사교육비조사의 기존 추정량에는 조사방법에 따른 차이가 별도로 반영되어 있지 않다. 사교육비조사뿐 아니라 여러 가지 조사방법으로 진행되는 다수의 혼합조사들의 추정에 기존의 추정 방법들이 그대로 적용되고 있는 것이 현실이다.

그러나 하나의 조사에 여러 가지 조사방법을 활용하면 조사방법의 선택 여부에 따른 선택효과와 조사방법에 따른 측정오차가 발생하게 되고, 이를 추정에 반영하지 않는다면 추정 결과에 편의가 발생하게 된다. 따라서 혼합조사로 인한 선택 및 측정효과 등을 파악하고, 이를 추정 결과에 반영함으로써 정도 높은 추정 결과를 도출할 수 있도록 해야 한다.

본 연구에서는 이와 같은 노력의 일환으로 사교육비조사를 대상으로 혼합조사에서의 선택효과와 측정오차에 관하여 살펴보고, 조사방법에 따른 특성을 추정 과정에 반영할 수 있는 방법에 대하여 살펴보았다. 두 가지 조사방법(종이 및 인터넷)으로부터 얻은 자료에 3가지 추정 방안을 적용하였으며, 그 결과를 기존 추정값과 비교하였다.

먼저 적용한 방안은 두 조사방법의 분산 크기를 고려한 최적결합방법이다. 최적결합 방법은 각각의 조사방법으로부터 얻어진 자료로부터 개별적으로 추정값을 도출한 후, 분산 크기에 반비례하는 가중값을 이용하여 두 추정 결과를 하나로 결합하는 방법이다. 최적결합추정량은 둘 이상의 조사방법으로부터 얻어진 조사 결과를 하나로 통합하는데 조사방법별 분산 크기만을 고려하게 되므로, 특정 성향의 조사대상자가 특정 조사방법을 선호하는 등의 선택효과를 반영하기에는 어려움이 있다. 그러나 상대적으로 조사방법에 따른 측정효과(분산의 차이)를 반영하는 것에는 적합한 방법으로 볼 수 있다. 예를 들어 인터넷조사를 통해 얻은 자료가 다른 조사방법에 비해 큰 분산을 갖고 있다면, 이는 인터넷조사가 해당 특성을 측정하는데 있어 다른 조사방법과는 다른 특성을 가지고 있다고 볼 수 있다. 소득, 소비, 정신 건강 또는 자살 등과 같이 다소 민감한 내용의 경우에 인터넷조사에서의 분산이 상대적으로 크게 나타나는 것으로 알려져 있다. 이는 응답자가 민감한 내용에 대한 답변에 있어 비교적 정직하게 답변하는 성향이 강하게 나타남과 동시에 다소 부정적이고 극단적으로 답변하는 성향도 함께 나타나기 때문이다.

두 번째로 적용한 방안은 성향점수를 이용한 방법으로, 조사방법의 선택에 조사대상자의 특정 성향이 영향을 미쳤을 것이라는 가정을 기반으로 한다. 만약 특정 성향을 가진 조사대상자가 특정 조사방법을 선호하는 경향이 강하다면, 성향점수를 추정에 반영하는 것은 정도 높은 추정값을 얻는데 매우 유용한 방법이 된다. 반면 조사방법을 선택하는데 별다른 성향이 발견되지 않거나, 선택 성향을 구분할 만한 적절한 보조변수가 확보되지 못할 경우에는 별다른 개선 효과를 내지 못하게 된다. 성향점수를 이용한 추정량은 조사

방법을 선택하는 조사대상자들의 성향을 이용하여 특정 조사방법을 선택할 확률을 가중값으로 이용하는 방법으로 선택효과를 반영하는 것에는 효과적이나, 조사방법에 따른 측정효과를 반영하는 데는 다소 어려움이 있을 것으로 예상된다.

