

중소기업 청년인턴 취업자의 재직기간 분석

박성익¹ · 류장수² · 김종한³ · 조장식⁴

¹경성대학교 국제무역통상학과 · ²부경대학교 경제학부 ·

³경성대학교 경제금융물류학부 · ⁴경성대학교 정보통계학과

접수 2016년 1월 11일, 수정 2016년 1월 19일, 게재확정 2016년 2월 20일

요약

취업자들이 재직기간이 경과하면서 이직 또는 실업 상태로 탈출확률 및 탈출요인 문제를 분석하는 것은 취업의 질을 측정할 수 있는 하나의 방법이다. 일반적으로 취업자들의 이직 또는 실업으로의 탈출확률은 취업자의 개인특성뿐만 아니라, 직종 특성에도 영향을 받는 복수의 분석단위를 가지게 된다. 복수의 분석단위를 가지는 위계적 (hierarchical) 자료구조에서는 직종별로 공유되는 특성이 존재하게 되어, 동일 직종 집단 내의 상관성이 발생할 수 있다. 따라서 본 연구에서는 취업자 개인특성 (1-수준) 과 직종 특성 (2-수준)의 위계적 자료구조 하에서 콕스의 비례위험 모형 (Cox's proportional hazard model)을 이용하여 중소기업 청년인턴사업에 참여한 취업자들의 재직기간 중 실직 및 이직으로의 탈출요인을 분석하였다. 분석결과를 요약하면 다음과 같다. 처리집단 (인턴집단)이 통제집단 (비인턴 집단)에 비해서 탈출확률이 통계적으로 유의하게 낮음을 알 수 있다. 또한 남자들이 여자들에 비해서 탈출할 확률이 높고, 연령이 높을수록 탈출할 확률이 더 낮아지는 것을 알 수 있다. 그리고 기업규모가 클수록 탈출확률이 낮으며, 직종별로는 관리사무 관련직에 비해서 전문 서비스 관련직의 탈출확률이 더 낮게 나타났다.

주요용어: 랜덤효과, 생존나무분석, 위계적 자료구조, 콕스 비례위험 모형, 혼합모형.

1. 서론

갈수록 심화되는 청년실업 문제 해소 방안의 하나로써, 정부는 중소기업 청년인턴제를 실시하여, 청년 실업자에게 중소기업에서의 인턴경험을 제공함으로써, 정규직 취업을 촉진하고 중소기업의 인력난을 해소하기 위하여 노력하고 있다. 그동안 많은 연구자들이 중소기업 청년취업인턴제도의 효과에 대해서 분석한 바 있다 (Lee, 2009; Ryu 등, 2012; Joo 등, 2013; Ryu 등, 2014). 이들은 특히, 중소기업 청년인턴 제도가 정부의 정책의도와 부합되게 청년들의 고용창출 및 정규직 전환에 긍정적인 효과를 미쳤다는 연구결과를 제시하고 있다. 그런데 청년인턴으로 취업한 사람이 정규직으로 취업을 하더라도, 오래 근무하지 않고 이직하거나 실직을 하게 되면, 중소기업 청년인턴제도의 도입취지는 제대로 달성되었다고 보기에는 무리가 있다. 이러한 관점에서 청년인턴제도의 성과를 측정하기 위하여 본 연구는 청년인턴 참가자의 재직기간을 결정하는 요인에 대하여 분석하고자 한다.

청년인턴 참가자들이 취업한 직장에서 재직기간이 경과하면서 이직 또는 실직 상태로 탈출하는 문제를 분석하는 것은 취업의 질을 측정할 수 있는 좋은 방법 중 하나이다. 좋은 직장이라면 재직기간이 길

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

² (48513) 부산광역시 남구 용소로 45, 부경대학교 경제학부, 교수.

³ (48434) 부산광역시 남구 수영로 309 번지, 경성대학교 경제금융물류학부, 교수.

⁴ 교신저자: (48434) 부산광역시 남구 수영로 309 번지, 경성대학교 정보통계학과, 교수.

E-mail: jscho@ks.ac.kr

게 될 것이고, 그렇지 않다면 재직 기간이 짧게 될 것이기 때문이다. 일반적으로 일정 기간 동안 재직을 계속하던 상태에서 이직 또는 실직상태로 탈출하게 될 확률은 해자드 모형 (hazard model)이라고 알려진 통계모형을 통해 추정할 수 있다. 해자드 모형은 생존과 관련된 독립변수가 있을 때 탈출에 영향을 미치는 여러 변수들을 동시에 알아보기 위해 사용되는 분석방법이다. 특히, 다중 회귀분석에 비해서 직장에서 계속 재직을 하여 이직 또는 실직상태로 탈출하지 않아서 재직기간을 관찰할 수 없는 우측 절단 (right censoring)이 있는 경우에도 일치추정량을 갖는 것으로 알려져 있다. 또한 생존시간에 대해 어떠한 분포형태도 가정하지 않으므로 비모수적인 분석이지만, 모형에 근거하여 회귀계수를 추정한다는 점이 모수적방법과 유사하여 준모수모형이라고 한다 (Song, 2005). 해자드 모형을 이용한 선행연구로 Park 등 (2012)은 중소기업의 업종별로 기업의 생존율과 생존요인을 콕스의 비례위험 모형을 이용하여 분석하였다. Chang과 Ho (2001)는 1997년 하반기 이후 발생한 실직자와 신규노동시장 진입자의 미취업 기간탈출의 경로를 정규직과 비정규직, 비 임금근로의 형태로 구분하여 분석하였다. Jun 등 (2009)은 누리사업단 대학졸업자들을 대상으로 미취업에서 취업으로의 탈출확률 및 탈출요인을 분석하였다. 그 외에도 Kiefer (1988), Yang과 Ha (2009), Jung (2012), Cho (2015) 등의 연구가 있다.

