

청년취업형태별 진입확률의 결정 요인 연구

정규직 및 대기업 취직 가능성에 스펙이 주는 영향을 중심으로

고병민, 김승현, 반준우, 성유지, 임경진

2021년 2월 5일

요약

취업난과 노동시장의 임금격차가 고착화되는 가운데 더 좋은 직장을 위해 많은 “스펙”을 획득하기 위해 노력하는 것이 당연시되고 있다. 본 연구의 목적은 청년구직자들의 기대소득과 학점, 해외수학경험 및 석박사 학위 등 다양한 직업능력의 척도, 그리고 성별과 유년기 성장지와 같은 개인적인 요인 사이의 관계를 규명하는 것이다.

본 연구는 청년패널조사(Youth Panel)에서 수집한 데이터를 단계별회귀를 사용해 분석하였다. 대기업 혹은 정규직 구직의 성공확률을 선형확률모형과 로지스틱 회귀 모형을 사용해 나타낸 뒤 각 모형의 매개변수를 추정하였다. 이후 이상의 분석에서 얻어진 확률들과 각각의 경우에 해당하는 고용자들의 평균소득을 사용한 가중평균으로 노동시장에 진입한 청년이 대기업이나 정규직에 취직했을 때의 기대소득을 추정하였다. 본고에서 제시하는 사실은, 다른 모든 조건이 일정할 때, 구직자들이 획득할 수 있는 다양한 직업능력의 척도 중 정규직 취직에 유의미한 영향을 미치는 것은 높은 학점과 해외수학경험, 그리고 업무 관련 자격증이며, 대기업 취직에는 높은 학점, 봉사활동 경험, 공인어학영어자격 획득 및 영어회화 능력 개발이 유의미한 영향을 주는 변수라는 것이다. 더 나아가 이런 척도를 보유한 구직자들은 더 높은 기대소득을 가진 것으로 나타났다. 이러한 발견은 구직희망자들이 구직경쟁에서 우위를 점하기 위해 이상의 자격들을 획득하는 동안 노동시장 진입을 늦출 유인이 있다는 것을 시사한다. 한편, 본 연구의 모형 선택과 다공성 문제도 논의되었다.

차 례

제 1 절 서론	2
1.1 연구 목적	2
1.2 선행연구	2
1.3 자료 수집 및 분석 방법	2
1.4 대상자의 일반적 특성 및 데이터 전처리 과정	3
제 2 절 방법론	6
2.1 선형확률모델	6
2.2 Index 모델	7
제 3 절 선형확률모델 추정 결과	12
3.1 정규직-비정규직 모델의 분석	12
3.2 대기업-중소기업 모형의 분석	13
3.3 공선성 분석	14
제 4 절 Index 모델 추정 결과	14
4.1 Index 모델의 선택	14
4.2 정규직-비정규직 데이터 분석	17
4.3 대기업-중소기업 데이터 분석	18
제 5 절 스페이 기대소득에 미치는 영향	19
제 6 절 결론 및 시사점	23

표 차례

1 회귀분석에 사용한 변수들의 설명	5
2 선형확률모델의 추정 결과	10
3 데이터, Index 모델 별 AIC	17
4 데이터, Index 모델 별 ΔAIC	17
5 정규직-비정규직 데이터의 로짓 모델 APE	20
6 대기업-중소기업 데이터의 로짓 모델 APE	21
7 정규직-비정규직 모델에 따른 개인의 스페에 따른 기대월급 변화량	22
8 대기업-중소기업 모델에 따른 개인의 스페에 따른 기대월급 변화량	22

제 1 절 서론

1.1 연구 목적

취업준비생들의 취업난과 노동 시장의 임금격차가 심화됨에 따라 취업경쟁 속에서 회사가 요구하는 스펙을 갖추는 것이 상당히 중요해지고 있다. 만약 특정한 자격증 유무에 따라 기대되는 취직 확률에 큰 차이가 존재할 경우에 취업준비생들은 그 자격증을 취득하기 위해 노동시장의 진입을 늦출 것이다. 마찬가지로 정규직-비정규직 또는 대기업-중소기업의 임금에 큰 격차가 존재할 경우에도 보다 높은 임금을 얻기 위해 각종 스펙을 취득하려 할 것이다. 본고의 목적은 스펙 쌓기 현상이 취직을 준비하는 개인의 합리적인 선택에서 비롯된 것임을 기대소득 모형을 사용해 검증하는 것이다. 우선 여러 스펙 중 정규직 및 대기업에 취직하는데 중요하게 작용하는 스펙을 확인하고, 각각의 스펙이 취직 확률과 기대 소득에 미치는 영향을 규명한다. 그 뒤 임금격차가 커질수록 스펙을 취득하는 것이 더욱 매력적임을 보인다.

1.2 선행연구

그동안 노동시장에 관한 연구는 대학 졸업자의 노동시장 성과를 평가하기 위해 대학 졸업자의 취업 후 소득을 비교하는 연구들이 주로 논의되었다. 최영섭(2003)은 졸업한 대학에서의 전공별 기대소득 격차를 분석하여 취업 자체의 가능성과 확률을 포함시켰다. 취업 확률의 차이와 취업 후 소득의 차이를 모두 감안하여 기대소득을 계산한 것이다. 선행 연구에서는 incidental truncation 모형과 선형확률모델을 사용했으나, 모형 설정에 따라 값 자체가 크게 변화하였다. 본고는 전공을 포함한 보다 포괄적인 범위의 스펙을 고려하여 장래의 기대 소득에 미치는 영향을 밝히고자 한다. 그리고 선형확률모델 뿐만 아니라 다양한 모델을 고려하여 결론이 특정 모델과 갖는 의존성을 줄일 것이다.

1.3 자료 수집 및 분석 방법

본고의 중요한 연구 질문은 현재 취업준비생이 취득하고 있는 다양한 스펙 중 정말로 정규직 및 대기업 취직에 유의미한 도움을 주는 변수가 무엇인지 가려내는 것이다. 이를 위해 제12차 청년패널조사(2018)에서 2007년 당시 만 15-29세였던 응답자들이 지금까지 보유한 스펙들과, 응답 당시 고용형태 및 고용된 기업의 규모, 그리고 학력, 성별, 14세 성장지 등 다양한 개인적 특성을 수집하였다. 그 외에도 스펙과 기대소득의 연관을 알아보기 위해 각 응답자의 응답 당시 월급을 수집하였다.

많은 변수 중 의미있는 변수를 골라내기 위하여 stepwise regression 기법을 아래 두 개의 선형확률모델(linear probability model)에 적용하였다.

정규직-비정규직 모델 $E(\text{정규직 취직 확률} | X) = X\beta + \varepsilon$, X 는 다양한 스펙변수 및 개인적 특성들.

대기업-중소기업 모델 $E(\text{대기업 취직 확률} | X) = X\beta + \varepsilon$, X 는 다양한 스펙변수 및 개인적 특성들.

표 1은 회귀분석에서 사용된 변수들이다. 설명변수 중 누구나 가지고 있는 ‘기초변수’로 성별, 생년, 학력 그리고 졸업한 학교와 성장지 및 주거형태를 고려하였고, 설명변수 중 취업을 위한 스펙을 나타내는 ‘특성변수’는 학점 관리와 자격증 그리고 인턴, 봉사활동, 해외경험과 같은 다양한

활동 및 경험 여부를 고려하였다. 이때 특성변수는 별도의 질적 구분 없이 유·무 여부만 수집하였다. 종속변수는 정규직과 비정규직 그리고 대기업과 중소기업으로 구분하였다. 비정규직과 정규직은 응답자 개인이 밝힌 자기선언적 비정규직이다. 따라서 본 연구에서 사용하는 정규직-비정규직의 정의는 경제활동인구조사 부가조사상의 정의와 다를 수 있다. 대기업-중소기업의 정의는 종사자 수 1000명 이상 여부를 기준으로 하였다. 우리는 이상의 자료를 활용하여 취득한 스펙에 따라 2007년에 만 15-29세의 나이였던 청년들이 2018년에 가진 직업의 고용형태와 직장의 규모에 어떤 차이가 있었는지 분석할 것이다.

한편 변수 선정 및 선형회귀모델의 OLS 추정이 끝난 뒤에는 그 결과를 다양한 이항종속변수 모델에서 제시하는 결과와 비교해 결과의 신뢰성을 높일 것이다. 모델별 비교분석의 결과는 절 4에 제시되었다. 본 연구의 자료 분석은 변수 선정을 위해 GNU R의 Rcmdr 패키지를 활용해 stepwise regression을 실시하였으며, 선형회귀모델을 활용하였다. Python 3의 Sklearn 패키지를 활용해 logistic regression을 실시하였다. 데이터 시각화에는 R에 내장된 시각화 패키지 및 ggplot2, ggpublisher 그리고 plotrix 패키지를 활용하였다.

1.4 대상자의 일반적 특성 및 데이터 전처리 과정

본 연구 대상자의 일반적 특성을 살펴보자. 학력은 인문계 고등학교를 졸업한 4년제 대졸이 대다수이며, 전공은 공학계열, 상경계열, 인문계열이 많은 부분을 차지한다. 취득한 사람이 많은 스펙부터 나열하면, 업무 관련 자격증(33%), 학점 관리(28%), 컴퓨터 관련 자격증(27%), 인턴 및 아르바이트 (23%), 영어 공인어학점수 취득(22%), 영어회화능력 개발(19%), 석사 이상의 학위, 제2외국어, 공모전 수상 경력, 한자검정시험 순이다.

준완전분리를 대한 절 2의 논의를 바탕으로, 본 연구에서 사용되는 청년패널데이터 중 준완전분리를 일으키는 독립변수는 제외하였다. GNU R의 detectseparation 패키지를 이용했을 때, 정규직-비정규직 모델에서는 대안학교, 해외고교를 나타내는 hsType7, 교대, 방통대, 사이버대를 나타내는 uniType4, uniType5, uniType6, 그리고 관사, 사택, 기숙시설을 나타내는 youthHouse5이 준완전분리를 일으키는 것으로 나타났고, 대기업-중소기업 모델에서는 hsType7, uniType5, uniType6와 youthHouse5가 정규직-비정규직 모델에서처럼 준완전분리를 일으키는 것으로 나타났고, 추가적으로 대학 전공 중 기타를 나타내는 uniMajor9과 고졸을 나타내는 newedu2, 대학졸업안한 사람들을 나타내는 uniType8이 준완전분리를 일으키는 변수로 지목되었다.

