

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

• 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건 을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer 🖃





경제학석사 학위논문

최상위 대학의 노동시장 성과 프리미엄 분석

2023년 8월

서울대학교 대학원 경제학부 경제학 전공 민희현

최상위 대학의 노동시장 성과 프리미엄 분석

지도교수 김 봉 근

이 논문을 석사 학위논문으로 제출함 2023년 7월

> 서울대학교 대학원 경제학부 경제학 전공 민 희 현

민희현의 석사 학위논문을 인준함 2023년 7월

위 원 장 <u>조성진 (인)</u> 부위원장 <u>김봉근 (인)</u> 위 원 <u>서명환 (인)</u>

국문초록

본 연구는 최상위 대학이 재학생들의 노동시장 성과에 유의미한 영향을 미치는지 파악하기 위한 분석을 수행하였다. 이때, 노동시 장 성과 지표로는 상위 임금에 속할 확률과 대기업 입사 확률이 사용되었다. 동 분석을 수행하기 위해 한국교육고용패널(KEEP)의 2004~2015년 자료를 사용하여 수능 성적과 대학명, 가정환경, 노동 시장 등을 변수로 활용했다. 그런데 최상위 대학에 입학하는 학생 은 대체로 인지 능력, 가정환경, 리더십 등에서 우수한 편이며, 이 는 대학 졸업 후 취업 시에도 뛰어난 취업 성과로 이어질 가능성 이 크다. 하지만 현실적으로 동 요인들을 모두 변수화하는 것이 불가능해 누락변수로 인한 편향이 발생하게 된다. 이에, 본 연구에 서는 AET(2005)의 조건부 이변량 프로빗 모형을 적용하여 누락변 수로 인한 편향을 고려하였다. 동 방법론은 누락변수로 인한 편향 의 크기를 일정한 범위로 가정하고, 성과 프리미엄에서 편향의 최 댓값을 제거한 뒤에도 성과 프리미엄이 남아있는지 확인하는 방법 이다. 동 방법론에 따라 편향의 크기를 조건으로 하는 이변량 프 로빗 모형을 분석한 결과, 누락변수의 효과를 고려 시 최상위 대 학의 성과 프리미엄은 사라지는 것으로 나타났다.

주요어 : 최상위 대학, 임금, 대기업, 프리미엄, 누락변수 편향, 이변 량 프로빗

학 번: 2018-22738

목 차

제 1 장 서론	1
제 2 장 선행연구	4
제 3 장 연구 방법	8
제 4 장 데이터	12
제 1 절 한국교육고용패널 자료	12
제 2 절 변수 구축 방법	14
제 5 장 분석 결과 및 해석	21
제 1 절 일변량 프로빗 모형	21
제 2 절 민감도 분석(Sensitivity Analysis)	23
제 3 절 누락변수의 효과 제거	25
제 6 장 결론	27
참고문헌	29
Appendix ·····	30
Abstract ·····	34

표 목 차

<翌	1>	최상위 대학 여부에 따른 집단별 구성	14
< 翌	2>	노동시장 성과변수에 따른 인원 구성	17
<弫	3>	설명변수의 정의	18
<弫	4>	최상위 대학 여부에 따른 집단별 기술 통계량	19
<丑	5>	일변량 프로빗 모형에서 최상위 대학의 효과	20
<斑	6>	ho에 대한 가정에 따른 최상위 대학의 효과	
		: 수능 점수 포함	24
<丑	7>	ho에 대한 가정에 따른 최상위 대학의 효과	
		: 수능 점수 미포함	24
<丑	8>	$ ho$ 의 최솟값과 최댓값에 따른 최상위 대학의 효과 \cdots	26

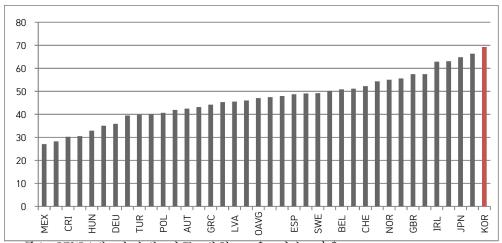
그림 목차

<그림	1>	2021년	OECD 국가별 대학 교육 이수율	1
<그림	2>	최상위	대학 여부에 따른 집단별 임금 분포	16
<그림	3>	최상위	대학별 월평균 임금	16

제 1 장서 론

한국은 대학 진학률이 높은 국가이다. <그림 1>의 2021년 OECD 국가별 대학 교육 이수율 데이터에 따르면, 한국의 25~34세 연령대 인구중에서 69.3%가 대학 교육을 이수한 것으로 나타났다. 동 수치는 OECD 국가 중에서 가장 높으며, OECD 소속 전체 국가의 대학 진학률 평균인 47.1%보다 20%p 이상 높다. 이처럼 유독 한국에서 대학 진학률이 높게나타나는 이유 중 하나는 높은 학업 성취를 통해 사회적으로 성공할 수있다는 믿음 때문이다. 우리 사회에는 대학을 졸업하면 더 좋은 기업에취업하고 더 높은 임금을 받을 것이라는 믿음이 있다. 실제로 대학 졸업을 지원자 필수 요건으로 내세우는 기업들이 많으니, 구직 시장에서 대졸자의 기업 선택 폭이 더 넓은 것은 사실이다. 그런데 대졸자가 69.3%에 이른다는 것은 구직자가 대학 졸업장만으로 자신의 능력을 증명하는 것이 어렵다는 것을 의미한다.

<그림 1> 2021년 OECD 국가별 대학 교육 이수율



주1. 25~34세 연령대 기준 대학 교육 이수 비율

자료. OECD

이에, 한국 사회에서는 '최상위 대학'에 입학하는 것 또한 중요한 요소로 여겨지고 있다. 한국 사회에서 '최상위 대학'이라고 여겨지는 대학을 졸업하면 더 좋은 기업에 또는 더 놓은 연봉으로 취업할 수 있을 것이라는 믿음이 존재한다. 취업포털 잡코리아 등이 2017년 시행한 설문조사¹⁾에서 학벌이 취업에 영향을 미친다고 생각하는지 묻는 질문에 구직자 478명 중 88.3%가 그렇다고 답하였다. 또한, 그렇다고 응답한 사람들에게 그 이유를 물었더니 55.9%가 '주요 대학 출신을 선호하는 사회적 분위기'를 꼽았다.²⁾

최상위 대학에 입학하는 것을 중요시하는 현상은 서울 소재 대학의 N수생 비율에서도 확인할 수 있다. 서울에는 최상위 대학이 다수 포진해 있어 서울 소재 대학은 '인서울'로 불리기도 한다. 입시학원 종로학원하늘교육에 따르면, 2011~2020년 서울 소재 4년제 대학 입학자 중 N수생비율은 29~34% 사이로, 전국 4년제 대학 입학자 기준 N수생 비율인 24.5%보다 5%p 이상 높은 것으로 나타났다. 입학 시기가 늦어지더라도 최상위 대학 입학을 희망하는 사람들이 상당히 많이 존재할 것으로 추정할 수 있다.

그런데 대학이 취업에 이토록 중요한 요소로 자리매김하게 된 이유는 무엇일까. Schultz(1961)는 인적자본(human capital)이라는 개념을 통해 교육이 노동생산성을 향상하는 효과가 있다고 주장했다. 인적자본이란, 장기간에 걸쳐 개인 내에 체화된 특성으로 개인이 보유한 능력(ability), 기술 숙련도(skill). 지식(knowledge)을 포괄하는 노동의 질적 측면을 나타낸다. 동 이론에 따르면, 사람들은 대학 교육 등의 학교 교육이나 직장내 훈련 등을 통해 노동의 질적 수준, 즉 노동생산성을 향상할 수 있다. 기업은 노동생산성이 더 높은 구직자를 선호하며 이들에게 더 높은 임금을 지급할 준비가 되어 있으므로 대졸자는 비대졸자보다 더 좋은 취업결과를 얻을 수 있을 것이다. 또한, 최상위 대학 졸업이 더 좋은 취업결과로 이어진다는 사람들의 믿음이 사실이라면, 최상위 대학은 전체 대학

¹⁾ 취업포털 잡코리아와 구인구직 플랫폼 알바몬은 신입직 구직자 478명을 대상으로 설문조사를 시행

²⁾ 동 문항에는 복수응답 가능

중에서도 노동생산성 향상 효과가 뛰어난 대학으로 볼 수 있다.

본 연구에서는 실증 분석을 통해 최상위 대학이 비최상위 대학 대비 인적자본을 효과적으로 향상하고 있는지 확인하였다. 최상위 대학 졸업 자가 노동시장에서 높은 임금을 받거나 사회적으로 선호도가 높은 대기 업에 입사할 가능성이 비최상위 대학 졸업자보다 큰지 확인이 이루어졌 다. 그러하다는 결론이 도출된다면 최상위 대학 졸업이 노동시장에서 구 직자에게 양(+)의 성과 프리미엄을 부여하는 것으로 이해할 수 있다.