한편 일반적으로 청년인턴 참가자들의 재직기간은 취업자의 특성뿐 만 아니라, 직종 특성에도 영향을 받는다. 즉 다수의 청년인턴 참가자 개인적 특성 (1-수준)외에도 동일한 직종 (2-수준)에 속하게 되어 복수의 분석단위를 가지게 되는 위계적 자료구조 (hierarchical data structure)를 갖게 된다. 이에 따라 직종별로 공유되는 특성이 존재하게 되어 집단 (직종) 내의 상관성이 발생할 수 있다. 이와 같이 위계적 자료구조를 가지는 경우에 콕스의 비례위험 모형 (Cox's proportional hazard model)과 같은 전통적인 통계모형은 더 이상 일치추정량을 제공하지 못한다. 따라서 위계적 자료구조에서 보다 타당한 분석을 위해서는 반드시 분석단위의 문제를 해결해야 한다는 지적이 Nasser와 Hagtvet (2006) 등에 의해 제기되어 왔다. Raudenbush과 Bryk (2002)은 위계적 선형모형이 횡단적 다층 자료구조의 통합모형, 다변량 모형, 잠재변수 모형, 베이저안 추론 모형 등에 폭넓게 적용될 수 있음을 제시하고 있다. Kim과 Jun (2010)은 초등학교 학부모의 자녀의 학교여건 만족도에 대한 영향을 분석하였고, Cho (2013)와 Park과 Cho (2015)는 위계적 선형모형을 이용한 강의평가점수 및 임금 결정요인 분석에 대한 연구를 하였다.

본 연구에서는 취업자 특성과 직종특성의 위계적 자료구조 하에서 콕스의 혼합모형 (Cox's mixed effects model)을 이용하여 청년인턴 참가자들의 실업 또는 이직으로 탈출요인을 분석하고자 한다. 제 2절에서는 분석자료를 소개하고 간단한 기술통계 결과를 제시하고, 제 3절에서는 독립변수들과 재직기간과의 관련성을 탐색하기 위해 생존나무분석 결과를 제시한다. 제 4절에서는 콕스의 혼합모형에 대한 결과를 소개하고, 제 5절에서는 결론을 제시한다.

2. 자료소개 및 기술통계

본 연구는 정부의 청년실업 대책인 중소기업 청년인턴사업에 참여하여 정규직으로 전환된 사람들의 재직기간 결정요인을 분석하는 것이다. 그런데 정부의 정책효과를 분석하기 위해서는 유사 실험설계 (quasi-experimental design)를 활용하는 것이 일반적이다. 즉, 인턴사업 참여자와 가능한 한 유사한 특성을 가진 비참여자 집단으로 통제집단을 구성한 후, 사업에 참여한 처리집단 (treatment group)과 참여하지 않은 통제집단 (control group)의 결과를 비교함으로써 정부 사업의 효과를 측정한다.

따라서 본 연구에서 사용한 데이터는 2009년부터 2012년까지 중소기업 청년인턴 사업에 참여한 모든 참여자 (청년인턴 DB)와 동 기간의 고용보험 가입자 (고용보험 DB)들을 대상으로 하였다. 그리고 청년인턴 DB와 고용보험 DB를 결합한 후, 성향점수매칭법 (propensity score matching method)을 활용하여 인턴사업 참여자 (처리집단)와 최대한 유사한 특성을 가진 비참여자 집단 (통제집단)으로 분석용 데이터를 구성하였다. 이에 대한 자세한 내용은 Ryu 등 (2014)의 자료를 참고하기 바란다.

본 연구에서는 분석을 위해 청년인턴사업 참여자 여부 (group), 성별 (sex), 연령 (age), 직장소재지 (area), 직장규모 (size), 직종 (job) 등을 독립변수로 하고, 재직기간 (duration)을 종속변수로 사용하였다. 여기서 직종은 관리 및 사무 관련직, 전문 서비스 관련직, 단순 서비스 관련직, 과학/공학 관련직 및 단순 생산 관련직 등 유사한 직종을 의미하며, 변수들에 대한 설명은 Table 2.1과 같다.

