

勞 動 經 濟 論 集  
第42卷 第3號, 2019.9. pp.39~74  
© 韓 國 勞 動 經 濟 學 會

## 대졸청년층의 취업지역에 대한 자기선택을 고려한 임금함수 추정\*

이 치 호\*\*

본 논문에서는 2007~2013년 「대졸자직업이동경로조사」 데이터와 Dahl(2002)의 준모수적 방법론을 사용하여 선택편의를 보정한 대졸자들의 취업 초기 임금함수를 추정하고, 각 계수의 지역 간 차이가 선택편의에 기인한 것인지 검증하고자 하였다. 분석 결과, 선택편의를 보정하더라도 성별에 따른 임금 프리미엄의 지역 간 차이는 큰 변화가 없으며, 학력, 전공에 따른 임금 프리미엄의 지역 간 차이는 각각 약 18%, 11% 감소하여 여전히 지역 간에 큰 차이가 존재하는 것으로 나타났다. 한편, 서울 및 경기 지역에 대해서는 선택편의가 미미한 것으로 나타나 대졸청년층의 수도권 선호 경향은 성별, 학력, 전공 등 노동자들의 특성에 별다른 관계없이 일반적인 것으로 보인다.

주제어 : 지역 노동시장, 취업지역 선택, 임금함수, 선택편의, 준모수적 추정

### I. 서론

한 국가 내 각 지역 노동시장에서는 임금수준이나 교육 수익률(return to education), 성별에 따른 임금격차 등 여러 측면에서 이질성이 관찰된다. 예를 들어 Dahl(2002)에

논문 접수일: 2019년 4월 15일, 논문 수정일: 2019년 8월 2일, 논문 게재확정일: 2019년 8월 19일

\* 본 논문은 2017년 2월 서울대학교 대학원 경제학부에 제출된 이치호의 석사학위 논문과 2019 경제학 공동학술대회에서 발표한 원고를 수정하여 작성되었다. 논문 작성을 지도해 주신 이정민 교수님과 논문의 완성도를 높이는 데 귀중한 조언을 해 주신 익명의 심사자들에게 깊은 감사의 말씀을 드린다.

\*\* 서울대학교 경제학부 박사과정 (chiholee@snu.ac.kr)

의하면 미국의 1990년도 Census 자료를 가지고 분석한 결과 고등학교 졸업 대비 대학교 졸업의 임금 프리미엄은 지역에 따라 와이오밍 주의 22.1%에서 텍사스 주의 51.8%까지 다양하게 분포하는 것으로 나타났다. 우리나라의 경우에도 이원호(2002)에 따르면 2000년 노동패널 자료 기준 광역시 단위 분석에서는 표본 수 부족으로 인한 신뢰도 부족이라는 한계는 있으나 중학교 졸업 이하 대비 대학교 졸업의 임금 프리미엄이 인천의 8.4%에서 대전의 54.3%까지 분포하는 것으로 나타나 미국과 상황이 유사하였다. 그리고 Moretti(2011)에 따르면 이러한 노동시장 간 이질성은 시간이 지나면 자연스럽게 사라지는 일시적인 현상이 아니었으며, 국내의 경우 김동수(2011)는 2002년과 2008년 사이에 임금수준으로 봤을 때 각 지역 노동시장 간 이질성이 오히려 확대 경향이었음을 확인하였다. 이처럼 지역 간 거주 및 이전이 자유로워 노동공급이 탄력적일 것으로 보이는 상황에서 각 지역 노동시장에 여전히 이질성이 존재하는 것은 Roback(1982)의 일반 균형모형의 관점에서 보면 토지와 같이 각 지역 노동시장별로 공급이 비탄력적인 재화가 존재함에 따라 노동자에게 각 지역 노동시장이 무차별해지는 균형에서 명목임금이 다르게 나타나기 때문으로 이해된다. Moretti(2011)는 노동자들이 각 지역에 대한 특이한(idiosyncratic) 선호를 갖고 있어서 노동자들의 지역 이동이 무한히 자유롭지 않고 토지 혹은 주택의 공급 역시 완전히 고정되어 있지 않은 것으로 모형의 가정을 완화함으로써 각 지역 노동시장 간 이질성을 좀 더 현실에 가깝게 분석할 수도 있음을 보이기도 하였다.

이렇듯 각 지역 노동시장 간 이질성은 실제로 존재하는 것으로 보이며 이론적으로도 어느 정도 분석이 이루어지고 있지만, 노동시장 간 격차를 측정하고 그 격차가 발생하는 원인을 실증적으로 분석하는 데 있어서는 여전히 세심한 주의가 필요하다. 노동자가 사냥꾼과 어부의 평균 소득 차이가 단순히 사냥꾼과 어부라는 직업적 특성에만 기인하는 것이 아니라 해당 직업을 어떤 능력을 지닌 사람들이 주로 선택하는지에 따라 서로 발생할 수 있음을 논한 Roy(1951)의 선구적인 연구와 유사하게, 성별, 학력, 전공 등 특성이 균일하지 않은 노동자들은 외부의 힘에 의해 무작위로 각 지역에 배치되는 것이 아니라 자신의 지역에 따른 비교우위를 검토해 보고 효용을 극대화시킬 수 있는 지역을 선택해 취업을 할 것이기 때문이다. 이러한 노동자들의 자기선택을 생각하면, 노동자들의 잠재적인 선택을 포함하지 않는 실제 노동시장 데이터로부터 추정되는 지역별 임금수준이나 각 노동자의 비교우위로 작용하는 노동자의 다양한 특성에 대한 지역별 가격차, 이를테면 성별에 대한 가격으로서의 성별 프리미엄, 학력에 대한 가격으

로서의 교육 수익률, 전공에 대한 가격으로서의 전공 프리미엄 등에는 선택편의(selection bias)가 포함되어 있을 가능성이 높다. 따라서 지역 노동시장 간 격차가 실제로 유의하게 존재하는지, 유의하다면 그 격차가 어느 정도인지, 이론적인 설명이 유효한지를 검증하는 데 있어 이러한 선택편의의 보정은 필수적이라고 할 수 있다. 또한, 김준영(2016)이 지적하였듯이 청년 노동자들의 수도권 집중 현상이 심각한 우리나라의 현실에서, 지역 노동시장 간 격차를 해소하기 위한 정책의 필요성을 평가하는 데 있어서도 선택편의의 보정은 중요하다.

미국의 경우, Dahl(2002)은 노동자들의 취업지역 자기선택이 교육 수익률에 대해 유의한 양의 선택편의로 이어지지만 선택편의를 보정하더라도 지역 노동시장 간 수익률 격차는 사라지지 않음을 보인 바 있다. 지역 노동시장 간 수익률을 직접 다룬 것은 아니지만 노동자들의 자기선택에 따라 공간적 숙련 집중 현상이 지역 노동시장 간 임금 격차의 원인에서 상당한 비중을 차지한다고 보고한 Combes et al.(2008)과 같은 연구도 존재한다. 하지만 국내의 경우 아직 노동자들의 취업지역 자기선택이 취업지역별 임금 함수에 미치는 영향을 체계적으로 연구한 사례는 없는 것으로 보인다.

이러한 문제의식에 따라, 본 논문에서는 지역 노동시장의 노동력 이동 상황을 개괄하고, Dahl(2002)의 방법론에 따라 선택편의를 보정할 경우 노동시장에서 중요한 변수로 작용하는 성별, 학력, 전공에 따른 임금 프리미엄이 어떻게 달라지는지를 살펴보고자 한다. 단, 본 논문에서는 Dahl(2002)와 달리 대졸청년층의 취업 초기 상황을 가정하여 그들의 취업 초기 임금함수를 추정함으로써 지역 선택의 효과를 좀 더 명확히 드러내고자 하였다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서는 노동자들의 지역 선택에 관한 선행연구 결과들을 소개하고, 본 논문의 의의에 대해 논하고자 한다. 그리고 제Ⅲ장에서는 연구에 사용된 분석 방법과 데이터에 관하여 소개한다. 제Ⅳ장에서는 선택편의 보정 시 임금함수 추정 결과가 어떻게 변화하는지 살펴보고, 그 의미와 시사점에 대해 논할 것이다. 마지막으로 제Ⅴ장에서는 결론 및 본 논문의 한계와 향후 연구 과제를 제시할 것이다.

## II. 선행연구

노동자들의 지역 이동은 지역 노동시장의 내생적 동작 메커니즘에서 가장 중요한 요소 중 하나이지만, 실증분석 시 각 지역 노동시장 특성에 관한 추정 결과에 선택편의를 초래하는 요인이 되기도 한다. 이러한 노동자들의 자기선택이 노동시장 분석 결과에 영향을 미칠 가능성이 있음은 Roy(1951)가 일찍부터 지적하였는데, 실제 관측되는 데이터에는 노동자들의 실제 취업지역 선택 결과만이 반영될 뿐 그들의 실현되지 않은(잠재적인) 선택의 결과는 포함되지 않기 때문이다. 이러한 선택편의를 보정하지 않을 경우 데이터로부터 추정되는 지역별 교육 수익률이나 성별에 따른 임금차이 등은 해당 지역 노동시장의 특성을 제대로 반영하고 있지 못할 가능성이 높다.

현실에서 노동자들은 각 지역에서의 스스로의 비교우위에 대해 검토해 보고는 스스로의 효용 수준을 극대화하는 취업지역을 합리적으로 선택할 것으로 생각되는데, 이러한 선택의 결과가 개별 노동자의 특성에 관계없이 각 지역에 무작위로 취업하는 경우와 차이가 큰 경우 유의한 선택편의가 발생할 수 있다. 그리고 지역 노동시장 간 차이에 대해 분석한 많은 선행연구에서 각 지역 노동시장의 차이에 반응해 어떤 특정한 특성을 가진 노동자들의 집중 현상이 발생하는 것으로 분석되고 있으므로, 유의한 선택편의의 발생 가능성이 높다고 할 수 있다. Black(2009)은 이론적으로 노동자들의 선호가 완전히 동질적이지 않은 한 교육 수익률이 지역에 관계없이 상수일 수는 없고, 실증적으로도 1980년에서 2000년 사이에 미국의 각 대도시 사이에 교육 수익률 차이가 줄어들지 않았으며, 샌프란시스코나 시애틀과 같이 살기 좋은 것으로 인식되는(어메니티가 높은) 지역에서 교육 수익률이 낮게 나타나는 결과를 얻었다. Diamond(2016)는 공간적 균형모형을 사용해서 좀 더 이론적으로 이 문제에 접근하여, 고등학교 졸업자와 대학교 졸업자 간 불평등이 단순히 교육 수익률에 더해 도시의 임금, 부동산 가격, 어메니티 등에 의해 더 커질 수 있음을 실증분석 결과를 포함해 보인 바 있다. Combes et al.(2008)은 지역별 임금격차의 상당 부분이 숙련 노동자의 편중에 의한 것으로 분석하였는데, 이는 노동자들의 집중 현상과 일맥상통하는 것이다. 또한, Lee(2010)는 비교역재화를 공급하면서 구직 가능성이 지역에 비교적 덜 영향을 받는 의료계 종사자들을

대상으로 분석한 결과를 통해 도시 지역 임금이 높은 것은 고숙련 노동자들이 그곳에 살기를 선택하기 때문이며, 시골 대비 도시 지역 임금 프리미엄은 숙련도가 높을수록 감소하며 심지어는 음의 값을 가지기도 한다는 분석 결과를 제시하였다. 위와 같은 해외 사례들을 살펴보면, 수도권 집중 현상이 현저한 우리나라에서도 이러한 노동자들의 자기선택에 따른 집중 현상이 나타날 가능성이 높을 것으로 추측할 수 있다. 김준영(2016)은 청년층의 지역 이탈과 수도권 유입에 초점을 맞춰 노동자들의 수도권 집중 현상을 개괄한 바 있다. 이상호(2012) 역시 공간적 요인이 대학 서열화 효과를 통제한 후에도 청년대졸자의 첫 일자리 하향취업에 통계적으로 유의미한 영향을 미친다는 점을 확인하였다. 이같이 노동자들의 집중 현상이 현저한 상황에서, 선택편의의 보정은 각 지역 노동시장의 특성을 정확히 파악하는 데 있어 중요할 것으로 추측할 수 있다.

이렇듯 지역 노동시장 간 격차를 추정하는 데 있어 선택편의 보정 여부가 중요함에도 불구하고, 실제로 지역 노동시장 분석에서 선택편의 보정을 시도한 연구는 많지 않다. Dahl(2002)에서 지적했듯이 선택편의는 Heckman(1979)의 2단계 추정법과 같이 이항 선택모형에서 추정하는 것이 일반적이지만 노동자들의 취업지역 선택 문제는 선택의 가능성이 상정하는 단위 지역의 개수만큼 발생하기 때문에 결국 다항선택 모형으로 다루어야 하는데, 이 경우 수많은 선택지가 모두 노동자가 지역을 선택하는 기준에 반영되어야 하므로 선택편의를 도출하는 것이 쉽지 않기 때문이다. 이에 Dahl(2002)은 Lee(1983)의 최대순서통계량 방법(maximum order statistic approach)과 Ahn and Powell(1993)의 준모수적(semiparametric) 선택모형 추정 방법을 결합한 방법론을 개발하고 1990년 미국 센서스 자료를 사용해 각 지역 노동시장의 교육 수익률이 선택편의 보정에 따라 어떻게 달라지는지를 분석하였다. 그 결과, 노동자들은 교육수준에 따라 상당히 다른 지역 이동 양상을 보였으며, 취업지역 자기선택이 교육 수익률에 대해 약 5% 정도의 상향 편의를 발생시켰으나, 선택편의 보정 후에도 지역 노동시장 간 교육 수익률 격차는 거의 줄어들지 않아 지역 노동시장 간에 상당한 특성 차이가 존재함을 확인하였다. 즉 숙련 인력에 대한 노동수요는 각 지역에 따라 현저한 차이가 있으며, 노동자들 역시 그에 반응해 이동하지만 그럼에도 불구하고 전체 노동시장이 평균화될 정도로 차이가 상쇄되지는 않는다는 것이다.

