

표본선택 편의를 반영한 임금결정요인 분석

박성익¹ · 조장식²

¹경성대학교 국제무역통상학과 · ²경성대학교 수학응용통계학부

접수 2016년 8월 12일, 수정 2016년 9월 13일, 게재확정 2016년 9월 23일

요약

본 연구에서는 한국고용정보원에서 실시한 「2013 고졸자 취업진로조사」 자료를 활용하여 특성화고 졸업자의 임금결정요인을 분석하였다. 일반적으로 임금은 개인의 취업여부와 임금의 크기에 대한 두 가지의 복합적인 정보를 담고 있다. 그러나 임금 결정요인분석의 많은 선행연구에서는 후자의 정보만을 대상으로 최소제곱법에 기초한 선형 회귀분석을 수행함으로써 표본선택에 의한 편이 (sample selection bias) 문제가 발생하게 된다. 본 연구에서는 임금결정요인분석에서 표본선택에 의한 편이의 문제를 극복하기 위해 Tobit 모형과 Heckman의 표본선택 모형을 분석에 활용하였다. 주요 분석 결과를 요약하면 다음과 같다. 먼저 Tobit 모형과 Heckman의 표본선택 모형에 대한 타당성은 통계적으로 유의함을 알 수 있었다. 성별은 취업확률과 임금의 크기에서 모두 통계적으로 유의한 것으로 나타났다. 마이스터고 졸업생은 취업확률과 임금의 크기 모두 기타고 졸업생에 비해서 높은 것을 알 수 있었으며, 부모소득이 높을수록 취업확률과 임금의 크기가 모두 통계적으로 유의하게 증가하였다. 부모학력이 고졸이하에 비해서 대졸이상이 취업확률은 통계적으로 유의하게 낮지만, 임금의 크기는 높게 나타났다. 고교성적은 높을수록, 고교 만족도가 높을수록, 그리고 자격증 수가 많을수록 취업확률과 임금의 크기 모두 통계적으로 유의하게 높은 것을 알 수 있다.

주요용어: 중단자료, 표본선택 편이, Heckman 표본선택 모형, Tobit 모형.

1. 서론

대졸자 과잉공급으로 인한 청년층 실업문제와 관련하여 정부는 대학 구조조정과 더불어 고졸채용 확대, 특성화고 및 마이스터고 육성 등 고졸자 취업 장려 및 유인 정책을 통하여 청년 고용문제를 완화하고자 노력하고 있다. 특히 정부는 최근 학력중심사회에서 능력중심사회로의 전환을 핵심적 정책과제로 선정하면서, 특성화고 학생들에 대한 지원을 강화하고 있다.

특성화고 졸업생들에 대한 노동시장 성과 연구와 관련하여, Lee와 Jung (2010)은 특성화 고등학교 학생들이 인식하는 취업장벽 수준을 측정하고 그와 관련된 변인과 그들의 설명력을 규명하는 연구를 했다. Jang과 Lee (2012)는 한국교육고용패널 자료를 활용하여 특성화고 학생의 취업선택을 결정하는 개인, 가정 및 학교 변인을 구명하고, 각 변인의 영향을 예측하였다. 그리고 Lee와 Joo (2013)는 한국 교육고용패널 조사 자료를 활용하여 특성화고 졸업생 취업에 대한 학교 효과분석을 하였으며, Cho와 Jung (2014)은 특성화고 졸업예정자의 취업결정요인분석을 하였다. Kang (2013)은 2011년도 ‘고졸자 취업진로조사’ 자료를 활용하여 특성화고 졸업자의 취업결정요인 및 노동시장 정착가능성에 대한 연구를 하였다. 이외에도 다수의 특성화고 졸업생들의 노동시장성과에 관한 연구들이 있으나, 주된 연구대

¹ (48434) 부산광역시 남구 수영로 309 번지, 경성대학교 국제무역통상학과, 교수.

² 교신저자: (48434) 부산광역시 남구 수영로 309 번지, 경성대학교 수학응용통계학부, 교수.

E-mail: jscho@ks.ac.kr

상은 취업실태 및 취업결정요인분석 중심이었으며 임금결정요인에 관한 연구는 별로 보고된 바가 없다. 물론 임금실태에 관한 기본통계분석 정도는 발견되고 있지만, 임金的 결정요인에 초점을 맞춘 통계적 분석은 부족한 것이 현실이다. 최근에 Ryu와 Cho (2016a)는 특성화고 졸업생의 임금 결정요인을 분석하기 위해 최소제곱법에 기초한 선형회귀모형과 의사결정나무분석을 활용하였다. 또한 Ryu와 Cho (2016b)는 혼합모형을 활용하여 특성화고 졸업생의 임금결정요인을 개인수준과 취업지역수준으로 분리하여 고정효과와 랜덤효과와의 관점에서 분석하였다.

