

기업 고용의 성장과 지브라의 법칙*

윤 종 인**

논문 초록

본 연구는 지브라의 법칙을 검정함으로써 우리나라 기업의 고용 증가에 대해 분석하였다. 기업의 성장과 규모의 관계를 분석하는 실증연구에서 표본선택편의는 실증분석결과에 큰 영향을 미치리라고 판단하였다. 이에 본 연구는 probit survival 방정식을 동시에 추정하는 표본선택모형을 이용하였다. 즉 Heckman의 모형과 분위회귀분석모형인 Arellano and Bonhomme(2017)을 이용하였다. Heckman 모형을 이용한 결과에 따르면 영세기업을 다수 포함한 검정결과에서는 대부분의 경우 지브라의 법칙이 기각되었다. 하지만 일정 수준(40% 분위값)보다 규모가 큰 기업들로 이루어진 표본에서는 지브라의 법칙이 기각되지 않는 결과가 더 많았고, 80% 분위값보다 큰 기업들로 이루어진 표본에서는 고용 규모가 큰 기업이 작은 기업보다 더 빠르게 성장하였던 것으로 나타났다. 한편 분위회귀분석 모형을 이용한 결과에 따르면 중위 기업의 고용 증가가 가장 컸다. 또한 최근 지브라의 법칙에 가까운 방향으로 변화가 나타나고 있는데, 이는 우리나라 기업 고용의 분포가 지프의 법칙이 성립하는 방향으로 변화하고 있음을 설명할 수 있다.

핵심 주제어: 지브라의 법칙, 표본선택편의(생존자편의), 표본선택모형, 표본선택 분위회귀분석, 지프의 법칙

경제학문헌목록 주제분류: D21, L22, L25, L26, M13

투고 일자: 2018. 10. 10. 심사 및 수정 일자: 2019. 2. 19. 게재 확정 일자: 2019. 3. 22.

* 본 논문을 심사해 주신 익명의 심사위원께 감사드리며, 여전히 남아있는 한계는 필자의 책임임을 밝혀둔다. 이 논문은 2019년도 백석대학교 대학연구비에 의하여 수행된 것임.

** 백석대학교 경상학부 부교수, e-mail: jiyeon@bu.ac.kr

I. 문제제기

기업의 성장은 여러 요인에 의해 영향을 받는다. 그중에서도 기업의 성장과 규모의 관계는 많은 관심을 끌어왔다. 지브라의 법칙(Gibrat's law)은 기업의 성장이 규모와는 독립적이며 기업의 성장은 순수하게 무작위적인(random) 현상이라고 말한다.

물론 기업의 규모는 여러 방식으로 정의할 수 있다. 예를 들어 자산규모, 매출액, 종업원수 등에 의해 대기업과 중소기업을 구분하곤 한다. 본 연구는 종업원수, 즉 고용 규모에 의한 구분에 관심을 갖는다. 따라서 지브라의 법칙이 성립한다면 기업 고용의 성장은 기존의 고용 규모와는 독립적이며 무작위적이다. 즉 대기업, 중기업, 소기업의 고용 증가는 확실적인 차이만 있을 뿐이며 기존의 고용 규모에 따른 차이가 없다는 것이다.

고용은 그 자체로 정책목표가 되어야 할 중요한 사안이기 때문에 지브라의 법칙이 성립하는가 여부는 정책적으로 중요한 시사점을 갖는다. 예를 들어 김두순(2016)은 지브라의 법칙과 중소기업정책을 연관시켰다. 만약 지브라의 법칙이 성립한다면 중소기업을 우선시하여 지원하는 정책은 고용 촉진의 측면에서 신중하게 재검토되어야 할지 모른다.

최초 Robert Gibrat의 논문은 1931년 제출되었으며, 이에 대한 실증연구 중 고전적인 것은 1962년 Edwin Mansfield의 논문이다. Mansfield(1962)는 지브라의 법칙을 지지하는 실증분석결과를 제시했다. 하지만 이후 이어진 연구들은 꽤 논쟁적이다.

