

준거 집단의 사교육비가 개인의 사교육비 지출에 미치는 영향 : 준 실험 상황(Quasi experiment)을 활용하여

박민선*

현재 한국의 사교육비 지출은 전 세계적으로 그 유례를 찾아볼 수 없을 만큼 높다. 평균적으로 한 가구가 가구 소득의 10~20%를 사교육비에 지출하고 있다. 이러한 현상은 역사, 사회경제적 요인, 개인적 요인, 학교 교육의 요인 등 다양한 부문에서 그 원인을 찾을 수 있다. 하지만 사람들이 서로의 사교육비를 의식하며 사교육비가 높아지는 것 역시 분명해 보인다. 그러나 이러한 측면은 이전의 연구들에서 도외시 되어왔다.

본 연구에서는 준거집단을 주로 지리적인 이웃으로 한정하여 사람들이 이웃의 사교육비에 맞춰 자신의 사교육비를 조절하는지를 살펴보았다. 특히 내생적으로 형성된 집단을 분석대상으로 할 경우 일반적인 회귀분석으로는 내생 편의를 극복하기 어렵기 때문에 본고는 노동패널에서 발견한 준 실험상황(Quasi experiment)을 활용하였다. 구체적으로 이중차분법(Difference in Difference)와 White의 표준오차를 고려한 OLS, 그리고 PSM(Propensity Score Matching)을 주요 방법론으로 하였다.

실증 분석 결과는 사람들이 자신의 이웃의 사교육비 수준에 맞춰 사교육비 수준을 조정한다는 것을 알 수 있었다. 사교육비가 높은 지역으로 이사 갔을 때에는 이사하지 않은 경우보다 사교육비가 약 5만원 상승하며, 반대로 낮은 지역으로 이사를 갔을 때에는 약 5만원 하락하는 것을 확인할 수 있었다.

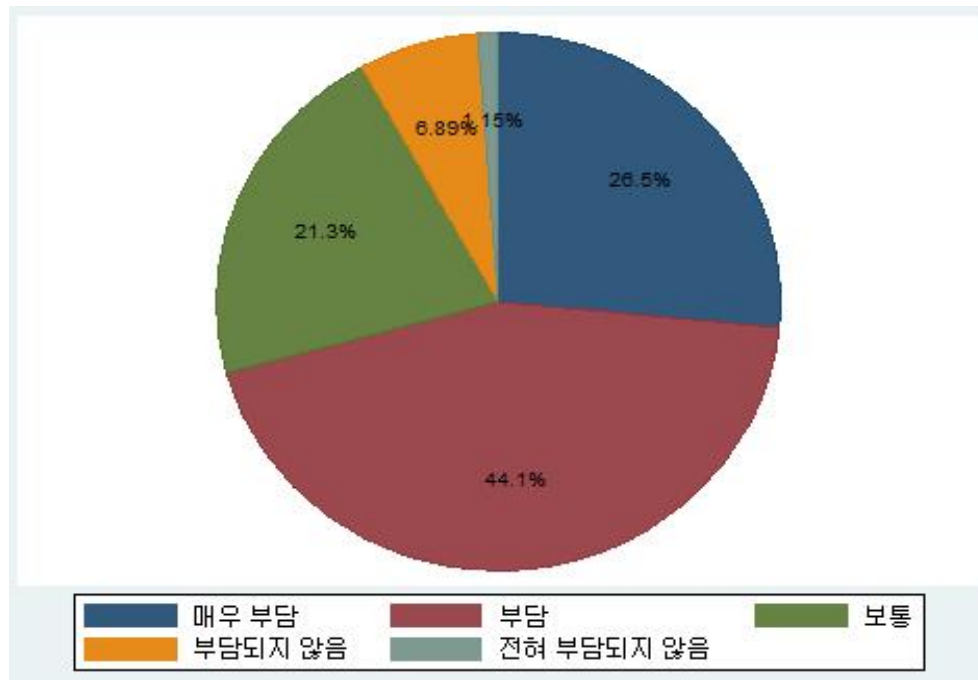
1. 들어가는 글

한국의 교육열이 세계에서 특기할 만큼 높다는 사실은 널리 알려져 있다. 특히 한국의 교육 투자는 공적인 부문이 아니라 사적인 부문에 집중되어 있다는 특징을 가지고 있다. 통계청이 2010년에 발표한 자료에 따르면 초, 중고생 중 73.6%가 사교육에 참여하고 있다고 응답하였다. 일인당 평균 사교육비는 24만 2천원이었으며 참여하는 학생들만을 대상으로 할 경우 평균적으로 32만 7천원의 사교육비가 지출되고 있었다. 일인당 GNI가 25,634,000원(한 달에 약 2백 1십만 원)이었음을 고려해 본다면 아이가 한 명 있는 한국의 대표가구는 소득의 10% 이상을, 아이가 두 명 있는 한국의 대표가구는 소득의 20% 이상을 과외, 학원, 인터넷 강의 등 사교육비에 사용하고 있었던 것이다. 위의 수치들에서도 짐작할 수 있지만 실제로 노동패널의 설문조사에 참여한 가구 중 다수의 가구

* 서울대학교 경제학부 석사과정, minsun813@snu.ac.kr

가 사교육비 지출이 부담스럽다고 대답하였다. [그림 1]은 노동패널의 2010년 조사를 활용하여 사교육비 지출이 가구에 어느 정도 부담이 되었는지를 도식화 한 것이다. 해당 항목에 응답한 1,828 가구 중 70.6%에 해당하는 가구가 사교육비가 부담된다고 대답하였으며 약 8%에 해당하는 가구만이 사교육비 지출이 부담되지 않는다고 응답하였다.

[그림 1] 사교육비 부담 정도



자료: 한국노동패널 13차 조사

더욱이 이러한 부담감은 가구의 소득분위와는 관계없이 모든 소득 분위에서 공통적으로 나타났다. <표 1>은 사교육비 부담 정도를 소득분위로 세분하여 보여준다. 모든 소득 분위에서 70% 정도의 가구가 현재 수준의 사교육비 지출이 가계에 부담이 된다고 응답하였고 오히려 소득분위가 낮은 가구에서 사교육비가 부담되지 않는다는 응답의 비율이 높은 것을 확인 할 수 있다. 이는 가구의 소득분위가 높아짐에 따라 사교육비 역시 급격하게 증가하기 때문인 것으로 보인다.

이처럼 다수의 가구가 부담감을 느끼면서도 소득의 큰 부분을 사교육비에 지출하는 현상은 다양한 관점에서 설명될 수 있겠지만 각 가구가 서로의 사교육비에 영향을 주며 전반적으로 사교육비가 높아지는 현상이 존재하는 것은 자명해 보인다. 다시 말해 교육이 위치재(positional good)¹⁾의 성격을 지닌다는 것이다. 하지만 이러한 관점에서 사교육비 지출을 설명하려는 시도는 이전의 연구들에서 도외시 되었거나 혹은 식별(identification)의 어려움으로 인해 다루어지지 않았다.

이 문제를 본격적으로 논의하기에 앞서 개인이 사교육비 지출을 결정할 때에 영향을 받게 되는

1) 자신의 절대적인 소비량뿐만 아니라 다른 사람들과 비교했을 시의 상대적인 소비에도 효용이 영향을 받는 재화를 위치재라고 한다.

“준거집단”이 누구인지에 대해 고민해볼 필요가 있다. 물론 가장 큰 범주에서는 한국에 거주하는 비슷한 나이의 자녀를 가진 모든 가구가 서로의 사교육비에 영향을 주고 있을 것이다. 이러한 설명은 특히 한국의 학생들이 정규 교육과정이 끝나면서 국가에서 전체적으로 실시하는 수학능력시험을 보게 되고 이것이 대학 입학에 크게 좌우하고 있다는 점에서 타당성이 있다.

<표 1> 소득 분위에 따른 사교육비 부담 정도

소득분위	사교육비 부담 정도					총합
	매우 부담	부담	보통	부담되지 않음	전혀 부담되지 않음	
상위 20%	25.69%	40.12%	28.14%	5.48%	0.58%	100%
20~40%	30.65%	46.33%	17.51%	4.66%	0.85%	100%
40~60%	28.97%	47.35%	16.18%	5.88%	1.62%	100%
60~80%	29.85%	41.34%	19.35%	8.23%	1.20%	100%
하위 20%	28.32%	41.99%	20.31%	7.23%	2.15%	100%

자료: 한국노동패널 13차 조사

하지만 Andrew E. Clark, Paul Frijters and Michael A. Shields(2008)에서 언급되었듯이 사람들은 보다 작은 “준거 집단”을 가지고 있다. 위의 연구에서는 준거집단을 이웃, 같은 구(district)에 사는 사람들, 직장 동료, 나라 전체 심지어 과거의 자기 자신에 이르기까지 다양하게 설정하여 개인의 상대적인 소득 비교를 통해 느끼게 되는 주관적 효용에 대해 연구하였다.