한편, 일반적으로 취업을 해서 임금을 받는 졸업자와 미취업으로 인해 임금이 없는 졸업자들은 인구통계학적 특성이 다르다고 예상할 수 있다. 즉 임금 결정요인을 분석할 때, 임금은 개인의 취업여부와 임金的 크기에 대한 두 가지의 복합적인 정보를 담고 있다. 특히 전자의 정보는 개인의 선택 (또는 역량)에 달려 있다는 점에서 자기선택성 (self-selectivity)을 내포하고 있다. 이러한 자기선택성에 대한 고려 없이 후자의 정보만을 대상으로 최소제곱법 (ordinary least squares; OLS)에 기초한 선형 회귀분석을 하게 되면 전형적인 표본선택에 의한 편익의 (sample selection bias)의 문제가 발생한다. 즉 임금결정요인을 분석할 때 현재 노동시장에 참여하여 임금을 받는 졸업자만을 대상으로 분석한다면, 이들 대상자들은 무작위로 추출된 표본이 아니며, 또한 비무작위적으로 미취업자 표본들이 배제됨으로써 표본선택 편익의 문제가 발생하게 된다. 앞의 Ryu와 Cho (2016a, 2016b)의 특성화고 졸업생의 임금결정요인을 분석한 연구뿐만 아니라, Cho (2011) 및 Park과 Cho (2015)의 연구도 노동시장에 참여하여 임금을 받는 졸업자만을 대상으로 분석함으로써, 비무작위적으로 미취업자들의 표본이 배제되어 표본선택 편익의 문제를 갖고 있다는 점에서 한계점을 갖고 있다.

본 연구에서는 종속변수의 값이 갖는 정보가 취업여부와 임金的 크기에 대한 복합적인 정보를 담고 있다는 점에서, 임금에 대한 자료는 취업자뿐 만 아니라 미취업자들도 분석에 포함하였다. 즉 종속변수인 임금이 '0'부터 시작해서 양의 값을 가지는 중단분포 (censored distribution)을 따른다. 또한 표본선택에 의한 편익의 문제를 극복하기 위해 임금이 갖는 복합적인 정보를 분해하여 분석한다. 이를 위해 Tobit 모형과 Heckman의 표본선택 모형을 분석에 활용하고자 한다. Tobit 모형은 Tobin (1958)이 종속변수인 지출금액이 음이 될 수 없다는 점에 착안하여 이를 고려한 회귀모형을 설계했으며, 이후 Goldberger (1964)가 프로빗 모형과 유사성을 이유로 Tobit 모형이라 지칭하였다. Tobit 모형은 독립변수들이 취업여부에 미치는 영향과 임金的 크기에 미치는 영향이 동일하다는 가정에 근거를 두고 있으며, Amemiya (1984)는 경험적 연구를 종합하여 Tobit 모형의 5가지 유형을 분류하였다. Heckman의 표본선택 모형은 전체 표본에서 임금이 없는 미취업자의 값들이 갖는 분포에 대한 고려가 반영된 것이며, 독립변수가 취업여부와 임金的 크기에 영향을 미치는 과정이 상이하다고 가정하고 각각의 경우에 적합한 분석모형을 적용하여 2단계로 나누어 추정하는 방법이다 (Heckman, 1976).

본 논문의 구성은 다음과 같다. 제 2절에서는 자료소개와 기술통계를 하고, 제 3절에서는 Tobit 모형과 Heckman 모형을 간단히 설명한다. 그리고 제 4절에서는 Tobit 모형과 Heckman의 표본선택 모형의 실증분석 결과를 제시한다. 마지막 제 5절에서는 결론을 제시한다.

2. 자료소개 및 기술통계

특성화고 졸업자의 임금결정요인을 분석하기 위해 「2013 고졸자 취업진로조사 (2013 High school graduates occupational mobility survey)」자료를 활용한다. 종속변수인 임금이 갖는 정보가 취업여부와 임金的 크기에 대한 복합적인 정보를 담고 있다는 점과 표본선택에 의한 편익의 문제를 해결하기 위해, 취업자뿐만 아니라 미취업자의 자료도 분석에 포함시켰다. 분석에 포함된 조사대상자는 2013년 특성화고, 마이스터고, 공업·상업 계열 졸업자로서, 총 43,703명이 분석 대상이며, 이 중에서 33,765는 취업자이고 9,938명은 미취업자이다. 여기서 결측 값을 갖는 표본은 제외시켰다.