하지만 국내의 경우 최근에서야 지역 노동시장을 어느 정도 들여다볼 수 있는 데이터가 나오기 시작한 만큼 관련 연구결과가 아직 충분히 축적되어 있지 않은 상황이다. 이원호(2002)는 한국노동패널(KLIPS) 데이터를 가지고 교육 수익률을 포함 우리나라 광

역대도시 지역 노동시장별로 임금함수를 추정하였으나, 노동패널 자료 특성상 표본 수가 제한적이므로 추정치의 신뢰도가 높지 않으며 선택편의에 대한 고려가 이루어지고 있지 않다. 김동수(2011)는 산업·직업별 고용구조조사(OES) 자료를 사용해 지역별 고용 수준과 지역 간 및 지역 내 임금격차를 분석하고 있으나 임금함수 등을 추정하는 분석은 하지 않았다. 김우영(2012)은 KLIPS, OES 외에 통계청의 지역별 고용조사(RES) 데이터를 사용하여 지역 간 임금격차와 지역 내 임금격차를 추정하였고, 취업 여부에 따른 선택편의를 고려하고 있으나 지역에 상관없이 동일 교육 수익률, 성별 프리미엄 등을 가정하고 지역 차이를 더미 변수만으로 처리하였기 때문에 노동자들의 취업지역 자기선택의 영향과 지역 간 임금격차의 원인에 대해서는 시사점을 얻을 수 없었다. 이상호(2012)는 2009년 「대졸자취업이동경로조사(GOMS)」 데이터를 이용해 공간적 요인이 청년대졸자의 첫 일자리 하향취업(어떤 취업자가 자신의 일자리에 요구되는 것보다 더 많은 교육을 받은 상태를 지칭)에 어떤 영향을 미치는지 분석하고 있는데, 노동시장의 공간적 효과를 고교 및 대학 소재지와 직장소재지 정보를 이용해 상세히 파악하고자 한 장점이 있다. 하지만 월평균 임금에 미치는 영향을 중심으로 분석하지는 않았으며, 역시 교육 수익률 등을 지역에 상관없이 동일하다고 가정하고 있다는 점에서 한계가 있다. 최광성 외(2018)는 본 논문과 유사한 문제의식을 가지고 선택편의를 포함한 내생성을 고려하여 성향점수매칭법(propensity score matching method) 및 이중차분매칭법(difference-in-difference matching method)을 사용하여 지역 노동시장 간 격차 문제를 분석한 중요한 예로서, 비수도권에서 수도권으로 이동할 경우 통계적으로 유의한 임금상승 효과가 나타난다는 사실을 보고하였다. 하지만 분석 방법의 차이와 함께 수도권과 비수도권 간 차이에 집중하였다는 점에서 본 논문과는 지향점에 차이가 있다.

이러한 상황을 고려하여 본 논문에서는 Dahl(2002)이 사용한 계량경제학적 분석 방법을 도입하여 각 지역 노동시장에 따른 임금함수를 추정하고 그것이 선택편의 보정 여부에 따라 어떻게 달라지는지 분석하고자 한다. Dahl(2002)에서는 분석대상을 25세에서 34세의 백인 남성으로 하고 주로 가족 관계를 가지고 집단을 나누어 선택 양상을 분석하였으나, 본 논문에서는 35세 미만 대졸청년층을 분석대상으로 한다. 이는 다양한 취업지역 선택의 계기 및 여러 번에 걸친 지역 이동의 가능성을 피하고 취업지역 선택의 계기를 대학 졸업이라는 공통된 사건으로 압축함으로써, 취업지역 선택의 효과를 좀 더 명확히 추정하기 위한 것이다. 또한, 성별이나 학력, 전공 등 노동시장에서 중요하게 취급되며 최초 취업 시 선택 양상을 잘 설명할 것으로 보이는 노동자의 다른 특성

들을 포함함으로써 취업지역 선택 과정을 현실에 가깝게 모사하고 실질적인 임금함수를 추정하려 하였다.

### III. 분석 방법

#### 1. 분석 모형

Dahl(2002)은 노동자들의 취업지역 선택 문제를 모형화할 때 발생할 수 있는 문제로 크게 다음과 같은 4가지를 거론하고 있다. 우선, 차원의 저주(curse of dimensionality) 때문에 계산해야 할 변수의 양이 폭증한다. 두 번째로, 선택확률의 분포에 대한 가정이 어렵다. 다항선택모형 중에서 자주 쓰이는 조건부 로짓 모형이나 그 확장형인 네스티드(nested) 로짓 모형(McFadden(1973, 1984), Trost and Lee(1984), Falaris(1987))은 선택확률이 다루기 편리한 형태의 함수로 나오지만, 그 때문에 특수한 선택 상황을 잘 반영하지 못하는 면이 있다. 세 번째로, 선택편의 크기의 분포에 대해서도 사전적으로 가정이 어렵다. 만약 실제 분포가 비선형이나 단조적이지 않은(nonmonotonic) 특성을 가진다면 실제 임금 회귀식에서의 선택편의의 보정이 부정확해질 가능성이 높다. 마지막으로, 관측되지 않거나 관측이 부정확한 변수들의 영향이 문제가 될 수 있다.

여기서 일단 문제를 명확히 하기 위해 다음과 같이 모형을 설정하고 나서 논의를 진행하기로 한다. 어떤 지역  $k$ 에 취업한 어떤 노동자  $i$ 의 로그임금  $y_i$ 가 다음 식 (1)과 같은 임금함수를 통해 추정된다고 하자.  $i$ 는 노동자  $i$ 가  $k$ 지역에 취업했을 경우에만 관측된다.

$$y_i = \alpha + \beta X_i' + \gamma SEX_i + \delta EDU_i + \theta MAJOR_i + u_i \quad (1)$$

식 (1)에서  $X_i$ 는 노동자  $i$ 의 나이, 학점 등으로 이루어진 개인의 특성을 나타내는 벡터이며,  $SEX_i$ 는 성별,  $EDU_i$ 는 학력,  $MAJOR_i$ 는 대학 전공을 나타내는 변수들이다.  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\theta$ 는 각각의 변수에 대한 회귀계수이다. 여기서 임금 데이터는 해당 노동

자가  $k$ 지역에 실제로 취업한 경우에만 관측되고 그가 다른 지역에 취업했을 경우의 가상적인 임금까지는 반영하고 있지 않으므로, 이 임금함수 내의 오차항  $u_i$ 에는 그 크기가 임금이므로 표현되는 선택편의가 포함되어 있다고 본다. 여기서 노동자의 능력이 취업지역에 대한 선호와 상관관계가 있다고 본다면, 선택편의가 각 지역을 취업지역으로 선택할 확률의 함수로 표현된다고 가정할 수 있을 것이다. 따라서 출신 지역이  $j$ 인 노동자  $i$ 가 취업지역으로  $k$ 지역을 선택할 확률을  $p_{ijk}$ 라고 가정하면, 오차항  $u_i$ 로부터 지역 선택확률의 함수인 선택편의 보정함수 부분  $\lambda_j$ (노동자의 출신 지역에 따라 달라진다고 가정)를 분리해 선택편의와 무관한 진정한 오차항  $\omega_i$ 를 구할 수 있을 것이다. 결국, 위의  $k$ 지역에서의 임금함수를 다음 식 (2)와 같이 나타낼 수 있다.

$$y_i = \alpha + \beta X_i' + \gamma SEX_i + \delta EDU_i + \theta MAJOR_i + \sum_j M_{ij} \lambda_j(p_{ij1}, p_{ij2}, \dots, p_{ijN}) + \omega_i \quad (2)$$

$$u_i = \sum_j M_{ij} \lambda_j(p_{ij1}, p_{ij2}, \dots, p_{ijN}) + \omega_i, \quad (j=1, \dots, N) \quad (3)$$

$$E(\omega_i | \text{노동자 } i \text{가 취업 지역 } k \text{를 선택}) = 0 \quad (4)$$

$$M_{ij} = 1 \text{ if 노동자 } i \text{가 } j \text{지역 출신이며 현재 } k \text{지역에 거주, } M_{ij} = 0 \text{ else} \quad (5)$$

여기서 선택편의 보정함수를 계산하려고 할 때 문제가 되는 부분은 크게 2가지이다. 첫째, 선택지(취업지역)가 다수 존재하는 상황에서 효용함수로 표현되는 노동자들의 취업지역 선택기준을 사전적으로 알기 어려우며, 따라서 선택확률을 계산하기 어렵다. 둘째, 그 선택확률의 함수로 생각되는 오차항의 선택편의 부분의 함수  $\lambda_j$ 의 형태를 사전적으로 알 수 없다. 결국, 이 2가지 문제는 위의 Dahl(2002)에서 언급한 4가지 문제점을 다른 방식으로 표현한 것이다.

위의 문제들을 회피하기 위해 Dahl(2002)에서는 다음과 같은 분석 방법을 개발하였다. 우선, 차원의 저주를 회피하기 위해서 Lee(1983)의 최대순서통계량 방법을 사용해 선택에 사용되는 식의 숫자와 개수를 줄인다. 즉 모든 지역의 정보를 가지고 복잡한 효용함수를 임의로 만들어내는 것이 아니라 가장 효용이 높은 하나의 지역에 대한 함수로 본다는 것이다. 나아가, 가장 높은 단 1개의 선택확률만으로도 오차항의 선택편의를 계산하는 데 필요한 정보는 충분하다고 가정한다. 이러한 가정들을 통해 선택편의를 보정하는 데 필요한 변수의 개수를 크게 줄일 수 있다. 또한, 그 단 1개의 선택확률



이 가리키는 취업지역은 실제로도 선택되었을 것이므로, 실제 관측된 지역 선택 양상은 선택편의 계산에 사용된 1개의 선택확률과 높은 상관관계를 보일 것이다. 물론 이러한 가정이 실제로 타당한지를 검증하기는 쉽지 않으나, 실제 사람들의 취업지역 선택 과정이 한 번에 모든 지역에 대해 취업을 검토하기보다는 1차로 특정 2, 3개 지역으로 선택지를 압축한 후 그중에서 다시 최종적으로 선택하는 방식에 더 가까울 것이므로 그렇게까지 비현실적인 가정은 아니다. 여기서 좀 더 현실에 가깝게 가정을 완화하여 실제 취업한  $k$  지역에 대한 선택확률  $p_{ijk}$ 에 더해 노동자 개인의 출신 지역  $j$ 에 그대로 취업할 확률  $p_{ijj}$ 의 2가지 확률을 선택편의 보정함수에 반영시킨다면, 선택편의 보정함수 부분  $\lambda_j$ 는 다음과 같이 단순화될 수 있다.

$$\lambda_j(p_{ij1}, p_{ij2}, \dots, p_{ijN}) \approx \lambda_j(p_{ijk}, p_{ijj}) \quad (6)$$

한편, 노동자들의 취업지역 선택기준 설정과 선택확률 계산을 위해서는 사전적으로 어떤 모형을 세우는 것이 아니라 실제로 관측되는 노동자들의 선택 결과를 가지고 역으로 유추하는 방식을 사용한다. 노동자들의 각 지역에 대한 선택확률은 데이터로 수집되어 있지 않지만, 유사한 특성(인종, 결혼 유무, 동거인들 등)을 가진 노동자들로 구성된 집단 내에서는 모든 노동자들이 동일한 선택확률  $p_{ijk}$ 를 공유한다고 가정한다면 집단에 속하는 노동자들의 선택 결과를 취합함으로써  $p_{ijk}$ 를 준모수적으로 (semiparametrically) 생성할 수 있다. 다시 말해 실제로 해당 노동자가 취업지역으로 선택한 지역의 선택확률뿐만이 아니라, 선택하지 않은 지역에 대한 선택확률까지 동일 집단 내 다른 사람들의 선택 결과를 마치 해당 노동자가 한 것처럼 가정함으로써 모든 취업지역에 대한 선택확률을 유추할 수 있다. 이렇게 얻은 각 노동자의 모든 취업지역에 대한 선택확률을 자기선택에 의한 선택편의 보정함수  $\lambda_j$ 를 계산하는 데 사용한다. 여기서 이해를 돕기 위해 아래와 같은 간단한 예를 살펴보자. 똑같이 서울 출신인 김씨(취업지역 서울), 이씨(서울), 박씨(서울), 최씨(대구), 정씨(전남)의 5명의 노동자로 이루어진 집단이 있다고 가정하자. 위에서 언급한 대로 가정하면, 이 5명의 노동자는 모두 다음과 같은 취업지역 선택확률을 공통적으로 가지고 있는 것으로 볼 수 있다.

$$P_{i, Seoul, Seoul} = 3/5, \quad P_{i, Seoul, Daegu} = 1/5, \quad P_{i, Seoul, Jeonnam} = 1/5, \quad P_{i, Seoul, etc} = 0 \quad (7)$$

물론 이 예시에서는 집단의 구성원이 5명뿐이라 조금 비현실적으로 보이지만, 집단의 구성 인원을 늘릴수록 선택확률을 실제와 가깝게 모사할 수 있을 것이다. 한편, 이렇게 준모수적으로 구한 취업지역 선택확률  $p_{ijk}$ 로 선택편의 보정함수  $\lambda_j$ 를 구할 경우, 회귀분석에서 추정된 계수가 일치추정량이 되는 것은  $p_{ijk}$ 가 연속일 때이다 (Newey(1997), Andrews(1991)). 그러므로 일치성을 확보하기 위해서는 집단의 개수를 늘림으로써  $p_{ijk}$ 의 가짓수를 충분히 늘려  $p_{ijk}$ 가 연속함수에 가까워지도록 해야 한다. 결국, 취업지역 선택확률을 현실에 가깝게 하려면 집단의 구성 인원을 늘리는 동시에 전체 모집단을 좀 더 다양한 집단으로 나누어야 한다는 두 가지 요건을 동시에 만족시켜야 하므로, 많은 표본이 요구된다. 이것은 Dahl(2002)의 방법의 단점이라고 할 수 있지만, 표본이 많을 경우에는 편리하고 강력하기도 하다.