Table 2.1 Variables explanation

variables	explanation	role
group	0=control, 1=treatment	independent
sex	0=female, 1=male	independent
age	age at the time of survey (1='≤20', 2='21~25', 3='26~30', 4='>30')	independent
area	1=Capital (CAP), 2=Chungcheong (CC), 3=Jeolla (JL), 4=Daekyung (DK), 5=Dongnam (DN), 5=Others (OT)	independent
job	1=administrative and office (ADM), 2=professional service (PS), 3=simple service (SS), 4=science and engineering (SE), 5=simple production (SP)	independent
size	number of workers (1='≤9', 2='10~49', 3='50~99', 4='100~299', 5='≥300')	independent
duration	duration periods worked in job	dependent

한편 재직기간은 연도별로 계산을 하는데, 2009년 사업에 참여한 대상을 분석하기 위해서는 2009년과 2010년의 고용보험 자료를 활용하고, 2010년에 사업에 참여한 경우에는 2010년과 2011년의 고용보험 자료를 활용하였다. 마찬가지로 2011년 사업에 참여한 경우 2011년과 2012년의 고용보험 자료를 활용하고, 2012년에 사업에 참여한 경우에는 2012년과 2013년의 고용보험 자료를 활용하였다.

또한 재직기간은 해자드 모형 (hazard model)을 활용하여 분석하는데, 해자드 분석을 하기 위해서는 서로 다른 연도에 고용된 사람들의 재직기간을 일치시키기 위하여 자료를 절단 (censoring)할 필요가 있다. 본 연구에서는 2009년에 고용된 취업자의 경우는 2011년 6월말을 기준으로 절단하고, 2010년에 고용된 취업자의 경우는 2012년 6월말을 기준으로 절단하였다. 그리고 2011년에 고용된 취업자의 경우는 2013년 6월말, 2012년에 고용된 취업자의 경우는 2014년 6월말을 기준으로 절단하였다.

분석에 사용된 조사대상자 수는 총 252,631명이고, 재직기간이 관찰된 자료는 181,298명 (71.8%)이며 절단된 자료는 71,333명 (28.2%)이다.

한편, 재직기간은 분석대상자들이 입직부터 실직이나 이직으로 인해 퇴사할 때까지 걸린 기간을 의미하며, 생존분석 (survival analysis)에서는 생존시간 (survival time)을 의미한다. 또한 입직으로부터 조사시점까지 실직이나 이직을 하지 않고 계속 근무하는 경우는 절단자료 (censored data)가 된다. 이와 같이 분석용 데이터에 절단자료와 같은 불완전 자료가 포함되는 경우 생존분석과 같은 준모수적 방법을 사용한다.

Table 2.2는 독립변수별 종속변수인 재직기간에 대한 기술통계와 각 독립변수 수준별 재직기간에 대한 생존함수 비교를 위한 로그-순위 검정 (log-rank test) 결과를 제시한 것이다.

Table 2.2의 결과에서, 통제집단에 비해서 처리집단의 재직기간이 긴 것을 알 수 있다. 또한 성별에 대해서는 재직기간이 평균값에서는 남자가 여자에 비해서 다소 길게 나타나지만, 중위수 값에서는 반대로 나타남을 알 수 있다. 그리고 연령이 많을수록 재직기간이 길고, 수도권과 충청권이 다른 지역에 비해서 재직기간이 긴 것을 알 수 있다. 그리고 전문서비스 관련직, 과학/공학 관련직 및 관리사무 관련직이 다른 직종에 비해서, 그리고 기업규모가 클수록 재직기간이 긴 것을 알 수 있다.

한편 로그-순위 검정결과에서 알 수 있듯이, 모든 독립변수들은 재직률 (생존률)에 통계적으로 유의한 차이를 보이고 있음을 알 수 있다.

Table 2.2 Descriptive statistic and log-rank test for duration

variables	categories	N	censoring rate (%)	mean	median	min	max	log-rank test
group	control	125,140	25.71	11.69	8.97	0.00	30.30	1,591.962**
	treat	127,491	30.71	13.18	12.23	0.03	30.30	
sex	female	88,358	26.18	12.39	10.53	0.00	30.30	69.895**
	male	164,273	29.34	12.47	10.37	0.00	30.30	
age	≤20	23,893	14.77	8.22	4.97	0.03	30.30	8,167.563**
	21~25	92,405	25.25	11.79	9.20	0.00	30.30	
	26~30	125,562	32.42	13.59	12.63	0.00	30.30	
	>30	10,771	35.01	14.09	13.33	0.00	30.30	
area	CAP	155,102	29.37	12.79	11.47	0.00	30.30	680.568**
	CC	19,822	28.42	12.50	10.47	0.00	30.30	
	JL	18,398	24.88	11.57	8.47	0.03	30.30	
	DK	26,642	25.41	11.54	8.43	0.03	30.30	
	DN	27,976	27.29	12.05	9.57	0.00	30.30	
	OT	4,691	24.69	11.68	9.10	0.03	30.30	
job	ADM	123,499	27.97	12.54	10.67	0.00	30.30	2,956.620**
	PS	16,557	28.52	13.27	12.17	0.03	30.30	
	SS	6,745	20.44	10.97	7.97	0.03	30.30	
	SE	71,536	32.77	13.27	12.47	0.00	30.30	
	SP	34,294	21.13	10.23	6.50	0.00	30.30	
size	≥9	71,700	15.88	10.25	7.37	0.00	30.30	8,379.382**
	10~49	83,378	29.66	12.92	12.00	0.00	30.30	
	50~99	40,486	34.43	13.49	12.82	0.00	30.30	
	100~299	43,158	36.92	13.73	13.27	0.00	30.30	
	≤300	13,909	38.45	13.80	13.63	0.03	30.30	

**p < 0.01

3. 생존나무분석

이 절에서는 독립변수들이 재직기간에 미치는 영향력을 분석하기 위해 의사결정나무분석 (decision tree analysis)을 수행하였다. 그런데 재직기간은 중도 절단자료를 포함하고 있기 때문에 회귀나무 (regression tree) 분석이 아닌 생존나무 (survival tree) 분석을 수행하였다. 생존나무분석을 위해 R 프로그램 version 3.2.2의 party 패키지를 사용하였다.