임금결정요인을 분석하기 위한 독립변수는 기존의 선행 연구 (Ryu와 Cho, 2016b)를 바탕으로 성별, 부모 학력과 부모소득, 그리고 고교 특성을 나타내는 마이스터고 여부, 공학계열 여부, 고교성적, 고교 만족도, 자격증 수, 고교 지역 등을 고려하였다.

분석에 사용된 변수와 변수에 대한 설명은 Table 2.1과 같다.

Table 2.1 Variables explanation

Variables	Explanation	Role
sex	0='female', 1='male'	independent
maester	0='the others', 1='maester'	independent
engineer	0='the others', 1='engineering'	independent
income	parent's income : 0='<100', 1='100~200', 2='200~300', 3='300~500', 4='>500'	independent
education	parent's education : 0='≤high school', 1='college', 2='≥university'	independent
gpa	graduation record : 0='middle', 1='very low', 2='low', 3='high', 4='very high'	independent
satisfy	degree of satisfaction with high school	independent
license	number of licenses : 0='none', 1='1~2', 2='≥3'	independent
region	high school region : 0='dongnam', 1='sudo', 2='chungchung', 3='junra', 4='daegyeong', 5='gangwon', 6='jeju'	independent
employment	employment or unemployment : 0='unemployment', 1='employment'	dependent
wage	wage of job	dependent

0=reference category.

Table 2.2는 취업여부와 취업자의 임금에 대한 기술통계 결과를 나타낸다.

Table 2.2 Cross table analysis for employment and descriptive statistic for wage

Variables		Employment			Wage		
		N	No (%)	Yes (%)	N	Mean	S.D.
sex	male	25,261	25.3	74.7	18,862	148.6	45.6
	female	18,442	18.2	81.8	15,092	150.0	45.4
maester	maester	2,769	10.4	89.6	2,480	175.1	42.9
	etc	40,934	23.1	76.9	31,474	147.2	45.1
engineer	engineer	24,530	24.0	76.0	18,640	149.1	44.6
	etc	19,173	20.1	79.9	15,314	149.4	46.7
income	<100	2,148	26.8	73.2	1,572	145.9	49.7
	100~200	10,591	25.4	74.6	7,904	145.4	46.3
	200~300	12,912	23.6	76.4	9,865	147.3	42.8
	300~500	15,264	19.3	80.7	12,317	151.5	45.5
	>500	2,788	17.6	82.4	2,296	161.5	48.3
gpa	very low	2,799	36.3	63.7	1,782	142.2	51.0
	low	8,047	30.3	69.7	5,608	138.6	42.4
	middle	16,071	24.6	75.4	12,117	144.3	44.1
	high	10,917	16.0	84.0	9,169	154.3	44.6
	very high	5,869	10.1	89.9	5,278	165.4	46.2
education	≤high	36,565	22.7	77.3	28,277	148.1	45.7
	college	2,177	19.3	80.7	1,756	150.0	43.3
	university	4,961	21.0	79.0	3,921	157.5	44.7
licenses	none	8,812	33.8	66.2	5,832	142.0	47.0
	1~2	23,122	23.5	76.5	17,698	146.1	44.2
	3~5	10,427	11.7	88.3	9,208	159.1	45.6
	≥6	1,342	9.4	90.6	1,216	155.7	43.1
regions	sudo	19,729	19.5	80.5	15,873	143.2	44.0
	chungchung	5,814	25.6	74.4	4,324	154.1	46.9
	junra	5,021	28.8	71.2	3,577	150.0	44.4
	daekyung	5,316	20.9	79.1	4,205	165.1	49.1
	dongnam	7,152	23.9	76.1	5,446	151.4	43.4
	gangwon	472	26.3	73.7	348	139.2	32.9
	jeju	199	9.0	91.0	181	137.4	47.6

Table 2.2의 결과에 따르면, 남자에 비해서 여자가 취업률과 임금의 크기가 모두 높음을 알 수 있다. 마이스터고 졸업자는 기타 졸업자에 비해서 취업률과 임금의 크기가 모두 높음을 알 수 있고, 공학계열은 비공학계열에 비해서 취업률과 임금의 크기가 모두 낮게 나타났다. 부모의 소득이 높을수록, 그리고 고교성적이 높을수록 취업률과 임금의 크기가 모두 높게 나타났다. 부모의 교육수준은 고졸이하에 비해

서 전문대졸과 대졸이상이 취업률과 임금의 크기 모두에서 높게 나타났으며, 또한 자격증 수가 많을수록 취업률과 임금의 크기 모두 높게 나타났다. 고교지역은 제주권, 수도권, 대경권, 동남권 순으로 취업률이 높으며, 임금의 크기는 대경권, 충청권, 동남권 순으로 크게 나타났다.

