

Heckman의 표본선택모형을 이용한 대졸자의 임금결정요인 분석[†]

조장식¹

¹경성대학교 수학과통계학부

접수 2017년 8월 21일, 수정 2017년 9월 16일, 게재확정 2017년 9월 18일

요약

본 연구에서는 한국고용정보원에서 실시한 「2014년 대졸자 직업이동 경로조사」 자료를 활용하여 대졸자의 임금결정요인을 분석하였다. 일반적으로 임금은 개인의 취업여부와 임금의 크기에 대한 두 가지의 복합적인 정보를 담고 있으나, 많은 선행연구에서는 임금의 크기에 대한 정보만을 활용하여 선형 회귀분석을 수행함으로써 표본선택에 의한 편향의 (sample selection bias) 문제가 발생하게 된다. 이런 문제점을 극복하기 위해 본 연구에서는 Heckman의 표본선택 모형을 분석에 활용하였다. 주요 분석 결과를 요약하면 다음과 같다. 먼저 Heckman의 표본선택 모형에 대한 타당성은 통계적으로 유의함을 알 수 있었다. 남자는 여자에 비해서 취업확률과 임금의 크기 모두 통계적으로 유의하게 높게 나타났으며, 연령이 증가하고 부모의 소득이 증가 할수록 취업확률과 임금의 크기 모두 높게 나타났다. 또한 대학만족도가 높아질수록, 그리고 취득한 자격증 수가 증가할수록 취업확률과 임금 모두 증가하는 경향이 있는 것으로 나타났다.

주요용어: 중단자료, 표본선택 편향, Heckman 표본선택 모형.

1. 서론

대졸자의 임금결정요인 분석은 그동안 많은 연구가 수행되어 왔다. 일반적으로 취업을 해서 임금을 받는 대졸자와 미취업으로 인해 임금이 없는 대졸자는 인구통계학적 특성이 다르다고 예상할 수 있다. 즉 임금결정요인을 분석할 때, 임금은 개인의 취업여부와 임금의 크기에 대한 두 가지의 복합적인 정보를 담고 있다. 특히 전자의 정보는 개인의 선택 (또는 역량)에 달려 있다는 점에서 자기선택성 (self-selectivity)을 내포하고 있으며, 자기선택성에 대한 고려 없이 후자의 정보만을 대상으로 최소제곱법 (ordinary least squares; OLS)에 기초한 선형 회귀분석을 하게 되면 전형적인 표본선택에 의한 편향의 (sample selection bias)의 문제가 발생한다. 즉 취업해서 임금을 받는 졸업자만을 대상으로 분석한다면 이들 대상자들은 무작위로 추출된 표본이 아닐 뿐만 아니라, 비무작위적으로 미취업자 표본들이 분석 대상에서 배제됨으로써 표본선택 편향의 문제가 발생하게 된다. Ryu와 Cho (2016a, 2016b)의 특성화고 졸업생의 임금결정요인을 분석한 연구뿐만 아니라, Cho (2011) 및 Park과 Cho (2015)의 연구도 취업하여 임금을 받는 졸업자만을 대상으로 분석함으로써, 비무작위적으로 미취업자들의 표본이 배제되어 표본선택 편향의 문제를 갖고 있다는 점에서 한계점을 갖고 있다.

본 연구에서는 임금이 갖는 정보가 취업여부와 임금의 크기에 대한 복합적인 정보를 담고 있다는 점에서, 취업자와 미취업자 모두 분석에 포함하였다. 종속변수인 임금이 '0'부터 시작해서 양의 값을 가지는

[†] 이 논문은 2017학년도 경성대학교 학술연구비지원에 의하여 연구되었음.

¹ (48434) 부산광역시 남구 수영로 309 번지, 경성대학교 수학과통계학부, 교수. E-mail : jscho@ks.ac.kr

중단분포 (censored distribution)을 따른다고 가정하며, 표본선택에 의한 편향 문제를 극복하기 위해 임금이 갖는 복합적인 정보를 분해하여 분석한다. 이를 위해 Heckman의 표본선택 모형을 분석에 활용한다. Heckman의 표본선택 모형은 전체 표본에서 임금이 없는 미취업자의 값들이 갖는 분포에 대한 고려가 반영된 것이며, 독립변수가 취업여부와 임금의 크기에 영향을 미치는 과정이 상이하다고 가정하고 각각의 경우에 적합한 분석모형을 적용하여 2단계로 나누어 추정하는 방법이다 (Heckman, 1976). 표본선택 편향을 반영한 선행연구로는 Park과 Cho (2016)는 특성화고 졸업생들의 임금결정요인을 분석하였으며, Kim (2011)은 가계소득이 학자금 대출 참여 및 누적 대출규모에 미치는 영향을 분석하였고, Jung, Kim (2009)는 대학생의 사교육 참여 및 사교육비 지출에 미치는 영향을 분석하였다. 또한 Lee 등 (2016)은 여성을 중심으로 휴학경험이 노동시장의 성과에 미치는 영향을 분석하였다.