마지막으로 적용한 추정 방안은 참값에 대한 보조변수와의 구조적 관계를 나타내는 구조 모형과 구조 모형 외에 관측값과 참값의 관계를 나타내는 측정 오차에 대한 측정 모형을 설정하고, 설정된 모형의 모수를 추정한 후 예측치를 계산하여 자료를 보정하는 방법이다. 모형을 이용한 추정 방법은 성향점수를 이용한 방법에서와 같이 조사방법의 선택에 영향을 미칠 것으로 예상되는 보조 변수들을 이용한 모형을 설정하고, 해당 모형을 이용하여 관심 변수의 값을 예측하게 된다. 즉, 실제 조사 결과 대신 예측값을 추정에 이용하게 되므로 일종의 대체(imputation) 방법으로 볼 수 있다.

이 때 모형 설정을 위한 참조조사는 모집단의 분포와 유사한 분포를 가지고 있어야 하며, 일반적으로 체계적인 방법에 의해 확률적으로 얻어진 조사 결과를 참조조사로 이용하게 된다. 만약 모형 설정에 이용하는 참조조사의 분포가 모집단의 분포와 같지 않으면, 어떤 보조 변수를 이용하든 정확성 높은 결과를 얻을 수 없다. 따라서 참조조사를 어떤 조사방법으로 설정하는가는 매우 중요한 결정 사항이며, 선택에 따라 추정값이 달라지게 된다. 본 연구에서는 별도의 참조조사가 진행되지 않았으므로, 기존 조사방법인 종이조사와 다수의 조사대상자가 선택한 인터넷조사를 각각 참조조사로 하여 추정값을 도출하고 이들을 비교하였다.

3가지 추정 방안에 의한 4가지 결과(결합, 성향, 모형(참조:종이), 모형(참조:인터넷))와 기존 추정값과의 비교 결과는 <표 1-48>과 같다.

조사방법별로 표본이 추출된 2011년도의 경우에는 조사대상자가 조사방법을 선택하지 못하므로, 조사 결과에 조사방법에 대한 선택효과가 포함되지 않으며, 조사방법에 따른 측정오차(측정효과)만이 포함된다. 이 때 사교육비 관련 주요 관심 변수에 대한 추정 결과, 측정효과만을 반영한 최적결합방법이 기존 추정값과 가장 가까운 추정값을 갖게 되는 것으로 확인되었다.

반면 조사대상자들이 조사방법을 선택했던 2012년도의 경우, 사교육에 참여 하지 않는 비율은 참조조사가 종이인 모형추정방안이, 사교육시간은 최적결합추정방안, 사교육비는 참조조사가 인터넷인 모형추정방안이 기존 추정값과 가장 가까운 추정값을 갖게 되는 것으로 나타났다. 이는 조사방법에 대한 선택효과가 사교육비에 보다 크게 영향을 미쳤기 때문인 것으로 보이며, 특히 인터넷조사를 선택한 조사대상자의 비중이 종이조사에 비해 더 크기 때문에 인터넷조사를 참조조사로 한 모형추정방안의 결과가 기존 추정값과 가장 가까운 값을 갖게 된 것이다. 이와 같이 선택효과의 유·무 및 참조조사의 선택에 따라 추정 방안에 따른 결과에 다소 간의 차이가 발생하는 것으로 확인되었다.



<표 1-49>는 각 추정 방안에 따른 분산 추정 결과이다. 표에 의하면, 분산이 큰 조사 결과에는 작은 가중치를, 분산이 작은 조사 결과에는 큰 가중치를 부여한 결합추정방안의 분산이 가장 작은 것으로 나타났다.

사교육비조사의 경우에는 관심변수에 대한 참값을 알 수 없다. 따라서 어떤 추정 방법이 참값에 가장 가까운가, 즉 정확성이 가장 높은 추정량이 무엇인가에 대한 의사결정이 쉽지 않다. 또한 추정값의 정확성을 확보하지 못한 상태에서 분산의 크기만으로 정도 높은 추정량이 어떤 것인가에 대하여 운운하는 것도 이치에 맞지 않다.