그리고 오차항 내 선택편의가 사전적으로 어떤 분포를 가지는지 알 수 없다는 문제에 대해서는 비모수적 회귀분석(nonparametric regression)을 사용함으로써 특정 분포를 가정하지 않고 직접 선택편의의 분포함수를 얻는 방법으로 회피한다. 이상의 논의들을 적용하면, 오차항에서 선택편의 보정함수 부분을 분리한  $k$ 지역에서의 임금함수는 다음과 같이 나타낼 수 있다.

$$y_i = \alpha + \beta X_i' + \gamma SEX_i + \delta EDU_i + \theta MAJOR_i + \sum_j M_{ij} \lambda_j^*(p_{ijk}, p_{ijj}) + \omega_i \quad (j = 1, \dots, N) \quad (8)$$

선택편의 보정함수는  $\lambda_1, \dots, \lambda_N$ 으로 표시되며, 지역  $j$ 에서  $k$ 로 이동한 경우에 대해 모두 선택편의 보정함수가 필요하므로 그 개수는 모두  $N$ 개가 되며 비모수적으로 추정된다. 일반적으로 비모수적 추정 방법에서 기저함수(basis function)로는 다항식 전개와 푸리에 전개가 주로 쓰이는데, 여기서는 다항식 전개를 각 선택확률의 2차까지 사용하기로 한다. 3차항을 포함시키더라도 결과는 크게 달라지지 않는다. 결과적으로 선택편의 보정함수  $\lambda_j^*$ 는  $p_{ijk}, p_{ijj}$ 의 1차, 2차항과 그 교차항들의 선형결합으로 표현되며, 임금함수와 함께 일반최소자승법(OLS)로 추정될 수 있다. 여기서 출신 지역과 동일 지역에 취업하는 노동자(정주자, stayer)의 경우  $j = k$ 이므로, 선택편의 보정함수의 변수는 출신 지역  $k$ 에서 그대로  $k$ 지역에 취업할 확률,  $p_{ikk}$ (<표 1> 및 <표 7>에서는 잔류확률

로 표기) 하나가 된다. 한편 고등학교 출신 지역을 떠나 다른 지역에 취업하는 노동자 (이주자, mover)에 대한 선택편의 추정함수의 변수는 고등학교 출신 지역  $j$ 에서 그대로  $j$ 지역에 취업할 확률  $p_{ijj}$ (<표 1> 및 <표 7>에서는 미실현 잔류확률)와  $p_{ijk}$ (<표 1> 및 <표 7>의 전입확률, 해당 노동자에 의해 실현)의 2가지가 된다. 이제 각 취업지역  $k$ 에 서의 임금함수는 다음 식 (9)와 같이 나타낼 수 있다.

$$\begin{aligned}
 y_i = & \alpha + \beta X_i' + \gamma SEX_i + \delta EDU_i + \theta MAJOR_i \\
 & + M_{i1}\lambda_1^*(p_{\text{노동자 } i \text{의 출신 지역이 } 1, \text{ 취업 지역이 } 1}, p_{\text{노동자 } i \text{의 출신 지역이 } 1, \text{ 취업 지역이 } k}) \\
 & + M_{i2}\lambda_2^*(p_{\text{노동자 } i \text{의 출신 지역이 } 2, \text{ 취업 지역이 } 2}, p_{\text{노동자 } i \text{의 출신 지역이 } 2, \text{ 취업 지역이 } k}) + \dots \\
 & + M_{ik}\lambda_k^*(p_{\text{노동자 } i \text{의 출신 지역이 } k, \text{ 취업 지역이 } k}) + \dots \\
 & + M_{iN}\lambda_N^*(p_{\text{노동자 } i \text{의 출신 지역이 } N, \text{ 취업 지역이 } N}, p_{\text{노동자 } i \text{의 출신 지역이 } N, \text{ 취업 지역이 } k}) + \omega_i \\
 & (j = 1, \dots, N)
 \end{aligned}
 \tag{9}$$

거기에 더해, Dahl(2002)은 현실적으로 각 취업지역마다 출신 지역 수(미국의 경우 51)만큼의 선택편의 보정함수를 추정하는 것은 계산상 현실적이지 않다고 생각하여, 다음과 같이 취업지역과 출신 지역이 다른 경우(이주자)는 출신 지역에 상관없이 동일한 선택편의 보정함수를 따른다고 가정하였다. 이렇게 되면 취업지역당 선택편의 보정함수는 정주자와 그렇지 않은 경우(타 지역 출신)에 대한 2가지만 남게 되어 추정이 더 간단해진다.

$$\lambda_j^* = \lambda_{\text{타 지역 출신}(mover)}^* \quad \forall j \neq k \tag{10}$$

이 논문에서 분석에 사용한 취업지역  $k$ 에 대한 선택편의 보정함수를 포함한 임금함수는 최종적으로 다음 식 (11)과 같다. 이 임금함수를 각 지역별로 추정할 것이다.

$$\begin{aligned}
y_i &= \alpha + \beta X_i' + \gamma SEX_i + \delta EDU_i + \theta MAJOR_i & (11) \\
&+ M_{i,j=k} \lambda_{stayer}^* (p_{ikk}) + M_{i,j \neq k} \lambda_{mover}^* (p_{ijj}, p_{ijk}) + \omega_i \\
&(j = 1, \dots, N)
\end{aligned}$$

## 2. 데이터 및 변수 설명

본 논문은 한국고용정보원의 「대졸자직업이동경로조사」(GOMS)의 2007-2013년 졸업생 데이터를 분석대상으로 한다. 「노동패널」이나 「지역별 고용조사」 등의 데이터를 사용할 수도 있지만, 굳이 GOMS 데이터를 사용하는 것은 다른 데이터들에 비해 몇 가지 장점이 있기 때문이다. 우선 2007년 졸업생 데이터 기준으로 전체 모집단이 498,827명, 목표표본이 18,301명(모집단의 약 3.7%)으로 표본 수가 절대적으로도 상대적으로 많은 편이며, 취업지역 선택에 영향을 미칠 것으로 생각되는 변수들이 GOMS에 더 풍부하게 포함되어 있다. 또한, 대학 졸업 1년 후에 데이터가 수집되므로 취업지역 선택의 계기가 대학 졸업이라는 사건으로 압축될 수 있으며, 노동자들의 나이 분포가 좁고 경력연수 역시 2년 이하로 통제된다는 점에서 순수하게 지역 선택에 따른 영향을 보는 데 있어 유리한 면이 있다.

한편 데이터를 실제로 분석에 사용하기 위해서는 여러 가지 정제 작업이 필요하다. 우선, 표본이 많이 필요한 분석대상과 분석 방법을 채택하고 있기 때문에 표본 수를 더 확보하기 위해 상기 기간 동안 대졸자들의 지역 선택 경향이 크게 달라지지 않았을 것이라는 가정하에 7개년의 데이터를 하나로 모아서 사용하며, 각 연도에 따른 임금의 차이 등은 연도 더미(dummy) 변수를 씌워서 통제하기로 한다. 또한, 대졸자 중 소득이 없거나 월평균 30만 원 이하이거나 2000만 원 이상인 노동자들을 결과를 왜곡시킬 수 있다고 판단하여 분석대상에서 제외하였다. 교육대학 졸업자의 경우 졸업 후 경로가 거의 확정적이고 대부분 공립초등학교에 근무하면서 동일한 임금체계를 적용받고 있으므로, 취업지역 선택이 임금 등에 영향을 미칠 가능성이 거의 없어 분석에서 제외하였다. 그리고 가상적으로 사회 초년생들을 분석대상으로 상정하고 있기 때문에 직장 재직 중 학위를 취득하는 경우 등을 제외하기 위해 나이가 35세 이상인 노동자들은 분석 대상에서 제외하였다. 또한, 기타 설명변수나 집단 분류 기준 등으로 사용되는 성별, 학점(만점이 4.0, 4.3, 4.5인 대학들이 혼재하므로 학점을 100점 만점으로 변환해서 사

용), 학력(2년제, 4년제), 전공(인문계열, 사회계열, 교육계열, 공학계열, 자연계열, 의약계열, 예체능계열), 출신 고등학교 지역, 출신 대학교 지역, 취업지역, 출생 지역, 부모님 교육수준 등이 누락된 표본은 제거하였다. 그리고 노동자들의 지역 이동을 다루는 연구 목적상 표본으로 채택된 데이터는 전체 표본에 대해 지역, 출신 대학교 등의 각 변수의 지역 비율을 그대로 유지하여야 한다. 「대졸자직업이동경로조사」 데이터에는 표본 추출에 따른 비중 변화를 보정하기 위한 가중치가 포함되어 있으므로, 그것을 이용한다. 위에서 논의한 각종 데이터 정제 작업을 거치면서 표본들이 더러 탈락하지만, 가중치에 대해서는 대체로 무작위로 탈락할 것이라고 가정한다.

한편, 분석대상의 지역 단위를 어떻게 설정할 것인지 역시 중요한 문제이다. 개념적으로는 실제로 하나의 지역 노동시장으로 기능하고 있는지에 따르는 것이 적절한 것으로 생각된다. 이상호(2008)는 통근권에 기초해 미국의 MSAs(metropolitan statistical areas), 유럽의 고용특구(employment zones)와 유사한 지역노동시장권을 설정하고 지역별 특성을 분석한 결과, 행정구역에 기반한 지역 구분에 비해 인구이동을 조금 더 잘 설명하는 것으로 보고한 바 있으나, 본 논문에서는 데이터의 한계로 일단 세종특별자치시를 충청남도에도 포함시킨 형태의 16개 광역 시도 기준으로 분석을 진행하기로 한다. 이는 노동자들이 취업지역을 선택할 때 생각하는 지역의 개념은 통근권이라는 세밀한 지역 구분보다는 일반적인 16개 시도에 더 가까울 것으로 보인다는 점에서는 타당성이 있다.

이상과 같이 임금함수에 필요한 설명변수들은 거의 포함되었다고 생각하지만, GOMS에 출신 대학교를 특정할 수 있는 데이터가 없어 출신 대학교 평판은 완전히 통제할 수 없었으며 출신 대학교 지역 변수를 통해 부분적으로만 통제할 수 있었다. 노동자들의 평균적인 능력이 출신 대학교에 따라 차이가 존재하며 기업들에게는 노동자의 출신 대학교가 능력을 판단하는 데 신호(signal)로도 작용할 수 있는 상황에서 이를 충분히 통제하지 못한 것은 누락변수에 의한 편의를 발생시킬 수 있다. 이는 본 논문의 주요한 한계 중 하나이다.

### 3. 집단 분류 기준에 대한 검토

집단을 어떻게 나눌 것인가를 설정하는 것은 노동자들이 취업지역을 선택할 때 어떤 기준으로 하는지를 가정하는 것과 같은 일이라고 할 수 있다. Dahl(2002)에서는 각 거주 지역에 대해 노동자들을 출신 지역과 동일 지역에 취업하는 경우와 그렇지 않은 경

우로 일단 나누는데 이때 출신 지역으로는 출생지를 사용한다. 거기에 추가로 교육수준, 일하는 배우자가 있는지 여부, 5세 이하 자녀가 있는지 여부, 5세에서 18세 사이의 자녀가 있는지 여부, 이혼 여부, 가족과의 동거 여부, 친구와의 동거 여부 등으로 집단을 나누었는데, 교육 수준을 제외하면 주로 가족관계에 근거해 나누었다고 할 수 있다.

본 논문에서는 Dahl(2002)과는 조금 다른 집단 구분을 사용한다. 실제 노동자의 교우관계나 지역 기반, 생활양식 등은 출생지보다는 어느 정도 성장한 후인 학창시절에 주로 형성될 가능성이 높다고 생각되기 때문에, 출신 지역은 출신 고등학교 지역으로 정의할 것이다. 즉 여기서 출신 지역과 동일 지역에 취업한다는 것은 출신 고등학교 지역과 같은 지역에 취업한다는 것을 의미한다. 또한 Dahl(2002)에서는 백인 남성으로 분석대상을 한정하였지만, 성별에 따라서도 취업지역 선택에 차이가 나타날 것으로 생각되기 때문에 성별을 집단 분류 기준에 포함하였다. 학력, 즉 교육대학을 제외하고 2-3년제 대학 출신인지, 4년제 대학 출신인지에 따라서도 취업지역 선택은 달라질 수 있으므로 분리한다. 이외에 출신 대학교 지역과 전공계열 역시 취업 시 중요한 역할을 할 것으로 보아 포함하였다.

이상과 같이 출신 지역마다 네 가지 분류 기준, 즉 성별, 학력, 전공, 출신 대학교 지역(10개 지역으로 분류)을 가지고 집단을 구성하였다. 기본적으로 성별과 학력의 영향은 핵심적으로 파악하고 싶은 목표에 해당하므로 성별과 학력은 기본적으로 집단 구성 기준으로 포함하였으며, 그 외에 다른 기준들은 집단의 가짓수와 집단 내 표본 수의 균형을 맞추기 위해 조정한 결과이다. 전공계열이나 출신 대학교 지역의 세분화를 달리하거나 부모님의 학력이나 대학교 학점 등의 다른 변수를 사용하는 방법도 검토해보았으나, 각 기준을 단독으로 집단 분류 기준으로 사용했을 때의 지역 선택 확률들을 보았을 때 전공계열, 출신 대학교 지역에 비해 비슷하거나 더 단조로운(표준편차가 작은) 지역 선택확률로 연결되었기 때문에 채택하지 않았다. 이는 <표 1>에 정리되어 있다.

결과적으로 집단의 총 개수는 (출신 고등학교 지역 : 광역시도 16) × (성별 : 2) × (학력 : 2-3년제, 4년제 이상 2) × (전공계열 : 인문, 사회, 공학, 자연, 의약, 예체능 6) × (대학권역 : 서울, 경기(인천 포함), 강원, 충북, 충남(대전 포함), 경북(대구 포함), 경남(부산, 울산 포함), 전북, 전남(광주 포함), 제주 10) = 3840 이 되어야 하지만, 동일 집단 내 표본 수가 너무 적으면 현실에 가까운 지역 선택확률의 모사가 어려워지기 때문에 표본이 8개 미만인 집단은 불가피하게 분석에서 제외하였다(전체 표본 중 약 6%

수준, Dahl(2002)의 경우 표본 수가 10개인 집단까지 제거하였다). 따라서 실제 분석에 사용된 집단의 개수는 3,840개보다 적다. 이러한 데이터 정제 과정에 따라 실제로 거의 일어나지 않는 이동 패턴의 기여분은 무시된다는 점은 분석 방법상의 한계이므로 주의해야 한다.