최상위 대학의 노동시장 성과 프리미엄을 다룬 국내 초기 연구에서는 주로 최소자승법(OLS)을 활용하여 성과 프리미엄을 추정하였다. 대표적인 예로, 장수명(2006)은 한국노동패널(KLIPS) 자료를 최소자승법(OLS)으로 분석하여 최상위 5개 대학에 임금 프리미엄이 집중되어 있음을 밝혔다. 이때, 대학 서열은 대학입시학원 진학사의 연도별 대학수학능력시험(이하 수능) 데이터베이스를 활용하여 6개의 구간으로 정의하고 상위101-150위권 대학을 기준집단으로 설정하였다. 동 연구에 따르면, 상위1~5위 대학 졸업자의 월 실질임금은 기준집단의 월 실질임금 대비33.1%, 6~10위 대학 졸업자는 24.1%, 11~30위 대학 졸업자는 14.4%, 31~50위 대학 졸업자는 13.5%, 51~100위 대학 졸업자는 9.2% 높게 나타난다.

다음으로, 김진영(2007)은 한국직업능력개발원의 2005년도 전문대·대학교 졸업생의 경제활동상태 추적조사 자료와 진학사의 학과별 평균 입학성적 자료를 이용하여 최소자승법(OLS)으로 대학 졸업 2년 후의 연소득을 분석했다. 그 결과, 상위 1~10위 대학 졸업생은 11위 이하 대학졸업생 대비 7% 높은 임금을 받는 것으로 나타났다.

그러나 동 분석에서는 개인의 인지 능력에 대한 고려가 되어 있지 않아 최상위 대학의 임금 프리미엄이 과대추정(over-estimate)되었을 가능성이 크다. 인지 능력이 높은 사람은 수능에서 높은 성적을 받아 최상위 대학에 입학할 가능성이 클 뿐만 아니라, 더 좋은 취업 결과를 얻을 가능성이 크다. 개인의 인지 능력을 고려하지 않으면, 높은 인지 능력에 따라 최상위 대학 입학 및 취업 결과 사이에서 나타난 상관관계를 두 변수

사이의 인과관계로 오인할 위험이 있다. 또한, 개인의 인지 능력 외에도 성취 동기, 가정환경 등 최상위 대학 진학과 취업 성과에 모두 양(+)의 영향을 미치는 공통 요인은 무수히 많이 존재한다. 하지만 동 요인들을 모두 객관적인 변수로 수치화하여 통제하는 것은 현실적으로 불가능하다. 이로 인해 분석에서 누락변수(omitted variables)들이 존재하면 최상위 대학의 성과 프리미엄은 양(+)의 편향을 포함하게 된다. 즉, 최상위대학 진학의 프리미엄이 없음에도 있다고 나타나거나 프리미엄의 크기를 더 크게 추정할 위험이 있다.

본 연구에서는 최상위 대학 성과 프리미엄의 과대추정 문제를 최소화하기 위해 Altonji, Elder, and Taber(이하 'AET')(2005) 방법론을 사용하였다. 동 방법론은 편향을 발생할 수 있는 요인을 가능한 한 모두 변수화하여 통제하되, 여전히 누락변수가 존재한다는 사실을 전제한다. 이때, 누락변수의 존재에 따른 성과 프리미엄의 편향은 통제변수가 최상위대학 입학에 미치는 영향을 지표로 삼아 추정된다. 성과 프리미엄에서 편향 크기의 추정치를 제거한 뒤에도 여전히 프리미엄이 존재하는지 확인하고자 한다.

본문의 구성은 다음과 같다. 제2장에서는 대학 서열이 노동시장 성과에 미치는 영향을 분석한 선행연구를 살펴본다. 제3장에서는 본 연구에서 사용한 AET(2005) 방법론을 소개한다. 제4장에서는 한국교육고용패널(KEEP) 자료와 이를 가공한 기준과 최종적인 분석 데이터의 특성을보이고, AET(2005) 방법론에 기한 분석 결과를 보여준다. 제5장에서는 분석에 따른 결론을 두 가지로 정리한다.

제 2 장 선행연구

최상위 대학의 성과 프리미엄을 추정한 연구 중에서 개인의 인지 능력을 통제한 국내외 연구는 다수 존재한다. 해외에서는 Brand and Halaby(2006)가 미국 자료를 통해 대학 서열에 따른 노동시장 프리미엄

이 존재함을 밝혔고, 김희삼(2007)은 국내 자료를 통해 임금 프리미엄이 존재함을 보였다. 그러나 Dale and Krueger(2002)는 개인의 특성에 대한 보이지 않는 요인들까지 통제 범위를 확장하면 최상위 대학의 노동시장 성과 프리미엄은 사라지게 됨을 보였다. 본 장에서는 종전의 선행연구에서 사용한 방법론과 결과를 살펴보며 한계점을 짚고, 이를 극복하기 위해 본 연구에서 선택한 AET(2005) 방법론을 소개하고자 한다.

먼저, Brand and Halaby(2006)는 IQ 지수(11학년 테스트 기준)와 고등학교 성적 백분위 등의 개인 능력 변수를 비롯해 부모의 소득, 어머니의 교육수준, 카톨릭 여부, 유대인 등 가정환경에 대한 변수를 통제하여분석을 실시하였다. 그 결과, 개인의 특성을 통제해도 최상위 대학 졸업이 노동시장 성과에 유의미한 양(+)의 영향을 미친다는 사실을 발견했다. 즉, 동 연구에서는 개인의 인지 능력을 통제했음에도 불구하고 여전히 최상위 대학의 노동시장 성과 프리미엄이 존재함을 보였다.

그런데 Dale and Kruger(2002)는 연구자들에게 관측되지 않는 누락변수들이 입학사정관들에게는 관측될 수 있다는 점을 이용해 누락변수로 인한 편향을 바로잡았다.³⁾ 그들은 대상 표본에서 대학 입시생의 지원 대학 리스트와 각 대학의 합격 여부를 바탕으로 대조군을 설정하였다. 예를 들어, Best Univeristy와 Mediocre University에 각각 입학 허가를 받은 학생 두 명이 있다고 하자. 두 학생의 인지 능력, 성취 동기, 리더십등의 최종 점수가 비슷할 것으로 추정할 수 있다. 이때, 한 학생이 Best Univeristy에, 다른 한 학생이 Mediocre Univeristy에 진학했다면, 두 학생의 대학 졸업 후 노동시장 성과를 비교하여 누락변수의 효과를 최소화할 수 있다. 동 방식으로 대조군을 설정하여 분석한 결과, 최상위 대학 재학이 임금에 미치는 양(+)의 효과가 유의하지 않음을 밝혔다.

다음으로, 국내 연구는 대체로 학벌에 따른 노동시장 프리미엄이 존재 한다는 결과가 주를 이루고 왔다. 김희삼, 이삼호(2007)는 한국노동패널

³⁾ 미국에서 입학사정관은 학교 교사의 평가 내용, 에세이 등을 바탕으로 학생의 지적 호기심, 창의성, 활동력, 자신감, 리더십, 다른 사람에 대한 배려, 서로 다른 것에 대한 존중심 등을 종합적으로 평가한다.

조사(KLIPS)를 사용하였으며, 최소자승법(OLS)을 활용해 대학 서열에 따른 노동시장 프리미엄을 추정하였다. 이때, 노동시장의 지표는 시간당임금과 사업체의 규모가 활용되었다. 동 연구에서는 개인의 인지 능력을통제하기 위해 한국노동패널조사(KLIPS) 5차년도 조사에 있는 대학입시성적 문항을 활용했다. 4) 그 결과, 대학 서열을 나타내는 지표인 학과 평균 수능 백분위 점수가 1%p 상승할 시, 시간당임금도 0.5% 증가하는 것으로 나타났다. 또한, 종업원 수를 기준으로 사업체 규모를 19개의 구간으로 나누었을 때, 서열이 높은 대학을 졸업한 사람일수록 규모가 큰사업체에 취업한 것으로 나타났다.

이외에, 동 연구에서 눈에 띄는 점은 변수 간 상관계수의 비교를 통해 개인의 능력이 중요함을 재확인했다는 점이다. 동 연구는 개인의 대학입시성적(12개 구간)과 개인이 진학한 대학 학과의 평균 점수(12개 구간)각각과 시간당 임금(로그값)의 상관계수를 구하여 비교하였다. 그 결과,후자의 상관계수는 0.1868로 나타났지만, 전자는 0.3243으로 전자보다 73.6% 높게 나타났다.