패키지 detectseparation에 의해서 준완전분리를 일으킨다고 지목된 변수는 대부분 이에 해당하는 표본 관측치의 수가 한 자리 수로 매우 적어서 준완전분리를 일으키는 것으로 보인다. 여기서 newedu2와 uniMajor8는 해당하는 표본 관측치의 수가 200개를 초과하는데도 준완전분리를 일으키는 변수로 지목되었는데, 이는 newedu2에는 속하지 않지만 uniMajor8에는 속하는, 고졸 미만을 나타내는 newedu4에 해당되는 표본 관측치의 수가 적어서 나타난 현상임을 발견하였다.

궁극적으로, 정규직-비정규직 모델에서는 hsType7, uniType4, uniType5, uniType6, youthHouse5, 그리고 대기업- 중소기업 모델에서는 hsType7, uniType5, uniType6, youthHouse5, uniMajor9, newedu4를 제외하였다. 이 모든 변수에 해당하는 표본 관측치의 수는 매우 적어서, 두 모델 모두

전처리 과정 이후에도 관측치가 2200개를 초과하였다.

변수명	설명	코드	비고
sex	성별	0: 남성, 1: 여성	
yob	생년		
newedu	학력	1: 대졸, 2: 고졸, 3: 전문대졸, 4: 고졸미만, 5: 석사학위 이상	
gpa	학점 관리 여부		
eng	토익, 토플 등 공인영어성적 여부		
engconv	영어회화 능력 개발 여부		
secfor	제2외국어 능력 개발 여부		
hanja	한자능력 개발 여부		
comp	컴퓨터 관련 자격증 취득 여부		
cert	업무(직무) 관련 자격증 취득 여부	0: 예, 1: 아니오	취업을 위해 해당 스펙을 준비했을 경우
abr	어학연수 등 해외경험 여부		
intern	인턴, 아르바이트 등 직무경험 여부		
grad	석사 및 박사학위 획득 여부		
vol	봉사활동 경험 여부		
cont	공모전 등 수상경력 소지 여부		
club	동아리 활동 여부		
hsType	졸업한 고등학교 형태	1: 일반계/인문계, 2: 특목고, 3: 상업계, 4: 공업계, 5: 기타 전문계, 6: 예체능계, 7: 대안학교, 해외고교, 8: 기타	
uniType	졸업한 대학 형태	1: 일반 4년제 대학, 2: 2-3년제 대학, 3: 산업대학, 4: 교육대학, 5: 방통대, 6: 사이버대학, 7: 폴리텍, 8: 대학졸업안함	
uniMajor	졸업한 대학 전공	1: 인문계열, 2: 사회, 상경계열, 3: 자연계열, 4: 공학계열, 5: 의/약학계열, 6: 교육계열, 7: 예체능계열, 8: 사관학교 및 경찰대, 9: 기타, 10: 모름, 11: 대학졸업안함	
youthLoc	만 14세 무렵 성장지	1: 서울, 2: 부산, 3: 대구, 4: 인천, 5: 광주, 6: 대전, 7: 울산, 8: 경기, 9: 강원, 10: 충북, 11: 충남, 12: 전북, 13: 전남, 14: 경북, 15: 경남, 16: 제주	
youthHouse	만 14세 무렵 주거형태	1: 자가, 2: 전세, 3: 월세, 4: 영구임대, 국민임대주택, 5: 관사, 사택, 기숙시설, 6: 기타	
perm	현재 고용형태	0: 정규직, 1: 비정규직	
corpsize	일하는 기업체의 종사자 수	0: 1000명 미만, 1: 1000명 초과	

표 1: 회귀분석에 사용한 변수들의 설명.

제 2 절 방법론

이항종속변수 모델이란 종속변수가 두 가지 값, 0과 1을 가지는 경우 종속변수를 설명하기 위해서 사용되는 모델이다. 이항종속변수 모델에서는 크게 두 가지 종류가 있는데, 선형확률모델(Linear Probability Model, LPM)과 Index 모델이다. 논의의 편의성을 위해 $(\Omega, \mathcal{F}, \mathbb{P})$ 를 확률공간(Probability Space)으로 놓고, 본 절에서 소개하는 모든 확률변수는 이 확률공간에서 정의된 변수이다.

2.1 선형확률모델

LPM 모델은 종속변수가 이항변수일 때의 고전적 선형회귀 모델(Classical Linear Regression Model)을 일컫는다. 구체적으로, LPM 모델은 다음과 같은 가정에 기반해있다고 생각하자:

선형성(Linearity) 종속변수와 독립변수는 선형적인 관계를 갖는다. 즉, 이항종속변수 \mathbf{y} 와 독립변수의 벡터 $(\mathbf{x}_1, \dots, \mathbf{x}_k) = \tilde{\mathbf{x}}$ 가 주어졌을 때, 실벡터 $(\beta_1, \dots, \beta_k) = \beta \in R^k$ 에 대해서 $\mathbf{y} = \tilde{\mathbf{x}}^t \beta = \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + \mathbf{u}$ 의 관계가 성립한다. 여기서 \mathbf{u} 는 오차항이며, 독립변수에는 상수 1이 포함될 수도 있다.

비정칙성(Non-Singularity) 주어진 데이터는 완벽한 공선성 (Perfect Collinearity) 를 보이지 않는다. 즉, 관측된 표본을

$$\begin{pmatrix} y_1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & x_{n1} & \cdots & x_{nk} \end{pmatrix} = \begin{pmatrix} Y & X \end{pmatrix} \in M_{n \times (k+1)}(R) \quad (1)$$

으로 나타낼 때, X 의 열벡터는 선형독립이다.

독립변수와 오차항의 독립성 주어진 독립변수는 오차항과 독립이다.

이 세 가지 가정 하에서 계수 β_1, \dots, β_k 의 최소제곱(Ordinary Least Squares, OLS) 추정량은 유일하며 관측된 표본이 $(Y \ X)$ 로 주어졌을 때 $\hat{\beta}(Y, X) = (X^t X)^{-1} X^t Y$ 로 계산된다. 추가적으로, 위의 독립성 가정 하에서 $\mathbb{E}(\mathbf{y} | \tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^t \beta$ 로 주어지는데, 여기서 종속변수 \mathbf{y} 는 이항변수이기 때문에 사건 $\{\mathbf{y} = 1\} \in \mathcal{F}$ 에 대해서 지시변수 (Indicator Variable) $\mathbf{y} = I_{\{\mathbf{y}=1\}}$ 의 형태로 나타낼 수 있다. 결국 독립변수 $\tilde{\mathbf{x}}$ 의 값이 주어졌을 때 사건 $\{\mathbf{y} = 1\}$ 이 일어날 조건부확률은 $\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}}) = \mathbb{E}(I_{\{\mathbf{y}=1\}} | \tilde{\mathbf{x}}) = \mathbb{E}(\mathbf{y} | \tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^t \beta$ 로 주어진다. 이는 관측된 표본 $(Y \ X)$ 을 이용해서 OLS 추정량을 계산하고 새로운 독립변수의 조합 \tilde{x}_0 이 주어졌을 때 사건 $\{\mathbf{y} = 1\}$ 이 일어날 조건부확률을 $\mathbb{P}(\widehat{\mathbf{y} = 1} | \tilde{x}_0) = \tilde{x}_0^t \hat{\beta}(Y, X)$ 로 추정할 수 있음을 의미한다. 앞으로는 조건부확률 $\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}})$ 을 기대확률, 이의 추정량 $\widehat{\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}})}$ 을 기대확률의 추정량이라고 한다.

LPM의 장점은 다음과 같다.

추정의 편의 최우추정법 (Maximum Likelihood Estimation)을 사용하지 않고 단순한 행렬 연산을 사용하여 바로 계수 추정량, 혹은 베타값을 구할 수 있다. 이는 LPM의 추정 속도가 다른 모델에 비해 현저히 뛰어나다는 것을 보여준다.

해석의 용이 LPM 결과의 해석은 여타 모델에 비해 뛰어나다. 일례로, LPM에서의 베타값은 그대로 독립변수가 기대확률에 미치는 효과로 해석할 수 있는 반면, 로짓 모델만 보더라도 베타값을 독립변수가 Odds Ratio에 미치는 효과라는, 직관적으로 이해하기 힘든 방식으로 해석해야한다.

이와 같은 장점에도 불구하고, LPM의 치명적인 약점은 기대확률의 추정량이 0과 1 사이에 국한되지 않는다는 것이다. 즉, 어떠한 독립변수값의 조합이 주어졌을 때, 기대확률의 추정량이 음수일 수도, 1 보다 클 수도 있다는 것이다. 이런 기대확률은 해석이 불가하므로, 다른 방식으로 기대확률을 추정해야 한다.

2.2 Index 모델

Index 모델은 LPM과 다른 방식으로 기대확률을 추정하는 이항종속모델을 일컫는다. Index 모델은 기대확률을 0과 1 사이에 국한시키기 위해서 고전 선형회귀에서의 적합치(Fitted Value) 부분을 0과 1 사이의 값을 갖는 함수 안에 넣음으로써 기대확률을 계산하게 되는데, 이때 이 함수를 link function이라고 부른다. 구체적으로, 이항종속변수 y 와 독립변수의 벡터 $(\mathbf{x}_1, \dots, \mathbf{x}_n) = \tilde{\mathbf{x}}$ 가 주어졌을 때, Index Model에서는 기대확률이 $\mathbb{P}(y = 1 | \tilde{\mathbf{x}}) = G(\tilde{\mathbf{x}}^t \beta)$ 로 주어지는데, 여기서 G 는 실선 \mathbb{R} 에서 정의되고 0과 1 사이 값을 갖는 link function이다.

Index model의 종류에는 표준정규분포의 누적분포함수(CDF)를 link function으로 이용하는 프로빗(Probit) 모델, 앞서서 설명했다시피 로지스틱 분포의 CDF를 link function으로 이용하는 로짓(Logit) 모델, 그리고 일반화 극단값(Generalized Extreme Value) 분포의 CDF를 link function으로 이용하는 Complementary Log-Log (CLL) 모델이 있다.

대부분의 경우 특정한 link function G 가 어떤 잠재변수 회귀(Latent Variable Regression)로부터 도출되었다고 가정한다. 어떤 실수값을 가지는 확률변수 y^* 가 존재하고, 사건 H 는 $H = \{y^* > 0\} \in \mathcal{F}$ 로 정의된다고 가정하자. 이때 새로운 이항변수 y 를 $y = I_H$ 로 정의하자. y 는 y^* 의 부호를 나타내는 지시변수로 볼 수 있는데, 이때 y^* 를 y 의 잠재변수라고 한다.

종속변수 y^* 를 독립변수들 $\mathbf{x}_1, \dots, \mathbf{x}_k$ 에 회귀시키자. 오차항을 e 로 표시한다면, 회귀식은 $y^* = \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + e = \tilde{\mathbf{x}}^t \beta + e$ 이 된다. $\tilde{\mathbf{x}}$ 가 주어진 e 의 조건부 분포는 다음을 만족시킨다고 하자.