<표 2>는 본 논문에서 사용된 각 설명변수들의 기술통계량을 정리한 것이다. 교육대학 졸업자와 표본 8개 미만인 집단에 속한 사람은 표본에서 제외되어 있다. 대졸자들의 취업 초기 월 평균임금은 지역별로 상당한 차이가 있으며, 가장 임금이 낮은 제주 지역은 가장 임금이 높은 울산 지역 대비 임금이 약 21% 낮음을 알 수 있다(단, <표 2>에 제시된 월 평균임금은 7개년도 임금을 물가나 지역별 성장률 등 별도의 보정 없이 동일하게 보고 계산한 것임에 유의해야 한다). 나이나 100점 만점 환산 GPA에는 지역별로 큰 차이가 없으나, 여성 비율, 4년제 대학 졸업자 비율, 취업자 중 정주자 비율은 각각 지역 간 최대 14.6%p, 17.3%p, 41.5%p까지 차이가 나고 있어 대졸자들이 취업 지역을 무작위로 선택하고 있지 않음을 알 수 있다.

참고로 취업자들의 속한 집단의 종류의 각 지역 간 총합은 집단의 총 개수와 일치하지 않는다. 각 집단은 출신 고등학교 지역이 동일하면서 서로 다른 지역으로 이동해서 취업하는 노동자들을 모두 포함하고 있기 때문이다.

## IV. 분석 결과

### 1. 전이행렬

노동자들의 취업지역에 대한 선택이 무작위로 이루어지는 것이 아니라 어떤 기준에 의거해서 이루어진다면, 그 기준에 따라 노동자들의 이동이 달라질 것이며 따라서 선택편의가 발생할 것이라는 것이 이 연구의 출발점이 되는 가설이다. 따라서 임금함수 추정에 앞서 인구이동 현황이 실제로 각 집단 분류 기준에 따라 다르게 나타나는지를 살펴볼 필요가 있다. 따라서 성별, 학력 등의 각 노동자들의 특성에 따라 출신 고등학교 지역으로부터 취업지역으로의 이동 비율을 나타내는 전이행렬(transition matrix)을 만

들어 봄으로써 그 원인을 어느 정도 분해해 보고자 한다. 이 절에서 제시하는 표들은 모두 데이터에 가중치를 적용하였다. 진한 음영은 행과 열의 지역이 같은 경우, 옅은 음영은 값이 5% 이상일 경우 표시하였다. 본 논문의 다른 분석 결과와는 달리, <표 3>에서 <표 6>은 전체적인 인구이동을 좀 더 정확히 살펴보기 위해 교육대학 졸업자와 표본 8개 미만인 집단에 속한 사람도 포함하였다.

<표 3>은 각 지역 고등학교 출신 노동자들이 어느 지역으로 취업하는지를 표로 정리한 것이다. 출신 고등학교 지역이 비수도권인 경우 해당 지역에서 취업하는 경우가 대략 50% 내외로 비율상 가장 높지만, 모든 출신 고등학교 지역에서 노동자들이 수도권, 즉 서울과 경기 지역으로 이동해 취업하는 흐름이 공통적으로 나타난다. 즉, 서울과 경기 지역은 노동자들에게 있어서 벗어나고 싶지 않거나(서울, 경기 지역 출신) 진입하고 싶은(타 지역 출신) 지역이라고 할 수 있으며, 이는 수도권 집중 현상이 잘 드러나는 대목이라고 할 수 있다. <표 3>의 결과를 각각 성별, 학력(대학 종류), 전공계열(인문, 공학계열만 뽑아서 표시)에 따라 다시 나누어서 살펴본 것이 <표 4>, <표 5>, <표 6>이다. 이에 따르면 남성에 비해서 여성, 2-3년제 대비 4년제 대학 졸업자, 공학계열 전공자 대비 인문계열 전공자의 서울 지역 선택확률이 상대적으로 높게 나타났다. 이러한 결과로부터 노동자들은 각 지역에 무작위로 취업하는 것이 아니라 출신 고등학교 지역, 학력, 성별, 전공계열 등 본인의 다양한 특성을 고려해서 최적의 취업지역을 선택하고 있음을 짐작할 수 있다.

## 2. 선택편의 보정 시 임금함수 추정 결과

제1절에서 취업지역 선택이 무작위적인 것이 아니라 노동자의 여러 특성에 따라 다르게 이루어짐을 확인할 수 있었으므로, 이번에는 그것이 임금함수의 각 회귀계수의 추정치에 얼마나 큰 선택편의를 발생시켰는지 제III장에서 설명한 방법론과 데이터에 따라 검증해보기로 한다. 먼저 선택편의 보정 시 각 취업지역(16개 광역시도)에 대한 임금함수 추정 결과는 <표 7>과 같이 나타난다. 추정 시 모든 데이터에 가중치를 적용하였다. 각 회귀계수들은 대체로 유의하게 추정되었으며, 각 지역에 따라 상당한 격차가 있음을 알 수 있다.<sup>1)</sup> 예를 들어 서울 지역의 경우 성별에 따른 임금차이가 전국에

1) 참고로 집단 구분 기준으로는 10개 지역 구분을 사용하였으나 회귀분석 시에는 대학권역 대신 16개 시도로 이루어진 출신 대학 지역 더미를 그대로 통제변수로 적용하였는데 이는 데이터상



서 가장 낮은 편이며(약 7.6%), 2-3년제 대비 4년제 이상 대학 졸업에 대한 프리미엄이 전국에서 가장 높은 편이다(약 13.7%).

<표 8>은 이 논문의 주된 목적에 해당하는 선택편의를 보정하지 않았을 경우와 보정했을 경우(<표 7>의 결과) 간에 주요 회귀계수(성별, 학력, 전공계열)들을 서로 비교한 결과이며, [그림 1], [그림 2], [그림 3]은 그것을 그래프로 나타낸 것이다. 성별의 경우 선택편의 보정 시 지역별 차이가 증가하지만 그 크기가 약 1% 정도로 미미하여, 실질적으로 선택편의 보정의 효과는 거의 없는 것으로 나타났다. 학력을 대변하는 2-3년제 대비 4년제 대학 졸업 프리미엄은 성별에 비해 선택편의의 크기가 크고, 대체로 보정 전 프리미엄이 작은 지역일수록 더 큰 음의 선택편의가 나타났으며 그 결과 [그림 1] 에서와 같이 보정 전 vs 보정 후 프리미엄의 선형 추세선의 기울기가 약 0.82로 나타났다. 이는 2-3년제 대비 4년제 대학 졸업 프리미엄의 지역 간 격차가 선택편의 보정 결과 약 18% 감소하는 것으로 해석할 수 있다. 2-3년제 대비 4년제 대학 졸업 프리미엄에 대해서는, 보정 전 프리미엄에 관계없이 비슷한 크기의 양의 선택편의가 나타났던 Dahl(2002)과는 조금 다른 결과이다. 지역 구분이 적은 점(미국의 51개 대비 16개)이나 표본수의 차이 등으로 인해 결과를 직접 비교하는 것은 어렵지만, 한국의 경우 노동자들의 자기선택에 기반한 지역 이동이 지역 노동시장 간 격차를 줄이는 방향으로 작동하였다고도 볼 수 있다. 한편, 인문계열 대비 공학계열의 프리미엄에 대해서도 프리미엄이 낮은 지역에서는 양의 선택편의, 높은 지역에서는 음의 선택편의로 지역에 따라서 방향이 다르게 나타나 선택편의 보정이 지역 간의 공학계열 프리미엄의 격차를 줄이는 방향으로 작용하였다. 그 결과 보정 전 대비 보정 후 프리미엄의 선형 추세선 기울기는 약 11% 감소하였다. 이러한 결과로 보아 학력이나 전공에 따라 노동자들의 취업지역 자기선택에 의해 발생한 선택편의가 지역 노동시장 간 차이를 더 크게 보이게 했을 것이라는 추측이 가능하지만, 선택편의를 보정하고 나서도 지역 간의 격차는 일부 감소할 뿐 여전히 존재한다. 따라서 결국 노동자들이 느끼는 매력도의 지역 노동시장 간 격차가 실질적으로 존재하며, 노동자들의 이동을 고려하더라도 그 격차의 대부분은 그대로 남아 있는 것으로 보인다.

---

사용할 수 있는 변수 중에 대졸자의 학력을 제대로 통제할 수 있는 변수가 달리 없는 상황에서 출신 대학 지역에 따라 대학교들의 평균적인 학력에 차이가 어느 정도 있는 점을 활용하고자 하였기 때문이다. 이를테면 부산 지역 대학 더미와 경남 지역 대학 더미의 계수는 많은 지역에서 유의하게 차이가 남을 알 수 있다.

흥미로운 것은, 서울과 경기 지역에서는 전체적으로 선택편의의 보정이 임금함수에 미치는 영향이 미미했다는 것이다. 서울과 경기 지역은 표본 수가 상대적으로 많은 만큼 선택편의의 보정 역시 더 유의하게 작동하였을 것으로 보이므로, 선택편의가 실제로 작을 가능성이 더 높다. 이러한 결과로부터 성별, 학력에 따라 서울과 경기 지역으로 이동할 확률의 절대값은 차이가 나지만, 그 이동하는 노동자들의 능력은 출신 지역이나 기타 특성에 관계없이 거의 무작위로 분포할 것이라는 추측이 가능하다. 즉 어떤 특성을 가지고 있는지에 관계없이 노동자들은 취업지역으로서 수도권 지역을 선호한다는 것이다.<sup>2)</sup> 결국, 대졸자들의 수도권 집중 현상은 대졸 노동자들에게 보편적으로 작용하는 수도권 지역 자체의 취업지역으로서의 높은 매력에 기인하는 것으로 볼 수 있으며, 이를 해소하기 위해서는 비수도권 지역의 취업지역으로서의 매력을 높이는 적극적인 정책이 필요할 것으로 생각된다.

<표 9>에는 선택편의 보정함수에 대한 왈드 검정(Wald test) 결과를 표시하였다. 여기서 출신 지역과 취업지역이 동일한 경우에 대한 보정함수는  $M_{i,j=k}\lambda_{stayer}^*(p_{ikk})$  이며, 동일하지 않은 경우에 대한 보정함수는  $M_{i,j\neq k}\lambda_{mover}^*(p_{ijk}, p_{ijj})$ , 전체 보정함수는  $M_{i,j=k}\lambda_{stayer}^*(p_{ikk}) + M_{i,j\neq k}\lambda_{mover}^*(p_{ijk}, p_{ijj})$  에 해당한다. 출신 지역과 취업지역이 동일한 경우에 대한 보정함수와 동일하지 않은 경우에 대한 보정함수의 유의성이 서로 다른 경향을 보이며, 전체 보정함수에 대해서는 16개 시도 중 7개 시도에서 1% 유의수준에서 유의하고 9개 시도에서 10% 수준에서 유의한 것으로 나타난다. 방법론상 표본 숫자가 많은 경우에 추정의 정확도가 높아질 것으로 추측되는데, 표본의 수가 훨씬 많았던 Dahl(2002)에서도 Wald Test 결과 51개 주 중에서 선택편의 보정함수가 1% 유의수준에서 유의한 주는 30개 주, 10% 유의수준에서 유의한 주는 35개 주였던 것을 생각하면 본 논문에서의 선택편의 보정함수 역시 유효하게 작동하고 있는 것으로 해석할 수 있다.

한편, 이상의 주요 결과는 보정함수에 각 선택확률의 2차항까지 사용해 얻은 것이다.

2) 그런데 만약 고소득 노동자들에게는 양의 선택편의가 발생하고, 저소득 노동자들에게는 음의 선택편의가 발생하는 등 평균적으로는 선택편의가 없는 것으로 나타나지만 선택편의가 소득분포에 따라 다르게 나타날 가능성이 있다. 결과는 생략하였으나 이를 검증하기 위해 임금분위별로 선택편의 보정 전후의 추정계수 차이를 살펴본 결과 서울, 경기 지역에 대해서는 전 분위에 걸쳐 거의 차이가 없었으며, 타 지역의 경우에도 고소득 분위와 저소득 분위 간에 경향이 뚜렷하게 달라지거나 하지는 않는 것으로 보인다.

결과는 생략하였으나, 보정함수에 선택확률의 3차항까지 포함시키더라도 성별, 학력, 전공계열에 대한 계수들이 2차항까지만 사용한 경우와 거의 차이가 없어, 차수를 3차로 늘리는 것에 별다른 이득은 없었다.

## V. 결 론

본 논문에서는 지역 노동시장 간 격차와 수도권 집중 현상이 두드러지는 우리나라에서 노동자들의 지역 이동을 대변하는 취업지역 선택 상황을 개괄하고, 각 지역 노동시장에서 중요한 변수로 작용하는 성별, 학력, 전공에 따른 임금 프리미엄이 선택편의 보정 전후에 어떻게 달라지는지 추정하였다. 노동자들의 취업지역 선택은 양자택일이 아니라 지역 수만큼 선택지가 존재하는 다항선택 문제이므로, 이를 분석하기 위해 최대순서통계량과 준모수적 방법을 결합한 Dahl(2002)의 방법론을 이용하였다.

본 논문의 중요한 결과들을 정리하면 다음과 같다. 성별의 경우 눈에 띄는 선택편의는 발견되지 않았으며 보정 후에 각 지역 노동시장 간 계수의 격차가 줄어들지도 않았다. 학력에 대해서는 2-3년제 대학 졸업 대비 4년제 대학 졸업 시 임금 프리미엄이 선택편의 보정 시 전반적으로 상승하는 경향이 나타나, 음의 선택편의가 있었음을 알 수 있었다. 이러한 선택편의는 프리미엄이 클수록 감소하는 경향으로, 결과적으로 지역 노동시장 간 임금 프리미엄 격차가 선택편의 보정 시 약 18% 감소하였다. 이는 같은 대학교 졸업 프리미엄에 대해 양의 선택편의와 지역 간 격차가 선택편의 보정 시에도 거의 그대로 유지되는 것으로 보고한 Dahl(2002)과는 조금 다른 결과이다. 전공계열의 경우 인문계열 대비 공학계열의 임금 프리미엄이 낮을수록 양의 선택편의가, 높을수록 음의 선택편의가 나타나 프리미엄의 지역 노동시장 간 격차가 약 11% 감소하는 결과가 나왔다. 이렇듯 선택편의가 학력과 전공계열에 대해서는 선택편의가 관측되어 그 보정이 정확한 프리미엄 추정에 중요한 역할을 함을 확인하였지만, 선택편의를 보정하고 나서도 학력, 전공계열에 따른 임금 프리미엄의 지역에 따른 차이는 여전히 80% 이상 남아 있으므로 지역 노동시장 간 이질성은 여전히 상당히 크다고 할 수 있다.