본 논문에서는 한국고용정보원에서 조사한 「2014년 대졸자 직업이동 경로조사」 자료를 이용하여 임금결정요인을 분석한다. 표본선택 편향문제를 해결하기 위해 Heckman의 표본선택모형을 활용하였으며, 논문의 구성은 다음과 같다. 제 2절에서는 자료소개와 기술통계를 하고, 제 3절에서는 Heckman 모형에 대해 간단히 소개한다. 그리고 제 4절에서는 Heckman의 표본선택 모형의 실증분석 결과를 제시하고, 마지막 제 5절에서는 결론을 제시한다.

2. 자료소개 및 기술통계

본 연구에서 대졸자의 임금결정요인을 분석하기 위해 「2014년 대졸자 직업이동 경로조사」 자료를 활용한다. 대졸자 직업이동경로조사는 대학졸업자의 경력개발 및 직업이동경로를 추적 조사하는 횡단면 조사로써, 매년 전국의 전문대졸 및 4년제 대졸자를 대상으로 졸업 후 약 18개월이 경과한 시점에서 경제활동상태, 일자리 경험, 교육 및 취업준비, 학교생활 등에 관하여 조사한 데이터이다. 종속변수인 임금은 취업여부와 임금의 크기에 대한 복합 정보를 갖고 있다는 점과 표본선택에 의한 편향 문제를 해결하기 위해, 취업자뿐만 아니라 미취업자의 자료도 분석에 포함시켰다. 분석에 포함된 조사대상자는 2014년 대졸자로서, 총 428,336명이 분석 대상이며, 이 중에서 취업자는 319,858명이고, 미취업자는 108,416명이다.

임금결정요인을 분석하기 위한 독립변수는 기존의 선행 연구인 Park과 Cho (2015, 2016)를 바탕으로 성별, 부모 학력과 부모소득, 전공계열, 대학유형, 국공립 여부, 복수전공 이수 여부, 자격증 수 등을 고려하였다.

분석에 사용된 변수와 변수에 대한 설명은 아래 Table 2.1과 같다.

Table 2.1 Variables explanation

Variables	Explanation	Role
sex	0='female', 1='male'	independent
age	the age at the time of survey	independent
university	whether or not it is university (0='college', 1='university')	independent
parent's education	0='under university graduation', 1='above university graduation'	independent
parent's income	parent's income : 0='<100', 1='100 ~ 300', 2='300 ~ 400', 3='400 ~ 500', 4='>500'	independent
university type	0='private', 1='public'	independent
department category	0='art', 1='humanity', 2='natural science'	independent
gpa	grade point average	independent
employment target	whether or not there was a job target at the time of enrollment	independent

double major	whether or not you completed a double major at the time of your studies	independent
major satisfaction	degree of satisfaction for major : 1='very low', 3='middle', 5='very high'	independent
university satisfaction	degree of satisfaction for university life : 1='very low', 3='middle', 5='very high'	independent
no. of licenses	number of licenses : 0='none', 1='1', 2='2', 3='≥3'	independent
labor union	whether or not there was labor union	independent
size	company size : 0='<10', 1='10 ~ 50', 2='50 ~ 100', 3='100 ~ 300', 5='≥300'	dependent
full-time worker	whether or not you are full-time worker	independent
company satisfaction	degree of satisfaction for company life : 1='very low', 3='middle', 5='very high'	independent
major match	degree of matching work and majors : 1='very low', 3='middle', 5='very high'	independent
region	0='metropolitan', 1='Chungchung', 2='Junra', 3='Daekyung', 4='Dongnam', 5='the others'	independent
employment	whether or not you are employed	dependent
wage	average monthly salary	dependent

아래 Table 2.2는 취업여부와 취업자의 임금에 대한 범주형 독립변수들의 기술통계 결과를 나타낸다.