대표적으로 Evans(1987)는 지브라의 법칙이 성립하지 않는다는 연구결과를 제시했다. 즉 그는 기업의 성장과 규모는 (-)의 관계(inverse growth-size relationship)를 가지며, 기업의 성장과 연령도 (-)의 관계(inverse growth-age relationship)를 갖는다는 결과를 얻었다. 이에 따라 전자는 지브라의 법칙과 불일치하는 것으로 후자는 Boyan Jovanovic의 기업이론과 일치하는 것으로 보았다. 그의 결론은 이후 연구에서 하나의 전형으로 받아들여진다.

비슷한 시기에 Hall(1987)은 조금 다른 결론을 얻었다. 규모가 작은 기업들로 이루어진 표본에서는 지브라의 법칙이 성립한다는 결과를 얻지 못했지만 규모가 큰 기업들로 이루어진 표본에서는 지브라의 법칙이 성립한다는 결과를 얻었다.

Hall (1987)의 실증분석결과도 이후 연구에서 반복되어 제시되어 왔고 이로 인해 분위회귀분석이 널리 이용되고 있다.

최근에도 논쟁은 진행 중이다. 최근의 것만 소개하면 다음과 같다. Stam (2010)은 이전의 연구성과를 정리하면서 지브라의 법칙이 성립하지 않는다는 연구결과가 많았음을 지적하였다. 또한 Neumark (2011)는 소규모 기업이 더 많은 일자리를 창출한다는 결과를 얻었으며, 이는 제조업과 서비스업 모두에서 확인된 것으로 산업별 차이는 없다고 보았다. 하지만 지브라의 법칙을 지지하는 결과도 적지 않다. Haltiwanger et al. (2013)는 기업의 연령을 통제한다면 기업의 성장과 규모 사이에는 체계적인 관계가 없다는 결과를 제시하였다. 심지어 Bentzen et al. (2012)는 기업의 성장이 규모와 (+)의 관계를 갖는다는 결과를 제시하기도 했다.

지브라 법칙을 지지하는 학자들도 기업의 연령이 중요하다는 점에 대해서는 동의하는 것으로 보인다. 예를 들어 Daunfeldt and Halvarsson (2015)는 한 번의 경이적인 성장(one-hit wonders)은 지속되지 않는다고 말했는데, 이는 기업의 성장과 연령이 (-)의 관계를 가짐을 의미한다. Lawless (2014)는 기업의 성장에 중요한 요인은 기업의 규모가 아니라 기업의 연령이라는 결과를 얻었다. 즉 기업의 성장은 규모와 무관하다는 것, 그리고 젊은 기업은 다이내믹하기 때문에 성장률이 높지만 일정한 연령에 이르게 되면 성장률은 낮아진다는 것이다. 그의 결과는 Jovanovic 가설과 지브라의 법칙을 모두 지지하는 것이다.

국내의 연구는 대체로 Evans (1987)와 비슷한 결과를 얻었던 것으로 보인다. 즉 Gibrat의 법칙은 성립하지 않으며, 기업의 성장과 연령은 (-)의 관계를 갖는다는 Jovanovic 가설을 지지한다는 것이다. 이인권 (2002)은 2000년 이전의 자료를 이용하여 이러한 결과를 얻었으며, 박현숙·김주한 (2013), 성낙일·김민창 (2013), 윤충한·손종철 (2014)은 비교적 최근의 자료를 이용하여 비슷한 결과를 얻었다. 분위회귀분석을 이용한 연구도 있었는데, 표한형·홍성철 (2014)과 김두순 (2016)도 지브라의 법칙에 대해서 비판적인 결과를 제시하였다. 다만 김두순 (2016)은 일부 분위에서 지브라의 법칙에 가깝다고 볼 수 있는 결과를 얻었다.

기존의 국내외 연구결과를 요약하면 다음과 같다. 첫째 기업의 성장과 연령이 (-)의 관계를 갖는다는 Jovanovic 가설에는 상당한 합의가 형성된 것으로 보인다. 둘째 기업의 성장과 규모가 무관하다는 지브라의 법칙에 대해서는 여전히 논쟁 중이다. 다만 규모가 큰 기업들로 이루어진 표본에서는 지브라의 법칙이 성립한다는

연구결과가 있었다.