고은미(2011)는 한국노동패널조사(KLIPS) 자료에서 1999~2008년 기준 26~28세 성인을 대상으로 대학 서열에 따른 임금 프리미엄을 추정하였다. 이때, 대학 서열은 대학입시학원 대성(1999년)과 진학사(2008년)의 자료를 바탕으로 평가되었다. 또한, 종속변수인 임금은 분석 대상 표본의월평균 임금을 로그화한 값이 사용되었다. 동 분석 표본을 최소자승법(OLS)으로 분석한 결과, 1999년 데이터에서 상위 13위 이내 대학 졸업자는 상위 14~50위 대학 졸업자 대비 7.24%의 임금 프리미엄을 갖는 것으로 나타났다. 또한, 1999년부터 2008년까지 3년 단위로 시기별 26~28세표본을 추출하여 분석했을 때, 시간이 지날수록 대학 서열에 따른 임금 프리미엄이 지속 상승했다.5)

^{4) 5}차년도 조사에는 대학입시성적을 묻는 문항이 존재한다. 다만, 동 문항은 응답자가 12개의 성적 구간이 주어진 상태에서 직접 자신이 성적 구간을 선택하는 방식이므로, 측정오차 등의 정보 손실이 발생할 수 있다.

⁵⁾ 고은미(2011)에서 상위 13위 이내 대학 졸업자는 상위 14~50위 대학 졸업자 와 비교하여 2002년에는 6.01%, 2005년에는 11.56%, 2008년에는 18.74%의 임금

이지영, 고영선(2019)은 한국노동패널조사(KLIPS)의 1~20차년도 자료를 활용해 대학 서열에 따른 임금 프리미엄을 추정하였다. 이때, 대학 서열은 대학입시학원 진학사(1998~2000년)의 대학별 및 학과별 평균 수능점수에 따라 5개 구간으로 정의되었다. 이를 바탕으로 로그화된 임금을 최소자승법(OLS)으로 분석한 결과, 25~29세 취업자 기준으로 서열이 가장 높은 5분위 대학 졸업자는 서열이 낮은 1분위 대학 졸업자 대비 14.0% 높은 임금을 받는 것으로 나타났다. 또한, 개인의 특성이 임금에 미치는 영향을 줄이기 위해 김희삼, 이삼호(2007)와 같은 방식으로 한국노동패널조사(KLIPS)의 5차년도 조사에 있는 대학입시성적 문항을 활용했다. 그 결과, 본 연구에서는 수능 성적의 영향을 통제한 후에도 임금 프리미엄이 여전히 존재함을 밝혔다.

이제까지 이루어진 선행연구에 따르면 최상위 대학 졸업자는 노동시장에서 성과 프리미엄을 갖는다. 하지만 동 선행연구에서 개인의 특성을 온전히 반영했다고 보기는 어렵다. 성취 동기, 리더십 등 개인의 성격 특성이 노동시장 성과 및 최상위 대학 입학에 상당한 영향을 미칠 것으로예상하나, 동 요인들을 정확하게 수치화하는 것에는 한계가 있기 때문이다. 또한, 한국노동패널조사(KLIPS)의 5차년도 조사에 있는 대학입시성적 문항은 답안이 구간으로 나누어져 있어 정보의 손실이 다소 발생했다는 점도 아쉬운 부분이다.

이에, 본 연구에서는 누락변수로 인한 편향을 최소화하기 위해 AET(2005) 방법론을 적용하였다. AET(2005)는 미국에서 카톨릭 고등학교 교육이 고등학교 졸업률 및 4년제 대학 입학률에 미치는 영향을 분석한 연구이다. 동 연구의 기술 통계량 분석에 따르면, 카톨릭 고등학교 학생은 공립 고등학교 학생보다 부모의 교육수준이 높고, 가구소득이 크다는 등의 특징을 보유하고 있다.6) 그런데 동 특징들을 가진 학생들은 고

프리미엄을 갖는 것으로 나타났다.

⁶⁾ AET(2005)는 공립 고등학교 10학년 집단과 카톨릭 고등학교 10학년 집단 간 가정환경, 지역 등의 변수가 통계적으로 차이를 보이는지 검토하였으며, 어머니의 교육수준(Mother's education), 아버지의 교육수준(Father's Education), 로그화된 가구소득(Log of family income) 등의 변수에서 통계적으로 유의미한

등학교 졸업 및 4년제 대학에 입학할 가능성이 크다. 즉, 카톨릭 고등학교 여부와 고등학교 졸업률 및 4년제 대학 입학률 사이에는 상관관계가 존재한다. 그런데 분석 과정에서 가정환경 등의 요인을 고려하지 않으면, 상관관계에 따른 효과를 인과관계로 분석하여 카톨릭 고등학교 입학 효과를 과대추정할 위험이 있다. 이에, 동 연구는 누락변수가 카톨릭 고등학교 입학에 미치는 영향의 범위를 가정하여 편향의 크기를 추정하였다. 그 결과, 누락변수의 효과를 고려한 뒤에도 카톨릭 고등학교 여부가 고등학교 졸업률 및 대학 진학률에 양(+)의 효과를 미친다는 사실이 드러났다.

제 3 장 연구 방법

본 연구에서는 AET(2005) 방법론을 이용해 누락변수에 의한 과대추정 문제를 최소화하고자 한다.

$$Y = 1(X'\gamma + \alpha D + \epsilon > 0)$$
$$D = 1(X'\beta + u > 0)$$

$$\begin{bmatrix} u \\ \epsilon \end{bmatrix} \sim N \begin{pmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \end{pmatrix} \tag{1}$$

Y는 월평균 임금 및 대기업 입사 여부를 의미하는 노동시장 성과를, D는 최상위 대학 교육 이수 여부를 나타내는 더미변수이다. X는 최상위 대학 입학과 노동시장 성과변수에 모두 영향을 미치는 요인의 집합이며, 각 식의 오차항은 u와 ϵ 으로 정의되었다. 이때, u와 ϵ 은 이변량 정규분

차이를 보였다.

포를 따르며, 두 오차항 간 공분산은 ρ이다. ρ는 최상위 대학 입학과 노동시장 성과에 모두 영향을 미치지만, 변수화하지 못한 요인으로, 지적호기심, 동기 부여 수준, 성실성 등의 학업 역량, 타인과의 소통 능력, 리더십 등의 공동체 역량을 포괄한다. 동 요인들은 객관적인 평가가 어려워 변수화하는 데 한계가 존재한다.7)

이제 노동시장 성과변수 Y에 대해 알아보겠다. 프로빗 모형은 Y가 1이 될 확률을 추정하므로, 동 확률을 잠재 변수 Y*로 표현할 수 있다. Y*에 대한 추정 식은 (2)로 표현할 수 있다.

$$Y^* = \alpha D + W' \Gamma = \alpha D + X' \Gamma_X + \xi \tag{2}$$

W는 변수화 가능 여부와 상관없이 Y*를 설명하는 모든 요인을 벡터로 나타낸 것이다. 이때, W는 개인의 인지 능력을 나타내는 수능 점수, 가정 환경을 나타내는 부모의 교육수준, 월평균 소득 등 객관적인 지표로 관측 가능한 변수(X)와 이를 통해 설명하지 못한 개인의 학업 역량, 공동체 역량 등의 누락변수 영역(ξ)으로 분리할 수 있다.

그런데 관측 변수와 누락변수 간에 상관관계가 존재할 가능성이 크다. 가령, 가정의 월평균 소득이 높으면 자녀의 사교육에 보다 적극적인 투자가 가능하며, 교육수준이 높은 부모를 보고 자란 아이는 동기 부여 수준이 보다 높게 나타날 수 있다. 이러한 가능성으로 인해 X의 계수인 Γ_X 에 편향이 발생할 수 있으므로, 동 가능성을 배제하기 위해 다음과 같은 식으로 재정의할 수 있다.

$$Y^* = \alpha D + X'\gamma + \epsilon, \text{ where } cov(X, \epsilon) = 0$$
(3)

⁷⁾ 본 연구에서는 부모 또는 선생님이 분석 대상의 성실성, 친구 관계, 리더십 등의 요소를 평가한 지표를 설명변수(X)로 추가하였다. 다만, 동 변수는 타인에 의한 주관적인 평가이므로, 개인의 의지와 내재한 성격을 100% 반영하였다고 보기에는 한계가 있다.

식 (3)에서 $X'\gamma$ 는 노동시장에서 좋은 성과를 거둘 확률 중에서 관측 변수에 의해 설명되는 부분으로, 누락변수에 의해 설명되는 부분인 오차 항 ϵ 과 관계가 없다고 가정하였다.

한편, 최상위 대학에 입학할 확률은 $D=1(D^*>0)$ 에 따라 잠재 변수 D^* 를 정의할 수 있다. (3) 식에서 D^* 가 나머지 항과 어떤 관계를 갖는지 파악하기 위해 식 (4)에서는 D^* 를 $X'\gamma$ 와 ϵ 에 linear projection하였다.

$$\operatorname{Proj}(D^*|X'\gamma,\epsilon) = \phi_0 + \phi_{X'\gamma}X'\gamma + \phi_{\epsilon}\epsilon \tag{4}$$

동 식에서 $\phi_{X'\gamma}$ 는 수능 점수, 부모의 교육수준 등 관측 변수와의 관계를 나타내며, ϕ_{ϵ} 는 노동시장 성과에 대한 설명력이 있으나 실제값을 변수화하는 데 한계가 있는 동기 부여 수준, 소통 능력 등 누락변수와의 관계를 나타낸다.