다양한 가능성들 중에 본 연구는 한국에서는 “교육열”이 주로 지역적 특성으로 환원된다는 점에 착안하여 지리적 이웃을 주요한 준거집단으로 설정하였다. 일례로 강남구, 대치동과 같은 지역은 한국 교육열의 상징과 같이 거론된다. 또한 통계청이 발표한 자료에 따르면 천 명 당 학원 수가 같은 서울 내에서도 강남구는 3.98명이었던 반면 강북구는 0.97에 불과하였으며 다른 도시인 대전에서도 유성구는 2.58개였던 반면 중구는 0.28개였다.

이러한 가능성에 비추어 본 연구에서는 개인의 사교육비가 이웃의 사교육비에 의해 상향 조정되거나 하향 조정되는지를 분석하였다. 하지만 이와 같이 집단이 각 표본의 내생적인 결정에 의해 구성되는 상황에서 일반적인 계량분석방법만으로 개인의 사교육비와 집단의 사교육비 간의 상관관계를 살펴보는 것은 그 내생성을 극복할 수 없다는 한계를 가지고 있다. Manski(1993)에 의해 지적되었듯이 연구자는 내생적으로 형성된 그룹에서 그룹의 영향과 그룹 내의 선천적인 기질의 동질성을 구분할 수 없기 때문이다. 예를 들어 강남구에 사는 가구들이 서로의 영향을 주고받으며 사교육비를 높인 것인지, 본래 교육에 대한 관심이나 견해가 유사하기 때문에 비슷하게 높은 수준의 사교육비를 지출하는 것인지를 알 수 없다는 것이다. 따라서 집단의 사교육비 평균과 개인의 사교육비 간의 관계를 살펴보는 등의 실증분석 방법은 준거집단의 영향을 “암시”할 수 있을 뿐이다.

이를 극복하기 위한 가장 이상적인 방식은 무작위로 샘플을 선정하여 일부는 사교육비를 많이 받는 지역으로, 일부는 사교육비를 적게 받는 지역으로 이주를 시키는 것이다. 하지만 이러한 방식은

사고 실험 상에서나 가능할 뿐 현실화될 수 없다.

따라서 본고는 데이터 내에서 이상적인 실험 상황과 유사한 준실험상황(quasi experiment)을 포착하여 연구를 진행하였다. 노동패널(KLIPS)의 1~17차 데이터 중 11차 이후의 조사에서는 지난 조사 이후 이사를 했는지 여부에 더해 이사한 이유가 무엇이었는지를 확인할 수 있다. 또한 “자녀 교육 때문에”가 다양한 선택지 중 하나였다. 만약 자녀교육문제 이외의 문제로 이사한 표본들이 외생적으로 형성된 준거집단을 가지게 된 것으로 볼 수 있다면 Manski(1993)의 비판을 피하면서 개인의 사교육비 지출에 집단의 사교육비 지출이 미친 영향을 살펴볼 수 있을 것이다.

본 연구는 다음과 같이 구성되어 있다. 2장에서는 선행 연구를 살펴볼 것이다. 우선 한국에서 사교육비 지출과 관련해 선행된 일군의 연구들을 살펴본 후 두 번째, 세 번째로 준거 집단 내의 소득과 소비지출에 관한 연구들과 학생들 사이에서의 동료 효과(peer effect)에 관한 연구들을 순서대로 살펴볼 것이다. 3장에서는 데이터에 대해 간략하게 소개하고 본고에서 활용한 준실험상황(quasi experiment)을 상술할 것이다. 마지막으로 4장에는 실증 분석 결과를 정리하였다.

II. 선행연구

한국에서 선행된 사교육비에 관한 연구들은 주로 어떤 인구학적 변수와 학교 요인 변수가 사교육비 지출과 밀접한 연관이 있는지를 분석하는데 그 중점이 있었다. 이혜정, 송종우(2014)는 어떤 인구학적 변인이 초, 중, 고 학생들의 사교육비 지출에 영향을 주는지 분석하여 기존의 연구에서와 마찬가지로 가구의 소득 및 부모의 학력이 사교육비 지출과 연관이 있음을 보였다.

박현정(2008)은 KELS(Korea Education Longitudinal Study) 자료를 활용하여 학교 단위의 변인 중 어떠한 것들이 학생의 사교육비 참여 및 사교육비 지출을 높이는지 연구하였다. 위계 선형모형을 활용하여 분석한 결과 위의 연구는 교사의 연구 시간과 학생들이 인식하는 교사의 열정이 학생들의 사교육비 참여율 및 사교육비 지출과 유의미한 부(-)적상관을 가지고 있음을 보였다. 김희삼(2009)도 마찬가지로 KELS를 활용하여 교사의 교육 수준이 높고 교과교실제가 실시되는 학교가 학생들의 사교육비가 낮다는 결론을 도출하였다.

사교육비 지출을 추동하는 심리적 동인에 관한 연구는 임성택과 여성민(2013)가 유일하다. 이 연구는 강원도에서 774명의 학생과 그들의 부모를 설문조사하였고 사교육비 지출의 동인을 내적 동기, 외적동기, 위탁동기, 경쟁 동기의 네 가지로 나누어 내적 동기, 외적 동기 및 경쟁 동기가 강한 부모일수록 아이의 사교육비에 높은 비용을 지출하고 있었음을 보였다. 이러한 동기들 중에 경쟁 동기가 본 연구에서 보고자 하는 준거집단에 의한 사교육비 지출과 큰 유사점이 있다고 하겠다.

한편 상대적 소비와 관련된 연구는 다양한 국가에서 진행되어왔다. R. H. Frank(1985)는 전통적인 효용함수에 위치재(positional good)의 소비 순위를 도입하여 x 가 위치재이고 y 는 일반적인 재화라 할 때 개인의 소비는 $U = U(x, y, R(x))$ 로 결정되는 것으로 보았다. 이 연구는 주로 개인 간 협조가 이루어진 상태일 때와 그렇지 않은 상태일 때의 결과를 이론적으로 보이는 것에 집중하고 있다.

앞서 설정한 효용함수가 모든 개인에게 공통적이기 때문에 각자 효용극대화 문제를 풀고 나면 각 개인의 순위 $R(x)$ 는 결국 자신의 소득에 의해 결정된다. 협조가 일어나는 상태에서는 $R(x) = R(\text{소득})$ 이 미리 합의되기 때문에 각 재화의 1원당 한계효용이 일치하는 지점에서 x 와 y 재의 소비를 결정하게 된다. 종합하면 비협조적인 상황과 협조적인 상황에서 결국 $R(x)$ 는 개인의 소득에 따라 결정된다는 점에서 동일한데 협조가 일어나는 경우 불필요한 노력을 들일 필요가 없어지는 것이므로 협조적인 상황에서 모든 개인의 효용수준이 높다. 이에 더해 비위치재 구매 비용이 전체 소득에서 차지하는 비중이 협조적인 상황에서 보다 높다.

Carlsson et al.(2008)은 실제 실험을 진행하여 위치재가 실제로 존재한다는 것을 다시 한 번 입증하였다. 스웨덴에서 무작위로 선정된 표본이 본 실험에 참여하였는데 참여 표본의 다수가 소비와 차 종류에서는 보다 상대적으로 높은 소비를 선호하였으며 차의 안전과 여가에 있어서는 절대적으로 높은 소비를 선호하였다. 이 실험의 결과는 상대적인 소비에 따라 선호가 결정되는 위치재가 존재한다는 담론을 뒷받침하였을 뿐만 아니라 어떤 재화는 다른 재화들 보다 더욱 위치재적인 성격을 지니고 있다는 것을 보여주었다.

상대적 소비와 관련된 계량분석 연구는 주로 자기 보고(self-report)에 기반을 둔 웰빙(self-reported subjective wellbeing)지수를 활용하여왔다. Clark et al.(2008)은 이러한 계통의 방대한 연구들을 정리하였다. 이러한 연구들은 준거집단을 다양한 차원(dimension)에서 규정하였는데 이웃, 동료, 친척, 주(state)단위 거주민, 국가 그리고 과거의 자신까지를 포함한다.

특히 John Knight과 Lina Song(2009)은 중국에서 국가적으로 실시하는 가구 설문 조사의 2002년 자료를 활용하여 개인 변수와 준거 집단의 변수가 개인의 주관적 행복에 미치는 영향을 분석하였다. 중국 시골 지역을 중점으로 한 본 연구에서는 각 개인이 자신의 소득이나 소비를 비교할 때에 마을 사람들을 살펴본다는 설문 결과에 착안하여 지리적 이웃을 준거집단으로 설정하였다. 분석 결과 마을 내에서의 상대적 소비와 과거 개인의 소득이 현재의 주관적인 행복을 결정하는데 주요한 요인이 됨을 확인하였다.