본 연구는 국내에서 지브라의 법칙이 성립하는가를 검정하고자 한다. 그 이유 중 하나는 이른바 지프의 법칙(Zipf's law)에 가까운 방향으로 기업 규모의 분포가 변해왔다고 판단하기 때문이다. 지프의 법칙은 우측 꼬리가 두터운 분포에서 성립한다. 우측 꼬리가 두터운 분포로 파레토분포(Pareto distribution)가 잘 알려져 있는데, 이 분포에서 지프승수(Zipf exponent)가 1이면 지프의 법칙이 성립한다고 말한다. 또한 지프승수가 1보다 크면 우측 꼬리는 덜 두터우며, 지프승수가 1보다 작으면 우측 꼬리는 매우 두텁다. 파레토분포를 이용하여 소득집중도(top income share)를 추정하였던 김낙년(2012)에서 지프승수는 1보다 훨씬 컸다.

Gabaix(2009)는 기업의 규모, 도시의 크기, 소득과 부, CEO의 연봉이 지프의 법칙을 따르는 잘 알려진 예이고, 이에 더하여 주식가격의 움직임과 거래량, 수출업자, 규제 등도 이 법칙을 따른다는 연구결과를 소개하고 있다. 기업의 규모는 지프의 법칙을 따른다고 알려진 가장 대표적인 변수인데, 국내에서도 최희갑(2006)은 2004년 자료를 이용하여 기업 규모에서 지프의 법칙이 성립함을 보인 바 있다. 이 하에서 제시하게 될 것인데, 흥미로운 것은 우리나라 기업 규모의 분포에서 지프승수의 추정치가 꾸준히 하락하여 1에 접근하고 있다는 점이다.

기업 규모의 분포가 지프의 법칙을 따르는 이유로는 여러 가지가 제시되어 왔다. 가장 기본적인 설명은 기업 규모의 성장이 지브라의 법칙을 따른다는 것이다. 이에 대해 Gabaix(2009)는 기업 규모의 성장이 지브라의 법칙을 따르면, 기업 규모의 분포는 우측 꼬리가 두터운 자연대수 정규분포(lognormal distribution)를 따르게 됨을 보였다. 또한 지브라의 법칙과 함께 마찰(friction)이 존재한다는 가정이 추가되면, 기업 규모의 분포는 지프의 법칙을 따르게 됨을 보였다. Gabaix(2009)가 말하는 마찰이란 '최소한의 고용인원이 존재한다'와 같은 것이므로 자명한 가정이다. 따라서 마찰의 존재를 자명한 것으로 받아들이고 기업 규모의 성장에서 지브라의 법칙이 성립하면 기업 규모의 분포에서 지프의 법칙이 성립한다.

본 연구는 기업 규모의 분포에서 지프의 법칙이 성립하는 방향으로 변화가 나타나고 있는 이유가 기업 규모의 성장에서 지브라의 법칙이 성립하였기 때문은 아닐까 추측한다. 이를 위하여 본 연구가 방법론적으로 주목하는 것은 표본선택편의(sample selection bias)이다. 개별 기업의 자료를 이용하는 연구에서 사망기업을 제외하고 생존기업만을 이용하는 경우가 많다. 사망기업의 규모는 결국 0이 될 것인

데, 규모가 작은 기업의 사망률이 높다면 사망기업을 제외한 연구는 규모가 작은 기업의 성장을 과대평가하게 된다. 이것이 표본선택편의이다. 본 연구는 이를 다루기 위한 방법으로 표본선택모형(sample selection model)을 이용하였다.