- 위 조건부 분포는 연속 분포이다
- 위 조건부 분포의 지지집합(Support)은 실선 전체이다
- 위 조건부 분포는 0에 대해 대칭이다. 즉, 모든 $c \geq 0$ 에 대해 $\mathbb{P}(e \leq -c | \tilde{\mathbf{x}}) = \mathbb{P}(e > c | \tilde{\mathbf{x}}) = 1 - \mathbb{P}(e \leq c | \tilde{\mathbf{x}})$.
- 위 조건부 분포는 $\tilde{\mathbf{x}}$ 의 값에 따라 변화하지 않는다.

$\tilde{\mathbf{x}}$ 가 주어진 e 의 조건부 분포의 CDF를 $F : R \rightarrow (0, 1)$ 이라고 하면¹, 다음과 같은 관계가 성립함을

¹ $\tilde{\mathbf{x}}$ 가 주어진 e 의 조건부 분포는 $\tilde{\mathbf{x}}$ 의 값에 따라 변화하지 않는다고 가정했으므로 각 $c \in R$ 에 대해서 $F(c) = \mathbb{P}(e \leq c | \tilde{\mathbf{x}})$ 는 상수이다

볼 수 있다.

$$\begin{aligned}\mathbb{P}(\mathbf{y} = 1 \mid \tilde{\mathbf{x}}) &= \mathbb{P}(\mathbf{y}^* > 0 \mid \tilde{\mathbf{x}}) = \mathbb{P}(\mathbf{e} > -\tilde{\mathbf{x}}^t \beta \mid \tilde{\mathbf{x}}) = 1 - \mathbb{P}(\mathbf{e} \leq -\tilde{\mathbf{x}}^t \beta \mid \tilde{\mathbf{x}}) \\ &= 1 - (1 - \mathbb{P}(\mathbf{e} \leq \tilde{\mathbf{x}}^t \beta \mid \tilde{\mathbf{x}})) = \mathbb{P}(\mathbf{e} \leq \tilde{\mathbf{x}}^t \beta \mid \tilde{\mathbf{x}}) = F(\tilde{\mathbf{x}}^t \beta)\end{aligned}$$

여기서 첫 등호는 잠재변수의 정의에 의해, 네 번째는 위 조건부 분포의 대칭성에 의해, 그리고 마지막은 F 의 정의에 의해 정당화할 수 있다. 결국 $\mathbb{P}(\mathbf{y} = 1 \mid \tilde{\mathbf{x}}) = G(\tilde{\mathbf{x}}^t \beta)$ 를 만족시키는 link function G 는 정확히 F 이다.

프로빗과 로짓 모델의 경우, 표준정규분포와 로지스틱 분포의 CDF는 대칭적이기 때문에 위에서 보인 것처럼 link function이 곧 그 분포들의 CDF이다. 만약 CLL 모델에서처럼 대칭성 가정이 성립하지 않는다면 $\tilde{\mathbf{x}}$ 가 주어진 \mathbf{e} 의 조건부 분포의 CDF는 link function G 와 일치한다고 할 수 없다. CLL 모델의 link function G 는 모든 실수 $t \in R$ 에 대해서 $G(t) = 1 - \exp(-\exp(-t))$ 로 주어져 있는데, 이는 $\mathbb{P}(\mathbf{y} = 1 \mid \tilde{\mathbf{x}}) = G(\tilde{\mathbf{x}}^t \beta) = 1 - \exp(-\exp(-\tilde{\mathbf{x}}^t \beta))$ 임을 의미한다. 여기서 종속변수 \mathbf{y} 가 반대로 주어져있을때, 즉

$$\mathbf{z} = \begin{cases} 1 & \text{if } \mathbf{y} = 0 \\ 0 & \text{if } \mathbf{y} = 1 \end{cases}$$

를 만족시키는 확률변수 \mathbf{z} 가 있을때, $\mathbb{P}(\mathbf{z} = 1 \mid \tilde{\mathbf{x}}) = \mathbb{P}(\mathbf{y} = 0 \mid \tilde{\mathbf{x}}) = 1 - \mathbb{P}(\mathbf{y} = 1 \mid \tilde{\mathbf{x}})$ 로 정의된다. CLL 모델에서 사용하는 일반화 극단값 분포의 CDF는 대칭적이지 않기 때문에

$$\mathbb{P}(\mathbf{z}) = 1 - \mathbb{P}(\mathbf{y} = 1 \mid \tilde{\mathbf{x}}) = 1 - G(\tilde{\mathbf{x}}^t \beta) = 1 - (1 - \exp(-\exp(\tilde{\mathbf{x}}^t \beta))) = \exp(-\exp(-\tilde{\mathbf{x}}^t \beta))$$

로, \mathbf{y} 에 대응되는 CLL link function 과 \mathbf{z} 에 대응되는 link function이² 같지 않다.

본 논문에서 다루는 Index 모델들은 다음과 같은 가정에 기반했다:

잠재변수 특성 (Latent Variable Specification) 이항종속변수 \mathbf{y} 와 독립변수의 벡터 $(\mathbf{x}_1, \dots, \mathbf{x}_k) = \tilde{\mathbf{x}}$ 가 주어졌을 때, \mathbf{y} 는 잠재변수 \mathbf{y}^* 를 잠재변수로 가지고 있는데, 독립변수들은 \mathbf{y}^* 와 선형 관계를 가지고 있으며, 오차 \mathbf{e} 의 분포는 각 모델의 link function G 가 유도될 수 있는 상기 조건들을 만족한다. 예를 들어, 프로빗 모델에서는 $\mathbf{e} \mid \tilde{\mathbf{x}} = \tilde{x} \sim N(0, 1)$ 이다.

표본 관측치의 독립성 독립변수의 표본 관측치는 서로 독립이다. 즉, n 개의 관측치가 있는 표본에서 독립변수의 i 번째 관측치를 확률벡터 $(\mathbf{x}_{i1}, \dots, \mathbf{x}_{ik}) = \tilde{\mathbf{x}}_i$ 로 나타낼 때, $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n$ 은 서로 독립이다.

\mathbf{y} 는 이항변수이기 때문에 특정한 $\tilde{\mathbf{x}}$ 값이 주어진 \mathbf{y} 의 조건부 분포는 베르누이 분포를 따르게 된다. 위의 두 가정 하에서, 관측된 표본의 종속변수 값이 y_1, \dots, y_n , 독립변수 벡터가 $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n$ 일

²o| link function $F(x) = \exp(-\exp(c))$ 는 편의상 Log-Log(LL) Link Function이라고 부른다

때, 관측된 표본의 log-likelihood function은

$$\begin{aligned} l(y_1, \dots, y_n | \tilde{x}_1, \dots, \tilde{x}_n, \beta) &= \log \left(\prod_{i=1}^n G(\tilde{x}_i^t \beta)^{y_i} (1 - G(\tilde{x}_i^t \beta))^{1-y_i} \right) \\ &= \sum_{i=1}^n [y_i \log(G(\tilde{x}_i^t \beta)) + (1 - y_i) \log(1 - G(\tilde{x}_i^t \beta))] \end{aligned}$$

으로 주어진다. 이를 극대화시키는 β 값을 수치해석적으로 구할 수 있는데, 이 값이 Index 모델의 최우추정량 $\hat{\beta}$ 이다.

위의 최우추정량 $\hat{\beta}$ 의 좌표들이 유한한 값을 가지기 위해서는 주어진 데이터에서 완전 분리 (Complete Separation) 혹은 준완전 분리(Quasi-complete Separation)이 일어나지 않아야 한다. 관측된 표본을

$$\begin{pmatrix} y_1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & x_{n1} & \cdots & x_{nk} \end{pmatrix} = \begin{pmatrix} y_1 & \tilde{x}_1^t \\ \vdots & \vdots \\ y_n & \tilde{x}_n^t \end{pmatrix} \in M_{n \times k}(R)$$

의 행렬로 나타낼 수 있다고 하자. 이때, 완전 분리란 모든 $1 \leq i \leq n$ 에 대해, $y_i = 0$ 이면 $c^t \tilde{x}_i < 0$ 이고 $y_i = 1$ 이면 $c^t \tilde{x}_i > 0$ 인 벡터 $c \in R^k$ 가 존재하는 상태를 일컫는다. 같은 맥락으로, 준완전 분리란 모든 $1 \leq i \leq n$ 에 대해, $y_i = 0$ 이면 $c^t \tilde{x}_i \leq 0$ 이고 $y_i = 1$ 이면 $c^t \tilde{x}_i \geq 0$ 인 벡터 $c \in R^k$ 가 존재하는 상태를 말한다. 즉, 관측된 종속변수 값을 0으로 만드는 독립변수 관측치와 1로 만드는 독립변수 관측치를 분리하는 R^k 상의 초평면 $c^t x = 0$ 이 존재하면 데이터가 분리되어있다고 하는데, 이때 초평면 상에 위치한 독립변수 관측치가 있으면 준완전 분리, 없으면 완전 분리라고 한다.

(Konis, 2003)의 Theorem 1과 Theorem 2에 의해 데이터가 완전 혹은 준완전 분리 되어있으면 최우추정량 $\hat{\beta}$ 의 좌표 중에 유한한 값으로 수렴하지 않는 좌표가 존재한다. 보통 (그리고 본 연구에서) 완전 혹은 준완전분리가 관측되는 이유는 관측치의 부족이므로, 완전 혹은 준완전분리를 야기하는 변수가 관측될 경우, 그 변수를 제외하고 연구를 진행한다.

마지막으로, 본 연구에서 비정규직에 들어가는 사건과 정규직에 들어가는 사건이 Sample Space 을 분할한다고 가정한다. 즉, 비정규직에 들어가지 않는다면 정규직에 들어가는 것이고, 정규직에 들어가지 않는다면 비정규직에 들어가는 것이라고 가정했습니다. 비슷한 방식으로, 대기업에 들어가는 사건과 중소기업에 들어가는 사건 또한 대응하는 Sample Space를 분할한다고 가정했다.

표 2: 선형확률모델의 추정 결과.