마지막으로 교육의 맥락에서 또래 효과(peer effect)를 살펴본 연구들은 주로 학생의 성취가 주변 학생의 성취와 어떻게 연관되어 있는지에 집중하였다. 본 연구는 학생/학부모가 주변의 준거집단에 의해 영향을 받는 상황을 탐구하고 있다는 점에서 이러한 연구들과 맥을 같이하지만 관심의 방점이 학생의 성취가 아닌 학생의 사교육비 지출이라는 점이 다르다고 할 수 있다.

이러한 연구들도 본 연구에서와 마찬가지로 Manski(1993)에 의해 밝혀졌듯이 내생적으로 형성된 그룹에서 위와 같은 분석을 하는 경우 준거 집단의 효과와 내생적인 동일성을 구분할 수 없다는 것을 고려하여 준실험 상황 혹은 RCT(Randomized Controlled Trial) 방법을 사용하여 집단이 외생적으로 주어지는 경우를 포착하였다.

강창희(2007)는 한국에서 학생들이 주로 지리적인 정보에 따라 무작위로(randomly) 학교에 배정된다는 사실을 활용하였다. 다시 말해 평준화 제도를 일종의 준실험상황으로 간주한 것이다. 지역과 학교의 고정효과(fixed effect)를 통제한 후 이 연구에서는 학생들의 시험 점수가 학급 평균 점수에 의해 유의하게 설명된다는 것을 밝혔다. 도구 변수로는 가구에서 첫째 아이인지 여부를 활용하였

고 이를 통해 준거집단의 시험점수가 높으면 개인의 점수도 높아지는 것을 보였다.

Duflo et al.(2008)은 케냐에서 RCT(Randomized Controlled Trial) 를 실행하였다. 총 61개의 학교가 무작위로 성적에 따라 반을 나누는 학교와 그렇지 않은 학교로 분류되었다. 학생들은 사전 성적에 따라 반 배정을 받았는데 사후 분석 결과 성적에 따라 반을 나누는 학교의 평균적인 성적 향상이 보다 높았다. 이러한 결과는 실험 1년 후에도 지속되었다. 이 연구에서는 학생들끼리 서로 주고받는 영향에 더해 학생들을 성적에 따라 분반한 경우 교사들이 학생들의 수준에 맞는 수업을 할 수 있기 때문에 이러한 결과가 발생한 것으로 보고 있다.

III. 데이터 및 준 실험(Quasi experiment)

1. 『한국노동패널』

본 연구에 사용된 자료는 한국노동연구원에서 1998년부터 조사한 『한국노동패널』의 11차~17차 자료이다. 본 자료는 농어촌 지역과 제주도를 제외한 지역에서 매년 대표 가구를 추출하여 각 가구의 사회경제적 사항에 대한 광범위한 설문결과를 수합하고 있다. 또한 각 가구에서 15세 이상인 가구원은 개인 설문의 대상이다.

총 17차의 자료 중 본 연구에서는 11~17차 자료를 활용하였는데 이는 10차 이전의 자료에서는 작년 설문 이후 이사하였는지의 여부는 포함되어 있으나 이사의 이유를 묻는 질문은 누락되어 있기 때문이다. 이외에도 노동패널에서는 자료를 가구 단위로 제공하고 있어 개별 자녀 단위로 데이터를 변형하여 사용하였고 11~17차 데이터를 통합(pooling)하여 사용하였음을 밝힌다. 이 결과 총 표본크기는 19,266이다.

2. 준 실험(Quasi experiment)

앞서 언급하였듯이 내생적으로 형성된 그룹을 대상으로 분석을 진행할 경우 준거 집단의 영향과 본래 형성되어 있던 동질성을 구분할 수 없다는 문제가 있다. 따라서 본고에서는 외생적으로 형성된 준거집단을 찾는데 중점을 두었으며 노동패널의 다음 두 가지 질문을 활용하여 준 실험 상황을 포착하였다.

<표 2>에는 질문의 번호와 내용 그리고 선지가 제시되어 있다. 두 질문에 대한 응답을 기반으로 표본들은 <표 3>에서와 같이 분류되었다. 먼저, 지난 조사 이후 이사를 하였는지 여부에 따라 집단을 크게 양분하였다. 이 때 이사는 다른 지역으로 옮겨간 것으로 한정하여 동일한 지역구 내에서 이사한 경우는 이사하지 않은 것으로 간주하였다. 왜냐하면 본 연구에서 준거집단을 지역구의 이웃으로 보고 있기 때문에 동일한 지역 내에서 이사한 경우 준거집단이 변한 것으로 볼 수 없기 때문이다. 예를 들어 서울시 서초구에서 서울시 서초구로 이사한 경우 해당 표본은 이사하지 않은

것으로 분류하였다.

표본을 세분하기 이전에 각 지역의 평균 사교육비 소비와 해당 년도 전국의 사교육비 평균의 차이를 계산하였다. 이후에 이사한 표본들은 다시 세 그룹으로 세분하였다. 그룹 H에는 사교육비가 상대적으로 높은 지역으로 이사한 표본들이 포함되었다. 예를 들어 서울시 관악구에서 서울시 서초구로 이사한 경우 두 지역의 사교육비 평균은 모두 전국 평균보다 높지만 서울시 서초구가 보다 더 높은 정도의 사교육비 평균 수준을 가지고 있었기 때문에 해당 표본은 그룹 H에 포함된다. 그룹 L은 반대로 사교육비 평균이 보다 낮은 지역으로 이사한 표본을 포함한다. Group H와 Group L은 공통적으로 준거집단이 외생적으로 주어진 경우를 나타낸다고 할 수 있는데 왜냐하면 이들은 자녀 교육 이외의 이유로 이사한 경우만을 포함하고 있기 때문이다. 마지막으로 이사를 하였지만 이사한 주요 이유가 자녀교육 때문이라고 대답한 경우에는 Group E로 분류하였다. 앞으로의 분석에서 Group E는 비교집단으로만 언급될 것이며 회귀 분석에서는 제외되었다.

<표 2> 이사 여부 및 이사 이유 설문 내용

질문번호	질문내용	선지
H**1401	지난 조사 이후 이사여부	(1) 있다 (2) 없다
H**1402	현재 집으로 이사 오게 된 동기	(1) 취업/ 창업 (2) 동일 직장 내 근무지 이동 (3) 현재 다니는 직장/학교의 통근 편의를 위해 (4) 자녀 교육 때문에 (5) 집값이 오를 것이라 여겨져서 (6) 평수를 넓히거나 줄이려고 (7) 집값이나 전세금이 저렴해서 (8) 보육 혹은 부모봉양 때문에 가족/친지와 가까이 살기 위해 (9) 환경 및 건강 때문에 (10) 인근의 편의시설 때문에(마트, 공원, 복지시설 등) (11) 신혼집 마련 (12) 내 집 마련 (13) 어려서부터 살아왔다 (14) 기타 (15) 계약만료 / 철거, 재개발 (16) 독립하고 싶어서 (17) 별거, 이혼, 가구불화 (18) 경제적 상황 악화

자료: 한국 노동패널 11~17차 조사

각 그룹에 속하는 샘플의 수는 <표 3>에 제시되어 있다. 그룹 H와 그룹 L은 각각 전체 표본의 1.34, 1.93%를 차지하고 있고 대부분의 표본은 이사하지 않은 그룹 C로 분류되었다. 그룹 간 표본

크기의 비대칭이 심한 상황인데 이는 연구자가 실제로 실험을 진행한 것이 아니기 때문에 생기는 한계로 볼 수 있다.

<표 3> 그룹 분류 및 해당 표본 수

		그룹 H	293(1.34%)
이사함	797(3.64%)	그룹 L	426(1.94%)
		그룹 E	78(0.36%)
이사 안함		그룹 C	21,141(96.37%)
총합			21,938(100%)

자료: 한국 노동패널 11~17차 조사

()안은 퍼센트

IV. 분석 결과

본 장에서는 세 가지의 주요한 결과를 제시한다. 첫째로 각 그룹의 사교육비의 차이의 차이가 (Difference in Difference)가 유효한지를 살펴볼 것이다. 만약 그룹 H가 그룹 C와 비교하였을 때 통계적으로 유의한 정도의 추가적인 사교육비 증가분을 보였다면 그룹 H가 사교육비를 상대적으로 많이 받는 지역으로 이사를 가면서 준거집단의 사교육비 수준에 따라 사교육비를 상향 조정할 것으로 볼 수 있을 것이다. 반대로 그룹 L의 사교육비 변화는 그룹 C의 것보다 작을 것으로 예상할 수 있다.

그러나 단순한 D_{inD} 는 다음의 중요한 두 가지 가정 위에서 정당화될 수 있다. 첫째로는 실제로 이사가 외생적으로 결정되었어야 한다는 것이며 다른 하나는 사람들이 실제로 지리적인 이웃을 준거집단으로 간주해야 한다는 것이다. 이러한 가정들에 의문을 제기해 볼 수 있기 때문에 다음 분석들은 위와 같은 가정을 완화하며 준거집단의 사교육비에 개인의 사교육비가 어떠한 영향을 받는지를 살펴보고자 한다.