3. 연구방법

일반적으로 임금 결정요인을 분석할 때 임금은 두 가지 정보를 담고 있다. 즉 특성화고 졸업생 개인의 취업여부와 임금의 크기에 대한 정보이다. 특히 후자의 정보만을 대상으로 회귀분석을 하게 되면 전형적인 표본선택에 의한 편이 (bias)가 발생하게 된다. 따라서 본 연구에서는 종속변수의 값이 복합적인 정보를 담고 있다는 점에서 이를 분해하여 분석한다. 이를 위해 Tobit 모형과 Heckman의 표본선택 모형을 분석에 활용한다. Tobit 모형은 독립변수들이 취업여부에 미치는 영향과 임금의 크기에 미치는 영향이 동일하다는 가정에 근거하고 있다. 반면 Heckman의 표본선택 모형은 취업여부와 임금의 크기에 영향을 미치는 과정이 상이하다는 가정에 근거하고 있으며, 취업여부와 임금의 크기를 결정하는 요인을 2단계에 걸쳐 분리해서 분석한다.

먼저 Tobit 모형에 대해 간단히 설명하면 다음과 같다. Tobit 모형은 종속변수가 취하는 값의 범위에 제약이 가해져서 종속변수의 일부가 중도 절단 (censored) 되는 경우에 활용된다. 즉 종속변수인 임금은 '0'을 포함한 양수 값만 가지며, 미취업자의 임금 자료는 '0'으로 중도 절단된 자료를 의미한다. Tobit 모형은 취업여부와 임금의 크기에 미치는 독립변수의 영향이 동일한 것으로 간주하여 다음과 같이 하나의 식으로 동시에 추정하는 방법이다.

$$y_i^* = x_i' \beta + \epsilon_i, \epsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n.$$

여기서 y_i 와 y_i^* 는 각각 관측된 임금과 잠재적인 임금을 의미한다. x_i 는 임금에 영향을 미칠 수 있는 독립변수들이며, ϵ_i 는 오차항으로 평균이 0이고 분산이 σ^2 인 정규분포를 따른다. β 는 추정해야 할 모수 벡터이며, 실제로 관측되는 임금인 y_i 는 다음 식 (3.1)과 같이 주어진다.

$$y_i = \begin{cases} y_i^*, & \text{if } y_i^* > 0 \\ 0, & \text{if } y_i^* \leq 0. \end{cases} \quad (3.1)$$

식 (3.1)에서 잠재적인 종속변수 y_i^* 의 기댓값은 다음과 같다.

$$E(y_i^* | y_i^* > 0) = x_i' \beta + \sigma \frac{\phi_i}{\Phi_i}.$$

여기서 ϕ_i 는 $x_i' \beta / \sigma$ 에 대한 표준정규밀도 함수값을 나타내며, Φ_i 는 $x_i' \beta / \sigma$ 에 대한 표준정규누적함수 값을 나타낸다. 따라서 회귀계수 β 와 분산 σ^2 의 값을 최대우도함수를 극대화시키는 방법으로 구한다.

다음으로 Heckman의 표본선택 모형을 간단히 소개하면 다음과 같다. 특성화고 졸업자 i 의 취업여부에 대한 관측변수 (z_i)와 임금의 크기에 대한 관측변수 (y_i)를 다음과 같이 2 단계로 분리하여 추정하는 모형이다.

$$z_i^* = w_i' \gamma + u_i, z_i = \begin{cases} 1, & \text{if } z_i^* > 0 \\ 0, & \text{if } z_i^* \leq 0 \end{cases} \quad (3.2)$$

$$y_i^* = x_i' \beta + \epsilon_i, y_i = \begin{cases} y_i^*, & \text{if } z_i^* > 0 \\ 0, & \text{if } z_i^* \leq 0. \end{cases} \quad (3.3)$$

여기서 $\begin{bmatrix} u_i \\ \epsilon_i \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right)$ 를 따르며, β 와 γ 는 추정해야 할 모수벡터, ϵ_i 와 u_i 는 오차항, x_i 와 w_i 는 각각 임금의 크기와 취업여부에 관련된 설명변수 벡터를 의미한다. 그리고 이변량 정규분포를 따르고 있는 오차항 u_i 와 ϵ_i 의 표준편차는 1과 σ 이며, 이들의 상관계수는 ρ 이다. 또한 z_i^* 는 취업여부에 대한 잠재변수 (latent variable)이고, y_i^* 는 임금의 크기를 나타내는 잠재변수이다.