이때, AET(2005)는 누락변수가 최상위 대학 입학 확률에 미치는 영향 (ϕ_{ϵ}) 을 파악하는 지표로 설명변수에 의한 설명력 $(\phi_{X'\gamma})$ 을 사용하였다. 동연구의 저자는 ϕ_{ϵ} 의 크기가 $\phi_{X'\gamma}$ 보다 작거나 같을 것이라고 가정했다. 그 이유는 일반적으로 연구자가 설명변수(X)를 구성할 때 종속변수에 대한 많은 영향을 미치는 주요 변수를 중심으로 구성하기 때문이다. $(\phi_{X'\gamma})$ 이를 식으로 표현하면, $(\phi_{X'\gamma})$ 이 기으로 나타낼 수 있다.

Condition 1 : $\phi_{\epsilon} = \phi_{X'\gamma}$ OLS condition : $\phi_{\epsilon} = 0$

반대로, OLS condition은 동기 부여 수준, 소통 능력 등의 누락변수가 최상위 대학 입학 확률 (D^*) 에 영향을 미치지 않는다는 가정으로, 다소

⁸⁾ 이때, 동 연구의 저자는 관측 가능한 설명변수의 집합인 X가 W로부터 무작위로 선택되며, X와 W 각각에 속한 변수가 매우 많아 어떤 하나의 변수가 지배적이지 않다는 가정하였다.

비현실적이지만 일반적인 OLS 가정을 전제하고 있다.

그런데 식 (3)에서 누락변수에 의해 설명되는 부분을 나타내는 오차항 ϵ 는 발생 시기에 따라 $\epsilon = \epsilon_1 + \epsilon_2$ 로 분리할 수 있다. 여기에서 ϵ_1 은 처치 이전에 발생한 누락변수이고, ϵ_2 는 처치 이후에 발생한 누락변수이다. 이를 바탕으로 Condition 1을 재정리하면, 식 (5)로 표현할 수 있다.

$$\phi_{X'\gamma} = \frac{Cov(D^*, X'\gamma)}{Var(X'\gamma)} = \frac{Cov(D^*, \epsilon_1)}{var(\epsilon_1)}$$
(5)

이때 ϕ_{ϵ} 와 $\phi_{X'}$ 는 분자의 크기가 같지만, 분모는 ϕ_{ϵ} 가 더 크다.

$$\phi_{\epsilon} \equiv \frac{Cov(D^*, \epsilon)}{Var(\epsilon)} \le \frac{Cov(D^*, \epsilon_1)}{Var(\epsilon_1)} = \phi_{X'\gamma}$$
(6)

 $Cov(D^*,X'\gamma)>0$ 이고 $\phi_{\epsilon}\geq 0$ 이므로 ϕ_{ϵ} 와 $\phi_{X'\gamma}$ 의 관계를 정리하면 Condition 2를 도출할 수 있다.

Condition $2: 0 \leq \phi_{\epsilon} \leq \phi_{X'\gamma}$

위의 조건들을 이용하면 α 는 $\phi_{\epsilon}=0$ 일 때 하한을 갖고 $\phi_{\epsilon}=\phi_{X'\gamma}$ 일 때 상한의 값을 갖는 구간으로 나타낼 수 있다. α 가 하한값을 가지게 하는 조건에서도 양(+)의 값을 가진다면 이는 양(+)의 효과에 대한 강력한 증거로 해석될 수 있다. 이제 동 조건들을 오차항 u와 ϵ 의 공분산인 ρ 에 적용해보고자 한다. ρ 는 다음과 같이 풀어쓸 수 있다.

$$\rho = cov(u, \epsilon) = cov(D^* - X\beta, \epsilon) = cov(D^*, \epsilon)$$
(7)

우선, $\phi_{\epsilon}=0$ 가 성립하면 식 (4)에서 D^* 와 ϕ_{ϵ} 두 변수 사이에 관계가 없음을 나타내며, (7)에서 $cov(D^*,\epsilon)$ 가 0임을 나타낸다. 이에 따라 α 의

하한은 $\rho=0$ 일 때 성립한다. 다음으로, $\phi_{\epsilon}=\phi_{X'\gamma}$ 가 성립하면 동 식을 $\frac{Cov(D^*,\epsilon)}{var(\epsilon)}=\frac{Cov(D^*,X'\gamma)}{Var(X'\gamma)}$ 로 풀어쓸 수 있다. 좌변에 $cov(D^*,\epsilon)$ 만 남기기 위해 식을 변형시키면, 식 (8)을 도출할 수 있다.

$$Cov(D^*, \epsilon) = \frac{Cov(D^*, X'\gamma)}{Var(X'\gamma)} *var(\epsilon)$$
(8)

이때, $var(\epsilon)$ 는 식 (1)에서의 가정에 따라 1의 상수값을 갖는다. 따라서 α 의 상한은 $\rho=\frac{Cov\left(D^*,X'\gamma\right)}{Var\left(X'\gamma\right)}$ 일 때 성립한다. 이를 정리하면 ρ 의 범위에 대해 Condition 3으로 정리할 수 있다.

Condition
$$3 : 0 \le \rho \le \frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)}$$

제 4 장 데이터

제 1 절 한국교육고용패널 자료

본 연구에서는 한국직업능력연구원의 한국교육고용패널(Korean Education & Employment Panel: KEEP)의 1차 조사 자료를 활용한다. 한국직업능력연구원은 2004년 기준 전국의 일반계 고등학교 3학년생 2,000명과 실업계 고등학교 3학년생 2,000명, 중학교 3학년생 2,000명을 조사한 것을 시작으로 2015년까지 동일한 대상에 대해 매해 추적 조사를 진행하였다. 또한, 2004년에는 조사 대상 학생의 보호자 가구 6,000명, 학생이 소속된 학교의 담임 1,112명도 추가 조사하여 가구 데이터와 학교생활 평가 데이터를 모두 보유하고 있다.

그중에서도 본 연구에서는 2004년 기준 일반계 고등학교 3학년 코호 트(2,000명)를 표본으로 사용하였다. 동 코호트는 고등학교 졸업 후 대학교에 진학하고 사무직 계열에 입사하는 비교적 동질적인 진로를 보유하고 있기 때문이다. 또한, 중학교 3학년 코호트 중 일반계 고등학교 진학자는 자기선택의 문제가 발생할 위험이 있어 포함하지 않았다. 성인 남성의 군대 복무기간을 고려하면 노동시장 변수를 관찰 가능한 기간은 단 2년으로, 동 집단을 포함하면 일찍 취업한 사람들이 선택적으로 포함되는 상황이 발생할 수 있기 때문이다.

한국교육고용패널(KEEP) 자료는 수능 점수, 대학명을 포함하여 가구 정보, 직업 정보 등을 포괄적으로 제공하고 있어 본 연구에 적합하다. 특 히, 수능 점수는 개인의 인지 능력을 파악하기 위한 대표적인 변수로 본 연구에서 필수적이다. 동 자료에서는 수능과 관련된 정보를 과목별 점수 와 백분율, 표준점수로 상세하게 제공하고 있다. 또한, 대학명을 제공하 고 있어 일부 대학을 최상위 대학으로 구분 짓기에 용이하다.

개인의 특성과 대학 졸업 후 직업 정보를 제공하는 패널 데이터로 잘 알려진 대졸자직업이동경로조사(GOMS)와 한국노동패널(KLIPS) 자료는 수능 점수와 대학명 정보 파악에 한계가 있어 사용하지 않았다. 동 자료들은 실제로 대학 교육과 관련된 연구에서 자주 사용되고 있지만, 대졸 자직업이동경로조사(GOMS)는 대학교 소재지에 대한 정보만 제공할 뿐 대학명과 수능 점수 정보를 제공하지 않는다. 또한, 한국노동패널 (KLIPS) 자료는 대학명 정보, 가구 정보, 직업 정보 등 필요한 정보를 다수 제공하고 있지만, 수능 점수를 제공하고 있지 않다.

본 연구에서는 일반 4년제 대학을 나온 사람들로 구성하였다. 이때, 노동시장에서 출신 대학의 영향이 크지 않은 특수계열을 제외하였다. 자료 내 대학 간 유사성을 유지하기 위해 교육대학교, 과학기술 대학교, 경찰대학교 등과 같은 특수계열의 학교도 제외하였다. 같은 이유로 의학, 치의학, 한의학, 약학과 같은 특수 학과에 재학 중인 학생도 표본에서 제외하였다. 또한, 본 연구에서는 노동시장에서 임금근로자로 근무하고 있는 인원만 포함하였다. 한국교육고용패널 자료에는 자영업자와 무급가족

종사자를 비롯한 비임금근로자의 데이터도 존재한다. 하지만 이들의 노동시장 성과는 가정환경, 네트워크, 성격 등 개인에 내재한 특성의 영향을 많이 받고 대학 교육을 통해 습득한 인적자본과는 다소 거리가 있기때문이다.