두 번째 분석은 이사가 외생적으로 결정되었다는 가정을 완화해보기 위해 사교육비가 높은/낮은 지역으로 이사하는 것에 영향을 미칠만한 변수들을 선정하여 PSM(Propensity Score Matching) 방법으로 이를 통제된 후 사교육비 차이의 차이를 분석해보고자 한다.

마지막으로 선행연구들에서처럼 준거집단이 다양할 수 있기 때문에 다른 사회 경제적 요인으로 구성된 집단의 사교육비 평균과 개인의 사교육비의 관계를 분석하고자 한다. 본 분석은 내생적으로 형성된 그룹을 이용한다는 점에서 Manski(1993)의 비판에서 벗어날 수 없으며 따라서 분석 결과 각 계수의 값을 절대적으로 해석하는 것 보다는 여러 가지 기준 집단 간의 상대적 영향을 살펴보는 것에 중점을 두도록 하겠다.

실제적인 분석을 이전에 <표 4>에는 참여율과 평균사교육비, 사교육비 상위/하위 5개 지역을 제

시해 놓았다. 매년 통계청에서 발표하는 사교육 참여율과 평균 사교육비를 함께 제시하여 노동패널의 자료가 사교육비 관련 분석에 적합한지 살펴보고자 하였다. <표 4>에서 볼 수 있는 바와 같이 두 데이터 사이의 수치는 크게 차이 나지 않는다는 것을 알 수 있다. 또한 노동패널의 자료를 기반으로 매년 사교육비 상위/하위 지역을 나누어본 결과 경남 하동군, 전남 해남군 등 해당 표본의 수가 매우 작아 한 두 학생의 사교육비에 지역 평균값이 결정되어 사교육비가 높게 나오는 지역도 있었으나 대부분의 경우에는 통념에 부합하는 것을 확인할 수 있었다.

<표 4> 노동패널, 통계청의 연도별 사교육비 평균 및 사교육비 높은/낮은 지역

	참여율 (%)	평균 사교육비(만원)		사교육비 상위 5개 지역**	사교육비 하위 5개 지역**
		전체 대상	참여 학생 대상		
2008	노동패널	70.0	21.4	30.6	대구 중구, 인천 강화군, 경기도 양평군, 전북 무주군, 충남 홍성군
	통계청	75.1	23.3	31.0	서울시 송파구, 양천구, 용산구, 종로구, 성남시 분당구
2009	노동패널	70.2	20.9	29.8	부산 동구, 인천 강화군, 경기도 과천시, 성남시 분당구, 용인시 수지구
	통계청	75.0	24.2	32.3	서울시 서초구, 부산 수영구, 성남시 분당구, 안양시 동안구, 경남 하동군
2010	노동패널	71.7	22.1	30.9	부산 중구, 인천 강화군, 충남 서천군, 홍성군, 전북 김제시
	통계청	73.6	24.0	32.7	서울시 서초구, 부산 금정구, 수원구, 성남시 분당구, 의왕시
2011	노동패널	69.2	22.3	32.2	강원 평창군, 충북 보은군, 진천군, 충남 계룡시, 전북 김제시
	통계청	71.7	24.0	33.5	서울시 서초구, 동작구, 서초구, 양천구, 종로구
2012	노동패널	72.2	22.1	30.6	부산 중구, 대구 중구, 인천 강화군, 강원도 철원군, 평창군
	통계청	69.4	23.6	34.0	서울시 서초구, 송파구, 양천구, 성남시 분당구, 전남 해남군
2013	노동패널	73.6	24.8	33.7	인천 강화군, 강원 평창군, 충북 음성군, 충북 진천군, 충북 옥천군
	통계청	68.8	23.9	34.7	서울시 마포구, 서초구, 노원구, 성남시 분당구, 고양시 일산 동구
2014	노동패널	75.5	25.9	34.3	경기도 양평군, 강원도 평창군, 충남 서천군, 전남 영암군, 전북 김천시
	통계청	68.6	24.2	33.5	

자료: 한국 노동패널 11~17차 조사, 통계청

**노동패널 자료만을 기준으로 함

1. 이중차분법(Difference in Difference)

<표 5>와 [그림 2]는 동일한 분석결과를 다르게 표현한 것이다. 예상한 대로 그룹 H의 경우 평균 사교육비가 약 8만원 증가하였다. 반면에 그룹 L은 사교육비가 한 달에 약 이만 원 감소하였다. 그룹 C와 E는 평균 사교육비 지출이 비교적 안정적이다. [그림 2]를 보면 사교육비 증감을 그래프의

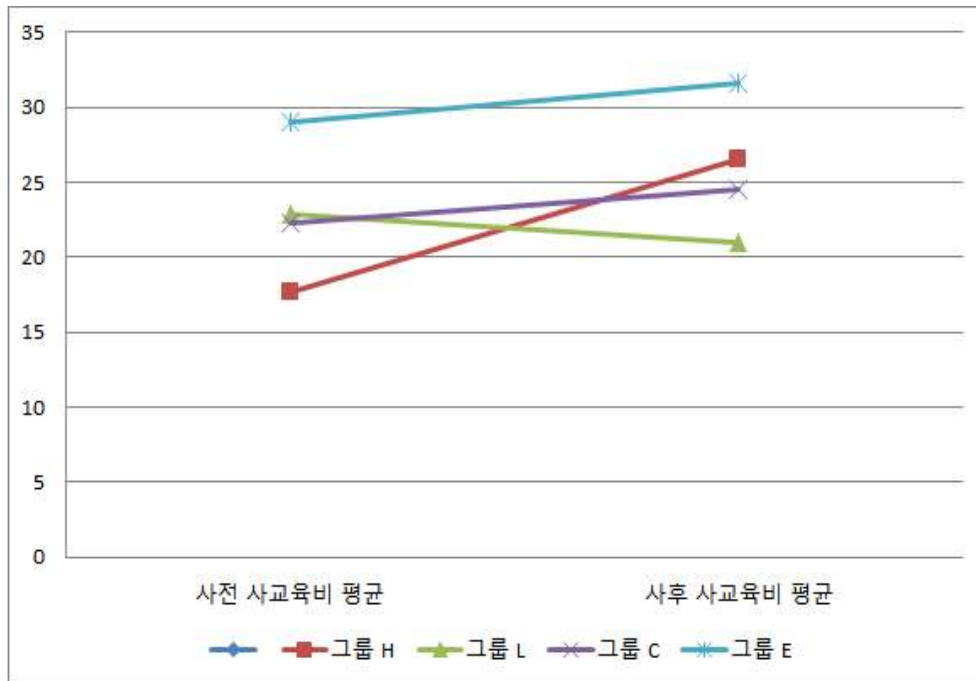
기울기로 보다 명료하게 확인할 수 있다. 그룹 H는 기울기(사교육비 증가)가 가장 가파른 반면 그룹 L은 기울기가 음의 방향으로(사교육비 감소) 가장 크다.

<표 5> 그룹별 단순 이중차분 결과

그룹	사전 사교육비 평균	사후 사교육비 평균	차이	(통제집단 대비) 차이의 차이	표본 수
그룹 H	17.66	26.53	8.87	6.6	293
그룹 L	22.88	20.98	-1.9	-4.17	426
그룹 C	22.23	24.50	2.27	-	19559
그룹 E	29.05	31.63	2.58	0.31	60

주 1) 모든 사교육비의 단위는 만원

[그림 2] 그룹별 단순 이중차분 결과



특기할만한 점은 그룹 E의 사전 사교육비 지출이 다른 그룹에 비해 상당히 높다는 것이다. 이는 해당 그룹의 교육열을 보여주는 것이라고 할 수 있다. 따라서 해당 그룹이 이사하게 된 이유는 보다 질이 좋은 사교육을 소비하기 위해서였을 것임을 예상할 수 있다. “자녀 교육 때문에”이사 했다고 대답한 집단이기 때문에 이러한 해석은 더욱 신빙성이 있다. 또한 그룹 H와 그룹 L의 사전 사교육비 평균은 그룹 E에 비해 낮은 수준인데 이것은 그룹 H, L이 그룹 E에 비해 상대적으로 낮은 수준의 사전 교육열을 가지고 있었다는 추측을 가능하게 한다. 만약 그룹 H가 상대적으로 사교육비가 높은 지역으로 이사를 간 것이 자녀 교육열에 의해 추동된 부분이 크다면 그룹 E와 마찬가지로 사전 사교육비 평균이 높았을 것이다. 이러한 점에서 본 연구에서 그룹 E는 비교 그룹이 역

할을 하고 있다.

이러한 차이의 차이가 사교육비 증감에 영향을 미칠만한 다른 변수들을 통제하고도 유의한지 살펴 보기 위해 <표 6>에 제시된 회귀식을 활용해 분석을 실시하였다.