Heckman 표본선택 모형의 1단계에서는 표본에 선택될 확률을 나타내는 식 (3.2)에 대해서 프로빗 모형 (probit model)을 이용하여 추정한다. Heckman 표본선택 모형의 2단계에서는 표본으로 선택된 사례만을 대상으로 하며 식 (3.3)으로부터 다음 식 (3.4)와 식 (3.5)가 계산된다 (Greene, 2003).

$$\begin{aligned} E[y_i|z_i^* > 0] &= x_i'\beta + E[\epsilon_i|u_i > -w_i'\gamma] \\ &= x_i'\beta + \rho\sigma \frac{\phi(-w_i'\gamma)}{1 - \Phi(-w_i'\gamma)} = x_i'\beta + \beta_\lambda \lambda(-w_i'\gamma). \end{aligned} \quad (3.4)$$

$$y_i|z_i^* = E[y_i|z_i^* > 0] + \nu_i = x_i'\beta + \beta_\lambda \lambda(-w_i'\gamma) + \nu_i. \quad (3.5)$$

여기서 $\phi(\cdot)$ 는 표준정규분포의 확률밀도함수를 나타내며, $\lambda(-w_i'\gamma)$ 는 역 밀의 비율 (inverse Mill's ratio)로서 각각의 관찰이 표본으로부터 배제될 순간적 확률(위험률)을 의미한다. 식 (3.4)에서 위험률 $\lambda(-w_i'\gamma)$ 는 새로 추가된 하나의 설명변수로 작용을 하고 $\rho\sigma$ 는 회귀계수(β_λ)의 역할을 한다. 즉 Heckman의 2단계 추정방법으로 먼저 프로빗 모형을 사용하여 $\hat{\gamma}$ 를 추정하고 $w_i'\hat{\gamma}$ 를 계산해서 $\hat{\lambda}(-w_i'\hat{\gamma})$ 를 구한다. 그리고 식 (3.5)에 $\lambda(-w_i'\gamma)$ 대신에 $\hat{\lambda}(-w_i'\hat{\gamma})$ 를 넣어서 선형회귀분석을 수행하여 β 와 β_λ 에 대한 일차추정량을 구한다. 만약 $\rho \neq 0$ 인 경우 관측자료, 즉 $z_i = 1$ 인 자료로만 모형을 적합하게 되면 $\rho\sigma\lambda(-w_i'\gamma)$ 만큼의 편의가 발생한다.

4. 분석결과

이 절에서는 실증분석을 위해 Table 2.1의 변수를 사용하여 Tobit 모형과 Heckman의 표본선택 모형을 활용하였다. 통계분석을 위해서 통계패키지 Stata SE 11.0을 사용하였다.

먼저 Tobit 모형의 결과는 Table 4.1에 제시하였다. OLS의 회귀분석 결과에서 계수값은 독립변수의 변화가 종속변수의 평균값의 변화량인 한계효과 (marginal effect)를 의미하지만, Tobit 모형에서 계수값은 독립변수의 변화가 종속변수의 평균값에 미치는 변화량과 그 값이 관찰될 확률의 변화량이라는 두 가지 효과를 동시에 가진다. 따라서 독립변수의 변화에 따른 종속변수의 평균값의 변화량을 의미하는 한계효과는 변화량과 관찰될 확률의 곱으로 결정되므로 (Green, 2003), 한계효과 (dy/dx)의 값도 함께 제시하였다.

먼저, 분석결과에서 σ 의 값을 통해 Tobit 모형을 채택한 것이 타당한지를 확인할 수 있다. 추정된 σ 의 값이 87.371로 통계적으로 유의한 것으로 나타나서 ($p < 0.001$), Tobit 모형이 타당함을 알 수 있다. 세부적인 독립변수의 효과는 통계적으로 유의한 결과를 보이는 독립변수 위주로 설명한다.