앞서 살펴본 바와 같이 여러 추정 방안을 통하여 추정값을 도출하고, 해당 결과 간에 차이가 어느 정도인지 확인하는 과정을 통하여, 우리는 혼합조사(사교육비조사)에 대한 최선의 혼합 추정 방법이 아닌, 조사에 참여한 방법과 조사방법의 선택 비중 그리고 관심을 갖는 변수의 특성 및 참조조사의 결정에 따라 추정 결과에 차이가 발생하게 된다는 사실을 발견하였다. 이는 두 개 이상의 조사방법을 활용한 혼합조사에서의 추정량이 조사방법의 선택 관련 특성, 주요 관심 변수, 참조조사 등 여러 요인에 따라 차이를 가질 수 있음을 의미하며, 이들의 효과와 특성을 면밀히 확인한 후에야 최적의 혼합추정 방법을 결정할 수 있을 것이다.

〈표 1-48〉 추정 방인별 추정값 비교 결과

	추정값					차이					최소 차이
	기준(a)	결함(b)	성형(c)	모형I(d)	모형II(e)	(a-b)	(a-c)	(a-d)	(a-e)		
2011	1차	사교육비참여율	34.746	33.702	34.298	34.878	-0.076	0.968	0.372	-0.208	결함
		사교육시간	5.717	5.704	5.399	5.202	0.013	-0.125	0.318	0.515	결함
		사교육비	35.642	35.666	35.137	35.48	-0.024	0.394	0.505	0.162	결함
	2차	시간당사교육비	1.313	1.309	1.289	1.358	0.004	0.024	0.028	-0.045	결함
		사교육비참여율	38.793	38.812	38.34	38.643	-0.019	0.453	0.15	-0.142	결함
		사교육시간	5.351	5.348	5.423	5.096	0.003	-0.072	0.255	0.478	결함
2012	1차	사교육비	35.531	35.58	35.299	35.343	-0.049	0.232	0.176	0.188	결함
		시간당사교육비	1.306	1.305	1.286	1.347	0.001	0.02	0.02	-0.041	결함
		사교육비참여율	35.905	36.111	37.338	35.811	-0.206	-1.433	0.094	-0.166	모형I
	2차	사교육시간	5.327	5.274	5.208	4.845	0.053	0.119	0.482	0.3	결함
		사교육비	36.766	35.546	34.771	36.074	1.22	1.995	0.692	-0.076	모형II
		시간당사교육비	1.485	1.427	1.424	1.392	0.058	0.061	0.093	-0.05	모형II
2차	사교육비참여율	40.076	40.224	40.612	40.113	-0.148	-0.536	-0.037	-0.18	모형I	
	사교육시간	4.964	4.94	4.917	4.624	0.024	0.047	0.34	0.243	결함	
	사교육비	36.225	35.244	34.661	35.904	0.981	1.564	0.321	0.022	모형II	
		시간당사교육비	1.404	1.356	1.359	1.324	0.048	0.045	0.08	-0.041	모형II