한편 서울과 경기도로 대표되는 수도권 지역의 경우 전이행렬상 비수도권 출신 노동자

들이 상당수 유입되는 것으로 나타나지만 선택편의 보정의 영향이 거의 없었다. 이러한 결과로 보아 우리나라 노동자들의 수도권 지역 선호 경향은 성별, 학력 등과 같은 본 논문에서 고려한 노동자들의 특성에 크게 관계없이 보편적인 것으로, 수도권 지역 자체의 취업지역으로서의 높은 매력에 기인한 것으로 보인다.

이상과 같은 결과들로부터, 각 지역 노동시장의 격차는 노동자들의 취업지역 자기선택으로 인해 보이는 착시가 일부 존재하기는 하지만 상당 부분은 실제로 존재하는 것이라는 점을 확인한 것이 본 논문의 중요한 의의이다. 다시 말해, 각 지역 노동시장 간 차이를 해소하기 위해서는 부동산 정책 등을 통해 노동자들의 지역 간 이동을 방해하는 요인을 제거하거나, 산업구조 개편 및 어메니티 투자 등으로 노동자들이 느끼는 각 지역 노동시장의 매력도의 격차를 실질적으로 줄이는 적극적인 정책적 노력이 필요하다는 것이다.

## 참고문헌

- 김동수. 「광역도시통계권별 소득격차 추이 분석」. 산업연구원 Issue Paper 2011-272 (2011. 8): 1-68.
- 김우영. 「한국의 지역간 임금격차: 지역별 고용조사(RES)를 중심으로」. 『노동정책연구』 12권 1호 (2012. 3): 1-28.
- 김준영. 「청년인구의 지방 유출과 수도권 집중: 5가지 특징」. 『지역고용동향브리프』 2016년 가을호 (2016. 8): 6-24.
- 이상호. 「지역노동시장권(LLMAs)의 측정과 적용 가능성에 관한 연구」. 『노동정책연구』 8권 4호 (2008. 12): 147-182.
- 이상호. 「공간적 요인이 청년 대졸자의 하향취업에 미치는 효과」. 『공간과 사회』 22권 2호 (2012. 6): 38-77.
- 이원호. 「우리나라 광역대도시 지역노동시장의 임금결정 과정과 소득격차」. 『한국경제지리학회지』 5권 2호 (2002.12): 187-207.
- 최광성, 강동우, 최충. 「지역이동이 대졸자의 임금 변화에 미치는 영향」. 『노동경제논집』 41권 2호 (2018. 6): 61-88.

- Ahn, Hyungtaik. and Powell, James L. "Semiparametric estimation of censored selection models with a nonparametric selection mechanism." *Journal of Econometrics* 58 (1-2) (July 1993): 3-29.
- Andrews, Donald W. K. "Asymptotic normality of series estimators for nonparametric and semiparametric regression models." *Econometrica* 59 (2) (March 1991): 307-345.
- Black. "Earnings functions when wages and prices vary by location." *Journal of Labor Economics* 27(1) (January 2009): 21-47.
- Combes, Pierre-Philippe., Duranton, Gilles., and Gobillon, Laurent. "Spatial wage disparities: sorting matters." *Journal of Urban Economics* 63 (2) (March 2008): 723-742.
- Dahl, Gordon. "Mobility and the return to education: Testing a Roy model with multiple markets." *Econometrica* 70 (6) (November 2002): 2367-2420.
- Diamond, Rebecca. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *The American Economic Review* 106 (3) (March 2016): 479-524.
- Falaris, Evangelos. "A nested logit migration model with selectivity." *International Economic Review* 28 (2) (June 1987): 429-443.
- Heckman, James J. "Sample selection bias as a specification error." *Econometrica* 47 (1) (January 1979): 153-161.
- Lee, Lung-Fei. "Generalized econometric models with selectivity." *Econometrica* 51 (2) (March 1983): 507-512.
- Lee, Sanghoon. "Ability sorting and consumer city." *Journal of urban Economics* 68 (1) (July 2010): 20-33.
- McFadden, Daniel L. "Conditional Logit Analysis of Qualitative Choice Behavior." in *Frontiers in Econometrics*. edited by P. Zarembka, pp.105-142. New York: Academic Press, 1973.
- McFadden, Daniel L. "Econometric analysis of qualitative response models." in *Handbook of Econometrics*. Vol. II, edited by Z. Griliches and M. D. Intriligator. Amsterdam: North Holland, 1984.
- Moretti, Enrico. "Local labor markets." *Handbook of Labor Economics* 4 (B) Chapter 14. (Jan 2011): 1237-1323.

- Newey, Whitney K. "Convergence rates and asymptotic normality for eries estimators." *Journal of Econometrics* 79 (1) (July 1997): 147-168.
- Roback, Jennifer. "Wages, rents and the quality of life." *Journal of Political Economy* 90(6) (December 1982): 1257-1278.
- Roy, A.D. "Some thoughts on the distribution of earnings." *Oxford Economic Papers New Series* 3 (2) (June 1951): 135-146.
- Trost, Robert. and Lee, Lung-Fei. "Technical training and earnings: A polychotomous choice model with selectivity." *The Review of Economics and Statistics* 66 (1) (February 1984): 151-156.

<표 1> 집단 분류 기준이 변수 단독일 때 지역 선택확률

(1)  $p_{ikk}$  : 정주자가 출신 고등학교 지역에서 그대로 취업할 확률 (잔류확률)

변수	관측치 수	가중치	평균	표준편차	최솟값	최댓값
<b>성별</b>	42882	1195584.03	0.147	0.135	0.012	0.371
<b>학력</b>	42882	1195584.03	0.148	0.135	0.011	0.377
<b>전공계열</b>	42882	1195584.03	0.151	0.143	0.007	0.447
거친 전공계열	42882	1195584.03	0.151	0.142	0.009	0.431
<b>대학권역</b>	42882	1195584.03	0.363	0.194	0.001	0.775
거친 대학권역	42882	1195584.03	0.280	0.201	0.001	0.700
아버지 학력	42882	1195584.03	0.153	0.146	0.010	0.453
어머니 학력	42882	1195584.03	0.153	0.147	0.009	0.497
학점 82점 이상	42882	1195584.03	0.147	0.134	0.011	0.359

(2)  $p_{ijj}$  : 이주자가 출신 고등학교 지역에서 그대로 취업할 확률 (미실현 잔류확률)

변수	관측치 수	가중치	평균	표준편차	최솟값	최댓값
<b>성별</b>	38317	985071.67	0.544	0.042	0.507	0.592
<b>학력</b>	38317	985071.67	0.536	0.070	0.491	0.645
<b>전공계열</b>	38317	985071.67	0.544	0.043	0.488	0.649
거친 전공계열	38317	985071.67	0.546	0.035	0.502	0.575
<b>대학권역</b>	38317	985071.67	0.543	0.044	0.493	0.759
거친 대학권역	38317	985071.67	0.545	0.036	0.493	0.591
아버지 학력	38317	985071.67	0.547	0.020	0.518	0.565
어머니 학력	38317	985071.67	0.548	0.017	0.509	0.557
학점 82점 이상	38317	985071.67	0.548	0.005	0.542	0.552

(3)  $p_{ijk}$  : 이주자가 출신 고등학교 지역을 떠나 다른 지역에 취업할 확률 (전입확률)

변수	관측치 수	가중치	평균	표준편차	최솟값	최댓값
<b>성별</b>	38317	985071.67	0.198	0.138	0.012	0.371
<b>학력</b>	38317	985071.67	0.200	0.143	0.011	0.377
<b>전공계열</b>	38317	985071.67	0.202	0.146	0.007	0.447
거친 전공계열	38317	985071.67	0.201	0.145	0.009	0.431
<b>대학권역</b>	38317	985071.67	0.290	0.217	0.001	0.775
거친 대학권역	38317	985071.67	0.267	0.214	0.001	0.700
아버지 학력	38317	985071.67	0.201	0.148	0.010	0.453
어머니 학력	38317	985071.67	0.200	0.148	0.009	0.497
학점 82점 이상	38317	985071.67	0.196	0.137	0.011	0.359

주 : 전공계열 - 교육계열 제외 인문계열, 사회계열, 공학계열, 자연계열, 의료계열, 예체능계열 6개 계열.  
 거친 전공계열 - 교육계열 제외 기존 6개 전공계열을 인문계열+사회계열, 공학계열+자연계열, 교육계열+의료계열, 예체능계열의 4가지 계열로 단순화함.  
 대학권역 - 출신 대학교 지역을 서울, 경기(인천 포함), 강원, 충북, 충남(대전 포함), 경북(대구 포함), 경남(부산, 울산 포함), 전북, 전남(광주 포함), 제주의 10개 권역으로 단순화함.  
 거친 대학권역 - 출신 대학교 지역을 서울권(서울), 경기권(인천, 경기, 강원), 충청권(대전, 충북, 충남), 경상권(부산, 대구, 울산, 경북, 경남), 전라권(광주, 전북, 전남, 제주)의 5개 권역으로 단순화함.

〈표 2〉 기술통계량 (가중치 반영)

취업지역	서울	부산	대구	인천	광주	대전	울산	경기
총 가중치 비중 (%)	34.8	6.2	4.7	3.7	2.8	3.6	2.2	17.8
월평균임금 (명목, 만원)	202.5	183.9	177.2	198.3	177.9	177.1	213.7	199.3
로그 월평균임금	5.21	5.12	5.07	5.18	5.08	5.08	5.28	5.20
로그 월평균임금 표준편차	0.47	0.44	0.46	0.45	0.45	0.45	0.41	0.44
나이 (세)	26.21	26.01	26.07	25.93	26.03	25.91	26.28	26.02
100점 만점 환산 GPA (점)	82.3	83.1	82.3	82.1	83.3	83.7	82.6	82.3
여성 비율 (%)	52.0	53.8	51.5	48.9	48.4	49.8	40.0	44.9
4년제 대학 졸업자 비율 (%)	69.5	62.7	58.8	52.2	63.5	69.3	62.4	59.5
취업자 중 정주자 비율 (%)	46.7	78.5	76.3	54.4	65.5	56.3	64.4	48.8
취업자들이 속한 집단의 종류	1143	339	266	534	222	439	269	1047
소수집단 포함 관측치 수주	29,675	4,377	3,157	3,041	2,159	2,807	1,697	14,857
소수집단 제외 관측치 수주	28,162	4,158	3,053	2,796	2,032	2,585	1,573	14,019

취업지역	강원	충북	충남	경북	경남	전북	전남	제주
총 가중치 비중 (%)	2.1	2.4	3.4	4.7	5.5	2.7	2.4	1.2
월평균임금 (명목, 만원)	169.6	189.2	201.0	191.8	204.4	173.4	187.9	168.7
로그 월평균임금	5.05	5.15	5.21	5.16	5.23	5.07	5.14	5.04
로그 월평균임금 표준편차	0.41	0.43	0.43	0.44	0.41	0.41	0.42	0.39
나이 (세)	26.04	25.83	26.07	26.15	26.12	26.19	26.16	25.83
100점 만점 환산 GPA (점)	83.0	83.0	83.3	82.9	82.8	83.7	83.2	84.6
여성 비율 (%)	39.2	44.3	40.0	40.0	43.4	49.6	46.4	44.5
4년제 대학 졸업자 비율 (%)	65.2	60.7	63.3	58.1	60.4	68.1	59.9	52.3
취업자 중 정주자 비율 (%)	66.1	64.3	45.3	58.9	67.0	85.6	71.1	86.8
취업자들이 속한 집단의 종류	362	398	532	398	395	243	255	128
소수집단 포함 관측치 수주	1,947	1,936	2,801	3,512	4,071	2,165	1,982	1,015
소수집단 제외 관측치 수주	1,724	1,721	2,550	3,326	3,839	1,998	1,856	916

주 : 여기서 관측치 수는 가중치를 반영하지 않은 원자료로부터 추출함. 교육대학 졸업자는 제외하였으며, “소수집단 포함 관측치 수” 항목을 제외하면 표본 8개 미만 집단 또한 제외하였음. 취업자 중 정주자 비율은 취업지역을 기준으로 해당 지역 취업자 중에서 같은 지역에서 고등학교를 졸업한 사람의 비율을 의미하며, 고등학교 졸업 지역 기준으로 계산된  $p_{ikk}$ 와는 다름에 유의.