Figure 3.1은 생존나무분석 결과를 나타낸 것이다.

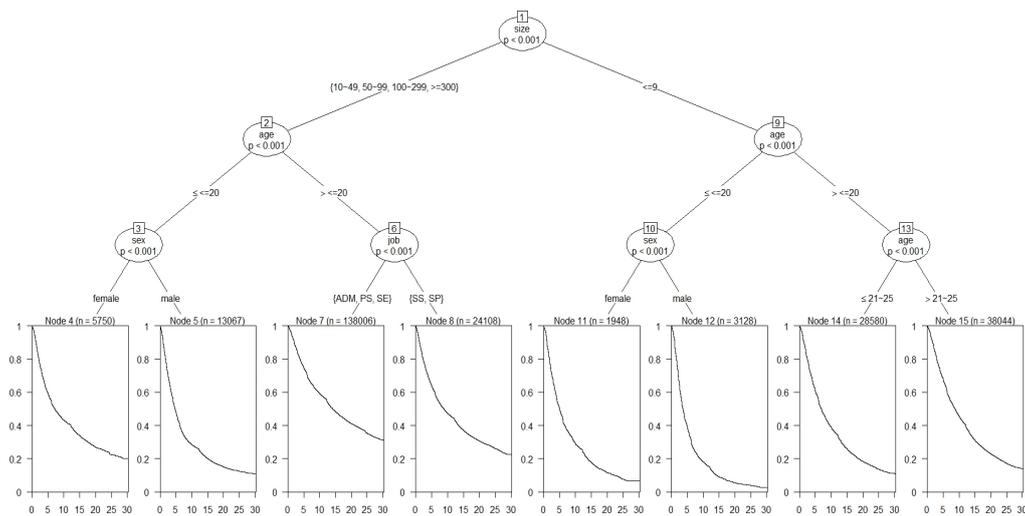


Figure 3.1 Result of survival tree analysis

Figure 3.1의 결과에서 재직할 확률 (survival rate)에 가장 많은 영향을 미치는 변수로는 기업규모이며, 그 다음으로 연령과 성별, 직종의 순으로 나타났다. 특히 시간이 지날수록 재직 (생존기간)할 확률이 완만하게 낮아지는 경우는 기업규모가 10인 이상이면서 연령이 20세 초과인 경우이며, 특히 직종이 관리, 전문서비스 및 과학/공학 관련직 (Node 7)은 재직할 확률이 가장 완만하게 낮아지는 것으로 나타났다. 또한 기업규모가 10인 이상이며, 연령이 20세 이하인 경우는 특히 여자 (Node 4)의 재직 확률이 완만하게 낮아짐을 알 수 있다. 반면, 기업규모가 9인 이하이면서 연령이 20세 이하이고 남자인 경우 (Node 12)이 재직할 확률이 가장 급격하게 감소하는 것으로 나타났다. 또한 Figure 3.1의 끝마디를 나타내는 생존함수에서 알 수 있듯이, 여자에 비해서 남자가, 관리나 전문서비스 및 과학/공학 관련직에 비해서 단순서비스나 단순생산 관련직이 연령이 낮을수록 재직할 확률이 더 급격하게 감소함을 알 수 있다.

4. 콕스의 혼합모형

이 절에서는 본 연구 분석대상자들의 재직기간 경과에 따른 실직 또는 이직으로의 탈출요인을 분석한다. 여기서 재직기간은 분석대상자들이 입직부터 실직이나 이직으로 인해 퇴사할 때까지 걸린 기간 (생존기간)이며, 조사시점까지 실직이나 이직을 하지 않고 계속 근무하는 경우 절단자료 (censored data)가 포함된다. 한편 실직 또는 이직으로의 탈출률에 영향을 미치는 독립변수들이 시간변화에 독립적이라고 가정한다. 이 경우 각각의 독립변수 값에 대해 탈출률이 비례적으로 되기 때문에 비례위험 모형 (Proportional hazard model)이라고도 불린다.

본 연구에서는 재직기간 결정요인 분석을 위해 콕스의 비례위험 모형 (Cox's proportional hazard model)을 사용한다. 일반적으로 재직기간은 취업자의 특성뿐 만 아니라, 직종에 따라 영향을 받는다. 즉 다수의 취업자들은 개인적 특성 (1-수준) 외에 동일한 직종 (2-수준)에 속하게 되어 복수의 분석단위를 가지게 되는 위계적 자료구조를 가지게 된다. 따라서 직종별로 공유되는 특성이 존재하게 되어 동일 직종 (집단) 내의 유사성과 다른 직종 간의 이질성이 발생할 수 있다.

Table 4.1은 직종 중분류 (job code)별 취업자의 재직기간에 대한 기술통계를 나타낸다.