먼저, 성별의 경우 여자에 비해서 남자의 임금이 통계적으로 유의하게 낮은 것으로 나타났다. 마이스터고 졸업생은 기타 졸업생에 비해서 임금이 유의하게 더 높게 나타났다. 한편 부모소득이 높을수록 임금이 더 높아지는 경향이 있으며, 부모의 교육수준은 고졸이하에 비해서 전문대졸인 경우 임금이 유의하게 더 낮게 나타났으나 대졸 이상에서는 더 높게 나타났다. 또한 고교성적은 보통에 비해서 매우 낮거나 낮은 경우는 임금이 통계적으로 유의하게 낮게 나타났으며, 고교성적이 보통에 비해서 높거나 매우 높은 경우는 임금이 통계적으로 유의하게 높게 나타났다. 고교만족도가 높을수록, 자격증 수가 많을수록 임금이 더 높아지며, 고교지역은 동남권에 비해서 수도권과 대경권, 그리고 제주권은 임금이 유의하게 높게 나타났으며, 충청권, 전라권 및 강원권은 동남권에 비해서 임금이 유의하게 낮게 나타났다. 이상의 독립변수 효과는 취업할 확률에도 동일하게 적용된다.

Table 4.1 Results of Tobit model

Variables	Tobit model		Marginal effects		
	<i>b</i>	s.e.	dy/dx	s.e.	
male	-13.741***	1.126	-12.189***	0.989	
maester	37.028***	1.868	33.897***	1.766	
engineer	-0.036	1.150	-0.031	1.018	
income	100~200	1.660	2.137	1.470	1.895
	200~300	5.389**	2.108	4.782**	1.876
	300~500	17.5667***	2.089	15.641***	1.870
	>500	19.377***	2.616	17.482***	2.401
education	college	-4.209**	1.992	-3.707**	1.746
	university	0.878	1.399	0.778	1.240
gpa	very low	-8.250***	1.912	-7.232***	1.660
	low	-7.251***	1.256	-6.379***	1.096
	high	20.121***	1.122	18.014***	1.015
	very high	37.422***	1.398	34.089***	1.304
satisfy	1~2	13.424***	0.527	11.880***	0.466
	≥3	16.096***	1.177	14.222***	1.041
no. licenses	≥3	31.671***	1.391	28.478***	1.267
	sudo	3.249***	1.254	2.876***	1.111
region	chungchung	-4.407***	1.601	-3.884***	1.405
	junra	-14.555***	1.674	-12.689***	1.435
	daegyeong	15.110***	1.635	13.552***	1.486
	gangwon	-12.485***	4.295	-10.870***	3.674
	jeju	15.663**	6.361	14.113**	5.826
constant	24.221***	3.171			
<i>n</i> (censored/uncensored)	43,703 (9,938/33,765)				
LR χ^2 (df)	5,303.90(22)***				
σ	87.371***				

p* < 0.10, *p* < 0.05, ****p* < 0.01.

다음으로 Heckman의 표본선택 모형의 결과는 Table 4.2와 같다.

Table 4.2 Results of Heckman sample selection model

Variables	Selection model (1-step)		Outcome model (2-step)		
	<i>b</i>	s.e.	<i>b</i>	s.e.	
male	-0.215***	0.018	-4.465***	0.995	
maester	0.417***	0.036	25.177***	1.562	
engineer	0.001	0.018	-1.138*	0.687	
income	100~200	0.014	0.033	-1.093	1.273
	200~300	0.062*	0.033	3.432***	1.273
	300~500	0.229***	0.033	8.650***	1.463
	>500	0.160***	0.042	13.647***	1.635
education	college	-0.012	0.034	-4.371***	1.152
	university	-0.046**	0.023	4.152***	0.829
gpa	very low	-0.082***	0.028	-1.726	1.261
	low	-0.068***	0.019	-3.725***	0.812
	high	0.262***	0.019	11.192***	1.070
	very high	0.507***	0.026	22.316***	1.653
satisfy	0.182***	0.008	5.953**	0.708	
no. licenses	1~2	0.235***	0.017	3.940***	1.179
	≥3	0.497***	0.023	11.200***	1.862
region	sudo	0.147***	0.020	-4.294***	0.884
	chungchung	-0.107***	0.025	1.224	1.011
	junra	-0.229***	0.026	-4.856***	0.271
	daegyeong	0.094***	0.027	13.780***	1.008
	gangwon	-0.068	0.067	-11.620***	2.555
jeju	0.667***	0.134	-6.363	3.965	
constant	-0.289***	0.049	101.005***	7.236	
<i>n</i> (censored/uncensored)	43,703 (9,938/33,765)				
LR χ^2 (df)	1,704.74(22)***				
σ	47.293				
ρ	0.627				
inverse Mill's ratio (λ)	29.633***				

Heckman의 표본선택 모형을 활용한 2단계 추정을 통해서, 각 단계에서 제시되는 계수 값은 정확하게 일치하는 것은 아니지만 Tobit 모형에서 나타나는 추정 계수 값을 종속변수 값이 나타날 확률과 종속변수의 평균값 변화로 분해한 것이라 볼 수 있다.