Table 2.2 Descriptive statistic for categorical independent variables

Variables		Employment			Wage		
		N	No (%)	Yes (%)	N	Mean	S.D.
sex	male	214,507	23.8	76.2	163,372	222.2	105.3
	female	213,829	26.8	73.2	156,504	180.3	77.3
university	university	276,487	26.8	73.2	202,452	213.3	98.6
	college	151,849	22.7	77.3	117,423	181.6	84.6
parent's education	<university	273,192	23.7	76.3	208,342	197.8	91.7
	≥university	155,144	28.1	71.9	111,533	208.9	100.4
parent's income	<100	16,357	23.6	76.4	12,494	215.5	143.8
	100 ~ 300	106,318	24.5	75.5	80,222	191.6	90.7
	300 ~ 400	99,296	24.7	75.3	74,810	198.1	82.3
	400 ~ 500	116,581	25.6	74.4	86,772	198.1	89.4
	≥500	89,785	27.0	73.0	65,579	220.3	105.5
sex	male	214,507	23.8	76.2	163,372	222.2	105.3
	female	213,829	26.8	73.2	156,504	180.3	77.3
university	university	276,487	26.8	73.2	202,452	213.3	98.6
	college	151,849	22.7	77.3	117,423	181.6	84.6
parent's education	<university	273,192	23.7	76.3	208,342	197.8	91.7
	≥university	155,144	28.1	71.9	111,533	208.9	100.4
parent's income	<100	16,357	23.6	76.4	12,494	215.5	143.8
	100 ~ 300	106,318	24.5	75.5	80,222	191.6	90.7
	300 ~ 400	99,296	24.7	75.3	74,810	198.1	82.3
	400 ~ 500	116,581	25.6	74.4	86,772	198.1	89.4
	≥500	89,785	27.0	73.0	65,579	220.3	105.5
university type	private	360,191	24.6	75.4	271,494	198.6	93.9
	public	68,145	29.0	71.0	48,382	219.1	99.1
department category	humanity	169,395	27.3	72.7	123,164	200.1	96.8
	natural	199,638	23.6	76.4	152,487	214.0	94.3
	art	59,303	25.4	74.6	44,224	163.8	81.0

employment target	no	236,183	25.7	74.3	175,378	196.0	95.2
	yes	192,153	24.8	75.2	144,497	208.6	94.2
double major	no	374,678	25.1	74.9	280,617	200.4	92.6
	yes	53,658	26.8	73.2	39,258	210.4	109.8
no. of licences	none	160,674	29.2	70.8	113,768	200.7	99.8
	1	142,246	21.8	78.2	111,254	203.7	93.5
	2	68,552	23.8	76.2	52,245	199.9	90.5
	≥3	56,866	25.1	74.9	42,609	200.9	90.6
labor union	no				256,361	193.6	91.7
	yes				63,515	234.1	100.7
size	<10				62,316	167.4	102.4
	10-50				62,888	184.4	77.3
	50-100				25,982	194.6	72.7
	100-300				34,978	203.0	69.2
	≥300				122,368	232.9	101.1
full-time worker	no				70,228	150.6	79.8
	yes				235,999	215.7	85.0
region	Metropolitan				174,048	203.7	95.1
	Chungchung				30,793	197.9	92.9
	Junra				24,349	194.5	101.3
	Daekyung				31,930	193.6	95.5
	Dongnam				45,491	209.2	92.4
	the others				12,864	189.8	91.1
Total		428,336	25.3	74.7	319,876	201.7	95.0

Table 2.2의 결과에 따르면, 여자에 비해서 남자가 취업률과 임금의 크기가 모두 높음을 알 수 있다. 전문대졸은 4년제 대졸자에 비해서 취업률은 높지만 임금은 낮게 나타났다. 부모의 교육수준이 4년제 대졸자 미만인 집단은 4년제 대졸자 집단에 비해서 취업률은 높게 나타났지만 임금은 낮게 나타났다. 부모의 소득이 높아질수록 취업률은 낮아지지만, 임금은 증가하는 경향을 보였다. 대학유형이 사립인 경우는 국공립에 비해서 취업률은 높게 나타났지만, 임금은 낮게 나타났으며, 자연계열이 인문계열 및 예체능계열에 비해서 취업률과 임금이 모두 높게 나타났다. 취업목표가 있는 경우는 없는 경우에 비해서 취업률과 임금 모두 높게 나타났으며, 복수전공을 이수한 집단은 그렇지 않은 집단에 비해서 취업률은 낮았지만 임금은 더 높게 나타났다. 취득한 자격증의 수는 없는 경우는 취업률이 가장 낮았으며, 자격증 수가 1개 이상인 경우는 많아질수록 다소 취업률이 낮아지는 경향이 있으며, 임금의 경우는 뚜렷한 경향이 없는 것으로 나타났다.