Table 4.1 Descriptive statistic for job code (2-digit)

job code	n	observed	censoring rate	mean	median	min	max
1	47,363	33,670	28.91	12.77	11.20	0.00	30.30
2	65,992	48,112	27.09	12.38	10.17	0.03	30.30
3	2,699	1,594	40.94	15.03	16.73	0.03	30.30
4	3,193	2,135	33.13	13.99	12.17	0.03	30.30
5	677	461	31.91	14.08	13.47	0.07	30.30
6	1,469	1,046	28.80	13.32	12.47	0.13	30.30
7	1,043	830	20.42	11.93	8.97	0.10	30.30
8	10,175	7,363	27.64	13.11	12.17	0.03	30.30
9	1,784	1,396	21.75	10.85	7.50	0.03	30.30
10	7,445	5,580	25.05	11.62	8.87	0.03	30.30
11	229	201	12.23	8.32	5.03	0.03	30.30
12	3,895	3,068	21.23	11.43	9.00	0.03	30.30
13	837	701	16.25	9.85	6.47	0.03	30.30
14	4,715	3,285	30.33	13.23	12.17	0.00	30.30
15	14,328	9,754	31.92	12.71	11.07	0.00	30.30
16	4,597	3,373	26.63	11.23	7.63	0.03	30.30
17	3,199	1,986	37.92	14.15	13.80	0.03	30.30
18	551	437	20.69	10.32	7.03	0.03	30.27
19	16,969	12,637	25.53	11.45	8.37	0.00	30.30
20	25,826	15,584	39.66	15.22	16.80	0.03	30.30
21	1,351	1,036	23.32	11.03	7.97	0.00	30.30
22	33,922	26,740	21.17	10.24	6.50	0.00	30.30
23	372	309	16.94	9.65	6.08	0.17	29.87
total	252,631	181,298	28.24	12.44	10.43	0.00	30.30

Table 4.1의 결과에서 알 수 있듯이 직종별 절단비율 (censoring rate)에 상당한 차이가 있으며, 평균 값과 중위수에서도 차이가 존재하는 것을 알 수 있다. 직종 중분류별 생존함수의 동일성을 검정한 결과 통계적으로 유의하게 차이가 있음을 알 수 있다 (log-rank (Mantel-Cox)=5,130.369, $p < 0.01$).

Figure 4.1은 직종 중분류별 취업자의 재직기간에 대한 생존함수를 나타낸다. 이로부터도 직종별 생존함수에 큰 차이가 존재하는 것을 알 수 있다.

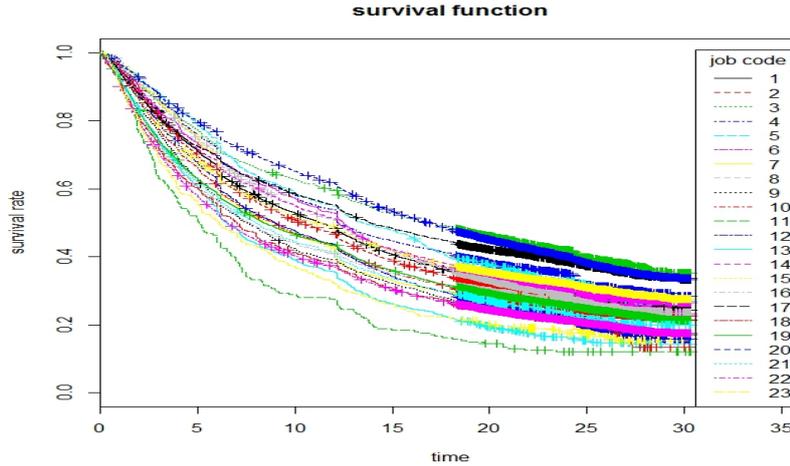


Figure 4.1 Survival function for job type

따라서 Table 4.1과 Figure 4.1의 결과를 통해서 알 수 있듯이, 직종별 특성을 고려한 위계적 자료구조를 갖는 모형으로 분석할 필요가 있다. 이 경우 콕스의 비례위험 모형 (Cox's proportional hazard model)과 같은 전통적인 통계모형은 더 이상 일치추정량을 제공하지 못한다. 따라서 보다 타당한 분석을 위해서는 고정효과 뿐만 아니라 직종특성을 반영한 랜덤효과까지 고려해야 하기 때문에, 본 연구에서는 콕스의 혼합모형 (Cox's mixed-effects model)을 이용하여 실업 또는 이직으로 탈출요인을 분석하고자 한다.

콕스의 혼합모형은 다음과 같은 과정으로 구성된다.