	정규직-비정규직 모델 Two-sided stepwise. [종속변수] 정규직: 0, 비정규직: 1	대기업-중소기업 모델 Two-sided stepwise. [종속변수] 중소기업: 0, 대기업: 1
나쁜 학점	0.07*** (0.02)	나쁜 학점 (0.026)
석사 미만	-0.11** (0.04)	영어 공인어학자격 없음 (0.033)
업무 관련 자격증 없음	0.03 (0.02)	영어 회화능력 없음 (0.03)
인턴 및 아르바이트 경험 없음	-0.05* (0.02)	공모전 수상경력 없음 (0.04)
해외 경험 없음	0.06 (0.03)	봉사활동 경험 없음 (0.032)
동아리 경험 없음	0.05 (0.03)	
사회, 상경계열 전공	-0.10*** (0.03)	사회, 상경계열 전공 (0.03)
자연계열 전공	-0.10** (0.04)	자연계열 전공 (0.04)
공학계열 전공	-0.13*** (0.03)	공학계열 전공 (0.03)
의/약학계열 전공	-0.17*** (0.04)	의/약학계열 전공 (0.04)
교육계열 전공	0.01 (0.05)	교육계열 전공 (0.05)
예체능계열 전공	-0.03 (0.04)	예체능계열 전공 (0.04)
사관학교 및 경찰대 졸업	0.00 (0.13)	사관학교 및 경찰대 졸업 (0.15)
기타 전공 졸업	0.15 (0.21)	졸업한 전공 모름 (0.08)
졸업한 전공 모름	-0.16* (0.06)	
대학졸업 안함	0.17 (0.16)	
특목고 졸업	0.00 (0.05)	2-3년제 대학 졸업 (0.02)
상업계 고등학교 졸업	-0.03 (0.03)	산업대학 졸업 (0.06)
공업계 고등학교 졸업	0.04 (0.03)	교육대학 졸업 (0.49)
기타 전문계고 졸업	-0.08 (0.06)	폴리텍 졸업 (0.08)
예체능계 고등학교 졸업	0.23** (0.08)	
기타 고등학교 졸업	0.25 (0.08)	
고졸	-0.15 (0.17)	만14세주거형태: 전세 (0.02)
전문대졸	0.06** (0.02)	만14세주거형태: 월세 (0.05)
석사학위 이상	-0.01 (0.04)	만14세주거형태: 영구임대, 국민임대주택 (0.11)
만14세성장지: 부산	0.03 (0.03)	

만14세성장지: 대구	-0.06 (0.03)		
만14세성장지: 인천	-0.01 (0.04)		
만14세성장지: 광주	0.01 (0.04)		
만14세성장지: 대전	-0.08* (0.04)		
만14세성장지: 울산	0.00 (0.06)		
만14세성장지: 경기	-0.03 (0.02)		
만14세성장지: 강원	0.01 (0.07)		
만14세성장지: 충북	-0.01 (0.06)		
만14세성장지: 충남	-0.14** (0.05)		
만14세성장지: 전북	0.05 (0.05)		
만14세성장지: 전남	0.08 (0.05)		
만14세성장지: 경북	-0.11* (0.05)		
만14세성장지: 경남	-0.06 (0.04)		
만14세성장지: 제주	-0.03 (0.13)		
여성	0.03 (0.02)	여성	-0.04* (0.02)
상수항	0.20*** (0.05)	상수항	0.48*** (0.05)
관측수	2188	관측수	2204
R ²	0.07342	R ²	0.1116
Adjusted R ²	0.05572	Adjusted R ²	0.1026
Residual Std. Error	0.362 (df = 2146)	Residual Std. Error	0.4068 (df = 2181)
F Statistic	4.148*** (df = 41; 2146)	F Statistic	12.45*** (df=22, 2181)

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

제 3 절 선형화률모델 추정 결과

본 절에서는 정규직 혹은 대기업 취직에 중요하게 작용하는 스펙들이 무엇인지 선형화률모델(LPM 모델)을 사용하여 살펴본다. 또한, 각각의 스펙이 정규직 혹은 대기업 취직 확률을 얼마나 높이는지 확인한다.

정규직이 0이고 비정규직이 1인 이진변수를 종속변수로 하는 ‘정규직-비정규직 모델’과 중소기업이 0이고 대기업이 1인 이진변수를 종속변수로 하는 ‘대기업-중소기업 모델’을 살펴보자. 표 1에 제시된 설명변수 중 어떤 변수를 사용한 모델이 Akaike information criterion(AIC)를 기준으로 가장 설명력이 높은지 알아보기 위하여 Forward stepwise regression, backward stepwise regression, 그리고 two-sided stepwise regression을 실행했다. 모델의 설명력이 가장 높은 설명변수들의 집합을 찾을 때, 변수를 하나씩 추가하면서 찾는 방식이 forward stepwise regression이고, 변수를 하나씩 줄여가면서 찾는 방식이 backward stepwise regression이다. Two-sided stepwise regression은 변수를 추가하고 빼는 과정을 여러 번 반복하며 예측력이 가장 좋은 모델을 찾는 방식이다.

3.1 정규직-비정규직 모델의 분석

먼저 여러 스펙과 개인적 특성에 따른 정규직 취직 확률을 분석해보자. 정규직-비정규직 모델에서 two-sided 단계별 회귀를 통해 선택된 변수들과 그 변수들의 계수 추정값은 표 2에 제시되어 있다. Two-sided 단계별 회귀분석 결과, 정규직에 취직할 기대확률에 중요한 영향을 끼치는 스펙은 좋은 학점, 석사 및 박사학위 획득 여부, 업무 관련 자격증 취득 여부, 인턴 및 아르바이트 경험 여부, 해외 경험 여부, 동아리 경험 여부인 것으로 나타났다. 그 밖에도 졸업한 대학의 전공과 졸업한 고등학교의 형태, 학력과 성별, 그리고 만14세 무렵 성장지도 정규직 취직 확률을 설명하는 변수로 선택되었다. 변수 선택 결과에 따라 ‘학점을 관리하지 않고 자격증과 직무경험, 해외 경험이 없으며 동아리 활동을 하지 않은, 일반계/인문계 고등학교를 졸업하고 대학에서 인문계열을 전공한, 만 14세 무렵 서울에서 성장한 대졸 남성’을 기준으로 하자. 선형화률모델을 추정한 결과, 이 개인이 정규직에 취직할 확률은 약 74.8%이다.

표 2에 제시되어 있는 계수 추정값들의 의미를 자세히 살펴보자. 기준이 되는 개인과 다른 특성은 동일하면서 학점을 관리하였을 경우에는 정규직 취직 확률이 약 7%p 증가한다. 업무 관련 자격증이 있으면 약 3%p, 해외 경험이 있으면 약 6%p, 동아리를 한 적이 있으면 약 5%p 만큼 정규직 취직 확률이 증가한다. 반면, 인턴 및 아르바이트 경험이 있을 경우 정규직에 취직할 확률은 약 5%p 낮아지는 것으로 나타났다. 석사 이상의 학위를 취득하는 것도 정규직 취직 확률을 약 10%p 하락시키는 것으로 나타났다. 많은 변수 추정값의 부호가 예상과 일치했지만, 인턴 및 아르바이트 경험과 석사 이상의 학위를 나타내는 변수는 직관에 반하는 결과를 보여줬다. 많은 청년들이 더 좋은 직장에 취직하기 위해 인턴 활동을 하고 석사 이상의 학위를 얻으려는 세태에 미루어 볼 때, 인턴 및 아르바이트, 석사 이상의 학위가 정규직 취직 확률을 감소시킨다는 것은 납득하기 어려운 결과이다. 이에 대한 설명은 소절 3.3에 있다.

대학 전공별 정규직 취직 확률을 살펴보자. 인문 계열 전공에 비하여, 사회, 상경계열과 자연계열은 약 10%p, 공학계열은 약 13%p, 의/약학계열은 약 17%p 만큼 정규직 취직 확률이 더 높았다.

졸업한 고등학교 형태에 따라서는, 일반계/인문계 고등학교에 비하여 예체능계 고등학교와 공업계 고등학교를 졸업했을 경우 정규직 취직 확률이 각각 약 23%p, 약 4%p 낮았다. 특목고를 졸업과 일반계/인문계고 졸업의 정규직 취직 기대 확률이 크게 다르지 않다는 점도 흥미롭다. 앞서 기준이 되는 개인과 다른 특성은 모두 동일하면서 최종학력이 ‘고졸’, ‘전문대졸’인 경우에는 정규직 취직 확률이 각각 2.3%p, 6%p 감소하는 것으로 나타났으며, 여성은 남성에 비해 정규직 취직 기대 확률이 약 3%p 낮았다. 마지막으로 지역에 따른 정규직 취직의 기대확률을 살펴보자. 만 14세 무렵 성장지를 정규직 취직 기대 확률이 가장 높은 곳부터 순서대로 나열하면, 충남, 경북, 대전, 경남, 대구, 경기, 제주, 충북, 인천, 울산, 서울, 강원, 광주, 부산, 전북, 전남 순이다. 특히 충남과 경북은 서울에 비해 정규직 취직 확률이 약 10%p 더 높았다.

3.2 대기업-중소기업 모형의 분석

여러 스펙과 개인적 특성에 따른 대기업 취직확률을 앞과 유사한 방식으로 선형확률모형과 two-sided 단계별 회귀를 사용해 분석하였다. 그 결과는 표 2에 제시되어 있다.

단계별 회귀분석 결과 대기업에 취직할 기대확률에 중요한 영향을 끼치는 스펙은 영어 공인어학 성적 취득 여부, 영어회화능력 개발 여부, 봉사활동 경험 여부, 공모전 수상경력 여부, 그리고 좋은 학점인 것으로 나타났다. 그 밖에도 졸업한 대학의 종류와 전공, 성별, 그리고 만14세 당시 주거형태도 대기업 취직 확률을 설명하는 변수로 선택되었다. 영어 공인어학성적을 취득하였을 경우 그렇지 않았을 때보다 대기업 취직의 기대확률은 약 12%p 높아지며, 영어회화 능력이 있을 경우 약 5.3%p, 좋은 학점을 유지했을 경우 약 5.7%p, 그리고 봉사활동 경험이 있을 경우 약 7.3%p 더 높아진다. 그러나 공모전 수상경력은 대기업 취직확률을 약 7.7% 낮추었는데, 이것은 직관적으로 납득하기 어려운 결과다. 이에 대한 설명은 소절 3.3에 있다.

대학 전공과 형태의 영향도 뚜렷했다. 사회(상경)계열은 인문계열에 비해 대기업 취직의 기대 확률이 약 3.6%p 더 높았고, 자연계열은 약 8.7%p 더 낮았다. 공학계열은 인문계열과 대기업 취직 확률에서 큰 차이를 보이지 않았다. 한편 2-3년제 대학과 산업대학, 그리고 기능대학을 졸업했을 경우 대기업에 취직할 확률이 각각 10.6%p, 5.4%p, 그리고 5.2%p 정도 더 낮았다.