한편 취업자를 대상으로 임금의 기술통계를 살펴보면, 노동조합이 있는 경우는 없는 경우에 비해서 임금이 더 높았으며, 기업규모는 커질수록 임금이 증가함을 알 수 있다. 정규직은 비정규직에 비해서 임금이 높았으며, 지역적으로는 동남권의 임금이 가장 높았으며, 그 다음으로 수도권, 충청권 순으로 나타났다.

3. 연구방법

일반적으로 임금 결정요인을 분석할 때 임금은 대졸자 개인의 취업여부와 임금의 크기에 대한 두 가지 정보를 담고 있다. 임금의 크기 정보만을 대상으로 회귀분석을 하게 되면 이들 대상자들은 무작위로 추출된 표본이 아닐 뿐만 아니라, 비무작위적으로 미취업자 표본들이 분석대상에서 배제됨으로써 전형적인 표본선택에 의한 편의 (bias)가 발생하게 된다. 따라서 본 연구에서는 종속변수의 값이 복합적인 정보를 담고 있다는 점에서 이를 분해하여 분석한다. 이를 위해 Heckman의 표본선택 모형을 분석에 활용한다. Heckman의 표본선택 모형은 취업여부와 임금의 크기에 영향을 미치는 과정이 상이하다는 가정에 근거하고 있으며, 취업여부와 임금의 크기를 결정하는 요인을 2단계로 분리해서 분석한다.

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

$$z_i^* = w_i' \gamma + u_i, \quad z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0. \end{cases} \quad (3.1)$$

$$y_i^* = x_i' \beta + \epsilon_i, \quad y_i = \begin{cases} y_i^* & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0. \end{cases} \quad (3.2)$$

여기서 $\begin{pmatrix} u_i \\ \epsilon_i \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right)$ 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 (3.1)에 대해서 프로빗 모형 (probit model)을 이용하여 추정한다. Heckman 표본선택 모형의 2단계에서는 표본으로 선택된 사례만을 대상으로 하며 식 (3.2)으로부터 다음 식 (3.3)와 식 (3.4)가 계산된다 (Green, 2003).

$$\begin{aligned} E[y_i | z_i^* > 0] &= x_i' \beta + E[\epsilon_i | u_i > -w_i' \gamma] \\ &= x_i' \beta + \rho\sigma \frac{\phi(-w_i' \gamma)}{1 - \Phi(-w_i' \gamma)} = x_i' \beta + \beta_\lambda \lambda(-w_i' \gamma), \end{aligned} \quad (3.3)$$

$$y_i | z_i^* = E[y_i | z_i^* > 0] + \nu_i = x_i' \beta + \beta_\lambda \lambda(-w_i' \gamma) + \nu_i. \quad (3.4)$$

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 (3.3)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 (3.4)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 Heckman의 표본선택 모형을 활용하였다. 통계분석을 위해서 통계패키지 Stata SE 11.0을 사용하였다.

Heckman의 표본선택 모형의 결과는 Table 4.1과 같다.

Table 4.1 The characteristic analysis for respondents

Variables	Selection model(1-step)		Outcome model(2-step)	
	b	s.e.	b	s.e.
sex	0.062***	0.031	28.764***	1.132
age	0.015***	0.005	5.935***	0.215
university	-0.093***	0.001	3.843***	1.467
parent's education	-0.100***	0.005	-9.356***	1.611
parent's income	100~300	0.150***	26.490***	2.808
	300 ~ 400	0.167***	33.738***	3.023
	400 ~ 500	0.124***	30.621***	2.535
	>500	0.132***	41.965***	2.655
university type	-0.132***	0.006	-13.167***	2.134
department category	humanity	-0.037***	6.585***	1.087
	natural	0.092***	33.486***	1.626
gpa	-0.004	0.005	5.124***	0.726
employment target	0.034***	0.004	8.744***	0.775
double major	0.014*	0.007	6.826***	0.929
major satisfaction	0.008*	0.003	-0.131	0.461
university satisfaction	0.017***	0.003	5.819***	0.515
no. of licenses	1	0.239***	31.971***	3.550
	2	0.178***	19.426***	2.792
	≥3	0.117***	15.081***	2.012
labor union	-	-	20.100***	0.347
size	10-50	-	13.897***	0.438
	50-100	-	22.963***	0.569
	100-300	-	29.865***	0.515
	≥300	-	54.918***	0.404
full-time worker	-	-	57.725***	0.319
company satisfaction	-	-	4.117***	0.173
major match	-	-	1.640***	0.119
region	Chungchung	-	-9.876***	0.472
	Junra	-	-15.270***	0.563
	Daekyung	-	-13.775***	0.496
	Dongnam	-	1.840***	0.403
	the others	-	-21.437***	0.720
constant	-0.049	0.031	-283.641***	23.489
n(censored/uncensored)			403,290(108,416/294,874)	
σ			188.095	
ρ			1.338	
inverse Mill's ratio(λ)			251.734***	