만약 k 개의 독립변수가 있는 콕스모형에서 j 번째 직종에 대해서 i 번째 취업자에 대한 독립변수 값이 $\mathbf{x}'_{ij} = (x_{1ij}, \dots, x_{kij})$ 이고 회귀모형 계수가 $\boldsymbol{\beta}_j = (\beta_{1j}, \dots, \beta_{kj})'$ 이라면 t 시점에서의 콕스의 비례위험 모형은 식 (4.1)로 표현된다.

$$h_{ij}(t) = h_0(t) \exp(\boldsymbol{\beta}'_j \mathbf{x}_{ij}) = h_0(t) \exp(\beta_{1j}x_{1ij} + \dots + \beta_{kj}x_{kij}). \quad (4.1)$$

식 (4.1)은 여러 독립변수들의 선형결합으로 식 (4.2)와 같이 표현할 수 있다.

$$\eta_{ij}(t) = \log \left(\frac{h_{ij}(t)}{h_0(t)} \right) = \beta_{1j}x_{1ij} + \dots + \beta_{kj}x_{kij}. \quad (4.2)$$

여기서 $h_0(t)$ 는 기저위험함수를 의미하며, 위험함수에 미치는 여러 독립변수들의 영향이 전혀 없을 경우를 가정한다. 또한 위험함수 (hazard function)인 $h_{ij}(t)$ 는 t 시점까지 재직한 j 번째 직종의 i 번째 취업자가 t 시점 바로 직후에 순간적으로 이직 또는 실직으로 탈출할 조건부 확률로 순간 위험률이라고 한다.

식 (4.2)에서 모형의 계수 $\boldsymbol{\beta}_j = (\beta_{1j}, \dots, \beta_{kj})'$ 가 고정된 상수라면 콕스의 비례위험 모형으로 분석을 하면 된다. 그러나 직종별로 $\boldsymbol{\beta}_j = (\beta_{1j}, \dots, \beta_{kj})'$ 가 변동을 가지는 확률변수가 된다면 고정효과 뿐만 아니라 랜덤효과까지 반영되는 콕스의 혼합모형을 적용한다.

따라서 본 논문에서는 (1) 1-수준의 고정효과만 고려한 모형 (Model I), (2) Model I에서 직종의 변동성만 추가한 모형 (Model II), (3) Model II에서 1-수준의 고정효과 및 2-수준에서 계수 $\beta_j = (\beta_{1j}, \dots, \beta_{kj})'$ 의 랜덤효과까지 추가한 혼합효과 모형 (Model III)을 설정하여 분석하고자 한다.

먼저 고정효과만 고려한 모형인 Model I (M_1)은 j 번째 직종의 i 번째 대상자에 대해서, 1-수준 변수만 포함한 고정효과 모형은 다음과 같이 설정한다.

$$\eta_{ij}(t) = \log \left(\frac{h_{ij}(t)}{h_0(t)} \right) = \beta_{1j}x_{1ij} + \dots + \beta_{kj}x_{kij}.$$

여기서 $\beta_{lj}(l = 1, \dots, k)$ 는 상수 값이다.

다음으로 고정효과 모형에서 직종의 랜덤효과를 추가한 Model II (M_2)는 다음과 같다.

$$\eta_{ij}(t) = \log \left(\frac{h_{ij}(t)}{h_0(t)} \right) = \beta_{1j}x_{1ij} + \dots + \beta_{kj}x_{kij} + u_j.$$

여기서 u_j 는 $N(0, \sigma_{u_j}^2)$ 를 따르며, 직종 간 효과로 전체 직종에서의 $\eta_{ij}(t)$ 차이를 의미한다. 그리고 $\sigma_{u_j}^2$ 는 u_j 의 분산으로 $\eta_{ij}(t)$ 에 대한 직종 간 분산이다.

마지막으로 Model II에서 $\beta_j = (\beta_{1j}, \dots, \beta_{kj})'$ 의 랜덤효과를 2-수준에 추가한 모형 Model III (M_3)은 다음과 같다.

$$\eta_{ij}(t) = \log \left(\frac{h_{ij}(t)}{h_0(t)} \right) = \beta_{1j}x_{1ij} + \dots + \beta_{kj}x_{kij} + u_j,$$

$$\beta_{lj} = \gamma_{l0} + u_{lj}, \quad u_{lj} \sim N(0, \sigma_{u_l}^2), \quad l = 1, \dots, k.$$

본 연구에서는 R 프로그램 version 3.2.2의 coxme 패키지를 활용하여 콕스의 혼합모형을 분석하였다. Table 2.1의 변수를 사용하여 위의 3가지 모형에 대한 콕스의 혼합모형 결과는 Table 4.2와 같다. 여기서 연령 (age)과 기업규모 (size)는 연속형 변수를 사용하였으며, Model III에서 2-수준의 랜덤효과는 집단, 성별, 연령, 기업규모 만을 고려하였다.

Table 4.2 Results of Cox's mixed model analysis

variables	Model I (M_1)		Model II (M_2)		Model III (M_3)			
	B	p-value	B	p-value	B	p-value	Exp(B)	
level-1								
Fixed effect								
group	-0.201	<0.001	-0.202	<0.001	-0.189	<0.001	0.827	
sex	0.095	<0.001	0.077	<0.001	0.106	<0.001	1.112	
age	-0.065	0.014	-0.061	<0.001	-0.059	<0.001	0.942	
area	CC	0.022	<0.001	-0.005	0.610	-0.005	0.570	0.995
	JL	0.122	<0.001	0.097	<0.001	0.093	<0.001	1.097
	DK	0.104	<0.001	0.076	<0.001	0.071	<0.001	1.074
	DN	0.069	<0.001	0.046	0.022	0.050	<0.001	1.052
	OT	0.077	<0.001	0.062	0.049	0.063	<0.001	1.065
size	-0.000	<0.001	-0.001	<0.001	-0.001	<0.001	0.999	
job	PS	-0.060	<0.001	-0.030	0.760	-0.917	0.009	0.400
	SS	0.176	<0.001	0.332	<0.001	0.328	0.390	1.388
	SE	-0.130	<0.001	0.003	0.970	-0.110	0.720	0.896
	SP	0.186	<0.001	0.331	0.009	0.122	0.790	1.129
level-2								
Random effect								
intercept			0.143			0.486		
group						0.192		
sex						0.133		
age						0.021		
size						0.001		
log-likelihood	-2,151,888		-2,149,308		-2,148,653			
χ^2	13,615**		15,175.90**		16,486.20**			
censoring rate	28.2		28.2		28.2			
n	252,631		252,631		252,631			
$H_0 : M_1$ v.s. $H_1 : M_2$				1,563.10**				
$H_0 : M_2$ v.s. $H_1 : M_3$				1,310.30**				