<표 6> 이중 차분 회귀식

식	종속변수	회귀식
(1)	사교육비 변화	$\Delta e_i = \beta D_i + \Delta u_i$
(2)		$\Delta e_i = \beta D_i + (\Delta X_i')\gamma + \Delta u_i$
(3)	사교육 개수 변화	$\Delta n_i = \beta D_i + (\Delta X_i')\gamma + \Delta u_i$
(4)		$\Delta n_i = \beta D_i + (\Delta X_i')\gamma + \Delta u_i$

Δe_i 는 개인의 사교육비 지출 변화분
 Δn_i 는 개인의 사교육비 개수 변화분
 $D_i = (D_h, D_l)$ 로 첫 번째 요소는 그룹 H에 속하면 1, 그룹 L 또는 C에 속하면 0이며 두 번째 요소는 그룹 L에 속하면 1, 그룹 H 또는 C에 속하면 0인 더미 변수
 $\Delta X_i'$ 는 학교 급 변화 여부 더미, 엄마와 동거 여부 변화 더미, 소득 변화분, 주변 사교육 기관의 개수 변화를 포함하는 벡터

종속변수는 사교육비 변화와 사교육 개수 변화 두 가지로 하였는데 사교육 개수 변화를 종속변수로 하여 분석을 실시한 이유는 사교육의 실제적인 양은 전혀 늘어나지 않았지만 이주 지역의 사교육비가 높거나 낮아서, 다시 말해 물가의 효과로 인해 사교육비에 증감이 생길 수 있기 때문이다. 따라서 식 (3)–(4)는 물가 효과를 배제하고 양적으로도 사교육 증감에 차이가 있었는지를 확인할 수 있도록 해준다.

위의 식에서는 통상적으로 고려되는 부모의 학력, 아이의 성별, 사전 소득 등은 사전-사후에 변화가 없는 변수이기 때문에 분석에 포함되지 않았다. 한편 기존의 선행 연구에서는 포함하지 않았던 거주 지역의 1000명당 사교육 기관의 수를 포함하였는데 이는 사교육 기관에의 접근성이 높아질 경우 주변의 사교육비 지출 수준과 상관없이 사교육을 더 많이 받게 될 수 있기 때문이다.

또한 위의 식 (1)–(4)는 눈에 보이지 않는 변수(unobservable variables)들이 사전 사후에 변화하지 않는다고 가정하면 이를 자연스럽게 통제하고 있다. 왜냐하면 그러한 변수들의 차이(differential)는 0이 되기 때문이다.²⁾

<표 7>에는 회귀분석 결과가 요약되어 있다. 분석 결과 식 (1)–(4)에서 모두 그룹 H, 그룹 D 더미의 계수가 통계적으로 유의미하다는 것을 알 수 있다. 특히 식 (3)–(4)에서는 그룹 C에 비해 그룹 H인 경우 사교육 개수가 유의하게 증가하고 그룹 L인 경우에는 사교육 개수가 유의하게 감소하여 사교육의 실질적인 양이 이주로 인한 준거 집단의 사교육 수준에 따라 조정된다는 것을 확인할 수 있다. 다시 말해 단순히 물가 수준의 변화로 인해 사교육비 변화가 야기된 것은 아니라는

2) 상기 식 (1)은 일종의 패널 분석이라고 할 수 있다. 일반적인 패널 분석은 준거집단의 사교육비 변동이 충분히 크지 않아 본 연구에서 사용할 수 없었다.

것이다.

<표 7> 이중 차분 회귀 결과

종속	사교육비 변화분		사교육 개수 변화분	
	(1)	(2)	(3)	(4)
그룹 H 더미	6.60*** (4.16)	5.78*** (3.58)	0.17** (2.17)	0.19*** (2.72)
그룹 L 더미	-4.15*** (-3.74)	-4.68*** (-4.13)	-0.21*** (-3.49)	-0.33*** (-5.00)
초등학교 진학		8.53*** (14.38)		0.41*** (18.18)
초등학교 고학년 진학		6.95*** (11.07)		0.21*** (8.68)
중학교 진학		2.10*** (3.43)		0.09*** (3.55)
고등학교 진학		-4.41*** (-6.37)		0.30*** (7.59)
엄마 동거여부 변화		7.15** (2.06)		0.32 (1.94)
소득 변화분		0.003*** (3.40)		0.00*** (3.74)
사교육 기관 수 변화분		0.97 (1.61)		0.02 (0.70)
표본크기	20278	20061	20288	20071
R^2	0.0020	0.0190	0.0010	0.0135

* p<0.10, ** p<0.05, *** p<0.01, 각 계수의 t값은 괄호 안에 표시.

*모든 식은 상수항을 포함하고 있으나 간결함을 위해 제시하지 않았다.

분석 결과에 따르면 다양한 통제변인을 통제했을 때 그룹 C에 속하는 경우를 기준으로 하여 그룹 H에 속하는 경우 평균적으로 사교육비가 한 달에 5만 7천원 더 상승하였으며 그룹 L에 속하는 경우 4만 6천원 감소하는 것을 확인할 수 있다. 본 분석에서 그룹 E는 그 수가 적어 제외하였다.

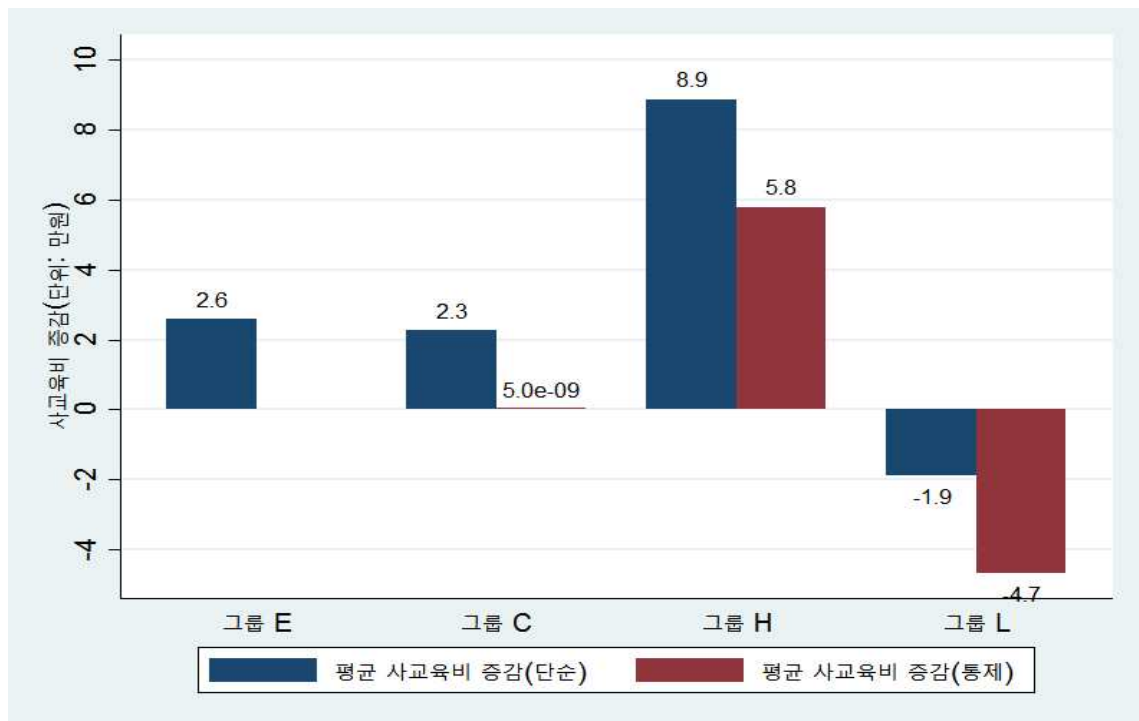
기타 통제변수의 계수를 살펴보면 자녀가 초등학교에 진학하거나, 초등학교 고학년이 되거나, 중학교에 진학하는 경우 사교육비가 유의하게 상승하는 것을 확인할 수 있었으나 고등학교에 진학하는 경우 사교육비가 유의하게 감소하는 것을 확인할 수 있다. 이는 통계청이 발표한 자료에서 확인할 수 있는 1인당 평균 사교육비가 자녀의 학교 급이 상승함에 따라 점차 높아져 중학교 급에서 최고 점을 기록하고 오히려 고등학생이 되면 감소하는 패턴과 일치한다.³⁾ 또한 자녀가 엄마와 함께 살

3) 2013년 발표 자료 기준 학교급 당 학생 1인당 사교육비 평균은 다음과 같다.

게 되는 경우 사교육비가 유의미하게 증가하였으며 가구의 수입이 늘어나는 경우에도 사교육비가 증가하였다. 거주 지역의 사교육 기관의 수 증감은 사교육비 증감과 통계적으로 유의미한 관계를 가지고 있지 않았다.