주1) 2011년 조사는 조사방법별로 표본이 추출(조사방법 선택 불가)되었으며, 2012년 조사는 조사방법을 조사대상자가 선택함

주2) 모형I은 참조조사가 종이조사이고, 모형II는 인터넷조사임



〈표 1-49〉 추정 방안별 분산 추정값 비교 결과

	분산 추정값				차이				최소 분산	
	기준(a)	결함(b)	성형(c)	모형(d)	모형II(e)	(a-b)	(a-c)	(a-d)		(a-e)
2011	사교육비참여율	0.226	0.113	0.223	0.225	0.227	0.113	0.003	0.001	-0.001
	사교육시간	37.985	18.975	38.510	33.169	32.339	19.010	-0.525	4.816	5.646
	사교육비	648.634	324.725	629.728	671.818	700.004	323.909	18.906	-23.184	-51.370
	시간당사교육비	1.314	0.644	1.262	0.840	1.064	0.670	0.052	0.474	0.250
2011	사교육비참여율	0.237	0.119	0.236	0.237	0.238	0.118	0.001	0.000	-0.001
	사교육시간	37.325	18.651	37.728	33.650	31.582	18.674	-0.403	3.675	5.743
	사교육비	693.975	347.039	685.047	722.829	740.589	346.936	8.928	-28.854	-46.614
	시간당사교육비	1.344	0.670	1.305	0.934	1.019	0.674	0.039	0.410	0.325
2012	사교육비참여율	0.230	0.115	0.234	0.230	0.231	0.115	-0.004	0.000	-0.001
	사교육시간	35.900	17.477	35.933	27.998	33.064	18.423	-0.033	7.902	2.836
	사교육비	689.740	311.276	604.913	733.554	734.626	378.464	84.827	-43.814	-44.886
	시간당사교육비	1.965	0.851	1.890	0.831	1.815	1.114	0.075	1.134	0.150
2012	사교육비참여율	0.240	0.120	0.241	0.240	0.241	0.120	-0.001	0.000	-0.001
	사교육시간	33.854	16.783	33.811	28.484	31.335	17.071	0.043	5.370	2.519
	사교육비	710.924	329.006	633.315	750.747	741.676	381.918	77.609	-39.823	-30.752
	시간당사교육비	1.543	0.676	1.489	0.669	1.449	0.867	0.054	0.874	0.094

주1) 2011년 조사는 조사방법별로 표본이 추출(조사방법 선택 불가)되었으며, 2012년 조사는 조사방법을 조사대상자가 선택함

주2) 모형I은 참조조사가 종이조사이고, 모형II는 인터넷조사임

따라서 보다 정밀하게 설계된 시험조사(Pilot survey)를 통하여 선택효과와 측정효과(나아가 순차효과⁴⁾) 등을 구분하고, 이들을 추정 과정에 반영하는 노력이 필요하다. 또한 추정 결과에 영향을 미칠 수 있는 참조조사를 기존 조사방법으로 할 것인지, 아니면 보다 많은 조사대상자가 선택한 조사방법으로 할 것인지 등에 대한 의사결정도 과학적인 방법에 의해 수집된 관련 정보를 통하여 내려야 한다. 이와 같은 다각적인 검토 결과가 바탕이 되어야 혼합조사에 대한 최적의 추정 방안을 결정할 수 있을 것이며, 이 때의 추정 방안은 혼합조사의 특성(조사방법의 종류 및 선택여부, 순차효과, 조사표효과, 조사원 개인 정도 등)에 따라 서로 다른 것이 될 것이다.

조사 환경의 악화로 조사방법, 조사표 등의 활용이 점차 다양해지고 있는 현 상황에서 혼합조사는 보다 더 확대될 것으로 예상된다. 또한 조사방법에 따른 조사표 적용 또는 조사원 개입 여부 등 조사방법은 더욱 다변화될 것이다. 그러나 빠르게 분화하고 있는 조사방법과는 달리 혼합조사의 추정 방법에 대한 고민은 그 속도를 따라가지 못하고 있는 것이 현실이다. 또한 선택효과, 측정효과 등 조사방법에 따른 속성을 파악하고 이를 추정에 반영하는 것이 모든 조사에서 가능한 것이 아니므로, 표준화된 추정 방법을 정하고, 이를 활용하는 것이 그리 녹록한 일은 아닐 것이다. 그러나 조사방법의 혼용과 조사 진행 과정의 차이가 조사 결과에 유의미한 영향을 미치고, 나아가 통계의 정합성을 떨어뜨리는 원인으로 작용하게 되므로, 보다 과학적인 방법을 통하여 이들 간의 관계를 면밀히 파악하고 이를 조사 결과에 반영하는 일은 더 이상 미룰 수 없는 상황에 직면해 있다.