이에 따라 분석 대상 표본은 최상위 대학 졸업자 67명(13.7%)과 그외 일반 4년제 대학 졸업자 423명(86.3%)으로 구성되어 총 490명의 데이터가 포함되어 있다.

<표 1> 최상위 대학 여부에 따른 집단별 구성

제 2 절 변수 구축 방법

본 연구에서 사용된 설명변수는 '최상위 대학' 졸업 여부이다. '최상위 대학'을 정의하기 위해 본 연구에서는 2009년 중앙일보 대학평가⁹⁾를 바탕으로 13개 대학을 선정하였다. 동 평가 기준으로 상위 15위 이내를 차지한 대학은 서울대, 연세대, 고려대, 서강대, 성균관대, 한양대, 이화여대, 중앙대, 경희대, 카이스트, 포스텍, 건국대, 한국외대, 부산대, 인하대이다. 이중 이화여대는 데이터가 존재하지 않아 제외되었으며, 카이스트와 포스텍은 종합대학이 아닌 이공계 중점 대학으로 대학 간 동질성을위해 제외하였다.¹⁰⁾ 또한, 아주대는 2009년 대학평가 기준 15위 안에 포

⁹⁾ 중앙일보 대학평가 외에도 수험생 사이에서 대학 서열로 일컬어지는 순위에 따라 상위 대학 13곳(서울대, 연세대, 고려대, 서강대, 성균관대, 한양대, 중앙대, 경희대, 한국외대, 서울시립대, 건국대, 동국대, 홍익대)을 '최상위 대학'으로 정의하여 분석한 때에도 결과는 유사했다. 프로빗 모형에서 최상위 대학의 한계효과는 0.113(0.073)으로 유의확률 10% 수준에서 유의했으며, 누락변수 고려 시한계 효과는 음수로 전환되어 프리미엄이 사라지는 것으로 나타났다.

함되어 있지 않지만, 2008년과 2010년 대학평가 기준으로 15위 안에 선정되어 최상위 대학으로 포함하였다. 이로써, 총 67명이 최상위 대학 졸업자로 선정되었다. 이에 따라 분석 대상 표본에서 최상위 대학 졸업 인원을 학교별로 정리하면, 서울대(3명), 건국대(7명), 경희대(4명), 고려대(4명), 서강대(2명), 성균관대(6명), 연세대(6명), 중앙대(5명), 한국외대(3명), 한양대(4명), 부산대(8명), 인하대(8명), 아주대(7명)로 나타난다.

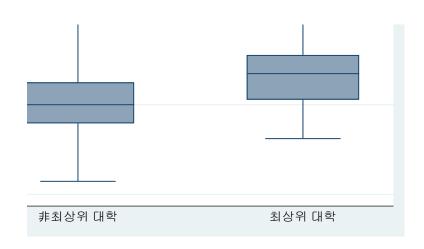
다음으로, 본 연구에서 사용된 노동시장 성과변수는 '상위 임금에 속 할 확률'과 '대기업 입사 확률' 두 가지를 더미변수로 구성하였다.

첫째, 상위 임금에 속할 확률은 대학 졸업 후 초임 기준으로 월평균임금이 상위 30%(241만 6,967원)에 속할 확률을 나타낸다. 우선, 노동시장 성과변수로 임금을 활용하는 것은 구직자들의 선호 직장 기준을 반영한 결과이다. 2021년 구인구직 플랫폼 사람인에서 2030 세대 1,865명을대상으로 좋은 직장의 기준에 대해 질문한 결과, 33.8%가 연봉 수준을택하였다. 또한, 2021년 구인구직 플랫폼 잡코리아와 알바몬이 성인 1,288명을 대상으로 직장 선택 기준에 대해 설문한 결과, MZ세대(1980~2000년 출생)와 X세대(1970년대 출생) 각 부문에서 연봉 수준이 1위를 차지했다. 이러한 이유로 종전의 선행연구에서도 임금은 대표적인성과변수로 활용되어왔다.

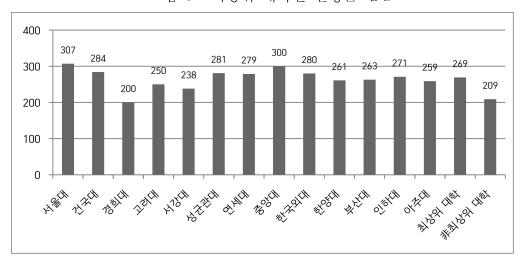
다음으로, 월평균 임금은 기본 임금, 초과근로수당, 특별급여의 합으로 정의된다. 이때, 연도별 인플레이션 효과를 제거하기 위해 소비자물가지수(2015년=100)를 기준으로 보정하였다. 동 기준에 따라 최상위 대학과비최상위 대학의 그룹별 평균 임금을 구하면, 두 집단 간 뚜렷한 차이가존재한다. 최상위 대학 졸업생의 월평균 임금은 269만 원, 그 외 일반 4년제 대학 졸업생의 월평균 임금은 209만 원으로, 두 집단 간 월평균 임금은 60만 원의 차이가 발견되었으며, 1% 유의수준에서 통계적으로 유의미한 차이가 존재한다.

¹⁰⁾ 분석 대상 표본에서 카이스트 졸업자는 0명, 포스텍 졸업자는 1명으로, 동기준에 따라 실제로 제외된 인원은 1명이다.

<그림 2> 최상위 대학 여부에 따른 집단별 임금 분포



<그림 3> 최상위 대학별 월평균 임금



또한, <그림 3>은 최상위 대학 집단 내에서 대학별 평균 임금을 보여준다. 서울대와 중앙대가 300만 원 이상으로 가장 높게 나타났으며, 경희대를 제외하고는 평균 임금이 250만 원 내외 또는 그 이상으로 높게 나타났다.

이때, 상위 임금의 기준을 상위 30%로 정의한 것은 상위 임금에 속할 확률 프리미엄이 프로빗 모형에서 통계적으로 유의하도록 설정하기 위함이다. 본 연구의 목적은 최상위 대학의 성과 프리미엄에서 누락변수에의한 편향을 제거한 뒤에도 통계적 유의성이 유지되는지 확인하는 것에 있다. 이는 AET(2005) 방법론을 적용하기 전 기본 모형에서 프리미엄이 존재해야 확인할 수 있다. 이에, 일변량 프로빗 모형에 따라 최상위 대학의 임금 프리미엄이 14%, 유의확률은 0.004로 도출되는 상위 30%를 기준으로 삼았다.11)

둘째, 본 연구에서 다루는 또 다른 노동시장 성과변수는 대기업 입사여부이다. 대기업은 연봉, 복지 등 구직자들이 선호하는 요소를 보유하고 있어 소규모 고연봉 집단을 제외하고는 대규모 기업이 상대적으로 선호되고 있다. 본 연구에서는 300인 이상의 종업원을 보유한 기업을 대기업으로 정의하여 더미변수로 활용하였다.

< 표 2>는 각 성과변수에 따른 최상위 대학 및 비최상위 대학의 인원을 나타내고 있다. 최상위 대학 졸업자 중 임금이 상위 30%에 해당하는 인원은 42명으로 62.3%에 이르렀으며, 대기업에 입사한 인원은 55명으로 82.1%였다.

¹¹⁾ 상위 임금의 기준을 상위 20%로 설정할 때에도 프로빗 분석 시 임금 프리미엄의 통계적 유의성이 존재한다. 다만, ρ 의 크기가 커질수록 프리미엄이 빠르게 줄어들어 분석 진행이 어렵다는 한계가 있다. 이에, 상위 30%를 상위 임금범위의 최소 기준으로 삼고 분석을 시행하였다.

<표 2> 노동시장 성과변수에 따른 인원 구성

	임금 상위 30% 이내	아닌 경우
최상위 대학	42명	25명
비최상위 대학	106명	317명
	대기업 입사	아닌 경우
최상위 대학	55명	12명
비최상위 대학	151명	272명

마지막으로, 본 연구에서 통제변수는 개인의 특성(수능 성적, 거주 지역, 여성 여부), 가구 정보(월 가구소득, 아버지 교육수준, 어머니 교육수준, 머리, 학교생활의 성실성), 학생 본인의 평가(가정생활 만족도, 건강상태, 수면시간), 담임의 학생 평가(공부, 리더십, 친구 관계, 학교생활)네 가지 기준에 따라 구성하였다.