성별과 만14세 무렵 주거형태가 대기업 취직 기대확률을 설명하는 변수로 채택된 것은 흥미롭다. 여성은 정규직 취직뿐만 아니라 대기업 취직 확률에 있어서도 남성에 비해 불리했으며, 그 크기는 약 4.1%p로 나타났다. 한편 만14세 무렵 월세의 주거형태를 가졌던 사람은 자가에 살고 있던 사람에 비해 대기업 취직확률이 약 6.5% 낮았고, 영구임대 및 국민임대주택에 살고 있던 사람은 약 24.4% 높았다. 만14세 무렵 주거형태를 설명변수로 삼은 모형은 주거형태에 따른 가정별 특성을 통제하는 것으로 해석할 수 있다. 월세 주택에 살고 있는 가정이 자가 주택에 살고 있는 가정보다 경제적 형편이 나빴던 것으로 생각한다면, 이상의 결과는 만14세 무렵 가정의 경제적 상황이 어려웠을수록 대기업 취직에 불리하다는 점을 시사한다. 그러나 영구임대 및 국민임대주택에서 유년기를 보낸 사람이 자가에 거주하던 사람에 비해 대기업 취직확률이 더 높았다는 사실은 저소득 가구 중 최저소득 계층에 해당했던 사람들에게는 대기업에 취직할 확률이 비교적 높음을 의미한다. 이것은 국내 주요 대기업들이 신입사원 공채시 저소득층 지원자들에게 가산점을 부여하는 사실과 연관이 있는 것으로

짐작된다.

3.3 공선성 분석

인턴 및 아르바이트, 석사 이상의 학위, 그리고 공모전 수상경력이 정규직과 대기업 취직 확률과 음의 관계를 가진다는 것은 납득하기 어렵다. 이것은 세 변수들이 다른 스펙변수들과 강한 연관이 있기 때문에 OLS 추정값에 오차가 생긴 것으로 볼 수 있다. 예컨대 인턴 경험이 있는 사람은 위에서 제시된 다른 변수들을 가졌을 개연성이 매우 크기 때문에 공변성 문제가 발생했을 수 있다. 실제로 그림 1, 2, 3, 4에서 보듯 석사 이상의 학위, 학점, 인턴 및 아르바이트 경험, 그리고 업무관련 자격증 취득의 네 변수는 높은 연관성을 보였다.

제 4 절 Index 모델 추정 결과

본 절에서는 Index 모델을 한 가지 선택한 이후, Index 모델과 LPM 추정의 결과를 비교하고 어떤 점에서 Index 모델이 LPM을 보완하는지 제시한다.

4.1 Index 모델의 선택

본 연구에서 고려한 Index 모델은 크게 로짓, 프로빗, CLL, 그리고 LL (종속변수의 방향을 바꾼 후의 CLL 모델) 네 가지이다.

우선 로짓 모델같은 경우는 다른 Index 모델을 선택할 이유가 없을 경우 흔히 선택되는 모델인데, 이는 로짓 모델의 계수는 log-odds에 미치는 효과라는, 비교적 직관적인 해석을 할 수 있기 때문이다. 로짓 모델은 link function으로 정규 로지스틱 분포의 CDF $G(x) = \frac{e^x}{1+e^x}$, $x \in R$ 를 사용한다. 즉, 종속변수가 \mathbf{y} , 독립변수의 벡터가 $(\mathbf{x}_1, \dots, \mathbf{x}_k) = \tilde{\mathbf{x}}$, 그리고 계수의 벡터가 $(\beta_1, \dots, \beta_k) = \beta \in R^k$ 일 때, $\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}}) = G(\tilde{\mathbf{x}}^t \beta) = \frac{e^{\beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k}}{1 + e^{\beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k}}$ 로 주어진다. 기대확률 $\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}})$ 을 π 로 표시한다면, 이는 $\log\left(\frac{\pi}{1-\pi}\right) = \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k$ 임을 의미한다. 여기서 $\frac{\pi}{1-\pi}$ 는 종속변수가 1의 값을 가질 승산을 나타내는 승산비(Odds Ratio) 임으로, 어떤 $1 \leq i \leq n$ 에 대해서 β_i 는 다른 변수들이 변하지 않을 때 i 번째가 변수의 증가가 log-odds인 $\log\left(\frac{\pi}{1-\pi}\right)$ 에 미치는 영향을 나타낸다.

만일 독립변수 값이 주어졌을 때 잠재변수가 정규분포를 따른다는 근거가 있으면 로짓 모델보다는 프로빗 모델을 쓰는 것이 더 적절하다. 예를 들어, 능력이라는 변수가 정규분포를 따르고, 지원자의 능력이 어느 기준치를 넘겼을 때 정규직에 배정된다면, 정규직 입사 여부를 나타내는 이항 변수는 정규분포를 따르는 변수인 능력을 잠재변수로 가지고 있는 것이고, 이 경우 프로빗 모델을 사용하는 것이 2절의 논의에 비추어보았을 때 적절할 것이다.

CLL 혹은 LL 모델과 로짓, 프로빗 모델의 가장 큰 차이점은 2절에서 언급한 link function의 대칭성이다. 만약 기대확률 $\mathbb{P}(\mathbf{y} = 1 | \tilde{\mathbf{x}})$ 이 대부분의 $\tilde{\mathbf{x}}$ 값에 대해 1 근방이라면 CLL 모델이, 0 근방이라면 LL 모델이 적절할 것이다. 구체적으로, 모집단에서 비정규직이나 대기업에 종사하는 인원이 더 많으면 CLL 모델이, 정규직이나 중소기업에 종사하는 인원이 더 많으면 LL 모델이 더 적절할 것이다. 때문에 데이터가 불균형적일 때, 즉 관측된 표본에서 종속변수 값의 0과 1 비율이 불균형적일

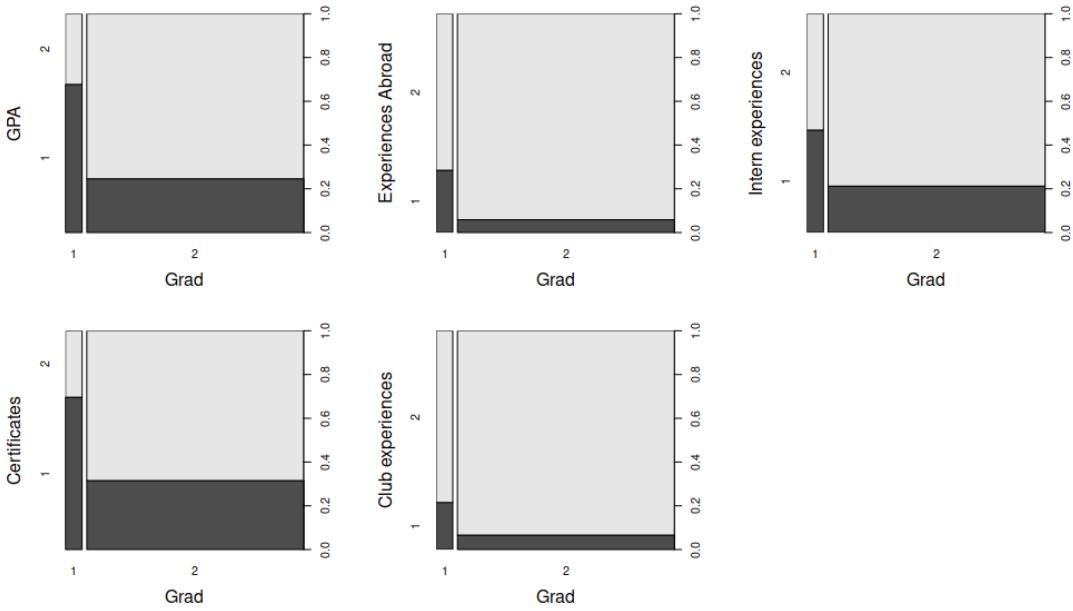


그림 1: 석사 이상의 학위(Grad)와 다른 변수들의 관계. 1은 해당 특성을 가지고 있음을, 2는 그 반대를 의미. 네 면적은 각각의 빈도를 뜻함. 순서대로 학점관리 여부, 해외경험 여부, 인턴 및 아르바이트 경험 여부, 업무관련 자격증 취득 여부, 동아리 활동 여부.

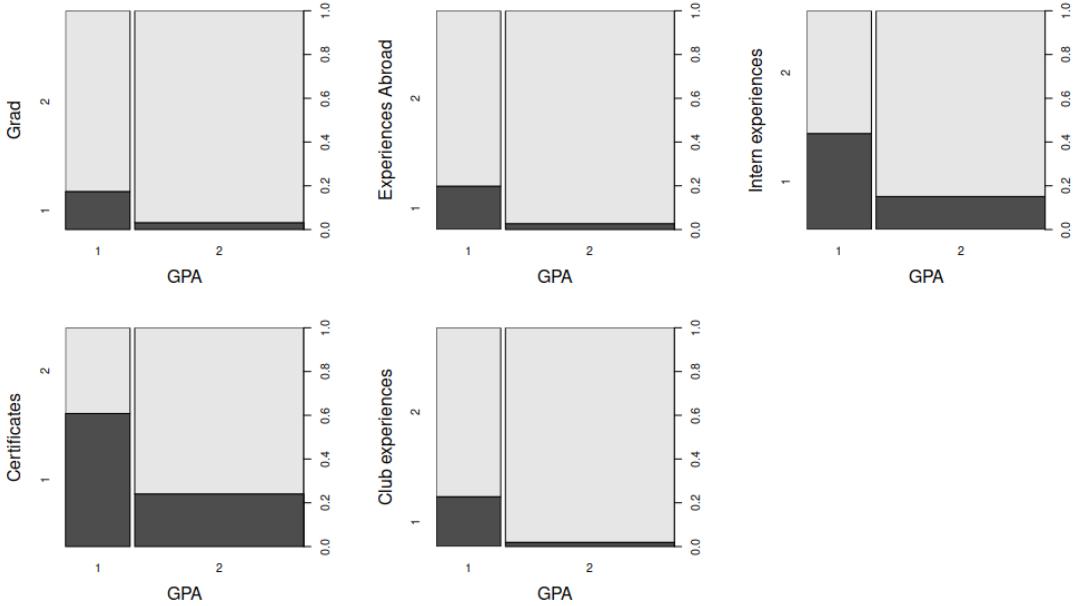


그림 2: 학점(GPA)과 다른 변수들의 관계. 1은 해당 특성을 가지고 있음을, 2는 그 반대를 의미. 네 면적은 각각의 빈도를 뜻함. 순서대로 석사 이상의 학위 여부, 해외경험 여부, 인턴 및 아르바이트 경험 여부, 업무관련 자격증 취득 여부, 동아리 활동 여부.