[그림 3]에 본 절의 분석 결과를 요약하였다. 진한 색 막대는 아무것도 통제하지 않은 상태에서 그룹별 사교육비 증감을 보여주며 연한 색 막대는 위의 회귀식에서 사용된 통제변수를 통제한 후 그룹별 사교육비 증감을 보여준다. 통제변수를 통제한 경우 그룹 H의 사교육비 증가는 5만 8천원이었으며 그룹 L의 사교육비 감소는 4만 7천원 이었고, 이와 그룹 C의 차이는 앞서 <표 7>에서 본 것과 같이 통계적으로 유의미하다.

[그림 3] 그룹별 사교육비 증감



2. PSM(Propensity Score Matching)

1절의 분석은 암묵적으로 이주가 RCT(Randomized Controlled Trial)에서처럼 외생적으로 결정되었다는 가정을 하고 있다. 이 절에서는 이러한 가정에 의문을 던지며 모델을 확장해보고자 한다.

학교급	초등학교	중학교	고등학교
1인당 월평균 사교육비(만원)	23.2	26.7	22.3

자료: 한국 노동패널 11~17차 조사, 통계청

이처럼 중학교 급에서 1인당 월평균 사교육비가 가장 높은 현상은 거의 매년 발견되었다.

[그림 2]에서 확인하였듯이 사전 사교육비 평균이 그룹 H과 L이 특기할만한 교육열을 가지고 있는 않았음을 보여주고는 있지만 이 절에서는 이주를 내생적인 결정으로 간주하여 교육열과 관련된 다른 변수들과 관련이 있는지를 살펴볼 것이다.

$$D_i = Z_i \rho + v_i \quad \text{식(1)}$$

$$\Delta e_i = \beta D_i + (\Delta X_i') \gamma + \Delta u_i \quad \text{식(2)}$$

Z_i : 이사를 결정하는 다양한 교육열에 관한 변수

다시 말해 각 개인은 사실상 위의 식 (1)을 통해 이사 여부를 결정하고 식 (2)를 통해 이사 후 사교육비 증감을 결정하게 된다는 것이다. 식 (2)는 1 절에서 사교육비 증감을 분석한 식과 동일하다. 본 절에서는 식 (1)을 도입하여 이사 여부를 결정하는 과정을 포함하여 분석을 확장하게 된다.

$$D_i = Z_i \rho + (D_i - Z_i \rho) \quad \text{식(1.2)}$$

$$\Delta e_i = Z_i \rho \beta + (D_i - Z_i \rho) \beta + \Delta X_i' \gamma + \Delta u_i \quad \text{식(3)}$$

식 (3)의 우변의 첫 번째 항: 개인적 특성

두 번째 항: 그룹 효과

식 (1)은 식 (1.2)와 같이 나타낼 수 있다. 식 (1.2)를 식(2)에 대입하면 식 (3)을 쉽게 도출할 수 있다. 식 (3)의 우변을 보면 D_i 와 D_i 의 효과가 두 가지 효과로 세분됨을 알 수 있다. 첫 번째 항은 사교육비 증감에 영향을 미치는 개인적 요인을 나타내며 두 번째 항은 개인적 특성으로는 설명할 수 없는 그룹의 효과를 나타낸다. 따라서 Z_i 가 1장에서 포함되지 않았음을 고려할 때, 위의 절에서 β 는 개인적 특성과 집단의 영향이 섞인 D_i 의 효과를 보여준다고 할 수 있다.

본 절에서는 이를 다루기 위해 두 가지 분석방법을 활용하였다. 첫째로 보다 다양한 변수를 통제하여 다중 회귀분석을 재 실시하였고 둘째로 PSM(Propensity Score Matching)을 실시하였다.

$$\Delta e_i = \beta D_i + Z_i \delta + \Delta X_i' \gamma + \Delta u_i \quad \text{식(4)}$$

$$D_i = Z_i \rho + v_i \quad \text{식(1)}$$

$$\Delta e_i = \beta \tilde{D}_i + (\Delta X_i') \gamma + \Delta u_i \quad \text{식(5)}$$

$$\tilde{D}_i \equiv D_i - Z_i \hat{\rho}$$

식 (4)는 식 (2)를 확장한 중회귀분석 모형이다. 식 (4)에서 β 는 Z_i , 즉 개인의 특성으로 설명할 수 없는 D_i 가 사교육비 증감에 미치는 영향을 의미한다. 다시 말해 기준 집단의 효과를 포착한다고 할 수 있다.

한편 PSM는 사후적으로 무선 표집을 하는 과정이다. 우선 식 (1)을 로짓분석을 실시하여 분석한

후 그룹 H, L과 그룹 C의 표본들을 $Z_i\hat{\rho}$, 즉 propensity score에 기초하여 쌍을 지어준다. 매칭 후 식(2)를 활용하여 OLS분석을 하는 것은 D_i 가 아니라 Z_i 에 의해 설명되지 않는 부분인(Z_i 와 orthogonal한) \tilde{D}_i 를 대입한 식 (5)로 OLS분석을 하는 것과 동일하다. 이미 매칭 과정에서 그룹 간 Z_i 의 차이를 조정한 상태이기 때문이다.

확장된 중다회귀 모델과 PSM 결과에 앞서, 이사 여부를 결정하는 로짓 회귀분석 결과를 <표 8>에 제시하였다.

<표 8> 로짓 분석 결과

종속	그룹 H, C 포함	그룹 L, C 포함	그룹 H, L, C 포함
	그룹 H 여부	그룹 L 여부	이사 여부
자녀 성별	-0.093 (-0.76)	0.011 (0.11)	-0.034 (-0.44)
자녀 나이	-0.030 (-0.51)	-0.062 (-1.32)	-0.050 (-1.36)
가구 소득	-0.248** (-2.36)	-0.262*** (-3.24)	-0.256*** (-3.96)
모 최종학력 연한	0.175 (2.26)**	0.332*** (5.17)	0.269*** (5.38)
부 최종학력 연한	0.075 (1.08)	0.039 (0.68)	0.054 (1.21)
모의 나이	-0.035 (-1.65)	-0.012 (-0.71)	-0.022 (-1.60)
부의 나이	-0.054*** (-2.75)	-0.032** (-2.04)	-0.040*** (-3.24)
모 취업여부	-0.228* (-1.76)	-0.253** (-2.36)	-0.245*** (-2.94)
부 취업여부	0.415 (1.62)	-0.314 (-1.95)	-0.066 (-0.49)
모 동거여부	-0.991 (-1.33)	-0.598 (-0.95)	-0.778 (-1.60)
사교육기관 수	-0.549*** (-4.09)	-0.165 (-1.60)	-0.314*** (-3.78)
교육비를 위한 자금 관리	-0.521** (-2.37)	0.344** (2.58)	0.056 (0.49)
교육으로 인한 이사 경험 여부	0.909*** (5.80)	0.515*** (3.46)	0.692*** (6.31)
표본크기	19172	19301	19576
R^2	0.0582	0.0390	0.0472

* p<0.10, ** p<0.05, *** p<0.01, 각 계수의 t값은 괄호 안에 표시.

*모든 식은 상수항을 포함하고 있으나 간결함을 위해 제시하지 않았다.

주 1) 모든 변수는 사전 년도의 것을 사용하였다.

주 2) 교육비를 위한 자금관리는 교육을 위해 저축을 하거나 빚을 낸 경험이 있으면 1, 둘 다 경험이 없으면 0으로 된 더미 변수이다.

주 3) 교육으로 인한 이사 경험 여부는 교육 문제로 이사한 적이 있으면 1 없으면 0을 가진 더미 변수이다.

일반적으로 선행연구에서 부모의 교육열과 관련되어 있는 것으로 간주되는 통제 변수들이 위의 로짓 분석에 포함되었다. 위의 로짓 분석을 활용한 매칭의 궁극적인 목적은 가능한 한 집단 간 교육열을 유사하게 만드는 것이다. 이에 본 연구에서는 기존 연구에서 사용되던 부모 학력, 소득, 취업 여부에 더해 *교육비를 위한 자금관리*와 *교육으로 인한 이사경험 여부*의 두 가지 변수를 추가하였다. *교육비를 위한 자금관리* 변수는 자녀 교육을 위해 빚을 지거나 저축을 한 경험이 있으면 1 그렇지 않으면 0의 값을 갖는 더미 변수인데 총 2,047개의 표본이 1의 값을 가졌다. 이 중 대다수(1,744)는 교육을 위한 저축을 하고 있다고 대답하였으며 비교적 소수인 329 표본이 교육을 위해 빚을 지고 있다고 답하였다. 두 가지 모두를 경험한 표본은 19개에 불과하였다.