지브라의 법칙에 초점을 맞추어 본 연구의 결과를 간략히 요약하면 다음과 같다. 첫째 규모가 작은 기업의 사망가능성이 높았던 것으로 나타났다. 둘째 사망기업을 제외하고 생존기업만을 이용한 회귀분석은 모형설정오류(specification error)를 범하는 것으로 나타났다. 셋째 규모가 일정 수준(예를 들어 48명) 이상 기업으로 이루어진 표본에서는 지브라의 법칙을 지지하는 결과가 많았다. 넷째 분위회귀분석결과도 고려할 때 최근 지브라의 법칙을 향한 변화가 나타나고 있는 것으로 보인다. 다섯째 국내 여러 실증연구와 마찬가지로 Jovanovic 가설을 지지하는 결과를 얻었다. 끝으로 최근 우리나라 기업 고용 규모의 분포는 지프의 법칙에 가까운 방향으로 변화하여 왔는데, 이것과 지브라의 법칙을 향한 변화와 관련이 있을 것으로 보인다.

이하의 논의 순서는 다음과 같다. 제Ⅱ장에서는 자료와 연구방법에 대해 설명한다. 본 연구는 통계청 자료와 KIS vlaue 자료를 모두 이용하였다. 또한 지브라의 법칙과 지프의 법칙을 검정하는 방법 그리고 표본선택모형에 대해 설명한다. 제Ⅲ장에서는 지브라의 법칙에 대한 추정 및 검정결과를 제시하되 분위별 추정 및 검정결과도 제시한다. 제Ⅳ장에서는 표본선택 하의 분위회귀분석모형을 통해 분위회귀분석을 수행한 이전 국내 연구결과와 비교한다. 제Ⅴ장에서는 지프의 법칙에 대한 추정결과를 제시한다. 제Ⅵ장에서는 결론을 맺고 시사점을 제시한다.

Ⅱ. 자료 및 연구방법

1. 자 료

기업의 규모를 정의하는 방식에는 여러 가지가 있지만 본 연구는 종업원수를 규모라고 정의한다. 두 가지 자료를 이용하였는데, 첫째는 통계청(MDIS)이 제공하는 기업활동조사 패널의 2006~2016년 자료이고, 둘째는 KIS value가 제공하는 1980~2017년 패널자료이다.

〈Table 1〉에는 두 자료의 고용 규모별 기업체 수가 제시되어 있다. 통계청 자료의 대상은 회사법인 중 상용근로자가 50인 이상, 자본금이 3억 원 이상인 기업이지

〈Table 1〉 Numbers of firms by employees

year	MDIS					KIS				
	~50	51 ~300	301 ~500	501~	sum	~50	51 ~300	301 ~500	501~	sum
1980						1	18	20	184	223
1981						1	21	17	184	223
1982						3	64	25	214	306
1983						15	144	46	279	484
1984						92	572	137	418	1,219
1985						118	702	181	442	1,443
1986						161	813	204	461	1,639
1987						194	955	228	499	1,876
1988						190	1,082	267	530	2,069
1989						228	1,173	269	540	2,210
1990						308	1,300	269	542	2,419
1991						433	1,472	276	548	2,729
1992						438	1,563	256	554	2,811
1993						443	1,570	258	550	2,821
1994						570	1,808	287	550	3,215
1995						1,224	2,182	318	577	4,301
1996						1,740	2,355	334	582	5,011
1997						2,555	2,709	320	569	6,153
1998						3,080	2,840	274	479	6,673
1999						3,564	3,193	321	516	7,594
2000						3,914	3,476	350	545	8,285
2001						4,165	3,877	390	566	8,998
2002						5,063	4,257	417	589	10,326
2003						5,693	4,583	439	625	11,340
2004						5,617	4,777	440	634	11,468
2005						5,603	4,993	420	675	11,691
2006	424	8,728	623	917	10,692	6,034	5,309	476	674	12,493
2007	631	8,413	698	922	10,664	6,599	5,594	483	696	13,372
2008	498	8,699	730	929	10,856	7,130	5,841	533	690	14,194
2009	676	8,415	755	957	10,803	7,610	6,100	533	703	14,946
2010	1,241	7,960	766	1,004	10,971	7,584	6,365	576	765	15,290
2011	1,041	8,769	773	1,056	11,639	7,768	6,685	610	822	15,885
2012	1,061	8,913	825	1,130	11,929	8,468	7,148	645	843	17,104
2013	1,333	8,792	893	1,123	12,141	9,117	7,406	643	786	17,952
2014	1,572	8,693	888	1,177	12,330	9,686	7,967	691	903	19,247
2015	1,750	8,530	916	1,190	12,386	9,990	8,156	727	947	19,820
2016	1,327	8,652	1,042	1,371	12,392	10,868	8,561	790	977	21,196
2017						12,522	8,908	796	1,032	23,258