따라서 조사방법에 따른 특성과, 해당 조사가 진행되는 과정에 대한 정보 등을 축적하고 이를 통하여 추정 방법을 개선하는 일은 앞으로도 꾸준한 연구가 필요한 분야임에 틀림이 없으며, 특히 고품질의 정도 높은 국가 통계를 생산해야 하는 통계청에서는 이에 대한 관심과 연구에 노력을 아끼지 말아야 할 것이다.

4) 일반적으로 순차적 혼합조사에 활용되는 다양한 방식의 조사방법들은 저렴한 비용의 조사방법을 시작으로 가장 비용이 많이 드는 조사방법으로 순차 적용된다(Griffin & Obenski, 2002). 이 때 조사방법에 따라 응답자 특성에 따른 접촉가능성에 차이가 발생하게 된다. 예를 들어 모바일을 활용한 조사의 경우 젊은층의 접촉 가능성이 노년층에 비해 월등히 높으며, 이들 자료를 하나로 묶어 단순 처리할 때 문제가 발생(Link et al., 2007) 하게 되므로, 순차효과가 발생하는 경우 추정 과정에 해당 효과를 반영할 필요가 있다.



참고문헌

- 강창구(2008) “결합추정기법을 이용한 분기 GDP 추정”, Quarterly National Accounts, 1st 2008, pp46~73.
- 김서영, 안다영, 권순필, 이승희 (2010) “사회조사에서 자원자표본 인터넷조사 추정”, 통계개발원 2010년 하반기 연구보고서 제I권, pp291~355.
- 김영원, 배예영 (2010) “인터넷 선거여론조사 가중치보정을 위한 성향점수의 활용”, 한국통계학회 논문집, 제17권, 제1호, pp.55-66.
- 박상언, 김재광, 김서영 (2012) “혼합조사에서의 추정 방법 개발”, 통계개발원 학술연구용역 결과 보고서.
- 이종원, 이상돈 (1999) “결합방식을 이용한 예측모형의 효율성 분석”, 한국응용경제학회, 제1권, 제1호, pp3~32.
- 홍영택(2011) “성향점수를 이용한 여론조사의 추정법”, 성균관대학교 석사학위논문.
- Marcel Das(2012), The 4th International Workshop on Internet Survey Methods.
- Adam S., Ramara B., and Rebecca S.(2001) "Using paradata to examine the effects of interviewer characteristics on survey response and data quality", Proceedings of the Annual Meeting of the American Statistical Association.
- Aneesah W.(2012) "Using Paradata to Improve Questionnaire Design and Operations" Censuses and Surveys of Governments: A Workshop on the Research and Methodology behind the Estimates.
- Barbara C. and Matthew E.(2011) "Designing a paradata application in a CAPI environment", O'Hare & Jans.
- Beth L.(2008) "The 2006 National Health Interview Survey (NHIS) Paradata File: Overview and Applications", JSM 2008.
- Chris S. and Jim D. "PANDA: Using Paradata for Quality and Performance Monitoring", U.S. Census Bureau.
- Frauke K., Mick C. and Lars L.(2010) "The Use of Paradata to Monitor and Manage Survey Data Collection", JSM2010.
- Jim O.(2009) "Paradata and Blaise: A Review of Recent Applications and Research"
- Nancy B., James D., Polly P., Adam S., and Lucilla T. (2010) "Assessing Contact History Paradata Quality Across Several Federal Surveys", JSM 2010.
- Teichman, Ari.(2009) "Panda: Using Paradata to Improve Data Quality", FedCasic Conference, Washington DC.
- William E. and Jr. (2004) "Contact History Instrument(CHI)", U.S. Census Bureau.