먼저 Heckman의 표본선택 모형에 대한 타당성을 보여주는 λ (inverse Mill's ratio)의 값이 29.633으로 통계적으로 유의한 것으로 나타나서 ($p < 0.001$), Heckman의 표본선택 모형이 타당함을 보여준다. 1단계의 분석결과는 취업여부를 결정하는 독립변수에 대한 효과를 추정한 결과이며, 2단계의 분석결과는 임금의 크기에 대한 독립변수의 효과를 추정한 것이다.

먼저 성별은 1단계와 2단계에서 모두 통계적으로 유의한 것으로 나타났다. 특히 남자의 취업확률과 임금의 크기가 여자에 비해서 유의하게 낮은 것으로 나타났다. 그리고 마이스터고 졸업생은 취업확률과 임금의 크기 모두 기타 졸업생에 비해서 높은 것을 알 수 있으며, 공학계열 출신은 비공학계열 출신에 비해서 유의하지는 않지만 취업확률은 높고, 임금의 크기는 통계적으로 유의하게 낮은 것으로 나타났다. 부모소득이 높을수록 취업확률과 임금의 크기가 모두 통계적으로 유의하게 높으며, 부모학력이 고졸이하에 비해서 대졸이상이 취업확률은 통계적으로 유의하게 낮지만, 임금의 크기는 높게 나타났다. 고교 성적이 높을수록 취업확률과 임금의 크기는 모두 높게 나타났다. 또한 고교 만족도가 높을수록, 자격증 수가 많을수록 취업확률과 임금의 크기 모두 통계적으로 유의하게 높은 것을 알 수 있다. 한편 고교지역은 동남권에 비해서 수도권은 취업확률은 높지만, 임금의 크기는 낮은 것으로 나타났고, 충청권은 동남권에 비해서 취업확률은 낮지만 임금의 크기는 높게 나타났고, 전라권과 강원권은 취업확률과 임금의 크기 모두 동남권에 비해서 낮게 나타났다. 또한 대경권은 동남권에 비해서 취업확률과 임금의 크기 모두 높게 나타났고, 제주권은 동남권에 비해서 취업확률은 높게 나타났지만, 임금의 크기는 낮은 것으로 나타났다. 이상의 결과는 대체로 Tobit 모형의 결과와 유사함을 알 수 있다.

5. 결론

본 연구에서는 한국고용정보원에서 실시한 「2013 고졸자 취업진로조사」 자료를 활용하여 임금결정요인을 분석하였다. 임금은 개인의 취업여부와 임금의 크기에 대한 두 가지의 복합적인 정보를 담고 있는데 초점을 두고, 표본선택 편의를 반영한 모형으로 Tobit 모형과 Heckman의 표본선택 모형을 활용하여 분석하였다. 또한 두 모형에 대한 분석결과는 대체로 유사한 패턴을 보였으며, Heckman의 표본선택 모형을 위주로 주요 분석결과를 요약하면 다음과 같다.

첫째, Tobit 모형과 Heckman의 표본선택 모형에 대한 타당성은 통계적으로 유의함을 알 수 있었다.

둘째, 남자는 여자에 비해서 취업확률과 임금의 크기에서 모두 통계적으로 유의하게 낮은 것으로 나타났으며, 마이스터고 졸업생은 취업확률과 임금의 크기 모두 기타 졸업생에 비해서 높은 것을 알 수 있었다.

셋째, 부모소득이 높을수록 취업확률과 임금의 크기가 모두 통계적으로 유의하게 높아짐을 알 수 있었다. 또한, 부모학력이 고졸이하에 비해서 대졸이상이 취업확률은 통계적으로 유의하게 낮지만, 임금의 크기는 높게 나타났다.