동 변수 중 가장 핵심적인 역할을 하는 변수는 수능 성적이다. 학생의실제 수능 점수인 표준점수가 가장 적합하나, 재수 또는 그 이상 수능을 치른 사람들 간의 차이를 배제하기 위해 본 연구에서는 백분위 점수를 사용하였다. 이외에도 대학 입시부터 취업까지 영향을 미치는 요소로 가정환경, 개인의 고유한 특성, 동기 부여 수준, 성격, 건강상태를 나타내는지표를 포함하였다. 변수에 대한 구체적인 정의는 아래의 <표 3>에 정리되어 있다.

<표 3> 설명변수의 정의

구분	변수명	정의				
개인	수능 성적	언어, 수리, 탐구 2과목 백분위 평균				
 특성	거주 지역	더미변수(서울/경기권 1, 지방 0)				
<u></u> 국성	여성	성별(여성 1, 남성 0)				
	아버지 교육수준	남자 보호자의 최종학력 기준 교육연수				
	어머니 교육수준	여성 보호자의 최종학력 기준 교육연수				
	월평균	 월평균 명목 임금액(만원)의 로그값				
-1 -7	가구소득	철평진 영국 담급역(한천/의 도그없 				
가구		더미변수(전혀 머리가 좋지 않다 1, 머리가 좋지				
조사	지능	않다 2, 보통이다 3, 머리가 좋다 4, 매우 머리가				
		좋다 5)				
	학교생활의	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
	성실성(가정)	통이다 3, 그렇다 4, 매우 그렇다 5)				
	가정생활	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
학생	만족도	통이다 3, 그렇다 4, 매우 그렇다 5)				
본인	7] 7L xL r]]	더미변수(매우 허약하다 1, 허약하다 2, 보통이다				
는 단	건강상태	3, 건강하다 4, 매우 건강하다 5)				
	수면시간	평균 수면시간(시간)				
	공부	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
	6 T	통이다 3, 그렇다 4, 매우 그렇다 5)				
	리더십	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
담임	되되省	통이다 3, 그렇다 4, 매우 그렇다 5)				
평가	친구 관계	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
	신丁 선계	통이다 3, 그렇다 4, 매우 그렇다 5)				
	학교생활의	더미변수(전혀 그렇지 않다 1, 그렇지 않다 2, 보				
	성실성(담임)	통이다 3, 그렇다 4, 매우 그렇다 5)				

주1. 수능 성적 변수에서 탐구 과목은 상위 백분율 2과목을 택한 것임 주2. 월평균 가구소득은 값의 범위가 크다는 점을 고려해 자연로그를 취한 값을 사용 <표 4>에서는 최상위 대학 여부에 따른 기술 통계량과 통계적으로 유의미한 차이의 존재 여부를 확인할 수 있다. 개인의 특성 부문에서는 수능 백분위 성적 평균에 24.754%p 차이가 나타나고 있으며, 수도권 지역 학 생들의 최상위 대학 진학률이 비수도권 지역 학생들의 진학률보다 높게 나타남 을 확인할 수 있다. 또한, 가구 조사와 담임 평가에 해당하는 모든 설명변수에 서 두 집단 간 유의한 차이가 존재함이 드러났다.

<표 4> 최상위 대학 여부에 따른 집단별 기술 통계량

	total (N=490)	최상위 대학 (N=67)	非최상위 대학 (N=423)	차이
	A. 통계	세변수		
	수능 성적(%)	55.85	80.61	24.754***
개인 특성	거주 지역 (수도권 여부)	0.32	0.51	0.186***
	여성	0.47	0.37	-0.093
	월평균 가구소득	5.64	6.01	0.366***
	아버지 교육수준	12.38	14.25	1.878***
가구	어머니 교육수준	11.56	12.7	1.141***
조사	지능	3.55	3.94	0.392***
	학교생활의 성실성(가정)	4.1	4.3	0.202**
충ե게	가정생활 만족도	3.78	4.01	0.237**
학생 본인	건강상태	3.57	3.52	-0.043
LL	수면시간	5.42	5.31	-0.107
	공부	3.19	4.16	0.970***
담임	리더십	3.13	3.42	0.286***
평가	친구 관계	3.78	3.93	0.150*
	학교생활의 성실성(담임)	4.04	4.39	0.346***
B. 노동시장 성과변수				
	상위 임금 여부	0.25	0.63	0.376***
<u>조</u> 1	대기업 입사 Significance levels: * < 10%	0.36	0.82	0.464***

주1. Significance levels: * < 10%, ** < 5%, *** < 1%

제 5 장 분석 결과 및 해석

제 1 절 일변량 프로빗 모형

본 연구에서는 일변량 프로빗 모형으로 최상위 대학의 노동시장 성과 프리미엄을 분석한 뒤, AET(2005) 방법론을 적용해 누락변수에 따른 편 향 제거 시 프리미엄이 어떻게 변화하는지 살펴보고자 한다.

<표 5>는 일변량 프로빗(univariate probit) 모형으로 최상위 대학의 노동시장 프리미엄 효과를 알아본 결과이다. 본 연구에서는 종속변수의 종류에 따라 패널 A, B로 나누어 한계 효과(marginal effects)를 살펴보았다. 패널 A는 상위 임금 해당 여부를 기준으로 분석한 결과이며, 패널 B는 대기업 입사 여부를 기준으로 분석한 결과이다.

<班 5>	일변량	프로빗	모형에서	최상위	대학의	효과
-------	-----	-----	------	-----	-----	----

	(1)	(2)	(3)	(4)
				수능 점수,
	None	기타 설명변수	수능 점수	기타 설명변수
		A. 상위	임금	
	.996	.797	.609	.600
probit	(000.)	(.000.)	(.001)	(.002)
	[.325]	[.240]	[.190]	[.175]
$PseudoR^2$.059	.128	.102	.157
		B. 대기약	겁 입사	
	1.285	1.143	.978	.989
probit	(000.)	(000.)	(.000.)	(.000.)
	[.460]	[.392]	[.340]	[.331]
$PseudoR^2$.079	.116	.105	.135

주1. ()은 유의확률, []은 한계 효과를 나타냄

이때, 동 분석 표본에서 수능 성적이 종속변수에 미치는 영향을 살펴 보기 위해 통제변수의 구성을 4가지로 설정하였다. 첫 번째 열인 (1)에 서는 아무것도 통제하지 않고 설명변수인 최상위 대학의 더미변수만을 사용하였으며, (2)는 수능 점수를 제외한 모든 설명변수를 통제하였다. (3)은 설명변수 중 수능 점수만 통제하였으며, (4)는 수능 점수를 포함한모든 설명변수를 통제하였다. 그 결과, 최상위 대학은 패널 A와 패널 B에서 모두 양(+)의 성과 프리미엄으로 나타났다.

먼저, 패널 A에서 최상위 대학 더미변수는 (1)에서 0.325의 한계 효과를 보였으나 (2)에서는 한계 효과가 0.240으로 감소하여 (1) 대비 26.2% 감소하였다. 또한, 수능 점수만 통제한 (3)에서는 한계 효과가 0.190으로, (1) 대비 약 41.5% 감소했다. 즉, 다른 설명변수에 비해 수능 점수가 상위 임금 여부에 미치는 영향이 크게 나타나고 있다. 마지막으로, 수능 점수를 포함하여 모든 요인을 통제한 (4)에서는 한계 효과가 0.175로 나타났다. 앞서, 상위 임금을 정의하면서 통계적 유의성이 성립하도록 설정하였으므로, (1)~(4) 모두 유의수준 5%에서 통계적으로 유의한 것으로 나타났다.

다음으로, 패널 B에서 최상위 대학 더미변수의 효과는 패널 A보다 뚜렷하게 나타났다. 아무것도 통제하지 않은 (1)에서는 한계 효과가 0.460로 나타났으며, 수능 점수를 제외한 설명변수를 통제한 (2)에서는 0.392로 감소하였다. 또한, 수능 점수를 통제한 (3)에서는 0.340로 나타났으며, 모든 설명변수를 통제한 (4)에서는 0.331로 (3)보다 소폭 감소하였다. 패널 B는 전반적으로 패널 A보다 한계 효과가 크게 나타났으며, (1)~(4)모두 5% 수준에서 유의한 것으로 나타났다.

한편, $Pseudo R^2$ 은 전반적으로 낮게 나타났다. 패널 A와 패널 B 모두 (1)에서는 $Pseudo R^2$ 이 0.1 이하로 나타났으며, (4)에서도 0.2 이내를 기록했다. 또한, 패널 A와 B 모두 (3)보다 (2)에서 $Pseudo R^2$ 이 더 큰 값을 갖는 것으로 나타났지만, (1) 대비 한계 효과의 감소분은 (3)에서 더 큰 것으로 나타났다.

이를 정리하면, 첫째, 수능 점수가 상위 임금에 속할 확률 및 대기업입사 확률 프리미엄에 미치는 영향력이 클 것으로 예상할 수 있다. 패널 A와 B 모두 수능 점수 제외 설명변수를 통제했을 때 모델의 전반적인설명력은 상승했지만, 이에 따른 프리미엄 감소 효과는 수능 점수만 통제했을 때 더 크게 나타났다. 수능 점수는 비교적 객관적 기준으로 개인

의 인지 능력을 평가한 변수로, 누락되었을 시 최상위 대학의 성과 프리 미엄에 상당한 크기의 편향을 발생시켰을 것으로 예상할 수 있다.