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

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

먼저 Heckman의 표본선택 모형에 대한 타당성을 보여주는 (inverse Mill's ratio)의 값이 251.734로 통계적으로 유의한 것으로 나타나서 (p -value < 0.001), Heckman의 표본선택 모형이 타당함을 보여준

다. 1단계의 분석결과는 취업여부를 결정하는 독립변수에 대한 효과를 추정한 결과이며, 2단계의 분석결과는 임금의 크기에 대한 독립변수의 효과를 추정한 것이다.

먼저 미취업자와 취업자 전체 조사대상자에 대한 결과를 살펴보면 다음과 같다. 성별의 결과는 남자의 취업확률과 임금의 크기가 여자에 비해서 유의하게 높은 것으로 나타났다. 그리고 연령이 증가할수록 취업확률과 임금의 크기 모두 증가하는 경향이 있음을 알 수 있으며, 전문대 졸업생에 비해서 4년제 대졸자들이 통계적으로 유의하게 취업확률은 낮지만 임금은 더 높음을 알 수 있다. 한편 부모의 교육수준은 4년제 대졸 미만인 집단에 비해서 4년제 대졸 이상인 집단이 통계적으로 유의하게 취업확률도 낮고 임금도 낮음을 알 수 있다. 반면 부모의 소득은 높아질수록 취업확률도 높아지고 임금도 높아짐을 알 수 있다. 대학유형이 사립에 비해서 국공립은 취업확률도 낮고 임금도 낮은 것으로 나타났다. 전공계열은 예체능에 비해서, 인문계열은 취업확률은 낮고 임금은 높게 나타났으며, 자연계열은 취업확률도 높게 나타났고 임금도 높게 나타났다. 또한 학점이 증가할수록 취업확률은 낮아지는 경향이 있지만, 임금은 증가하는 경향이 있다. 취업목표와 복수전공이 없는 집단에 비해서 있는 집단이 취업확률도 높고 임금이 증가하는 경향이 있는 것으로 나타났다. 그리고 대학만족도가 높아질수록 취업확률과 임금 모두 증가하는 경향이 있는 것으로 나타났다. 취득한 자격증 수가 증가할수록 취업확률과 임금 모두 증가하는 경향이 있는 것으로 나타났다. 이상의 Heckman 모형의 분석 결과는 대체로 Table 2.2의 기술통계의 결과와 유사하나 일부의 결과는 다소 상이하게 나타났는데, 그 이유는 Heckman 모형에서는 여러 독립변수들의 통계 하에서 주효과를 해석하기 때문에 나타나는 차이로 볼 수 있다.

다음으로 전체 조사대상자 중에서 취업자들에 한해서 임금결정요인의 결과를 살펴보면 다음과 같은 결과를 얻을 수 있다. 먼저 노동조합이 없는 집단에 비해서 있는 집단의 임금이 통계적으로 유의하게 높은 것을 알 수 있고, 기업규모가 클수록 임금이 유의하게 높아짐을 알 수 있다. 비정규직에 비해서 정규직의 임금이 더 높은 것으로 나타났으며, 기업 만족도와 전공일치도가 높을수록 임금이 더 높아짐을 알 수 있다. 기업소재지의 측면에서 결과를 살펴보면, 수도권에 비해서 동남권만이 유의하게 임금이 높게 나타난 반면에 충청권, 전라권, 대경권 및 기타 지역은 수도권에 비해서 임금이 유의하게 낮게 나타났다.

5. 결론

본 연구에서는 한국고용정보원에서 실시한 「2014년 대졸자 직업이동 경로조사」 자료를 활용하여 대졸자들의 임금결정요인을 분석하였다. 임금은 개인의 취업여부와 임금의 크기에 대한 복합적인 정보를 담고 있다는데 초점을 두고, 표본의 선택편의를 반영한 Heckman의 표본선택 모형을 활용하여 분석하였다. 분석 결과를 요약하면 다음과 같다.