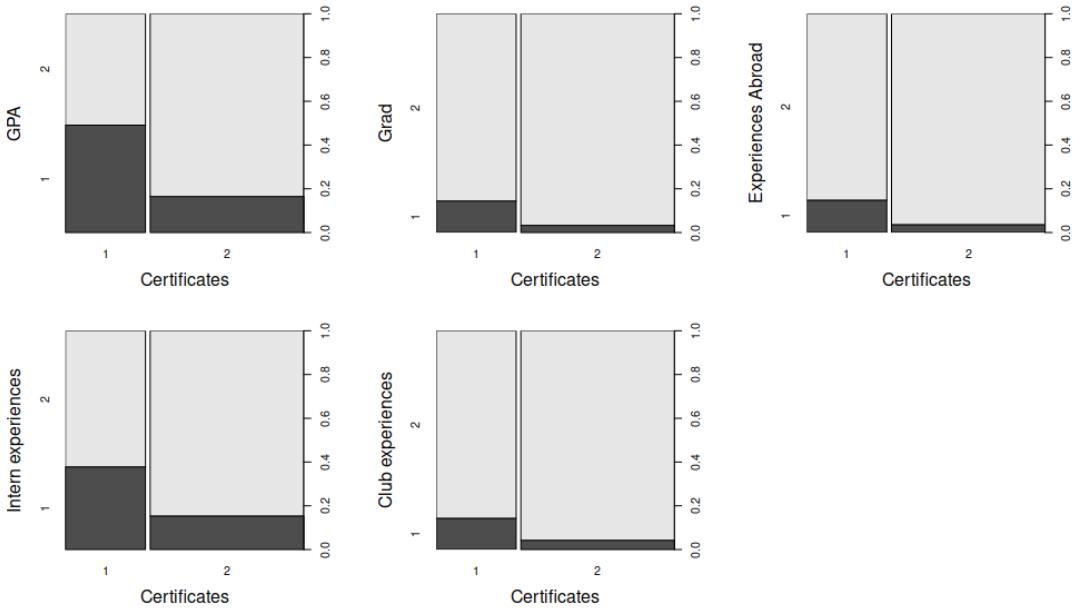


그림 3: 자격증과 다른 변수들의 관계. 1은 해당 특성을 가지고 있음을, 2는 그 반대를 의미. 네 면적은 각각의 빈도를 뜻함. 순서대로 학점관리 여부, 석사 이상의 학위 여부, 해외경험 여부, 인턴 및 아르바이트 경험 여부, 동아리 활동 여부.

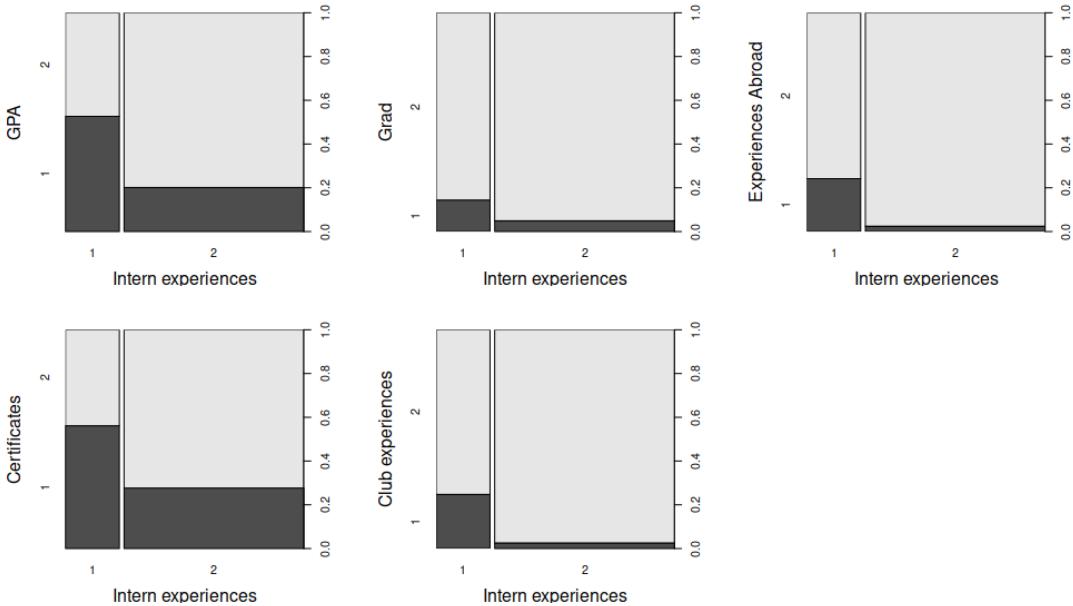


그림 4: 인턴 활동 여부와 다른 변수들의 관계. 1은 해당 특성을 가지고 있음을, 2는 그 반대를 의미. 네 면적은 각각의 빈도를 뜻함. 순서대로 학점관리 여부, 석사 이상의 학위 여부, 해외경험 여부, 업무관련 자격증 취득 여부, 동아리 활동 여부.

Binary Response Model	정·비정규직 자료	중소·대기업 자료
Logit	1902.11	2290.81
Probit	1902.01	2292.03
Complementary Log-log	1902.3	2295.01
Log-log	1901.97	2290.65

표 3: 데이터, Index 모델 별 AIC

ΔAIC	정·비정규직 자료	중소·대기업 자료
Logit	0.14	0.16
Probit	0.04	1.38
Complementary Log-log	0.33	4.36
Log-log	0	0

표 4: 데이터, Index 모델 별 ΔAIC

때 LL 혹은 CLL 모델과 같은 비대칭적 Index 모델의 적합도(fit)가 높게 나온다.

(Chen & Tsurumi, 2010)에 의하면, 데이터가 불균형적일수록 AIC가 이항변수모델³ 선택의 지표로써 더 정확하다. 정규직-비정규직 데이터에서는 1의 비율이 (비정규직에 종사한다고 답변한 인원의 비율이) 전체 관측치 대비 대략 16.6%, 그리고 대기업-중소기업 데이터에서는 1의 비율이 (대기업에 종사한다고 답변한 인원의 비율이) 전체 관측치 대비 대략 24.4%로, 두 데이터 모두 어느정도 불균형성을 띤다고 볼 수 있다.

(Burnham & Anderson, 2004)를 바탕으로, ΔAIC⁴를 사용해 모델을 평가한다. 각 모델의 AIC은 표 3와 같이, Δ AIC는 표 4와 같이 계산된다. 정규직-비정규직 데이터가 주어졌을 때 네 모델의 ΔAIC는 모두 2 이하이기 때문에 이 경우 모델 선택에 있어서 AIC는 유의미한 기여를 하지 않는다. 대기업-중소기업 데이터가 주어졌을 때 CLL 모델의 ΔAIC만 2보다 크다. 결론적으로 AIC를 바탕으로 로짓, 프로빗, 그리고 LL 모델 간 유의미한 차이가 있다고 결론을 내기는 어렵다.

한 가지 눈여겨볼만한 사항이라면, 정규직-비정규직 데이터와 대기업-중소기업 데이터가 주어졌을 때 모두 LL 모델의 AIC가 가장 작고, CLL 모델의 AIC가 가장 크다는 점이다. 이는 LL 모델이 (미세하게나마) CLL 모델보다 적합도가 좋다는 것을 의미하는데, 이는 앞서 언급했던 것처럼 모집단에서의 불균형성을 보여준다고 해석할 수도 있다.

4.2 정규직-비정규직 데이터 분석

소절 4.1에서는 AIC 비교를 통해서 로짓, 프로빗, LL모델에 유의미한 차이가 없다고 결론지을 수 있었다. 그렇기에 비교적 해석이 용이한 로짓 모델을 본 연구의 Index 모델로 사용한다. 로짓 모델에서의 변수선택은 3절에서의 LPM 모델처럼 AIC를 최소화하는 stepwise regression을 통해서

³(Chen & Tsurumi, 2010)에서는 로짓과 프로빗 모델만을 비교했지만, 본 연구에서는 CLL과 LL 모델 또한 비교 대상에 포함한다.

⁴네 가지 모델 중에서 가장 작은 AIC를 가지는 모델의 AIC를 AIC_{min} 이라고 하면, AIC 값이 AIC_m 인 모델 m의 ΔAIC는 $\Delta_m = AIC_i - AIC_{min}$ 으로 주어진다.

이루어졌다. Stepwise regression의 결과, 정규직-비정규직 데이터가 주어졌을 때는 ‘학점 관리 여부’, ‘자격증 취득 여부’, ‘어학연수 등 해외경험 여부’, ‘인턴, 아르바이트 등 직무경험 여부’, ‘석사 및 박사 학위 획득 여부’, ‘동아리 활동 여부’, ‘졸업한 대학 전공’, ‘만 14세 무렵 성장지’, ‘학력’ 9가지 변수가 독립변수로 선택되었다.

정규직-비정규직 데이터가 주어졌을 때 변수의 기대확률 추정량에 미치는 평균부분효과 (Average Partial Effect, APE)는 표 5에 나타나있다. 기준이 되는 개인은 학점관리를 하지 않았고, 자격증을 소지하지 않고, 해외경험이 없고, 인턴경험이 없고, 대학원 진학을 하지 않았고, 동아리 활동을 하지 않았고, 대학에서 인문계열을 전공했고, 만 14세 무렵 서울에서 거주한 개인이며, 변수 i 의 APE는 기대확률 추정량의 차 $APE_i = \hat{\pi}_{\text{standard}} - \hat{\pi}_i$ 로 계산하였다⁵ 이 정의 하에서 각 변수의 APE는 그 변수를 변화시켰을 때 정규직에 입사할 기대확률 추정량의 변화로 해석할 수 있다.

표 5를 참고하면, 기준이 되었던 사람과 다른 특성은 동일하면서 학점 관리를 했을 때는 정규직이 될 확률이 약 7.6%p만큼 증가함을 볼 수 있다. 마찬가지로, 사회·상경 계열을 전공하면 정규직일 확률이 8%p 가량 증가하는 것으로 추정된다. 이와 같이 표 5를 해석하면, 로짓 모델 하에서 스펙으로는 학점 관리가 가장 중요하고, 해외경험, 자격증 취득, 동아리 활동은 모두 정규직이 될 확률을 높이는 것으로 나타났다. 인문계열보다 사회·상경 계열, 자연과학계열, 공학계열과 의학계열이 정규직에 진출하는 데에 있어서 유리하고, 고졸의 경우 4년제 졸보다 정규직에 속할 확률이 2.3%p 가량 낮다. 부산과 전북을 제외한 지역에서 만 14세 무렵 성장했다면 서울에서 성장한 것에 비해 정규직이 될 확률이 높아지고, 스펙이 하나도 없다면 대졸과 석사 및 박사 학위 취득자가 정규직이 될 가능성성이 가장 높은 것으로 나온다.