〈표 3〉 출신 고등학교 지역별 취업지역 선택 비율 (가중치 적용, %)

각 행 (→ 방향) 총합100(%)	취업지역																
	서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	경북	경남	전북	전남	제주	계
서울	74.0	0.3	0.2	2.0	0.3	0.7	0.3	17.3	0.8	0.8	1.5	0.6	0.6	0.3	0.3	0.1	100
부산	15.7	55.8	1.0	0.8	0.2	0.8	3.2	6.3	0.4	0.6	1.2	2.3	10.9	0.3	0.3	0.3	100
대구	15.0	1.5	50.3	1.3	0.2	1.0	1.5	7.1	0.8	0.7	1.8	14.9	3.1	0.2	0.2	0.4	100
인천	34.7	0.5	0.2	41.1	0.1	0.9	0.3	17.2	1.0	0.7	2.2	0.3	0.5	0.3	0.2	0.1	100
광주	19.5	0.5	0.2	1.6	50.5	1.1	0.3	8.0	0.7	0.6	1.2	0.7	1.2	2.1	11.6	0.2	100
대전	17.8	0.5	0.9	1.2	0.7	54.9	0.3	8.0	1.1	3.1	8.2	1.0	1.0	0.7	0.4	0.1	100
울산	15.7	4.7	2.1	0.4	0.2	0.7	55.6	7.1	0.3	0.6	1.2	5.0	5.5	0.5	0.3	0.1	100
경기	38.9	0.3	0.2	3.0	0.2	1.0	0.2	50.3	1.2	1.0	1.9	0.5	0.6	0.4	0.2	0.2	100
강원	27.6	0.9	0.8	2.3	0.2	1.0	0.2	15.1	46.1	1.3	1.7	0.9	1.2	0.3	0.6	0.2	100
충북	21.8	0.4	0.4	1.5	0.2	6.4	0.1	9.6	1.5	50.5	4.6	1.3	0.7	0.5	0.4	0.3	100
충남	23.8	0.6	0.5	2.3	0.4	12.1	0.1	11.8	0.9	2.7	42.1	0.7	0.9	1.1	0.2	0.1	100
경북	18.6	3.4	11.0	1.4	0.2	2.7	1.9	10.1	1.1	1.6	2.1	43.3	2.0	0.4	0.3	0.1	100
경남	16.2	10.8	1.9	1.1	0.4	1.1	1.9	7.4	0.7	0.7	1.4	2.9	52.3	0.3	0.7	0.3	100
전북	20.0	0.8	0.2	1.7	2.6	3.0	0.4	10.0	0.6	1.2	3.3	0.6	0.9	52.7	1.8	0.2	100
전남	19.9	1.3	0.2	1.9	15.5	1.4	0.6	10.1	0.4	1.0	2.2	1.0	1.7	2.4	40.1	0.3	100
제주	16.3	1.0	0.2	1.4	0.3	0.6	0.6	7.0	0.5	0.3	0.6	0.8	1.1	0.2	0.6	68.5	100

〈표 4〉 학력에 따른 각 출신 고등학교 지역별 취업지역 선택 비율 (가중치 적용, %)

각 행 (→ 방향)	총합 100(%)	취업지역																	
		서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	경북	경남	전북	전남	제주	합	
출신 고등 학교 지역	서울	SC	77.7	0.2	0.2	1.9	0.3	0.3	0.2	16.2	0.4	0.5	1.0	0.3	0.3	0.2	0.2	0.1	100
		CG	72.3	0.4	0.2	2.0	0.3	1.0	0.3	17.7	0.9	0.9	1.7	0.8	0.7	0.3	0.3	0.2	100
	부산	SC	6.3	69.4	0.8	0.7	0.1	0.2	3.6	4.6	0.2	0.5	1.2	2.3	9.6	0.2	0.2	0.0	100
		CG	20.2	50.2	1.0	0.8	0.3	1.0	3.1	6.7	0.4	0.5	1.3	2.1	11.5	0.2	0.3	0.5	100
	대구	SC	7.6	1.2	62.5	1.2	0.2	0.6	1.1	5.5	1.1	0.5	2.2	13.6	2.1	0.0	0.0	0.7	100
		CG	19.1	1.6	43.9	1.3	0.2	1.3	1.8	7.9	0.7	0.9	1.5	15.4	3.6	0.3	0.3	0.2	100
	인천	SC	28.2	0.4	0.1	51.6	0.0	0.2	0.2	16.2	0.7	0.4	1.6	0.1	0.2	0.1	0.1	0.1	100
		CG	40.8	0.6	0.3	31.5	0.2	1.4	0.3	18.1	1.3	1.0	2.6	0.5	0.7	0.4	0.2	0.1	100
	광주	SC	9.6	0.1	0.4	1.6	64.0	0.7	0.2	6.3	0.4	1.1	0.9	0.6	1.7	1.6	10.7	0.2	100
		CG	24.1	0.6	0.1	1.6	44.8	1.4	0.4	8.6	0.9	0.4	1.2	0.7	1.1	2.1	11.8	0.2	100
	대전	SC	7.1	0.3	0.2	1.1	0.7	71.8	0.2	5.7	2.0	2.5	7.2	0.5	0.4	0.3	0.0	0.2	100
		CG	22.3	0.6	1.1	1.2	0.8	48.2	0.4	9.1	0.8	3.3	8.1	1.3	1.2	0.9	0.6	0.1	100
	울산	SC	5.8	5.8	1.2	0.1	0.0	0.0	73.6	4.3	0.4	0.6	0.9	2.4	4.4	0.5	0.1	0.0	100
		CG	21.2	4.2	2.5	0.5	0.2	0.8	46.6	8.6	0.3	0.7	1.3	6.2	6.1	0.4	0.3	0.2	100
	경기	SC	31.8	0.1	0.1	3.2	0.1	0.5	0.1	60.6	0.7	0.5	1.2	0.3	0.4	0.2	0.1	0.2	100
		CG	44.4	0.4	0.3	3.0	0.2	1.4	0.3	42.3	1.6	1.3	2.5	0.7	0.7	0.5	0.3	0.2	100
	강원	SC	19.8	1.1	1.2	2.1	0.4	0.9	0.2	15.1	54.2	1.3	1.5	0.8	1.1	0.2	0.2	0.0	100
		CG	32.6	0.8	0.5	2.5	0.1	1.1	0.2	15.0	41.1	1.3	1.8	0.9	1.3	0.2	0.3	0.3	100
	충북	SC	12.1	0.4	0.6	1.0	0.0	4.5	0.0	8.2	1.6	65.3	4.1	1.3	0.3	0.0	0.3	0.4	100
		CG	27.8	0.4	0.4	1.8	0.3	7.6	0.2	10.4	1.4	41.4	5.0	1.3	0.8	0.7	0.4	0.2	100
	충남	SC	15.6	0.3	0.4	2.3	0.2	11.8	0.0	10.2	1.0	3.0	52.9	0.2	1.0	0.9	0.3	0.0	100
		CG	29.5	0.8	0.5	2.3	0.4	12.3	0.1	12.6	0.9	2.4	34.8	0.9	0.8	1.3	0.2	0.1	100
	경북	SC	9.9	3.5	10.9	0.9	0.1	1.2	1.4	8.5	0.6	1.3	2.0	57.6	1.6	0.1	0.3	0.2	100
		CG	24.6	3.3	10.9	1.7	0.2	3.6	2.2	10.9	1.3	1.8	2.1	34.4	2.2	0.5	0.4	0.1	100
	경남	SC	8.2	11.3	2.0	0.6	0.5	0.4	1.8	5.6	0.6	0.3	1.2	3.1	63.9	0.1	0.4	0.1	100
		CG	21.4	10.6	1.8	1.4	0.4	1.6	2.0	8.4	0.8	0.8	1.3	2.7	45.1	0.4	0.9	0.4	100
	전북	SC	12.3	1.2	0.2	2.4	2.4	2.3	0.4	9.4	0.3	1.0	3.4	0.4	0.6	61.4	2.3	0.1	100
		CG	23.9	0.7	0.2	1.3	2.8	3.3	0.5	10.1	0.8	1.2	3.0	0.6	1.1	48.8	1.6	0.2	100
	전남	SC	13.0	1.7	0.1	2.5	16.0	1.0	0.2	9.7	0.3	0.7	1.8	1.6	1.3	1.6	48.1	0.5	100
		CG	24.3	1.1	0.2	1.6	15.3	1.7	0.7	10.5	0.5	1.1	2.4	0.7	1.8	2.8	35.0	0.2	100
	제주	SC	9.0	1.2	0.2	0.7	0.3	0.9	0.5	7.9	0.4	0.3	0.0	0.4	0.7	0.0	0.3	77.2	100
		CG	22.4	0.9	0.2	1.9	0.2	0.5	0.8	6.2	0.5	0.4	1.1	1.2	1.5	0.3	0.6	61.4	100

주 : 'SC'는 2-3년제 대학, 'CG'는 4년제 대학을 의미

<표 5> 성별에 따른 각 출신 고등학교 지역별 취업지역 선택 비율 (가중치 적용, %)

각 행 (→ 방향) 총합 100(%)	취업지역																	
	서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	경북	경남	전북	전남	제주	합	
서울	남	67.5	0.4	0.3	2.4	0.4	1.0	0.5	20.8	1.1	1.0	2.1	0.9	0.8	0.4	0.4	0.2	100
	여	81.1	0.2	0.1	1.6	0.2	0.4	0.1	13.5	0.5	0.5	0.8	0.3	0.3	0.1	0.1	0.1	100
부산	남	14.7	48.5	0.9	0.9	0.3	0.8	4.4	8.2	0.5	0.8	1.7	3.3	14.0	0.3	0.5	0.3	100
	여	16.8	63.7	1.0	0.7	0.2	0.7	1.9	4.4	0.3	0.3	0.7	1.2	7.6	0.2	0.1	0.4	100
대구	남	15.0	1.5	42.7	1.2	0.4	1.2	2.0	8.1	1.1	1.1	2.2	18.5	4.2	0.2	0.3	0.5	100
	여	15.0	1.5	59.5	1.5	0.0	0.8	1.0	5.9	0.4	0.3	1.3	10.6	1.7	0.2	0.1	0.2	100
인천	남	32.6	0.6	0.3	37.5	0.1	1.1	0.3	20.0	1.2	1.2	3.3	0.4	0.9	0.3	0.2	0.0	100
	여	36.5	0.3	0.1	44.3	0.1	0.6	0.3	14.6	0.9	0.3	1.1	0.1	0.2	0.3	0.2	0.2	100
광주	남	18.1	0.5	0.3	1.7	46.0	1.3	0.4	9.2	1.2	1.1	1.8	0.7	1.4	2.6	13.4	0.4	100
	여	21.0	0.4	0.1	1.5	55.6	0.9	0.2	6.7	0.1	0.2	0.4	0.6	1.1	1.5	9.6	0.0	100
대전	남	16.4	0.6	1.0	1.3	0.8	48.8	0.5	10.7	1.1	4.0	10.0	1.7	1.4	1.0	0.6	0.2	100
	여	19.3	0.4	0.7	1.1	0.7	61.1	0.1	5.3	1.1	2.1	6.4	0.3	0.5	0.4	0.3	0.1	100
울산	남	14.9	3.5	2.2	0.3	0.3	0.5	52.7	8.9	0.4	1.0	1.6	6.1	6.6	0.6	0.2	0.2	100
	여	16.6	6.1	1.9	0.5	0.0	0.9	59.0	5.1	0.3	0.2	0.7	3.7	4.2	0.4	0.4	0.0	100
경기	남	36.9	0.3	0.3	3.5	0.3	1.1	0.3	49.0	1.8	1.4	2.7	0.8	0.8	0.4	0.3	0.1	100
	여	40.6	0.3	0.1	2.6	0.1	0.9	0.1	51.3	0.7	0.6	1.2	0.3	0.4	0.4	0.1	0.3	100
강원	남	26.1	0.9	0.4	2.2	0.3	1.4	0.3	17.0	44.3	1.4	1.8	1.1	1.4	0.3	1.0	0.2	100
	여	29.3	0.8	1.2	2.4	0.1	0.5	0.1	12.9	48.1	1.2	1.5	0.6	1.0	0.2	0.0	0.2	100
충북	남	21.1	0.3	0.5	2.1	0.2	5.9	0.2	12.1	2.2	45.3	5.3	2.0	1.0	0.7	0.7	0.5	100
	여	22.5	0.6	0.4	0.9	0.1	7.0	0.0	7.2	0.7	55.4	4.0	0.6	0.4	0.2	0.1	0.1	100
충남	남	22.3	0.6	0.6	2.7	0.5	13.0	0.1	13.7	1.2	3.4	38.1	0.9	1.4	1.3	0.4	0.1	100
	여	25.4	0.6	0.3	1.9	0.2	11.1	0.1	9.7	0.6	2.0	46.4	0.5	0.4	1.0	0.1	0.0	100
경북	남	17.3	3.5	9.9	1.6	0.2	2.8	2.2	12.0	1.5	1.7	2.4	41.2	2.7	0.5	0.5	0.2	100
	여	20.1	3.2	12.3	1.1	0.1	2.6	1.5	7.9	0.7	1.4	1.8	45.7	1.1	0.2	0.1	0.1	100
경남	남	15.0	9.3	2.3	1.1	0.6	1.4	2.3	8.0	1.1	1.0	1.9	4.0	50.5	0.3	1.0	0.4	100
	여	17.4	12.2	1.5	1.1	0.3	0.8	1.5	6.8	0.3	0.5	0.9	1.9	54.0	0.3	0.4	0.2	100
전북	남	19.1	0.8	0.0	1.6	3.2	2.3	0.7	12.0	1.2	1.5	3.8	0.8	1.4	49.3	1.9	0.2	100
	여	21.0	0.8	0.3	1.8	2.1	3.6	0.2	7.9	0.1	0.8	2.7	0.4	0.4	56.2	1.7	0.1	100
전남	남	19.2	1.1	0.2	2.0	15.0	1.7	0.8	11.2	0.6	1.2	3.0	1.4	1.6	2.6	38.1	0.4	100
	여	20.5	1.6	0.1	1.9	15.9	1.2	0.3	9.1	0.2	0.9	1.4	0.6	1.8	2.2	42.2	0.3	100
제주	남	16.7	0.9	0.4	1.8	0.3	0.5	1.1	6.4	0.5	0.3	0.9	1.0	1.5	0.2	0.7	67.0	100
	여	15.9	1.1	0.0	1.0	0.2	0.8	0.1	7.6	0.4	0.4	0.4	0.6	0.8	0.2	0.4	70.2	100

출신고등학교지역

〈표 6〉 전공에 따른 각 출신 고등학교 지역별 취업지역 선택 비율 (가중치 적용, %)