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

둘째, 남자는 여자에 비해서 취업확률과 임금의 크기에서 모두 통계적으로 유의하게 높게 나타났으며, 연령이 증가할수록 취업확률과 임금의 크기 모두 높게 나타났다.

셋째, 부모의 교육수준은 4년제 대졸 미만인 집단에 비해서 4년제 대졸 이상인 집단이 통계적으로 유의하게 취업확률도 낮고 임금도 낮음을 알 수 있다. 반면 부모의 소득은 높아질수록 취업확률도 높아지고 임금도 높아짐을 알 수 있다.

넷째, 대학유형이 사립에 비해서 국공립은 취업확률도 낮고 임금도 낮은 것으로 나타났으며, 학점이 증가할수록 취업확률은 낮아지는 경향이 있지만, 임금은 증가하는 경향이 있는 것으로 나타났다.

다섯째, 대학만족도가 높아질수록, 그리고 취득한 자격증 수가 증가할수록 취업확률과 임금 모두 증가하는 경향이 있는 것으로 나타났다.

여섯째, 노동조합이 없는 집단에 비해서 있는 집단의 임금이 통계적으로 유의하게 높은 것을 알 수 있고, 기업규모가 클수록 임금이 유의하게 높아짐을 알 수 있다. 또한 비정규직에 비해서 정규직의 임금이

더 높은 것으로 나타났으며, 기업 만족도와 전공일치도가 높을수록 임금이 더 높아짐을 알 수 있다.

마지막으로, 기업소재지의 측면에서는 수도권에 비해서 동남권만이 유의하게 임금이 높게 나타난 반면에 충청권, 전라권, 대경권 및 기타 지역은 수도권에 비해서 임금이 유의하게 낮게 나타났다.

한편 본 연구에서 사용한 독립변수는 선행연구를 바탕으로 제한적으로 선택했다는 점에서 연구결과에 대한 지나친 일반화는 무리가 있음을 밝혀 둔다. 따라서 더 많은 독립변수를 포함하는 연구와 「2014년 대졸자 직업이동 경로조사」 데이터뿐만 아니라 다양한 패널조사 데이터에 대한 임금결정요인 분석은 향후 과제로 남겨둔다.

References

- 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 edition)*, Prentice hall, New Jersey.
- 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.
- Jung, J. S. and Kim, H. H. (2009). Analysis on the influential factors on private tutoring expenditure of Korean college students. *The Journal of Economics and Finance of Education*, **18**, 89-122.
- Kim, H. H. (2011). An analysis of impacts of family income on financing experience and cumulative amount of student loan : Focusing on private junior college graduate. *The Journal of Economics and Finance of Education*, **20**, 183-203.
- Lee, J. H., Kim, D. J. and Kim, J. (2016). Leave of absence and labor market performance: Focus on female college graduates under the age of 39. *The Journal of Women and Economics*, **13**, 1-20.
- 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.
- Park, S. I. and Cho, J. S. (2016). The wage determinants applying sample selection bias. *Journal of the Korean Data & Information Science Society*, **27**, 1317-1325.
- 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 of college graduates using Heckman's sample selection model[†]

Jangsik Cho¹

¹Division of Mathematics and Applied Statistics, Kyungsoong University

Received 21 August 2017, revised 16 September 2017, accepted 18 September 2017

Abstract

In this study, we analyzed the determinants of wages of college graduates by using the data of 「2014 Graduates Occupational Mobility Survey」 conducted by Korea Employment Information Service. In general, wages contain two complex pieces of information about whether an individual is employed and the size of the wage. However, in many previous researches on wage determinants, sample selection bias tends to be generated by performing linear regression analysis using only information on wage size. We used the Heckman sample selection models for analysis to overcome this problem. The main results are summarized as follows. First, the validity of the Heckman's sample selection model is statistically significant. Male is significantly higher in both job probability and wage than female. As age increases and parents' income increases, both the probability of employment and the size of wages are higher. Finally, as the university satisfaction increases and the number of certifications acquired increased, both the probability of employment and the wage tends to increase.

Keywords: Censored data, Heckman's sample selection model, Sample selection bias.

[†] This research was supported by Kyungsoong University Research Grants in 2017.

¹ Professor, Division of Mathematics and Applied Statistics, Kyungsoong University, Busan, 48434, Korea.
E-mail : jscho@ks.ac.kr