4.3 대기업-중소기업 데이터 분석

대기업-중소기업 데이터가 주어졌을 때도 변수 선택은 stepwise regression을 통해 진행하였고, 이때 ‘성별’, ‘학점 관리 여부’, ‘공인영어성적 여부’, ‘영어회화 능력 개발 여부’, ‘공모전 등 수상경력 소지 여부’, ‘봉사활동 경험 여부’, ‘졸업한 대학 형태’, ‘졸업한 대학 전공’, ‘만 14세 무렵 주거형태’ 9가지 변수가 독립변수로 선택되었다.

대기업-중소기업 데이터가 주어졌을 때의 APE는 표 6에 나타나있다. 이때 변수 i 의 APE는 $APE_i = \hat{\pi}_i - \hat{\pi}_{\text{standard}}$ 로 계산했으며, 기준이 되는 개인은 남성이고, 학점 관리를 하지 않았고, 영어 점수를 가지고 있지 않고, 영어 회화 실력 개발을 위해 노력하지 않았고, 공모전 수상경력이 없고, 봉사활동 경험이 없고, 4년제 대학에서 인문계열을 전공했고, 만 14세 무렵 자가에서 거주한 개인이다. 이 APE 정의 하에서 각 변수의 APE는 그 변수를 변화시켰을 때 대기업에 입사할 기대확률 추정량의 변화로 해석할 수 있다.

표 6의 결과를 소절 4.2에서처럼 해석하면, 스펙으로는 영어공인영어학점수, 학점, 봉사활동, 영어 회화가 중요하고, 여성의 경우 남성보다 대기업에 갈 확률이 8%p 낮다. 4년제 졸보다 대기업에 갈 확률이 높은 대학 형태는 교대밖에 없으며, 사회계열이 거의 유일하게 인문계열보다 대기업에 갈

⁵ $\hat{\pi}_{\text{standard}}$ 는 기준이 되는 개인의 기대확률 추정량, 그리고 $\hat{\pi}_i$ 는 다른 변수들을 모두 기준으로 고정시켜놓은 채 변수 i 만 변화시킬 때의 기대확률 추정량이다.

확률이 높은 것으로 나타난다.⁶ 고졸의 경우 4년제 졸보다 대기업에 갈 확률이 4%p 가량 낮다.

제 5 절 스펙이 기대소득에 미치는 영향

본 절에서는 각 스펙에 따른 정규직 또는 대기업 취직 확률의 변화에 따른 기대소득의 증감을 살펴본다. 개인이 취업시장에 진입했을 때 직업을 얻을 확률이 100%라고 가정하고, 취직을 하지 않았을 때 주어지는 소득이 0이라 가정하자. 이런 가정 아래에서 모든 직업이 정규·비정규직으로 구분된다면 직업을 가졌을 때 각 개인의 기대소득 \bar{w} 는

$$\begin{aligned}\bar{w} &= p_{비} w_{비} + p_{정} w_{정} \\ p_{비} &: \text{비정규직 취직 확률}, \quad w_{비} : \text{비정규직일 때 얻는 임금}, \\ p_{정} &: \text{정규직 취직 확률}, \quad w_{정} : \text{정규직일 때 얻는 임금}\end{aligned}\tag{2}$$

처럼 쓸 수 있다. 이와 유사하게 모든 직업을 대·중소기업으로 구분한다면 기대소득은

$$\begin{aligned}\bar{w} &= p_{소} w_{소} + p_{대} w_{대} \\ p_{소} &: \text{중소기업 취직 확률}, \quad w_{소} : \text{중소기업에서 얻는 임금}, \\ p_{대} &: \text{대기업 취직 확률}, \quad w_{대} : \text{대기업에서 얻는 임금}\end{aligned}\tag{3}$$

와 같다. 주목할 만한 점은 공인어학성적 취득과 같은 스펙 x 의 추가가 두 식 속의 확률 p 와 임금 w 들을 모두 바꾼다는 것이다. 전미분을 활용해 개인의 특성 x 에 따른 식 (2)의 변량을 살펴보면

$$\Delta\bar{w} = w_{비} \frac{\Delta p_{비}}{\Delta x} \Delta x + w_{정} \frac{\Delta p_{정}}{\Delta x} \Delta x + p_{비} \frac{\Delta w_{비}}{\Delta x} \Delta x + p_{정} \frac{\Delta w_{정}}{\Delta x} \Delta x\tag{4}$$

인데, 변수 x 의 값을 스펙 x 를 취득했을 때 1, 그리고 취득하지 않았을 때 0으로 둔다면 스펙의 획득은 $\Delta x = 1$ 으로 나타난다. 이러한 분석은 식 (3)에 대해서도 동일하게 이루어진다.

우리는 각 스펙이 주어진 기초특성을 가진 개인들의 기대소득에 어떤 영향을 주는지 계산할 수 있다. 예를 들어 아무런 스펙이 없는 서울 거주 인문계열 4년제 대졸자의 대기업 취직 확률은 절 3에서 보았듯 약 48%인데, 이 사람이 영어 공인어학자격을 취득한다면 중소·대 기업 취직 확률의 변화는 $\Delta p_{소}/\Delta x_{영어} = -0.118$, 그리고 $\Delta p_{대}/\Delta x_{영어} = 0.118$ 로 나타난다.⁷ 한편 앞선 논의에서 사용했던 것과 같은 자료에 따르면 영어 공인어학자격 취득에 따른 응답자의 월급은 평균적으로 아래와 같았다.

대기업 취득시 323만원, 미취득시 282만원

중소기업 취득시 271만원, 미취득시 236만원

⁶표 6에 따르면 군사 계열의 APE가 0.4558로 대기업 취직에 있어서 가장 유리하나, 이는 본 연구에서 대기업의 기준을 피고용자 1000명 이상인 기업으로 설정했기 때문에 나타난 결과로 보여진다.

⁷정의에 따라 모든 스펙변수 x 에 대해 항등식 $\frac{\Delta p_{소}}{\Delta x} + \frac{\Delta p_{대}}{\Delta x} := 0$ 이 성립한다.

	APE
이진 독립변수	
학점관리 여부	0.0758
업무관련 자격증 소지 여부	0.0426
해외경험 여부	0.0303
인턴 및 아르바이트 경험 여부	-0.05
석사 이상의 학위 여부	-0.0541
동아리 활동 여부	0.0542
대학 전공에 관한 독립변수	
사회·상경계열	0.0798
자연과학계열	0.0508
공학계열	0.0872
의약학계열	0.0941
교육계열	0.0012
예체능계열	0.0011
사관학교 및 경찰대	0.0443
기타	0.0483
모름	0.0667
대학졸업 안함	-0.0228
만14세 성장지에 관한 독립변수	
부산	-0.0483
대구	0.0593
인천	0.0306
광주	0.0208
대전	0.0686
울산	0.0571
경기	0.029
강원	0.0046
충북	0.0423
충남	0.0718
전북	-0.0004
전남	0.0067
경북	0.0629
경남	0.0294
제주	0.0301
학력에 관한 독립변수	
고졸 미만	-0.041
고졸	-0.0715
전문대졸	-0.069
석사학위 이상	-0.0089

표 5: 정규직-비정규직 데이터의 로짓 모델 APE

	APE
이진 독립변수	
성별	-0.0814
학점관리 여부	0.0647
공인영어자격 여부	0.0785
영어회화능력 여부	0.0397
공모전 수상경력 여부	-0.0494
봉사활동 경험 여부	0.0609
대학 종류에 관한 독립변수	
2-3년제	-0.119
산업대학	-0.1597
교육대학	0.5028
폴리텍	-0.1554
대학졸업 안함	-0.0414
대학 전공에 관한 독립변수	
사회·상경계열	0.0438
자연과학계열	-0.1163
공학계열	-0.0357
의약학계열	-0.0588
교육계열	-0.002
예체능계열	-0.1047
사관학교 및 경찰대	0.4558
모름	-0.109
대학졸업 안함	-0.0414
만14세 주거형태에 관한 독립변수	
전세	-0.0037
월세	-0.1075
영구임대 및 국민임대주택	0.47

표 6: 대기업-중소기업 데이터의 로짓 모델 APE

스펙	$w_{정}$	$w_{비}$	$\Delta w_{정}/\Delta x$	$\Delta w_{비}/\Delta x$	$\Delta p_{정}/\Delta x$	$\Delta p_{비}/\Delta x$	$p_{정}$	$p_{비}$	$\Delta \bar{w}$
학점관리	257.6	199.7	30.6	-6	0.07	-0.07	0.8	0.2	27.333
대학원	262.7	192.8	61.4	78.4	-0.11	0.11	0.8	0.2	57.111
해외경험	265	195.5	23.7	71.1	0.06	-0.06	0.8	0.2	16.648
인턴	263.3	198.2	15.2	2.4	-0.05	0.05	0.8	0.2	9.385
자격증	265.2	192.9	4.5	22.8	0.03	-0.03	0.8	0.2	10.329
동아리	265.9	199.2	11.1	-11.6	0.05	-0.05	0.8	0.2	9.895

표 7: 정규직-비정규직 모델에 따른 아무런 스펙이 없는 서울 출신 인문계열 4년제 대졸 개인의 스펙에 따른 기대월급 변화량($\Delta \bar{w}$). 단위는 만원.

스펙	$w_{대}$	$w_{소}$	$\Delta w_{대}/\Delta x$	$\Delta w_{소}/\Delta x$	$\Delta p_{대}/\Delta x$	$\Delta p_{소}/\Delta x$	$p_{대}$	$p_{소}$	$\Delta \bar{w}$
영어공부	281.7	236.4	41.2	35	0.12	-0.118	0.48	0.52	43.412
학점관리	284.4	237.5	31.4	20.7	0.06	-0.06	0.48	0.52	28.65
봉사활동	297.4	242.7	2	-5.4	0.07	-0.73	0.48	0.52	1.981
공모전	294.7	241.2	54	20.8	-0.08	0.08	0.48	0.52	32.456
영어회화	283.8	237.7	43	32.5	0.05	-0.053	0.48	0.52	39.845

표 8: 대기업-중소기업 모델에 따른 아무런 스펙이 없는 인문계열 4년제 대졸 개인의 스펙에 따른 기대월급 변화량($\Delta \bar{w}$). 단위는 만원.

따라서 우리는 영어 공인어학자격 미취득자의 월급을 $w_{소} = 236$, $w_{대} = 282$, 그리고 자격 취득에 따른 월급의 변화를 $\Delta w_{소}/\Delta x_{영어} = 271 - 236 = 35$, $\Delta w_{대}/\Delta x_{영어} = 323 - 282 = 41$ 으로 어림할 수 있다.⁸ 이때 무스펙 서울 거주 인문계열 4년제 대졸자가 영어 공인어학자격을 취득할 때 기대소득의 변화량은

$$\Delta \bar{w} = (236 \times -0.118) + (282 \times 0.118) + (0.53 \times 35) + (0.47 \times 41) \approx 43$$

으로, 약 43만원의 기대소득 증가가 생김을 알 수 있다.