넷째, 고교성적은 높을수록 취업확률과 임금의 크기는 모두 높게 나타났으며, 고교 만족도가 높을수록, 자격증 수가 많을수록 취업확률과 임금의 크기 모두 통계적으로 유의하게 높은 것을 알 수 있다.

다섯째, 고교지역은 동남권에 비해서 수도권, 대경권 및 제주권은 취업확률이 높게 나타났고, 충청권, 전라권, 강원권은 취업확률이 낮게 나타났다. 한편 임금의 크기는 동남권에 비해서 대경권이 통계적으로 유의하게 높게 나타났으며, 수도권, 전라권, 강원권 등은 낮게 나타났다.

한편 본 연구에서는 특성화고 졸업생들 중 공업계열과 상업계열만을 대상으로 했으며, 사용한 독립변수는 선행연구를 바탕으로 제한적으로 선택했다는 점에서 연구결과에 대한 지나친 일반화는 무리가 있

음을 밝혀 둔다. 따라서 더 많은 독립변수를 포함하는 연구와 「2013 고졸자 취업진로조사」 데이터뿐만 아니라 다양한 패널조사 데이터에 대한 임금결정요인분석은 향후 과제로 남겨둔다.

References

- Amemiya, T. (1984). Tobit models : A survey. *Journal of Econometrics*, **24**, 3-61.
- Cho, K. H. and Jung, C. Y. (2014). Factors affecting the determinants of employment decision making of vocational high school senior students. *The Society of Korean Agricultural Education*, **46**, 73-102.
- Cho, J. S. (2011). Determinants of job finding using student's characteristic information. *Journal of the Korean Data & Information Science Society*, **22**, 849-856.
- Green, W. (2003). *Econometrics*, 5th Ed., Prentice hall, Upper Saddle river, New Jersey.
- Goldberger, A. S. (1964). *Economic theory*, Wiley, New York.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, **5**, 475-492.
- Jang, H. J. and Lee, J. H. (2012). Determinants of the specialized vocational high school student's career choice using hierarchical generalized linear model. *Journal of Vocational Education & Training*, **15**, 51-74.
- Kang, S. H. (2013). The employment determinants and labor market attachment possibility of vocational high school graduates. *Quarterly Journal of Labor Policy*, **13**, 91-123.
- Lee, J. H. and Jung, C. Y. (2010). Variables associated with employment barriers of specialized vocational high school students. *The Society of Korean Agricultural Education*, **42**, 25-47.
- Lee, S. C. and Joo, H. J. (2013). The effects of specialized vocational high school characteristics on student employment: Utilizing HGLM model. *Journal of Employment and Skills Development*, **16**, 59-82.
- Park, S. I. and Cho, J. S. (2015). Determinants of employee's wage using hierarchical linear model. *Journal of the Korean Data & Information Science Society*, **26**, 65-75.
- Ryu, J. S. and Cho, J. S. (2016a). The employment path and employment determinants of the vocational high school graduates. *Journal of Regional Studies*, **24**, 199-218.
- Ryu, J. S. and Cho, J. S. (2016b). The wage determinants of the vocational high school graduates using mixed effects model. *Journal of the Korean Data & Information Science Society*, **27**, 935-946.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24-36.

The wage determinants applying sample selection bias

Sungik Park¹ · Jangsik Cho²

¹Department of International Trade and Commerce, Kyungsoong University

²Division of Mathematics and Applied Statistics, Kyungsoong University

Received 12 August 2016, revised 13 September 2016, accepted 23 September 2016

Abstract

The purpose of this paper is to explain the factors affecting the wage of the vocational high school graduates. We particularly examine the effectiveness of controlling sample selection bias by employing the Tobit model and Heckman sample selection model. The major results are as follows. First it is shown that the Tobit model and Heckman sample selection model controlling sample selection bias is statistically significant. Hence all the independent variables seem to be statistically consistent with the theoretical model. Second, gender was statistically significant, both in the probability of employment and the wage. Third, the employment probability and wage of Maester high school graduates were shown to be high compared to all other graduates. Fourth, the higher parent's income, the higher are both the employment probability and the wage. Finally, parents education level, high school grade, satisfaction, and a number of licenses were found to be statistically significant, both in the probability of employment and wages.

Keywords: Censored data, Heckman sample selection model, sample selection bias, Tobit model.

¹ Professor, Department of International Trade and Commerce, Kyungsoong University, Busan 48434, Korea.

² Corresponding author: Professor, Division of Mathematics and Applied Statistics, Kyungsoong University, Busan 48434, Korea. E-mail: jscho@ks.ac.kr