**p < 0.01

Table 4.2의 결과로부터 Model I (M_1)과 Model II (M_2)를 비교한 결과, Model II가 통계적으로 유의함을 알 수 있다 ($\chi^2=1,563.10, p < 0.01$). 즉 직종 랜덤효과를 추가한 모형이 고정효과만 고려한 모형보다 타당하다는 것을 의미한다. 또한 Model II (M_2)와 Model III (M_3)를 비교한 결과, Model III가 통계적으로 유의함을 알 수 있다 ($\chi^2=1,310.30, p < 0.01$). 즉 재직기간이 경과하면서 이직 또는 실직으로 탈출요인을 분석함에 있어서 직종의 랜덤효과와 개인특성변수의 랜덤효과까지 함께 고려하는 것이 타당함을 의미한다. 그리고 Model III이 통계적으로 유의할 뿐만 아니라 ($\chi^2=16,486.20, p < 0.01$), log-likelihood의 값 (-2,148,653)이 다른 모형에 비해서 절대 값이 가장 작은 것으로 나타나서, 모형의 적합도가 다른 모형에 비해서 높은 것으로 나타났다. 따라서 Model III을 중심으로 결과를 살펴보면 다음과 같은 사실을 알 수 있다.

먼저 분석 대상자들의 재직기간 중 이직 또는 실직으로 탈출확률에 미치는 영향을 살펴보면, 처리집단이 통제집단에 비해서 탈출확률이 통계적으로 유의하게 낮음을 알 수 있다. 그러나 성별로 결과를 살펴보면, 남자들이 여자들에 비해서 이직 또는 실직으로 탈출할 확률이 높은 것으로 나타났다. 그리고 연령이 많아질수록 이직 또는 실직으로 탈출할 확률이 더 낮아지는 것을 알 수 있고, 지역별 결과를 살펴보면, 충청권을 제외한 모든 지역이 수도권에 비해서 이직 또는 실직으로 탈출할 확률이 더 높은 것으로 나타났다. 또한 기업규모가 클수록 이직 또는 실직으로 탈출확률이 낮으며, 직종별로는 관리사무직에 비해서 전문 서비스 관련직에서 이직 또는 실직으로의 탈출확률이 더 낮게 나타났다.

한편 직종별 $\beta_j = (\beta_{1j}, \dots, \beta_{kj})'$ 에 대한 랜덤효과를 살펴보면, 집단에 대한 표준편차는 0.192, 성별은 0.133, 연령은 0.021, 기업규모는 0.001로 각각 나타났다.

5. 결론

본 연구에서는 청년인턴 참가자들의 재직기간에 미치는 영향을 분석하기 위해 청년인턴 DB와 고용보험 DB를 결합하여 성향점수매칭법으로 통제집단을 구성하였다. 생존나무분석 및 혼합효과를 갖는 콕스의 비례위험 모형을 활용하여 재직기간을 분석하였다. 주요 분석결과를 요약하면 다음과 같다.

첫째, 생존나무분석 결과에서 재직기간에 가장 영향을 미치는 변수는 기업규모이며, 그 다음으로 연령과 성별, 직종의 순으로 나타났다. 특히 시간이 지날수록 재직 (생존기간)할 확률이 완만하게 낮아지는 경우는 기업규모가 10인 이상이면서 연령이 20세 초과인 경우이며, 특히 직종이 관리, 전문서비스 및 과학/공학 관련직은 재직할 확률이 가장 완만하게 낮아지는 것으로 나타났다.

둘째, 기업규모가 10인 이상이며, 연령이 20세 이하인 경우는 특히 여자의 재직 확률이 완만하게 낮아짐을 알 수 있다.

셋째, 고정효과만 고려한 콕스의 비례위험 모형보다 혼합효과를 반영하는 콕스의 비례위험 모형이 통계적으로 유의한 것으로 나타났다.

넷째, 인턴 참여자 집단이 미참여자 집단에 비해서 이직이나 실직으로 탈출할 확률이 통계적으로 유의하게 낮음을 알 수 있었다. 또한 남자들이 여자들에 비해서 이직이나 실직으로 탈출할 확률이 높음을 알 수 있었다.

다섯째, 연령이 높을수록, 기업규모가 클수록 이직이나 실직으로 탈출확률이 더 낮게 나타났으며, 관리 및 사무 관련직에 비해서 전문 서비스 관련직이 이직이나 실직으로 탈출할 확률이 낮은 것으로 나타났다. 마지막으로 충청권을 제외한 모든 지역이 수도권에 비해서 이직이나 실직으로 탈출할 확률이 더 높은 것으로 나타났다.