각 행 (→ 방향) 총합 100(%)	취업지역																		
	서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	경북	경남	전북	전남	제주	합		
출신 고등 학교 지역	서울	인	80.0	0.0	0.2	1.5	0.3	0.4	0.1	14.4	0.8	0.3	1.0	0.4	0.5	0.1	0.1	0.0	100
		공	60.2	0.5	0.2	2.7	0.4	2.0	0.7	25.0	0.6	1.0	2.6	1.6	1.4	0.6	0.3	0.3	100
	부산	인	21.2	59.1	0.6	0.7	0.2	0.5	1.4	5.1	0.2	0.2	0.7	1.9	7.6	0.4	0.1	0.1	100
		공	17.8	36.6	0.7	1.1	0.4	1.4	5.3	10.4	0.4	0.6	2.2	4.2	18.5	0.2	0.4	0.1	100
	대구	인	20.6	1.1	56.9	1.7	0.0	0.0	1.9	4.1	0.5	0.1	0.6	11.4	0.8	0.3	0.0	0.0	100
		공	18.0	1.0	27.6	1.7	0.2	1.7	3.4	11.4	0.8	1.5	2.9	22.0	7.3	0.2	0.5	0.0	100
	인천	인	48.2	0.8	0.0	32.0	0.1	1.3	0.0	14.0	0.8	0.0	1.9	0.3	0.0	0.5	0.2	0.0	100
		공	38.1	1.2	0.1	26.2	0.3	2.2	0.7	23.2	0.8	0.5	3.7	1.1	1.7	0.1	0.3	0.0	100
	광주	인	33.5	0.0	0.0	0.6	44.6	0.0	0.2	7.0	0.9	0.0	0.0	0.6	1.9	1.3	9.4	0.0	100
		공	20.0	1.1	0.4	2.2	37.2	2.3	0.9	14.7	0.7	0.9	2.0	1.3	2.2	2.8	11.3	0.0	100
	대전	인	24.1	0.2	0.0	1.0	0.1	55.3	0.0	6.1	0.8	3.6	6.0	1.3	0.7	0.4	0.6	0.0	100
		공	20.1	1.9	1.7	1.6	1.0	40.7	1.2	13.3	0.5	2.8	8.3	2.0	2.7	0.9	1.2	0.0	100
	울산	인	25.3	3.8	2.0	1.4	0.0	0.7	48.3	6.4	0.7	0.0	0.0	4.9	6.0	0.5	0.0	0.0	100
		공	17.0	3.5	1.1	0.3	0.7	0.5	42.1	10.5	0.1	1.0	3.0	8.5	10.7	0.6	0.4	0.1	100
	경기	인	48.4	0.5	0.2	2.2	0.2	0.8	0.2	42.9	0.6	0.4	2.3	0.2	0.2	0.6	0.0	0.2	100
		공	39.6	0.5	0.3	3.7	0.5	2.2	0.6	42.5	1.5	1.8	3.7	1.0	1.3	0.4	0.4	0.1	100
	강원	인	33.3	2.3	0.0	2.5	0.0	0.5	0.0	13.6	44.6	1.8	0.3	0.0	1.1	0.0	0.0	0.0	100
		공	30.9	1.1	0.2	2.6	0.0	1.3	0.3	20.0	32.5	2.2	3.6	2.0	2.6	0.4	0.5	0.0	100
	충북	인	34.4	0.8	0.0	1.9	0.0	10.1	0.4	7.4	0.9	40.1	3.4	0.3	0.0	0.2	0.2	0.0	100
		공	22.3	0.6	0.9	3.2	0.4	6.5	0.4	15.4	1.4	37.3	6.3	2.0	1.4	1.0	0.7	0.2	100
	충남	인	33.8	0.6	0.4	1.6	0.3	8.9	0.0	8.7	1.2	2.4	38.7	1.6	0.4	1.4	0.0	0.0	100
		공	28.8	1.1	0.4	2.3	0.2	10.5	0.2	18.4	1.1	2.5	29.7	1.1	1.6	1.5	0.5	0.2	100
	경북	인	27.4	4.4	13.8	2.2	0.0	2.9	0.5	7.6	0.3	1.0	2.1	35.2	2.0	0.0	0.4	0.3	100
		공	19.9	2.4	9.3	2.3	0.1	3.9	4.1	16.7	1.1	2.2	3.1	30.4	3.6	0.3	0.5	0.1	100
	경남	인	28.0	13.3	1.2	0.5	0.0	1.3	1.7	6.3	0.2	0.7	0.4	3.5	41.8	0.0	0.7	0.4	100
		공	18.7	9.3	1.8	0.8	0.8	1.4	3.2	12.0	0.6	0.8	1.4	3.8	43.0	0.4	1.3	0.6	100
	전북	인	25.4	1.1	0.4	0.8	1.4	3.8	0.0	9.0	1.1	0.8	0.3	0.1	0.0	55.5	0.3	0.0	100
		공	21.3	0.9	0.0	1.9	2.5	2.9	1.3	14.9	0.8	1.3	4.4	1.3	3.2	41.8	1.4	0.2	100
	전남	인	30.0	0.6	0.4	1.8	19.6	1.0	1.0	7.0	0.5	0.3	1.9	1.1	0.6	3.2	31.1	0.0	100
		공	22.3	2.4	0.2	2.3	12.3	1.8	1.6	15.5	1.0	0.5	3.8	1.0	2.3	3.0	30.1	0.1	100
	제주	인	27.2	0.6	0.0	1.2	0.0	0.0	0.0	8.0	1.1	1.1	0.0	1.2	1.2	0.0	0.0	58.4	100
		공	23.8	2.7	0.0	2.6	1.0	1.2	2.1	12.5	1.6	0.9	1.6	2.8	3.4	0.0	1.3	42.5	100

주: '인'은 인문계열을, '공'은 공학계열을 의미

<표 7> 선택편의 보정 시 임금함수 추정 결과

변수	서울	부산	대구	인천	광주	대전	울산	경기
	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금
남성 대비 여성	-0.076*** (0.008)	-0.198*** (0.019)	-0.204*** (0.025)	-0.146*** (0.022)	-0.195*** (0.028)	-0.127*** (0.024)	-0.232*** (0.027)	-0.112*** (0.010)
나이	0.065*** (0.020)	0.105** (0.043)	0.069 (0.050)	0.109** (0.052)	0.004 (0.063)	0.164*** (0.061)	0.023 (0.067)	0.014 (0.024)
나이 <sup>2</sup>	-0.000 (0.000)	-0.002** (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	-0.002** (0.001)	-0.000 (0.001)	0.000 (0.000)
학점 10점 (100점 만점)	0.003*** (0.000)	0.001 (0.001)	0.000 (0.001)	0.005*** (0.001)	0.001 (0.001)	-0.001 (0.001)	0.005*** (0.001)	0.002*** (0.000)
2~3년제 대비 4년제	0.137*** (0.008)	0.082*** (0.023)	0.082*** (0.028)	0.066*** (0.024)	0.082** (0.032)	0.060* (0.032)	0.132*** (0.032)	0.095*** (0.010)
서울 대비 부산 대학	-0.054** (0.022)	-0.244*** (0.069)	0.063 (0.104)	-0.031 (0.069)	-0.011 (0.142)	-0.025 (0.113)	-0.200*** (0.064)	-0.050* (0.030)
서울 대비 경남 대학	-0.124*** (0.030)	-0.236*** (0.073)	-0.041 (0.124)	-0.046 (0.094)	-0.039 (0.155)	-0.066 (0.128)	-0.227*** (0.070)	-0.078** (0.037)
(대학 지역 더미 일부 생략)								
인문계 대비 사회계	0.101*** (0.009)	0.127*** (0.023)	0.058* (0.030)	0.095*** (0.031)	0.143*** (0.041)	0.060* (0.035)	0.076** (0.036)	0.093*** (0.014)
인문계 대비 교육계	-0.030* (0.016)	0.027 (0.035)	0.031 (0.043)	-0.009 (0.050)	0.089* (0.052)	0.009 (0.048)	-0.011 (0.053)	0.050** (0.022)
인문계 대비 공학계	0.077*** (0.010)	0.107*** (0.032)	0.038 (0.039)	0.129*** (0.032)	0.095** (0.045)	-0.038 (0.038)	0.120*** (0.037)	0.174*** (0.014)
인문계 대비 자연계	-0.001 (0.011)	0.017 (0.029)	-0.062* (0.037)	0.027 (0.037)	0.021 (0.049)	-0.062 (0.039)	0.115*** (0.041)	0.067*** (0.016)
인문계 대비 의약계	0.295*** (0.013)	0.307*** (0.030)	0.234*** (0.039)	0.296*** (0.040)	0.290*** (0.049)	0.207*** (0.044)	0.207*** (0.047)	0.297*** (0.018)
인문계 대비 예체능계	-0.110*** (0.010)	-0.100*** (0.028)	-0.128*** (0.034)	-0.045 (0.036)	0.024 (0.047)	-0.112*** (0.040)	-0.136*** (0.043)	-0.053*** (0.015)
연도 더미 : 2008	-0.084*** (0.009)	-0.092*** (0.022)	-0.037 (0.028)	-0.026 (0.028)	-0.034 (0.034)	-0.091*** (0.032)	-0.072*** (0.031)	-0.079*** (0.012)
연도 더미 : 2009	-0.038*** (0.009)	-0.008 (0.022)	0.005 (0.030)	-0.049 (0.030)	-0.009 (0.032)	-0.045 (0.032)	-0.032 (0.033)	-0.033*** (0.013)
(연도 더미 일부 생략)								
정주자 : 잔류확률	-0.244 (0.152)	-0.069 (0.272)	-0.351 (0.337)	-0.351 (0.267)	-0.434 (0.558)	-0.057 (0.300)	-0.650** (0.315)	0.044 (0.158)
정주자 : (잔류확률) <sup>2</sup>	0.122 (0.117)	0.100 (0.230)	0.398 (0.299)	0.348 (0.280)	0.324 (0.502)	0.114 (0.260)	0.604** (0.266)	0.098 (0.157)
이주자 : 미실현 잔류확률	-0.232 (0.156)	0.299 (0.268)	-0.536 (0.343)	-0.153 (0.229)	-0.930** (0.414)	-0.241 (0.300)	-0.447 (0.323)	0.354*** (0.126)
이주자 : (미실현 잔류확률) <sup>2</sup>	0.207 (0.134)	-0.377 (0.289)	0.594 (0.390)	0.250 (0.227)	1.285*** (0.474)	0.228 (0.322)	0.539* (0.323)	-0.323*** (0.109)
이주자 : 전입확률	-0.095 (0.169)	-0.633 (0.622)	-0.537 (0.855)	1.338 (0.944)	1.003 (1.026)	-0.627 (0.686)	0.218 (1.115)	0.734*** (0.245)
이주자 : 전입확률 × 미실현 잔류확률	0.101 (0.207)	0.606 (0.684)	0.952 (0.891)	-1.928 (1.181)	-1.647 (1.012)	1.432* (0.812)	0.515 (1.888)	-0.874*** (0.259)
이주자 : (전입확률) <sup>2</sup>	-0.026 (0.136)	1.270 (0.966)	1.847 (1.658)	-1.901 (3.437)	-0.859 (1.930)	1.104 (1.252)	-2.665 (2.628)	-0.938*** (0.363)
상수항	3.731*** (0.271)	3.755*** (0.590)	4.094*** (0.699)	3.143*** (0.717)	4.834*** (0.882)	2.888*** (0.844)	4.740*** (0.933)	4.425*** (0.330)
관측치(집단) 수	28,162	4,158	3,053	2,796	2,032	2,585	1,573	14,019
R-squared	0.191	0.210	0.206	0.200	0.173	0.162	0.349	0.215

Standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

잔류확률 :  $p_{ijj}$ (정주자의 경우), 미실현 잔류확률 :  $p_{ijj}$ (이주자의 경우), 전입확률 :  $p_{ijk}$ 에 해당 (III장 참조)

〈표 7〉 선택편의 보정 시 임금함수 추정 결과 (계속)

변수	강원 로그임금	충북 로그임금	충남 로그임금	경북 로그임금	경남 로그임금	전북 로그임금	전남 로그임금	제주 로그임금
남성 대비 여성	-0.095*** (0.025)	-0.144*** (0.028)	-0.118*** (0.022)	-0.166*** (0.020)	-0.130*** (0.017)	-0.169*** (0.026)	-0.163*** (0.025)	-0.102*** (0.035)
나이	0.076 (0.060)	-0.124* (0.069)	0.094* (0.054)	0.098** (0.050)	0.005 (0.039)	0.108* (0.059)	0.019 (0.058)	0.157* (0.082)
나이 <sup>2</sup>	-0.001 (0.001)	0.003** (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	-0.002 (0.001)
학점 10점 (100점 만점)	0.002** (0.001)	0.000 (0.001)	-0.000 (0.001)	0.002*** (0.001)	0.004*** (0.001)	0.002** (0.001)	0.001 (0.001)	0.006*** (0.001)
2~3년제 대비 4년제	0.077*** (0.029)	0.076** (0.032)	0.030 (0.024)	0.032 (0.025)	0.094*** (0.020)	0.056* (0.032)	-0.006 (0.026)	0.062* (0.035)
서울 대비 부산 대학	-0.033 (0.104)	-0.005 (0.098)	-0.170*** (0.063)	-0.013 (0.055)	-0.264*** (0.052)	-0.245 (0.181)	-0.268** (0.120)	-0.111 (0.138)
서울 대비 경남 대학 (대학 지역 더미 일부 생략)	-0.033 (0.104)	-0.085 (0.121)	-0.248*** (0.078)	-0.062 (0.068)	-0.316*** (0.052)	-0.466 (0.298)	-0.151 (0.104)	-0.075 (0.241)
인문계 대비 사회계	0.126*** (0.040)	-0.010 (0.042)	0.050 (0.034)	0.010 (0.031)	0.047* (0.025)	0.110*** (0.037)	0.151*** (0.043)	0.011 (0.047)
인문계 대비 교육계	0.110** (0.054)	0.012 (0.054)	0.101** (0.045)	0.055 (0.041)	0.011 (0.034)	0.088* (0.047)	0.133** (0.059)	0.118** (0.059)
인문계 대비 공학계	0.102** (0.044)	0.087* (0.044)	0.147*** (0.035)	0.122*** (0.032)	0.164*** (0.026)	0.139*** (0.041)	0.191*** (0.044)	-0.013 (0.053)
인문계 대비 자연계	0.061 (0.044)	-0.013 (0.046)	-0.014 (0.038)	-0.005 (0.035)	0.010 (0.029)	0.095** (0.043)	0.139*** (0.047)	-0.062 (0.054)
인문계 대비 의약계	0.307*** (0.048)	0.131** (0.052)	0.200*** (0.041)	0.111*** (0.040)	0.189*** (0.032)	0.270*** (0.045)	0.245*** (0.048)	0.195*** (0.067)
인문계 대비 예체능계	0.022 (0.047)	-0.122** (0.050)	-0.092** (0.041)	-0.174*** (0.037)	-0.054* (0.031)	-0.076* (0.044)	0.007 (0.049)	-0.103* (0.055)
연도 더미 : 2008	-0.065* (0.036)	-0.096*** (0.035)	-0.063** (0.027)	-0.065*** (0.024)	-0.095*** (0.021)	-0.032 (0.031)	-0.075** (0.031)	-0.119*** (0.045)
연도 더미 : 2009	-0.001 (0.037)	-0.040 (0.037)	-0.033 (0.029)	-0.034 (0.026)	-0.040* (0.021)	0.047 (0.032)	-0.046 (0.031)	-0.020 (0.047)
(연도 더미 일부 생략)								
정주자 : 잔류확률	-0.018 (0.330)	-0.368 (0.281)	0.326 (0.282)	0.067 (0.316)	-0.101 (0.229)	0.239 (0.346)	-0.241 (0.365)	-0.386 (0.553)
정주자 : (잔류확률) <sup>2</sup>	0.034 (0.311)	0.483** (0.237)	-0.283 (0.304)	-0.072 (0.319)	0.089 (0.217)	-0.042 (0.266)	0.063 (0.353)	0.422 (0.417)
이주자 : 미실현 잔류확률	0.428 (0.290)	0.205 (0.288)	0.109 (0.200)	0.129 (0.242)	0.459** (0.196)	0.702* (0.399)	0.685** (0.335)	0.574 (0.695)
이주자 : (미실현 잔류확률) <sup>2</sup>	-0.451 (0.309)	-0.204 (0.313)	-0.045 (0.206)	-0.039 (0.267)	-0.610*** (0.211)	-0.612 (0.470)	-0.929** (0.401)	-1.024 (0.774)
이주자 : 전입확률	-2.376*** (0.878)	0.248 (0.797)	0.180 (0.695)	0.632 (0.724)	-0.162 (0.521)	-0.552 (0.805)	-0.645 (1.091)	0.245 (2.667)
이주자 : 전입확률 × 미실현 잔류확률	-0.076 (1.237)	-1.766 (1.118)	-0.070 (0.766)	-1.455* (0.762)	0.544 (0.640)	-0.051 (0.966)	0.071 (1.291)	1.520 (3.492)
이주자 : (전입확률) <sup>2</sup>	5.698*** (1.660)	0.230 (1.466)	0.236 (1.631)	-0.358 (1.504)	0.030 (0.906)	1.698 (1.351)	0.139 (2.440)	-5.562 (8.187)
상수항	3.532*** (0.843)	6.740*** (0.948)	3.862*** (0.751)	3.561*** (0.689)	4.775*** (0.548)	3.250*** (0.837)	4.847*** (0.829)	2.298** (1.131)
관측치 수	1,724	1,721	2,550	3,326	3,839	1,998	1,856	916
R-squared	0.175	0.213	0.256	0.249	0.307	0.191	0.238	0.196