둘째, 최상위 대학 졸업자는 상위 임금 30%에 속할 확률과 대기업 입사 확률에서 프리미엄을 갖는 것을 확인했다. 개인의 인지 능력뿐만 아니라 기타 개인의 특성, 가정환경, 주변의 평가 등 다양한 요인을 통제한뒤에도 여전히 프리미엄이 존재한다는 사실을 발견했다.

제 2 절 민감도 분석(Sensitivity Analysis)

누락변수에 의한 편향을 고려 시에 최상위 대학의 노동시장 성과 프리미엄이 어떻게 변화하는지 확인하기 위해 민감도 분석을 시행하였다. A 동 분석은 이변량 프로빗(bivariate probit) 모형에 파라미터 ρ 에 대한 제약을 추가하여 진행하였다. ρ 는 최상위 대학에 대한 프로빗 모형과 노동시장 성과변수에 대한 프로빗 모형의 잔차 간 공분산으로, 동 민감도 분석에서 ρ 는 0부터 0.4까지 0.1 단위로 나누어 상수로 가정하였다. 이때, ρ =0은 누락변수로 인한 편향을 고려하지 않은 경우이므로, 일변량 프로빗 모형과 동일하다.

< 표 6>는 노동시장 성과변수인 두 개의 종속변수를 패널 A와 패널 B로 나누어 민감도 분석 결과를 제시하였다. 동 분석에서는 일변량 프로빗 모형의 (4)와 같이 수능 점수를 포함한 모든 설명변수가 통제되었다. 분석에 따른 결과는 최상위 대학 더미변수의 계수, p-value, 한계 효과가차례로 나타나 있다.

이때, ρ 는 최상위 대학 더미와 노동시장 성과변수 더미에 모두 영향을 미치지만 누락된 변수로 인한 편향을 나타낸다. 예컨대, 모델에서 설명변수로 충분히 포함하지 못한 동기 부여 등 최상위 대학 더미와 노동시장 성과변수 더미에 동일한 방향의 영향을 미치는 양의 편향(+)이 포함될수 있다. 따라서 ρ 가 0일 때는, 즉, 누락변수에 의한 편향이 0일 때는 최상위 대학 더미변수의 한계 효과에서 양(+)의 편향을 제거할 필요가 없어지므로, 한계 효과가 가장 크게 나타난다. 또한, ρ 가 증가하면 누락변

수로 인한 편향의 크기가 증가하고, 편향을 제외하고 남은 최상위 대학 더미변수의 한계 효과는 감소하게 된다. 이를 풀어 말하면 ρ 가 커짐에 따라 최상위 대학의 노동시장 프리미엄은 감소하는 것으로 이해할 수 있 다.

동 분석에서는 전통적인 가설검정 방법을 적용하여 양(+)의 한계 효과가 유의하게 나타나는 ρ 의 값을 살펴보고자 한다. 여기에서 양(+)의 한계 효과로 특정 지은 이유는 직관적으로 생각했을 때 최상위 대학 졸업이 노동시장 성과변수에 부정적 영향을 미칠 가능성은 지극히 낮기 때문이다. 가설검정은 5% 유의수준 상에서 최상위 대학 더미변수의 효과가존재하지 않는다는 귀무가설을 검정하는 방식으로 이루어졌다.

먼저, 패널 A에서 한계 효과는 ρ =0.1일 때까지 양의 한계 효과가 유의한 것으로 나타났으며, ρ =0.2일 때부터 유의수준 5%에서의 통계적 유의성을 상실했다. 또한 ρ =0.4일 때, 한계 효과가 음수로 전환되었다. 반면, 패널 B에서는 최상위 대학의 한계 효과는 ρ =0.3인 경우에도 한계 효과가 0.167(0.016)으로 5% 유의수준에서 유의한 것으로 나타났다. 그러나 ρ =0.4부터는 통계적 유의성을 상실했으며, ρ =0.6일 때부터 한계 효과가음수로 전환되었다.

<표 6>
ho에 대한 가정에 따른 최상위 대학의 효과 : 수능 점수 포함

		오차항들의 상관관계							
	$\rho = 0$ $\rho = 0.1$ $\rho = 0.2$ $\rho = 0.3$ $\rho = 0.4$								
	A. 상위 임금								
	.600 .440 .278 .111059								
	(.002)	(.024)	(.154)	(.564)	(.755)				
수능	[.175]	[.129]	[.082]	[.033]	[017]				
점수	 B. 대기업 입사								
포함	.989	.828	.661	.486	.303				
	(.000.)	(.000.)	(.001)	(.016)	(.130)				
	[.331]	[.280]	[.225]	[.167]	[.105]				

주1. ()은 유의확률, []은 한계 효과를 나타냄

<표 7>은 수능 점수를 포함하지 않았을 때 한계 효과의 변화를 나타 낸 것이다. 동 분석에 따르면 패널 A와 B 모두에서 ρ가 0.4에 이를 때 까지도 한계 효과가 양수로 나타났다. 특히, 패널 B에서는 0.030의 한계 효과가 통계적으로 유의하 것으로 나타났다. 다시 말해, 수능 점수를 통 제하지 않았다면 누락변수로 인한 편향을 고려할지라도 <표 5>와는 다 른 결론이 도출되었을 것이다. 이는 주요 변수가 누락될 경우 최상위 대 학의 성과 프리미엄이 과대추정되는 것을 넘어 최종적인 분석 결과가 바 뀔 수 있음을 시사한다.

<표 7> ρ에 대한 가정에 따른 최상위 대학의 효과 : 수능 점수 미포함

	오차항들의 상관관계								
	$\rho = 0$ $\rho = 0.1$ $\rho = 0.2$ $\rho = 0.3$ $\rho = 0.4$								
	A. 상위 임금								
	.797 .627 .454 .277 .096								
	(.000.)	(.001)	(.014)	(.133)	(.596)				
수능	[.240]	[.191]	[.140]	[.086]	[.030]				
점수	B. 대기업 종사								
미포함	1.143	.974	.797	.613	.421				
	(.000.)	(000.)	(.000.)	(.002)	(.030)				
	[.392]	[.338]	[.281]	[.218]	[.151]				

주1. ()은 유의확률. []은 하계 효과를 나타냄

제 3 절 누락변수의 효과 제거

AET(2005) 방법론에서 Condition 3를 통해 구한 ρ 의 값의 최솟값과 최댓값을 적용해보았다. 동 방법론에서 잔차 간 상관관계인 ρ 의 최솟값은 $\rho = \frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)}$ 으로 가정했다.

<표 8>에서 최상위 대학의 노동시장 프리미엄, 즉, α 의 크기를 범위로 추정한 결과 상위 임금에 대한 프리미엄은 존재하지 않으나 대기업입사 확률에 대한 프리미엄은 존재하는 것으로 나타났다. 다만, 잔차 간 상관관계인 ρ 의 값이 최댓값을 가질 때는 프리미엄이 존재하지 않는다고 해석할 여지도 있다.

구체적으로 살펴보면, 누락변수가 최상위 대학에 입학할 확률과 상위임금에 해당할 확률에 동시에 양(+)의 영향을 미치지 않는다고 가정하면 $(\rho=0)$, 최상위 대학 졸업자는 비최상위 대학 졸업자 대비 상위 30% 임금에 속할 확률이 17.5%p 높다. 또한, ρ 는 0.542에서 최댓값을 갖는데,이때는 최상위 대학의 프리미엄이 음수(-)로 10% 수준에서 유의하게 나타나 프리미엄이 존재하지 않는 것으로 해석할 수 있다.

다음으로, 대기업 입사 비율에 대한 분석이다. 누락변수가 최상위 대학에 입학할 확률과 대기업에 입사할 확률에 동시에 양(+)의 영향을 미치지 않는다고 가정하면(ρ =0), 최상위 대학 졸업자는 비최상위 대학 졸업자 대비 대기업에 입사할 확률이 33.1%p 높은 것으로 나타났다. 그러나동 효과는 ρ 의 최댓값인 0.626에서 음수로 전환되었다.

<표 8>
ho의 최솟값과 최댓값에 따른 최상위 대학의 효과

수능 점수 미포함	A. 상위 임금		B. 대フ	업 입사
	^	^	^	^
	ho	α	ho	α
$\rho = \frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)}$.542	308	.626	149
		(0.086)		(0.416)
		[092]		[052]
$\rho = 0$	0	.600	0	.989
		(.002)		(.000.)
		[.175]		[.331]

주1. ()은 유의확률, []은 한계 효과를 나타냄

이처럼 ρ 의 최댓값에서 한계 효과가 음수로 나타났다는 점은 누락변수에 의한 편향이 클수록 프리미엄이 존재하지 않다고 해석할 여지가 있

음을 시사한다. 즉, ρ =0일 때, 선행연구와 같이 통계적으로 유의한 프리미엄이 나타났지만, 누락변수 효과가 발생함에 따라 프리미엄이 상실될 위험이 크다.