<표 8>의 분석 결과를 보면 *교육비를 위한 자금관리*가 D_h 와 유의한 부적 관계를 가지고 있는 것을 확인할 수 있다. 이는 매우 강력한 교육열을 가지고 있는 표본의 경우 이미 사교육비가 높은 지역에 거주하고 있거나 교육으로 인해 이사했다고 답한 그룹 E에 속해있었기 때문인 것으로 보인다. 일반적으로 가구의 교육열과 관련되어 있다고 보는 가구 소득이 D_h 를 유의하게 낮추는 것 역시 같은 맥락에서 해석할 수 있을 것이다.

본 절의 주요한 결과는 <표 9>에 제시되어 있다. PSM은 다시 ATET(Average Treatment Effect on the Treated)와 ATE(Average Treatment Effect) 두 가지로 나누어 두 방법 모두를 활용하였다. 간단하게 설명하자면 ATET는 이사를 간 집단, 즉 그룹 H와 그룹 L이 이사를 가지 않았을 때를 가상적으로 설정하여 이사를 갔기 때문에 변화한 사교육비가 얼마나 되는지를 살펴보는 것이다. 그룹 H와 그룹 L이 이사를 가지 않은 경우는 사고 실험 속에서 존재하는 것이기 때문에 그 대신 로짓 분석에서 Propensity score가 비슷했던 이사를 가지 않은 그룹 C의 표본으로 그 수치를 같음하게 된다. 따라서 총 표본의 수가 1344개로 이전의 분석에 비해 그 수가 적은 것을 확인할 수 있다. 개념상으로 그룹 H와 L의 표본에 각각 그룹 C의 표본에서 한 명 씩을 매칭하면 되기 때문이다.

한편 ATE는 전체 표본에서 이사로 인한 효과를 살펴보는 것으로 이사를 간 집단과 이사를 가지 않은 집단에게 각각 반대편 집단에서 Propensity score가 비슷한 표본을 한 명 씩 매칭한 후 그룹 간 사교육비 변화분의 차이가 통계적으로 유의한지 보는 방법이다. 모든 그룹의 표본에게 다시한 명씩 매칭이 되므로 총 표본 크기가 앞서의 분석에서보다 두 배가량 커진 것을 확인할 수 있다.

분석 결과 어떤 방법을 사용하더라도 이사를 하지 않은 집단과 비교했을 때, 사교육을 상대적으로 많이 받는 지역으로 이사를 간 집단(그룹 H)은 사교육비를 유의미하게 증가시키고 적게 받는 지역으로 이사를 간 집단(그룹 L)은 사교육비를 유의미하게 감소시켰다는 사실을 확인할 수 있었다.

다만 세부적으로는 약간의 차이가 있었는데, PSM을 사용한 결과 그룹 H의 사교육비 증가가 중다회귀분석에서 보다 컸고 그룹 L이 사교육비 감소는 오히려 중다회귀 분석에서 가장 뚜렷하게 나타났다. 보다 구체적으로 보면 사교육비가 보다 높은 지역으로 이사한 경우 중다회귀 분석에서는 한 달 평균 교육비가 5만 2천원 정도 상승한데에 반해 ATET에서는 6만 3천원, ATE에서는 7만 원원

이 상승한 것으로 나타났다. 사교육비가 낮은 지역으로 이사한 결과 중다 회귀 분석에서는 한 달 평균 사교육비가 5만 4천원 감소하였고 ATET에서는 4만 2천원, ATE에서는 4만 6천원이 감소한 것으로 나왔다.

<표 9> 중다 회귀 및 PSM 분석 결과

종속	중다 회귀분석	PSM	
		ATET	ATE
그룹 H 더미	5.26*** (3.18)	6.30*** (33.55)	7.12*** (4.02)
그룹 L 더미	-5.47*** (-4.77)	-4.27*** (-3.10)	-4.65*** (-4.21)
사교육 증감 관련 공변인(X'_i)	직접 통제	직접 통제	직접 통제
이사 관련 공변인(Z'_i)	직접 통제	매칭으로 통제	매칭으로 통제
표본크기	19516	1344	40464
R^2	0.0243	0.0573	0.0186

* p<0.10, ** p<0.05, *** p<0.01, 각 계수의 t값은 괄호 안에 표시.

*모든 식은 상수항을 포함하고 있으나 간결함을 위해 제시하지 않았다.

주1) 사교육 증감 관련 공변인은 1장의 분석에서도 통제되었던 초등학교 진학, 초등학교 고학년 진학, 중학교 진학, 고등학교 진학 더미와 소득 변화, 엄마 동거여부 변화, 사교육 기관 수 변화분을 포함한다. 모든 분석에서 통제하였으나 간결함을 위해 제시하지 않았다.

주 2) 이사 관련 공변인은 위의 <표 8>에서 로짓 분석을 위해 사용된 변수들을 말한다.

주 3) 교육으로 인한 이사 경험 여부는 교육 문제로 이사한 적이 있으면 1 없으면 0을 가진 더미 변수이다.

이러한 차이는 표본을 자녀의 나이에 따라 세분하여 보면 설명할 수 있다. 비록 그룹 C가 전체 표본에서 압도적으로 큰 비중을 차지하고 있었으나 나이로 세분한 그룹에서 그룹 H, L, C가 차지하는 비중이 미세하게 달랐다. 나이별 그룹 H, L, C의 분포는 <표 10>의 상단에 제시되어 있다. ATET는 처지 집단(treated group), 다시 말해 본 연구에서는 이사를 간 집단에서의 나이별 분포에 따라 나이 집단별 사교육비 차이에 가중치를 부여한다. 한편 ATE는 전체 표본에서의 나이별 분포에 따라 나이 집단별 사교육비 차이에 가중치를 준다. 마지막으로 중다회귀분석은 그룹 H, L, C에 속할 확률의 표준 편차가 가장 커지는 나이집단의(다시 말해 가장 균일하게 H, L, C가 분포되어 있는 나이집단에) 사교육비 차이에 가장 큰 가중치를 부여한다.

<표 10>의 하단에는 나이 그룹별로 H 그룹과 C 그룹의 평균 사교육비 차이, L 그룹과 C 그룹의 평균 사교육비를 계산하고 어떠한 방법에서 해당 나이그룹의 결과에 높은 가중치를 부여하는지를 명시하였다. 그룹 H의 효과는 초등학교 고학년과 중학생에게서 가장 극명하게 나타났으며 그룹 L의 효과는 초등학교 고학년과 고등학생에게서 가장 두드러졌다. ATET와 ATE는 상대적으로 고등학생 그룹보다 중학생 그룹에게 높은 가중치를 부여하므로 위의 <표 9>에서 그룹 H의 효과가 다중 회귀 분석보다 높았던 것이다. 반면에 다중 회귀분석은 고등학생 그룹에게 높은 가중치를 부여

하기 때문에 그룹 L의 계수의 값이 다른 분석방법에서 보다 컸다는 것을 확인할 수 있다.

<표 10> 나이 그룹 별 그룹C, H, L의 분포 및 그룹별 사교육비 변화 차이

		그룹 C		그룹 H		그룹 L		총합		표준편차
미취학 아동	0~3세	4344	(95.10)	101	(2.21)	123	(2.69)	4568	(100)	53.49
	4~6세	3392	(95.63)	56	(1.58)	99	(2.79)	3547	(100)	53.95
초등학생	1~3학년	3374	(96.59)	50	(1.43)	69	(1.98)	3493	(100)	54.79
	4~6학년	3363	(96.72)	44	(1.27)	70	(2.01)	3477	(100)	54.90
중학생		3581	(97.90)	34	(0.93)	43	(1.18)	3658	(100)	55.91
고등학생		2553	(98.84)	8	(0.31)	22	(0.85)	2583	(100)	56.73
고등학생 이상		103	(100)	0	(0.00)	0	(0.00)	103	(100)	57.74
총합		20710		293		426		21249		

()안은 비율을 나타냄

		그룹 H 평균 사교육비 변화-그룹 C 평균 사교육비 변화	그룹 L 평균 사교육비 변화-그룹 C 평균 사교육비 변화	가중치 높은 방법
미취학 아동	0~3세	3.43	-0.08	ATET, ATE
	4~6세	1.74	-4.17	ATET, ATE
초등학생	1~3학년	4.45	-7.33	ATET
	4~6학년	6.94	-9.16	
중학생		16.9	-6.54	ATE, 다중회귀분석
고등학생		0.03	-12.84	다중회귀분석

3. 다양한 준거집단의 사교육비 평균과 개인 사교육비 간의 관계비교

앞서의 1,2절의 분석에서는 개인이 사교육비를 지출할 때 참고하는 준거집단이 지리적 이웃이라는 암묵적인 가구가 있었다. 하지만 Clark et al(2008)에서도 확인할 수 있듯이 각 개인의 준거집단은 상황에 따라, 혹은 개인의 성향에 따라 다양할 수 있다. 따라서 본 절에서는 다양한 준거집단의 평균 사교육비와 개인 사교육비 간의 관계를 살펴보고자 한다. <표 11>는 본 절에서 사용한 분석이 식(1)-(4)로 제시되어있다. 식(1)-(4)는 회귀식 자체는 동일하지만 준거 집단이 다르다. 식 (1)은 이전의 분석들에서 활용하였던 인근 지역의 이웃들을 준거 집단으로 활용하였으며 (2)-(4)는 17개 광역시 및 도에 따라 집단을 대별하고 그 안에서 10개 소득분위, 부의 직업, 모의 직업 분류로 각각 세분하여 준거 집단을 사후적으로 형성하였다.