Standard errors in parentheses \*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

잔류확률 :  $p_{ijj}$  (정주자의 경우), 미실현 잔류확률 :  $p_{ijj}$  (이주자의 경우), 전입확률 :  $p_{ijk}$  에 해당 (III장 참조)

<표 8> 임금함수의 주요 계수의 선택편의 보정 전후 비교

변수	서울	부산	대구	인천	광주	대전	울산	경기
	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금
남성 대비	-0.087***	-0.208***	-0.198***	-0.171***	-0.188***	-0.140***	-0.234***	-0.119***
여성	(0.007)	(0.016)	(0.019)	(0.020)	(0.023)	(0.021)	(0.025)	(0.009)
보정	-0.076***	-0.198***	-0.204***	-0.146***	-0.195***	-0.127***	-0.232***	-0.112***
후	(0.008)	(0.019)	(0.025)	(0.022)	(0.028)	(0.024)	(0.027)	(0.010)
2-3년제	0.152***	0.094***	0.074***	0.099***	0.075***	0.055***	0.108***	0.099***
대비 4년제	(0.007)	(0.014)	(0.019)	(0.020)	(0.022)	(0.021)	(0.022)	(0.008)
보정	0.137***	0.082***	0.082***	0.066***	0.082**	0.060*	0.132***	0.095***
후	(0.008)	(0.023)	(0.028)	(0.024)	(0.032)	(0.032)	(0.032)	(0.010)
인문 대비	0.087***	0.094***	0.018	0.125***	0.101**	-0.059	0.123***	0.190***
공학계열	(0.009)	(0.024)	(0.032)	(0.032)	(0.042)	(0.037)	(0.036)	(0.014)
보정	0.077***	0.107***	0.038	0.129***	0.095**	-0.038	0.120***	0.174***
후	(0.010)	(0.032)	(0.039)	(0.032)	(0.045)	(0.038)	(0.037)	(0.014)

Robust standard errors in parentheses

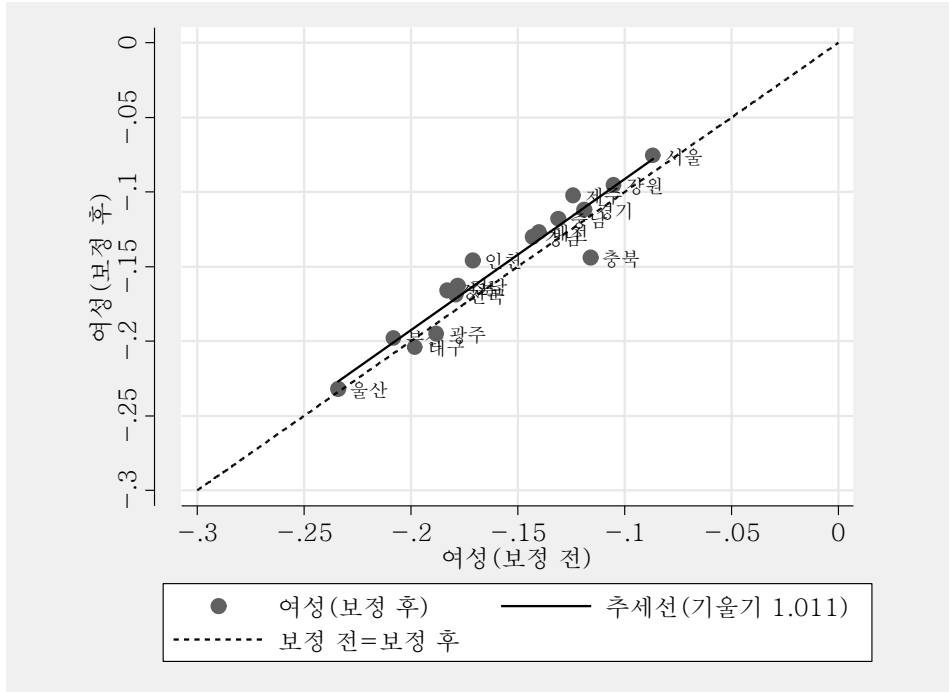
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

변수	강원	충북	충남	경북	경남	전북	전남	제주
	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금	로그임금
남성 대비	-0.105***	-0.116***	-0.131***	-0.183***	-0.143***	-0.179***	-0.178***	-0.124***
여성	(0.023)	(0.026)	(0.020)	(0.018)	(0.016)	(0.022)	(0.022)	(0.030)
보정	-0.095***	-0.144***	-0.118***	-0.166***	-0.130***	-0.169***	-0.163***	-0.102***
후	(0.025)	(0.028)	(0.022)	(0.020)	(0.017)	(0.026)	(0.025)	(0.035)
2-3년제	0.078***	0.019	0.031*	0.056***	0.109***	0.046**	0.029	0.071**
대비 4년제	(0.022)	(0.023)	(0.018)	(0.016)	(0.014)	(0.020)	(0.020)	(0.029)
보정	0.077***	0.076**	0.030	0.032	0.094***	0.056*	-0.006	0.062*
후	(0.029)	(0.032)	(0.024)	(0.025)	(0.020)	(0.032)	(0.026)	(0.035)
인문 대비	0.104**	0.089**	0.140***	0.130***	0.187***	0.112***	0.198***	-0.003
공학계열	(0.042)	(0.044)	(0.034)	(0.031)	(0.025)	(0.038)	(0.044)	(0.053)
보정	0.102**	0.087*	0.147***	0.122***	0.164***	0.139***	0.191***	-0.013
후	(0.044)	(0.044)	(0.035)	(0.032)	(0.026)	(0.041)	(0.044)	(0.053)

Robust standard errors in parentheses

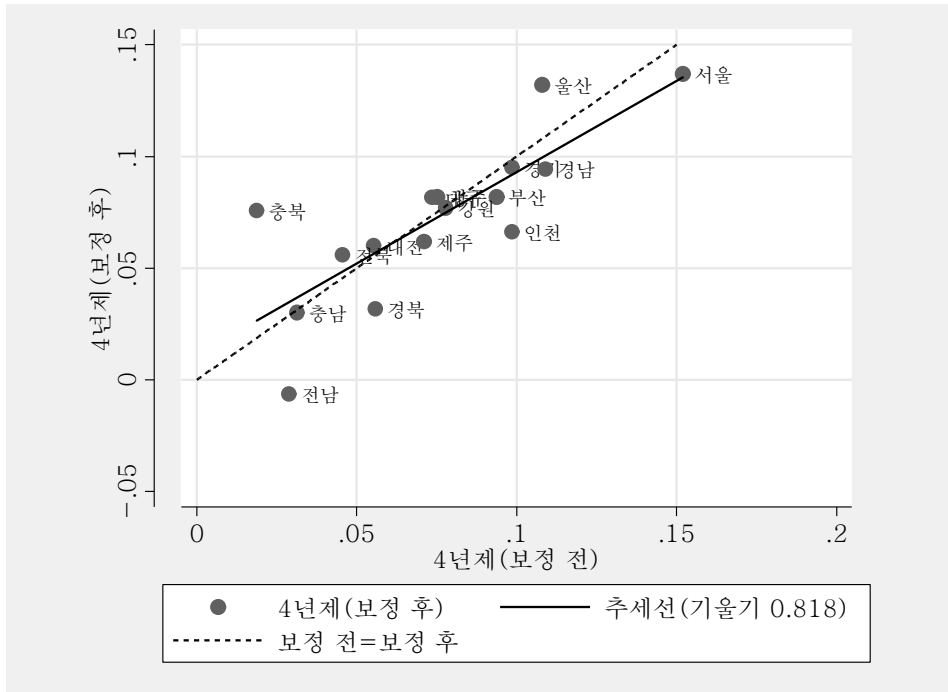
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

[그림 1] 임금함수 중 성별(남성 대비 여성) 계수의 선택편의 보정 전후 비교

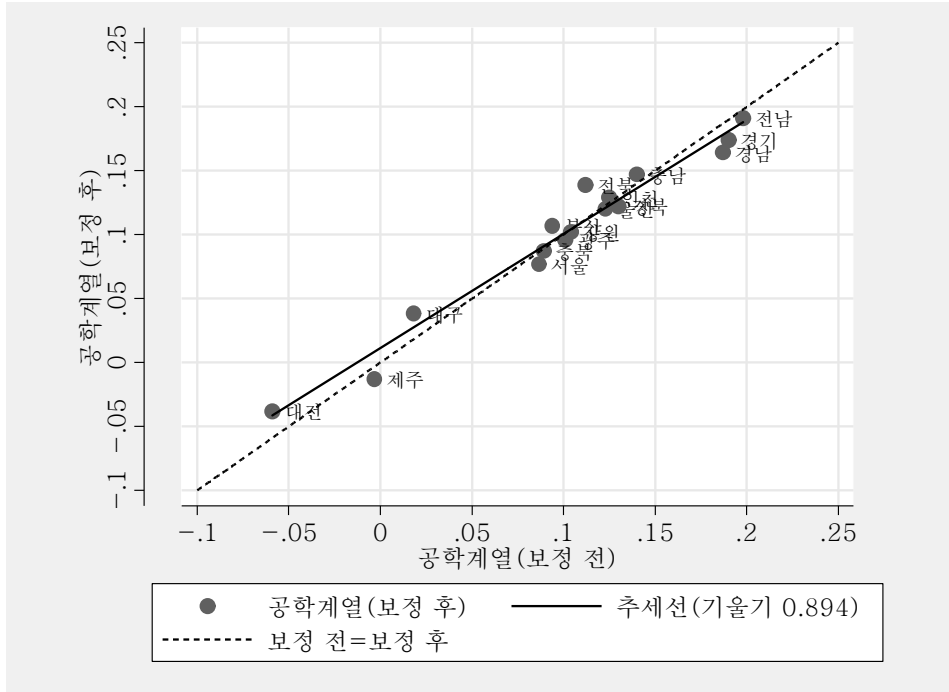




[그림 2] 임금함수 중 학력(2-3년제 대비 4년제) 계수의 선택편의 보정 전후 비교



[그림 3] 임금함수 중 전공계열(인문 대비 공학계열) 계수의 선택편의 보정 전후 비교



〈표 9〉 보정함수에 대한 Wald Test 결과

	서울	부산	대구	인천	광주	대전	울산	경기
Wald Test : 정주자 에 대한 선택편의 보정함수								
F-stat	3.013	0.194	1.078	0.867	0.391	0.335	2.615	4.666
p-value	0.049	0.824	0.341	0.420	0.676	0.715	0.074	0.009
Wald Test : 이주자 에 대한 선택편의 보정함수								
F-stat	2.504	0.876	1.822	1.824	1.966	1.070	1.251	6.664
p-value	0.028	0.496	0.105	0.105	0.081	0.375	0.283	0.000
Wald Test : 전체 선택편의 보정함수								
F-stat	6.265	1.129	1.402	6.348	1.514	1.190	4.330	11.02
p-value	0.000	0.341	0.200	0.000	0.158	0.305	0.000	0.000

	강원	충북	충남	경북	경남	전북	전남	제주
Wald Test : 정주자 에 대한 선택편의 보정함수								
F-stat	0.017	3.324	0.783	0.026	0.098	1.037	1.167	1.051
p-value	0.983	0.036	0.457	0.975	0.907	0.355	0.312	0.350
Wald Test : 이주자 에 대한 선택편의 보정함수								
F-stat	4.681	0.792	0.385	0.840	2.475	1.821	1.630	0.963
p-value	0.000	0.555	0.860	0.521	0.030	0.105	0.149	0.440
Wald Test : 전체 선택편의 보정함수								
F-stat	4.119	2.376	0.558	0.831	6.223	1.726	3.465	1.102
p-value	0.000	0.020	0.790	0.561	0.000	0.099	0.001	0.359

abstract

---

## Estimation of Wage Equation for College Graduates with Correction for Selection Bias upon Working State

Chiho Lee

In this paper, the wage equations of local labor markets for college graduates in Korea are estimated by Dahl(2002)'s methodology to correct for selection bias. The results suggest that the variations of coefficients in wage equations across the local labor markets are mostly remained after correcting for selection bias. The gender wage gap is hardly affected by selection bias. The variations of return to education and the major premium are reduced about 18% and 11% respectively. Meanwhile, the selection bias is negligible in the national capital region, which suggests that college graduates prefer the national capital region regardless of their gender, level of education, and major.

Keywords : local labor markets, self-selected migration, wage equation, selection bias, semiparametric estimation