위의 분석결과는 청년인턴 사업 참여자의 재직기간이 직종별로 랜덤효과가 존재한다는 것과 이를 고려한 통계모형을 설정해서 분석했다는 데에 의미가 있으며, 재직기간 분석뿐만 아니라, 입직할 때까지 걸리는 기간 등 다양한 분야에도 적용할 수 있다는 점에서 의의가 있다. 그러나 본 연구에서는 수집

가능한 제한된 독립변수만을 사용했다는 점과, 2-수준 변수에서 직종에 영향을 미치는 다양한 변수들을 고려하지 못했다는 점에서 연구의 한계점을 가진다. 또한 랜덤효과로 직종만을 고려했는데, 지역, 업종 등 다양한 랜덤효과를 고려한 모형에 대한 분석은 향후 과제로 남겨두고자 한다.

References

- Chang, J. E. and Ho, J. H. (2001). Change of labor market and re-employment of women after the economic crisis. *Korean Sociological Association*, **35**, 159-188.
- Cho, J. S. (2013). Determinants of student course evaluation using hierarchical linear model. *Journal of the Korean Data & Information Science Society*, **24**, 1-12.
- Cho, J. S. (2015). Determinants of the number of job experience and duration to first job in college. *Journal of the Korean Data Analysis Society*, **17**, 183-194.
- Joo, M. H., Kim, D. S., Kim, B. W., Choi, J. I. and Kim, J. H. (2013). *The study on the employment effect of the youth-intern project*, Korea Employment Information Service, Seoul.
- Jun, H. J., Ryu, J. S., Cho, J. S., Park, S. I. and Kim, J. H. (2009). An analysis on the employment determinants of university graduates in NURI project teams-the case of the universities in Busan region. *Korean Journal of Labor Economics*, **32**, 31-56.
- Jung, D. I. (2012). Cox model with non-proportional hazards and its application to organization studies. *Journal of the Korean Data Analysis Society*, **14**, 657-668.
- Kim, H. I. and Jun, H. J. (2010). Studying on parents' satisfactory factor to elementary school which their children go to - focusing on Anyang city. *Journal of the Korean Data & Information Science Society*, **21**, 1009-1020.
- Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, **26**, 646-679.
- Lee, K. Y. (2009). *The evaluation of the youth-intern project*, The monthly labor review, The Korea Labor Institute, Seoul.
- Nasser, F. and Hagtvet, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, **47**, 559-590.
- Park, J. K., Oh, K. H. and Kim, M. S. (2012). Survival analysis on the business types of small business using Cox's proportional hazard regression model. *Journal of the Korean Data & Information Science Society*, **23**, 257-269.
- 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.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models : Applications and data analysis methods*, SAGE publication, San Francisco.
- Ryu, J. S., Park, S. I., Cho, J. S., Kim, J. H., Ha, B. C. and Kim, J. H. (2012). *The current state and the employment effect of the youth-intern project*, Human Resource Development Institute at Pukyong National University, Busan.
- Ryu, J. S., Park, S. I., Cho, J. S. and Jung, H. J. (2014). *The performance evaluation of the youth-intern project*, Human Resource Development Institute at Pukyong National University, Busan.
- Song, H. Y. (2005). *Survival analysis*, Cheungmungak, Seoul.
- Yang, W. Y. and Ha, H. T. (2009). Analyze cervical cancer relapse using Cox proportional hazards model. *Journal of the Korean Data Analysis Society*, **11**, 2291-2301.

Study on the determinants of employment duration in the youth-intern project

Sungik Park¹ · Jangsoo Ryu² · Jonghan Kim³ · Jangsik Cho⁴

¹International Trade and Commerce, Kyungsoong University

²Division of Economics, Pukyong National University

³Division of Economics, Finance and Logistics, Kyungsoong University

⁴Department of Informational Statistics, Kyungsoong University

Received 11 January 2016, revised 19 January 2016, accepted 20 February 2016

Abstract

In general, employment duration is influenced by the individual characteristics (level-1) as well as type of the occupational characteristics (level-2). That is, the data has hierarchical structure in the sense that individual employment duration is influenced by the individual-level variables (level-1) and the job-level (level-2) variables. In this paper, we study the determinants of the employment duration of youth-intern in the SMEs (small and medium enterprises) using Cox's mixed effect model. Major results at level-1 variables are as followings. First, the hazard rate of treatment group is lower than that of control group. Second, the hazard rate of woman is lower than that of man. Also, the hazard rate is lower, for the older and the workers working in the bigger company. Investigation of level-2 variables has shown that random effect for job-level is statistically significant..

Keywords: Cox's proportional hazard model, hierarchical data structure, mixed effect model, random effect, survival tree analysis.

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

² Professor, Division of Economics, Pukyong National University Busan 48513, Korea.

³ Professor, Division of Economics, Finance and Logistics, Kyungsoong University, Busan 48434, Korea.

⁴ Corresponding author: Professor, Department of Informational Statistics, Kyungsoong University, Busan 48434, Korea. E-mail: jscho@ks.ac.kr