<부 록>

1. 학교급별 분산 추정 및 비교 결과

<부표 1-1> 사교육 비참여율에 대한 분산 추정 및 비교 결과

		기존 (a)	결합 (b)	성향 (c)	모형I (d)	모형II (e)	차이				최소 분산	
							(a-b)	(a-c)	(a-d)	(a-e)		
2011	1차	초등학교	0.139	0.070	0.140	0.134	0.137	0.069	-0.001	0.005	0.002	결합
		중학교	0.200	0.100	0.201	0.197	0.199	0.100	-0.001	0.003	0.001	결합
		일반고	0.242	0.120	0.242	0.242	0.243	0.122	0.000	0.000	-0.001	결합
		특성화고	0.188	0.094	0.190	0.181	0.181	0.094	-0.002	0.007	0.007	결합
	2차	초등학교	0.141	0.070	0.142	0.140	0.136	0.071	-0.001	0.001	0.005	결합
		중학교	0.214	0.107	0.215	0.212	0.214	0.107	-0.001	0.002	0.000	결합
		일반고	0.250	0.125	0.250	0.250	0.250	0.125	0.000	0.000	0.000	결합
		특성화고	0.165	0.083	0.167	0.158	0.158	0.082	-0.002	0.007	0.007	결합
2012	1차	초등학교	0.153	0.078	0.159	0.152	0.153	0.075	-0.006	0.001	0.000	결합
		중학교	0.200	0.100	0.208	0.198	0.200	0.100	-0.008	0.002	0.000	결합
		일반고	0.243	0.122	0.247	0.243	0.244	0.121	-0.004	0.000	-0.001	결합
		특성화고	0.173	0.087	0.169	0.170	0.165	0.086	0.004	0.003	0.008	결합
	2차	초등학교	0.154	0.079	0.159	0.149	0.153	0.075	-0.005	0.005	0.001	결합
		중학교	0.218	0.109	0.224	0.218	0.218	0.109	-0.006	0.000	0.000	결합
		일반고	0.250	0.125	0.250	0.250	0.250	0.125	0.000	0.000	0.000	결합
		특성화고	0.157	0.078	0.153	0.146	0.152	0.079	0.004	0.011	0.005	결합



〈부표 1-2〉 주당 사교육시간에 대한 분산 추정 및 비교 결과

		기존 (a)	결합 (b)	성향 (c)	모형I (d)	모형II (e)	차이				최소 분산	
							(a-b)	(a-c)	(a-d)	(a-e)		
2011	1차	초등학교	37.214	18.706	37.418	31.812	32.938	18.508	-0.204	5.402	4.276	결합
		중학교	44.754	22.309	44.993	40.909	37.371	22.445	-0.239	3.845	7.383	결합
		일반고	29.771	14.864	30.161	25.45	24.850	14.907	-0.390	4.321	4.921	결합
		특성화고	19.298	9.283	19.861	13.082	16.118	10.015	-0.563	6.216	3.180	결합
	2차	초등학교	35.864	17.933	36.132	33.513	30.860	17.931	-0.268	2.351	5.004	결합
		중학교	44.178	22.049	44.324	40.089	42.390	22.129	-0.146	4.089	1.788	결합
		일반고	28.503	14.215	28.831	24.947	23.085	14.288	-0.328	3.556	5.418	결합
		특성화고	18.387	8.159	19.044	10.607	16.210	10.228	-0.657	7.780	2.177	결합
2012	1차	초등학교	35.987	17.47	36.537	26.866	33.239	18.517	-0.550	9.121	2.748	결합
		중학교	40.025	19.811	39.955	33.75	36.704	20.214	0.070	6.275	3.321	결합
		일반고	30.658	14.14	29.701	21.988	28.835	16.518	0.957	8.670	1.823	결합
		특성화고	19.039	9.673	18.344	16.719	14.900	9.366	0.695	2.320	4.139	결합
	2차	초등학교	31.776	15.919	32.237	25.114	30.243	15.857	-0.461	6.662	1.533	결합
		중학교	38.808	19.318	38.623	35.219	35.972	19.490	0.185	3.589	2.836	결합
		일반고	29.048	14.084	28.212	22.356	27.078	14.964	0.836	6.692	1.970	결합
		특성화고	14.883	7.344	14.096	9.497	12.938	7.539	0.787	5.386	1.945	결합