하지만 앞서 이야기 하였듯이 이와 같이 내생적으로 형성된 그룹에서 그룹의 평균값이 개인의 값에 미치는 영향은 그룹의 영향과 내생적인 동질성을 연구자가 구분할 수 없다는 한계(Manski, 1993)가 있기 때문에 본 절의 분석에서는 주로 식 (1)-(4)에서 β 의 상대적인 크기에 집중하는 것에 그치려고 한다.

실제 분석 시에는 보다 다양한 기준으로 다양한 기준집단을 형성해보았으나 여기서는 의미 있는 결과만을 제시하려고 한다. 식 (2)~(4)에서 넓은 지리적 이웃을 유지한 이유는 준거집단의 효과와 내생적 동질성을 본 분석에서 구분하는 것이 불가능 하다고 할지라도, 지리적 접근성을 기준으로 준거집단을 한정함으로써 분석 결과의 계수가 정보의 교환 등을 포함한 실제의 상호작용을 최소한 “암시”할 수 있도록 하는데 목적이 있었다.

또한 소득 분위는 매 년의 상대적인 소득을 기준으로 산정하였으며 직업 분류는 통계청의 직업분류표에 따랐다. 이에 더해 부모가 실업 상태에 있는 경우도 하나의 그룹으로 하였다. 부모가 실업 상태에 있는 가구를 제외하는 경우도 분석결과는 거의 유사하였다.

<표 11> 나이 그룹 별 그룹C, H, L의 분포 및 그룹별 사교육비 변화 차이

식	지리적 집단	세부집단	회귀식
(1)	시군구		$e_i = \beta e_{r(-i)} + X'_i \gamma + \Delta u_i$
(2)	17개 광역시/도	10개 소득분위	
(3)		부의 직업	
(4)		모의 직업	

e_i 는 개인의 로그 사교육비

$e_{r(-i)}$ 는 자기 자신을 제외한 준거집단의 로그 사교육비 평균

X'_i 는 아이의 성별, 나이, 부모의 학력, 부모의 나이, 부모의 취업여부, 로그 소득, 로그 금융소득, 로그 빚, 사교육 기관의 수, 연도 더미를 포함한다.

<표 12> 나이 그룹 별 그룹C, H, L의 분포 및 그룹별 사교육비 변화 차이

	(1)	(2)	(3)	(4)
집단의 사교육비	0.307*** (7.68)	0.368*** (6.96)	0.192*** (4.55)	0.173*** (3.93)
표본크기	5961	5993	5935	5941

* p<0.10, ** p<0.05, *** p<0.01, 각 계수의 t값은 괄호 안에 표시.

주1) 본 분석에서는 11~15차 데이터만을 사용하였다.

주 2) <표 11>에 제시된 바와 같이 모든 분석에는 아이의 성별, 나이, 부모의 학력, 부모의 나이, 부모의 취업여부, 로그 소득, 로그 금융소득, 로그 빚, 사교육 기관의 수, 연도 더미를 포함하였지만 간결함을 위해 제시하지 않았다.

제시된 식 (1)~(4)에서 계수의 값은 독립변수와 종속변수가 모두 로그 변환되었으므로 탄력성으로 해석할 수 있다. 계수의 절대적인 값은 17개 시도에 소득 분위로 그룹을 나눈 식 (2)에서 가장 높았지만 t값은 시군구 단위로 그룹을 나눈 (1)의 경우가 가장 높아 앞의 1,2절에서의 분석의 타당성이 확보되었다고 할 수 있겠다.

V. 나가는 글

높은 교육열로 인한 다양한 교육적 현상이 발생하여왔기 때문에 한국에서는 사교육비 지출에 관한 다수의 연구가 존재하였다. 하지만 이러한 일련의 연구들은 주로 어떤 인구학적/ 학교 변수가 사교육비 지출에 영향을 미치는지에 관해서만 집중되어있다. 그러나 현재와 같은 높은 수준의 사교육비에는 국가 간 군비경쟁과 마찬가지로 서로가 서로의 비용에 영향을 주고받는 측면이 분명히 존재한다. 이 효과는 기존의 연구에서 도외시되거나 식별의 어려움으로 인해 다루어지지 않았다.

이를 식별하기 위해 본 연구는 데이터에서 준실험상황(Quasi experiment)을 활용하였다. 지난 조사 이후 이사를 간 적이 있는 가구 중 자녀 교육 이외의 이유로 이사를 했다고 응답한 가구들을 사교육비를 상대적으로 많이 받는 지역으로 이사 간 표본과 적게 받는 지역으로 이사 간 표본으로 다시 세분하였다. D_in_D 분석을 통해 각 개인이 이사한 지역의 평균적인 사교육비에 맞추어서 자신의 사교육비를 재조정하는 것을 확인할 수 있었고, 이러한 차이는 통계적으로도 유의하였다.

하지만 데이터 내의 이주는 연구자에 의해 할당된 완벽한 무작위 실험이 아니기 때문에 사교육비가 높은/ 낮은 지역으로 이사를 가는 데 영향을 미칠만한 내생적 요인을 통제하기 위해 보다 확장된 다중회귀 분석과 PSM(Propensity Score Matching)을 활용하였다. 각 방법에서 그룹 H와 그룹 L의 계수는 조금씩 달랐으나 모든 분석에서 이사 가지 않은 집단과 비교해보았을 때 사교육비가 높은 지역으로 이사를 간 그룹은 약 5만원 사교육비를 높이고, 사교육비가 낮은 지역으로 이사를 간 그룹은 약 5만 원 정도 사교육비를 낮춘다는 것을 발견할 수 있었다. 이에 더해 4장의 3절에서는 다양한 준거집단의 사교육비 평균이 개인의 사교육비 평균과 어떤 관련이 있는지 살펴보았다.

사교육비 지출을 결정하는 것은 게임이론의 죄수의 딜레마 상황과 유사하다고 할 수 있다. 죄수의 딜레마 게임에서 유일한 균형이 두 죄수 모두 자백을 하는 것인 것처럼 사교육비에 높은 지출을 하는 것은 많은 학생과 학부모에게 최적 전략(best response)인 것이다. 하지만 사교육비가 상대적으로 낮은 지역으로 이사를 간 표본에서 사교육비가 통계적으로 유의하게 낮아지는 것은 사교육비가 많은 가구에게 실제적인 부담이 되고 있음을 방증한다.

『참고문헌』

A. E. Clark, P. Frijters, M. A. Shields, 2008. “Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles”, *Journal of Economic Literature*, Vol.24 No. 1, 95-144.

Changhui Kang, 2007. “Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea *Journal of Urban Economics*”, Vol. 61, 458-495.

Charles F. Manski, 1993. “Identification of Endogenous Social Effects: The Reaction Problem”. *Review of Economic Studies*, Vol. 60, 531-542.

F. Carlsson, O. J. Stenman, P. Martinsson, 2007. “Do You Enjoy Having More than Others? Survey Evidence of Positional Goods”, *Economica*, Vol. 74, 586-598.

Esther Duo, Pascaline Dupas, Michael Kremer 2008. “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”, *NBER Working paper*, No. 1447.

Guido W. Imbens 2000. “The role of propensity score in estimating dose-response functions”, *Biometrika*, Vol. 87, No. 3, 706-710

John Knight, Lina Song, Ramani Gunatilaka 2009. “Subjective Well-being and Its Determinants in Rural China”. *China Economic Review*, Vol. 20, 635-649.

Joshua D. Angrist 1995. “Using social security data on military applicants to estimate the effect of voluntary military service on earnings”. *NBER*, Vol. 5192.

R.H. Frank, 1985. “The Demand for Unobservable and Other Nonpositional Goods”, *AER*, Vol.75, No.1, 101-116.

Joshua D. Angrist, Jörn-Steen Pischke 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*.

김희삼, 2009, 「사교육비 지출에 영향을 주는 학교 특성의 분석」, 노동경제 연구, 제 32권, pp.27~59

박현정, 2008, 「학교교육의 질과 사교육 참여의 관계 분석」, 아시아 교육연구 9권 4호

이혜정 송종우, 2014, 「초,중,고 사교육비 영향요인 분석」 The Korean Journal of Applied Statistics, Vol. 27, No. 7, 1125-1137.

임성택 어성민, 2013, 「학부모의 사교육 동기와 부담감의 관계: 사교육 투자의 매개효과」, 교육학 연구, 제 51권 제 2호, pp. 63-87.