제 6 장 결론

본 연구에서는 한국교육고용패널(KEEP)의 1차 조사 자료 중에서 일 반계 고등학교 3학년생 코호트를 이용하여 대학 서열에 따른 노동시장 프리미엄을 분석하였다. 기존의 선행연구에서는 누락변수에 의한 편향을 고려하지 않았다.

그런데 노동시장 성과와 상위 대학 여부 모두에 영향을 미치는 개인의 성격과 역량을 모두 변수화하여 분석에 반영하는 것은 현실적으로 불가능에 가깝다. 따라서 누락변수의 영향을 인정하여 최상위 대학의 프리미엄에 편향이 존재할 가능성이 크다는 점을 고려해야 한다. 동 연구에서는 선행연구에서 고려하지 않은 누락변수에 의한 편향을 배제하기 위해 AET(2005) 방법론을 사용해 분석했다.

이에 따른 본 연구의 임금 프리미엄 분석 결과는 다음과 같이 요약할수 있다. 첫째, 누락변수로 인한 편향을 고려하지 않았을 때는 최상위 대학 졸업자가 비최상위 대학 졸업자보다 높은 임금을 받거나 대기업에 입사할 확률이 더 큰 것으로 나타났다. 둘째, 누락변수로 인한 편향을 최대치로 가정할 경우 최상위 대학의 노동시장 성과 프리미엄은 사라진다. 이때, 프리미엄이 사라지는 속도는 상위 임금에 속할 확률 프리미엄에서 더 높게 나타났다. 셋째, 개인의 인지 능력에 대해 고려하는 것이 중요함을 다시 재확인했다. 개인의 인지 능력은 최상위 대학 입학 및 노동시장성과와 밀접한 관련이 있어 분석 결과를 뒤바꿀 수 있을 정도의 편향을 발생시킨다.

다만, 본 연구는 다음과 같은 한계점을 갖고 있다. 첫째, 최상위 대학에 대한 정의에 따라 결과가 달라질 수 있다. 본 연구에서는 중앙일보

대학평가를 기준으로 13개 대학을 선정했지만, 선행연구처럼 대학별 입학성적 점수에 따라 정의하는 것도 가능하다. 이때, 몇 년도를 기준으로 하는지에 따라 대학 서열은 달라진다. 이러한 차이는 분석 결과에도 영향을 미칠 수 있다. 둘째, 연속형 변수인 월평균 임금을 더미변수로 전환하여 분석했다. 이는 정보의 손실을 발생시켜 분석 결과에도 영향을 미칠 수 있다. 본 연구에서는 Oster(2019) psacalc 패키지를 사용해 분석결과의 일치 여부를 확인하였다. 12) 하지만 여기서 더 나아가 민감도 분석을 시행하여 프리미엄이 변화하는 과정을 확인할 수 있다면, 보다 정확하고 신뢰할 수 있는 결과를 도출할 수 있을 것으로 생각된다.

¹²⁾ psacalc 패키지로 분석한 결과, 누락변수 고려 시 최상위 대학의 한계 효과는 음수로 전환되어 노동시장 성과 프리미엄이 통계적으로 유의하지 않다는 결과가 도출되었다.

참고문헌

- Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of political economy, 113(1), 151–184.
- Brand, J. E., & Halaby, C. N. (2006). Regression and matching estimates of the effects of elite college attendance on educational and career achievement. Social science research, 35(3), 749–770.
- Dale, S. B., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. The Quarterly Journal of Economics, 117(4), 1491–1527.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187–204.
- Schultz, T. W. (1961). Investment in human capital. The American economic review, 51(1), 1–17.
- Ko, E. (2011). Changes in wage differentials among college graduates in South Korea, 1999–2008. *Journal of Labour Economics*, *34*(1), 103–138.
- 김희삼, & 이삼호. (2008). 고등교육의 서열과 노동시장 성과. 한국개발연구원.
- 김진영, 「대학 서열화와 노동시장」, 『한국경제의 분석』, 제 13권 3호, 2007.
- 이지영, & 고영선. (2019). 대학 서열과 생애임금격차. *KLI 페널 워킹페이퍼*
- 장수명. (2006). 대학서열의 경제적 수익 분석. [KEDI] 한국교육, 0-0.

Appendix

1. 변수 정의

(종속변수) g2m1: 대기업 종사, g2m3: 상위 30% 임금 (설명변수) top: 최상위 대학

(통제변수) p_sat: 수능 성적, i_region: 거주 지역(수도권 여부), female: 여성, h_mon_income_ln: 월평균 가구소득, h_m_edu: 아버지 교육수준, h_w_edu: 어머니 교육수준, h_iq: 지능, h_attitude: 학교생활의 성실성(가정), i_family: 가정생활 만족도, i_health: 건강상태, i_sleeptime: 수면시간, t_study: 공부, t_leadership: 리더십, t_friendship: 친구 관계, t_schoollife: 학교생활의 성실성(담임)

2. 이변량 프로빗 모형 및 민감도 분석

global ylist1 g2m1 global ylist3 g2m3

global xlist4 p_sat h_mon_income_ln female h_m_edu h_w_edu h_iq h_attitude i_* t_*

```
forvalues i = 0(0.1)0.4{
local athrho=1/2*ln((1+(`i'))/(1-(`i')))
constraint define 1 _b[/athrho] = `athrho'
biprobit ($ylist3 = top $xlist3) (top = $xlist3), const(1)
margins, dydx(top) predict(pmarg1) force
}
```

3. 누락변수의 효과를 고려한 이변량 프로빗 모형

gen rhoold=1

local rhoold=rhoold

```
biprobit ($ylist3 = top $xlist4) (top = $xlist4)
predict xb, xb1
gen xb1 = xb - b[top]*top
egen sdxgamma = sd(xb1)
gen vxgamma=sdxgamma*sdxgamma
gen e1 = \$ylist3 - xb
egen mean_e1 = mean(e1)
egen sd_e1 = sd(e1)
gen var_el = sd_el*sd_el
predict xb2, xb2
egen sdxbeta = sd(xb2)
gen vxbeta=sdxbeta*sdxbeta
gen e2 = top - xb2
egen mean_e2 = mean(e2)
egen sd_e2 = sd(e2)
gen var_e2 = sd_e2*sd_e2
corr xb1 xb2, cov
gen covxbxg = r(cov_12)
gen rhonew = covxbxg/vxgamma * var_e1
sum rhonew
gen iter = 1
```

```
while abs(rhoold-rhonew)>0.001 {
replace rhoold=rhonew
local rhoold=rhoold
display "iteration 'iter'"
local iter = 'iter'+1
local athrho = 1/2*ln((1+('rhoold'))/(1-('rhoold')))
constraint define 1 _b[/athrho] = 'athrho'
biprobit ($ylist3 = top $xlist4) (top = $xlist4), constraint(1)
margins, dydx(top) predict(pmarg1) force
drop xb* sdx* vx* covx* rhonew *e1 *e2
predict xb, xb1
gen xb1 = xb - b[top]*top
egen sdxgamma = sd(xb1)
gen vxgamma=sdxgamma*sdxgamma
gen e1 = \$ylist3 - xb
egen mean_e1 = mean(e1)
egen sd_e1 = sd(e1)
gen var_e1 = sd_e1*sd_e1
predict xb2, xb2
egen sdxbeta = sd(xb2)
gen vxbeta=sdxbeta*sdxbeta
```

```
gen e2 = top - xb2
egen mean_e2 = mean(e2)
egen sd_e2 = sd(e2)
gen var_e2 = sd_e2*sd_e2

corr xb1 xb2, cov
gen covxbxg = r(cov_12)

gen rhonew = covxbxg/vxgamma * var_e1
sum rhoold rhonew
}
```

Abstract

Labor Market Performance Premium for Selective Schools

Min Heehyun
Economics
The Graduate School
Seoul National University

This study set out to analyze whether the selective universities have a significant impact on the labor market performance of the students. The probability of belonging to the upper wage and the probability of joining a large company were used. However, students entering selective universities are generally excellent in cognitive ability and have been grown up in good family environment, which is likely to lead to excellent employment performance after graduation. However, it is difficult to include all of these factors in the model, resulting in 'omitted variable bias'. Therefore, the conditional bivariate probit model of AET (2005) is applied in the study. This methodology assumes the magnitude of the bias due to the omitted variables in a certain range and checks whether the performance premium remains even after removing the maximum value of the bias from the performance premium. As a result, it was found that the performance premium of the selective universities disappears when

considering the bias due to the omitted variables.

keywords: selective school, wage, large companies, omitted

variable bias, bivariate probit model

 $Student\ Number: 2018-22738$