〈부표 1-3〉 월평균 사교육비에 대한 분산 추정 및 비교 결과

		기존 (a)	결합 (b)	성향 (c)	모형I (d)	모형II (e)	차이				최소 분산	
							(a-b)	(a-c)	(a-d)	(a-e)		
2011	1차	초등학교	353.16	175.93	361.21	367.53	374.014	177.23	-8.05	-14.37	-20.85	결합
		중학교	511.50	252.50	510.96	541.29	575.861	259.00	0.54	-29.79	-64.36	결합
		일반고	890.57	445.64	877.63	941.06	971.956	444.93	12.94	-50.49	-81.39	결합
		특성화고	497.67	248.76	497.66	501.69	520.230	248.91	0.01	-4.02	-22.56	결합
	2차	초등학교	350.43	175.15	361.08	373.00	362.763	175.28	-10.65	-22.57	-12.33	결합
		중학교	525.36	258.33	530.42	559.10	802.190	267.03	-5.06	-33.74	-276.83	결합
		일반고	1,027.81	513.91	1,033.18	1,101.01	1,106.11	513.90	-5.37	-73.20	-78.30	결합
		특성화고	506.63	243.75	510.49	523.28	546.882	262.88	-3.86	-16.65	-40.25	결합
2012	1차	초등학교	288.40	143.77	278.16	316.62	309.133	144.63	10.24	-28.22	-20.73	결합
		중학교	556.18	266.97	521.87	615.05	615.217	289.21	34.31	-58.87	-59.04	결합
		일반고	968.50	416.01	877.11	1,071.62	1,026.43	552.49	91.39	-103.12	-57.93	결합
		특성화고	473.77	233.91	466.68	896.68	534.607	239.86	7.09	-422.91	-60.84	결합
	2차	초등학교	280.46	135.19	271.64	300.32	297.930	145.27	8.82	-19.86	-17.47	결합
		중학교	537.69	258.75	519.84	618.59	578.547	278.94	17.85	-80.90	-40.86	결합
		일반고	1,089.99	502.10	999.71	1,162.22	1,144.13	587.89	90.28	-72.23	-54.14	결합
		특성화고	439.96	219.44	415.24	461.21	463.385	220.52	24.72	-21.25	-23.43	결합



〈부표 1-4〉 월평균 시간당 사교육비에 대한 분산 추정 및 비교 결과

		기존 (a)	결합 (b)	성향 (c)	모형I (d)	모형II (e)	차이				최소 분산	
							(a-b)	(a-c)	(a-d)	(a-e)		
2011	1차	초등학교	0.505	0.251	0.516	0.322	0.384	0.254	-0.011	0.183	0.121	결합
		중학교	0.987	0.451	0.997	0.517	0.845	0.536	-0.010	0.470	0.142	결합
		일반고	1.862	0.922	1.826	1.228	1.474	0.940	0.036	0.634	0.388	결합
		특성화고	1.269	0.582	1.256	0.683	1.164	0.687	0.013	0.586	0.105	결합
	2차	초등학교	0.420	0.210	0.418	0.291	0.301	0.210	0.002	0.129	0.119	결합
		중학교	1.046	0.473	1.063	0.526	0.922	0.573	-0.017	0.520	0.124	결합
		일반고	2.009	1.004	2.002	1.504	1.420	1.005	0.007	0.505	0.589	결합
		특성화고	1.145	0.572	1.174	0.816	0.880	0.573	-0.029	0.329	0.265	결합
2012	1차	초등학교	0.986	0.430	1.016	0.388	0.879	0.556	-0.030	0.598	0.107	모형I
		중학교	1.794	0.731	1.863	0.635	1.671	1.063	-0.069	1.159	0.123	모형I
		일반고	2.567	1.135	2.492	1.098	2.394	1.432	0.075	1.469	0.173	모형I
		특성화고	1.487	0.667	1.538	0.761	1.402	0.820	-0.051	0.726	0.085	결합
	2차	초등학교	0.646	0.303	0.654	0.284	0.566	0.343	-0.008	0.362	0.08	모형I
		중학교	1.312	0.546	1.340	0.487	1.230	0.766	-0.028	0.825	0.082	모형I
		일반고	2.130	0.953	2.147	0.909	2.011	1.177	-0.017	1.221	0.119	모형I
		특성화고	1.905	0.636	1.740	0.519	1.964	1.269	0.165	1.386	-0.059	모형I