위에서 소개한 방법을 활용해 절 3에서 유의하다고 판단된 변수들이 기대소득을 어떻게 변화시키는지 계산한 결과가 표 7과 8에 있다. 이는 앞에서의 예시와 마찬가지로 스펙이 없는 서울 거주 인문계열 4년제 대졸자가 주어진 스펙 중 하나를 취득할 때 기대소득의 변화를 나타낸 것이다. 여기서 소득은 월급(만원)을 뜻한다. 정규직-비정규직 모델에서 학점 관리, 석사 이상의 학위, 해외 경험, 그리고 업무 관련 자격증 취득은 모두 기대월급을 각각 약 27만원, 57만원, 17만원 그리고 10만원 가량 증가시킨다. 그러나 인턴 및 아르바이트 경험과 동아리 활동은 기대소득에 큰 영향을 주지 못하는 것을 볼 수 있다. 또한, 대기업-중소기업 모델에서 영어 공인어학성적 취득, 학점 관리, 공모전 수상, 그리고 영어회화 공부는 모두 기대월급을 증가시키는 것이 확인되었다. 그러나 봉사 활동 경험은 기대월급에 큰 영향을 주지 않았다. 기대소득(월급)을 크게 증가시키는 변수는 석사 이상의 학위(57만원), 영어 공인어학자격(43만원), 영어회화공부(40만원) 등이 있다. 높은 학점의 경우 정규직-비정규직 모델과 대기업-중소기업 모델 모두에서 약 30만원 정도 임금을 상승시키는 것으로 나타났다.

⁸이렇게 계산한 월급과 월급의 변화량은 주어진 기초특성을 가진 개인들에 한정해 계산한 것이 아니라 자료 속 모든 개인의 월급을 단순평균한 값임에 유의해야 한다. 이는 각 개인이 가질 수 있는 기초특성의 종류가 너무나 다양해 모든 경우에 따른 임금변수를 자료의 부족으로 계산할 수 없었기 때문이다.

한편 식 (4)를 변형하면 임금 격차가 기대소득 증가폭에 미치는 영향을 알 수 있다. 항등식 $\frac{\Delta p_{\text{소}}}{\Delta x} + \frac{\Delta p_{\text{대}}}{\Delta x} := 0$ 을 활용해 위 식을 정리하면

$$\Delta \bar{w} = (w_{\text{정}} - w_{\text{비}}) \times \frac{\Delta p_{\text{정}}}{\Delta x} + \left(p_{\text{비}} \frac{\Delta w_{\text{비}}}{\Delta x} \Delta x + p_{\text{정}} \frac{\Delta w_{\text{정}}}{\Delta x} \Delta x \right) \quad (5)$$

을 얻는데, 이때 $\Delta p_{\text{정}}/\Delta x$ 가 많은 스펙에 대하여 양수이다.⁹ 따라서 스펙이 추가될 때 기대임금의 변화량 $\Delta \bar{w}$ 를 정규직과 비정규직 사이의 임금격차 $w_{\text{정}} - w_{\text{비}}$ 의 함수로 본다면 $\Delta \bar{w}$ 는 임금격차와 양의 관계를 가지고 있음을 알 수 있다. 즉, $w_{\text{정}} > w_{\text{비}}$ 를 가정했을 때 정규·비정규직 또는 대·중소기업 사이의 임금 차이가 커지면 커질수록 스펙을 추가했을 때 기대소득이 커지는 것이다. 여기에는 두 가지 이유가 있는데, 하나는 정규직의 임금 $w_{\text{정}}$ 이 비정규직의 임금 $w_{\text{비}}$ 보다 크기 때문이고, 나머지 하나는 스펙 x 를 추가했을 때 정규직이 될 확률의 변화량 $\Delta p_{\text{정}}/\Delta x$ 가 양수이기 때문이다.

제 6 절 결론 및 시사점

비정규직-정규직 및 중소기업-대기업 사이의 임금격차가 고착화되는 가운데 좋은 직장을 위해 더 많은 “스펙”을 모으려고 노력하는 풍조가 심해지고 있다. 본고에서는 제12차 청년패널조사(2018)을 사용해 현재 청년들이 주로 취득하는 스펙 중 대기업과 정규직 취직에 유의미한 도움을 주는 것이 무엇인지 알아보았다. 첫번째로 선형확률모델과 로지스틱 회귀모델을 사용해 좋은 직장을 가질 확률과 여러 스펙 변수들 사이의 관계를 살펴보았다. 그리고 선형확률모델에서의 측정치를 사용해 스펙이 기대소득에 미치는 영향을 규명한 뒤 고용 형태 및 기업 규모 사이의 임금격차와 스펙을 취득했을 때 기대소득 증가폭과의 관계를 대수적인 방법으로 제시하였다.

Stepwise regression을 통해 표 1에 제시된 여러 스펙 중 정규직 또는 대기업 취직의 기대확률과 관련이 있는 변수로 학점, 공인영어성적, 어학연수 등 해외경험, 업무관련 자격증, 봉사활동 경험, 인턴 및 아르바이트 등 직무경험 여부, 석사 이상의 학위, 영어회화 공부, 그리고 공모전 등 수상경력 소지 여부가 선택되었다. 선형확률모델을 사용해 각 변수의 효과를 OLS로 추정한 결과, 다른 스펙과 출입한 대학의 형태 및 전공, 그리고 14세 성장지가 동일한 개인에 대해 학점을 관리한 개인은 정규직 취직 기대확률이 7%p 높고, 해외경험이 있을 시 6%p, 그리고 업무관련 자격증이 있을 시 3%p 높았다. 인턴 및 아르바이트 경험과 석사 이상의 학위는 정규직 취직 확률을 각각 5%p 와 11%p 만큼 낮췄다. 대기업 취직 확률은 다른 스펙과 대학 형태 및 전공을 고정시켰을 때 공인영어성적이 있을 시 128%p, 학점을 관리했을 때 6%p, 봉사활동 경험이 있을 시 7%p, 그리고 영어회화 능력을 개발했을 시 5%p가 높았으며, 공모전 수상경력이 있을 시 8%p가 낮았다. 한편 이러한 결과는 LPM 대신 다양한 이항종속변수 모델을 사용해도 바뀌지 않았다.

위에서 언급한 스펙들을 취득했을 시 기대되는 월급의 금액은 모두 상승하였다. 월급이 가장 크게 상승하는 스펙은 석사 이상의 학위(약 57만원)이고, 그 뒤를 공인영어성적(약 43만원)과 영어회화공부(약 40만원)가 이었다. 월급 상승이 가져다주는 생애소득의 증가에 비해 스펙 취득에 드는 비용은

⁹예외적인 변수들로는 석사 이상의 학위와 인턴 및 아르바이트 경험이 있다. 이들은 정규직이 될 확률을 낮추는 경향을 보였다. 또한 공모전 수상 경력은 대기업에 취직할 확률과 음의 관계를 보였다.

일시적이고 작기 때문에 취업준비생에게는 앞서 언급한 스펙 취득에 시간과 비용을 투자하는 것이 합리적이다. 한편 스펙을 취득했을 시 기대소득의 증가폭은 고용 형태 및 기업 규모 사이의 임금격차와 양의 관계가 있었다. 즉, 대기업-중소기업 및 비정규직-정규직 사이의 임금격차가 커질수록 대기업 및 정규직 취직에 긍정적인 영향을 주는 변수를 취득했을 시의 기대소득 증가폭이 커진다.

본고는 현재의 낮은 청년고용률에 대해 한가지 시사점을 제시한다. 지난 5년간 청년고용률은 약 26% 수준으로, 30% 수준을 윗돌았던 2000년도 초반의 수준을 회복하지 못하고 있다. 본 연구는 취업준비생들이 기대소득을 증가시키기 위해 노동시장에 곧바로 진입하기 보다 스펙쌓기를 택한 현상이 여기에 영향을 주었을 개연성을 제시한다. 또한, 정규직-비정규직 및 중소기업-대기업 사이의 임금 격차가 줄어든다면 스펙획득 시 기대소득의 증가폭이 감소하기 때문에 취업준비생이 스펙을 쌓기 위해 시간을 투자할 유인을 줄일 것이라는 사실도 눈여겨 볼만 하다.

참고 문헌

- [1] 한국고용정보원. (2017). *유저가이드: YP2007 1~12차 조사자료*.
- [2] 최영섭. (2003). 대학 이상 졸업자의 계열별 기대소득 격차에 대한 분석, 노동경제논집, 제26호 (2).
- [3] Burnham, K. P., & Anderson, D. R. (2004). *Multimodel Inference: Understanding AIC and BIC in Model Selection*. Sociological Methods & Research, 33(2), 261–304.
- [4] Guo Chen & Hiroki Tsurumi. (2010). *Probit and Logit Model Selection*. Communications in Statistics—Theory and Methods, 40:1, 159-175.
- [5] Konis, K. (2007). *Linear programming algorithms for detecting separated data in binary logistic regression models*. [PhD thesis]
- [6] R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- [7] Rossum, G. V., & Drake, F. L. (2009). *Python 3 Reference Manual*. United States: SohoBooks.
- [8] Wooldridge, J.M. (2011). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- [9] Wooldridge, J.M. (2020). *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning.

저작권

본고의 조판에는 LATEX와 한국출판인회의에서 배포하는 공개글꼴인 KoPubWorld 서체가 사용되었다.

영문 초록

This paper seeks to elucidate the relationship between job applicants' expected income, various measures of career skills, including but not limited to one's GPA, experience studying abroad, academic qualifications (in the form of Master's degrees and Doctorates) and internship opportunities, as well as more personal characteristics such as gender and place of origin (defined as one's location at the age of 14). In pursuit of this goal, the paper utilizes data from the Youth Panel (YP) and adopts a stepwise approach. Initially, the probabilities of finding employment in a permanent position or a sizable firm are modelled with the linear probability model (LPM) and the logistic regression model, whose parameters are estimated afterwards. Then, the expected income of entrants to the job market is computed using the probabilities thus obtained as weights on the average incomes of employees, which are categorized according to the permanence of their vocation and the size of their employer. The results suggest that, all else being equal, applicants with a higher GPA, experiences studying abroad, and relevant certifications are more likely to land a permanent position than their peers, while applicants with an exemplary GPA, a record of volunteer work, and experience taking standardized English tests and honing their English conversation skills are more likely to be accepted into a conglomerate than others. By extension, applicants with these qualifications face higher expected incomes. This finding lends credence to the argument that prospective employees may be incentivized to postpone entry into the job market in order to obtain the above qualifications, which could confer on them an advantage over their competitors. Issues pertaining to model selection and multicollinearity are also addressed.