2. 성향점수를 이용한 추정 방안에서 $C=5$ 인 경우의 주요 특성에 대한 추정 결과

〈부표 1-5〉 주요 특성에 대한 평균 추정 및 비교 결과($C=5$)

			종이	인터넷	성향점수 보정(a)	기존 추정값(b)	차이 (a-b)
2011	1차	사교육참여율	31.940	37.780	33.699	34.670	-0.971
		사교육시간	5.959	5.440	5.842	5.717	0.125
		사교육비	34.875	36.600	35.185	35.642	-0.457
		시간당사교육비	5.058	5.497	1.287	1.313	-0.026
	2차	사교육참여율	37.910	39.730	38.301	38.770	-0.469
		사교육시간	5.433	5.265	5.428	5.353	0.075
		사교육비	34.789	36.436	35.239	35.557	-0.318
		시간당사교육비	5.067	5.404	1.283	1.306	-0.023
2012	1차	사교육참여율	36.690	35.540	37.225	35.900	1.325
		사교육시간	5.159	5.405	5.216	5.327	-0.111
		사교육비	33.721	38.164	34.989	36.766	-1.777
		시간당사교육비	5.415	6.182	1.431	1.485	-0.054
	2차	사교육참여율	40.550	39.900	40.684	40.080	0.604
		사교육시간	4.891	4.991	4.912	4.964	-0.052
		사교육비	33.817	37.084	34.578	36.225	-1.647
		시간당사교육비	5.195	5.769	1.356	1.404	-0.049



〈부표 1-6〉 주요 특성에 대한 분산 추정 및 비교 결과

			기준 (a)	성향 <C=4> (b)	성향 <C=5> (c)	차이	
						(a-b)	(a-c)
2011	1차	사교육비율	0.226	0.223	0.223	0.003	0.003
		사교육시간	37.985	38.51	38.502	-0.525	-0.517
		사교육비	648.634	629.728	628.350	18.906	20.284
		시간당사교육비	1.314	1.262	1.263	0.052	0.051
	2차	사교육비율	0.237	0.236	0.236	0.001	0.001
		사교육시간	37.325	37.728	37.719	-0.403	-0.394
		사교육비	693.975	685.047	684.021	8.928	9.954
		시간당사교육비	1.344	1.305	1.300	0.039	0.044
2012	1차	사교육비율	0.230	0.234	0.234	-0.004	-0.004
		사교육시간	35.900	35.933	35.909	-0.033	-0.009
		사교육비	689.740	604.913	616.172	84.827	73.568
		시간당사교육비	1.965	1.89	1.894	0.075	0.071
	2차	사교육비율	0.240	0.241	0.241	-0.001	-0.001
		사교육시간	33.854	33.811	33.793	0.043	0.061
		사교육비	710.924	633.315	627.903	77.609	83.021
		시간당사교육비	1.543	1.489	1.482	0.054	0.061