만 소매업 및 기타 서비스업은 상용근로자가 50인 미만, 자본금 10억 원 이상인 기업을 포함한다. 따라서 50인 미만 기업은 50인 이상 기업과 매우 다른 특성을 가질 수 있다. 다만 표본기간 중 고용 규모별 기업의 비중은 안정적인 편이다. 반면에 KIS value 자료의 경우 고용 규모별 기업의 비중에는 큰 변동이 있었다. 고용규모가 50명 이하 기업의 비중은 1990년대 중반까지 매우 적었으나 최근에는 50%를 넘고 있으며, 51~300명인 기업의 비중은 40%, 301~500명인 기업의 비중은 3~4%, 끝으로 500명 이상인 기업의 비중은 최근 4~5%이다. KIS value 데이터베이스가 상장기업 자료 중심에서 외감기업 또는 일반기업으로 확대되어 왔기 때문인 것으로 보인다.

MDIS 자료의 경우 표본기간이 짧고 얻을 수 있는 설명변수가 많지 않다는 것이 단점이다. KIS value 자료의 경우 표본기간이 길고 얻을 수 있는 설명변수가 많다는 점이 장점이다. 다만 이 자료는 모집단을 추적할 수 있도록 표본이 선택된 것이 아님을 고려해야 한다. 특히 최근에는 고용 규모가 영세한 기업의 비중이 너무 커졌는데, 이를 고려하여 종업원수 10인 이하의 기업은 표본에서 제외하였다. 또한 1980년대 자료의 경우 표본이 빈약하기 때문에 추정결과를 얻기가 쉽지는 않았다. 따라서 실질적으로는 1987년 이후 자료를 주로 이용하였다.

추정방정식에 포함된 설명변수는 다음과 같이 구하였다. 우선 MDIS 자료를 이용한 경우 설명변수는 다음과 같다. 산업별 차이를 반영하기 위하여 산업더미변수 4가지를 이용하였다. 첫째 농림, 임업 및 어업의 경우 industry-agriculture, 둘째 광업, 제조업, 전기, 가스, 증기 및 수도사업, 하수·폐기물 처리, 원료재생 및 환경복원업의 경우 industry-manufacturing, 셋째 건설업의 경우 industry-construction, 넷째 도소매업의 경우 industry-wholesale/retail의 더미변수를 이용하였다. 따라서 벤치마크가 되는 산업은 운수업, 숙박 및 음식점업, 출판, 영상, 방송통신 및 정보서비스업, 부동산업 및 임대업 등으로 크게 보면 서비스산업이라고 볼 수 있을 것이다. 또한 관계회사 여부를 나타내는 더미변수 relational은 관계회사가 있는 경우 1이고 그렇지 않은 경우 0이다. 한편 IPR/asset은 총자산 대비 지적재산권 보유건수인데, MDIS는 특허권, 실용신안권, 디자인권, 상표권 등 지적재산권의 보유건수를 제공한다. 끝으로 재무적 안정성을 나타내는 debt/asset는 총자산 대비 부채이다.

다음으로 KIS value 자료를 이용한 경우 설명변수는 다음과 같이 구하였다. 우선

산업더미변수와 관계회사 여부를 나타내는 relational는 MDIS 자료와 같은 방식을 이용하였다. listed는 유가증권시장 및 코스닥시장에 상장된 경우 1의 값을 갖는 더미변수이다. 기업의 연령을 나타내는 변수 age는 회사 설립 이후 연수를 이용하였다. IPR/asset은 총자산 대비 지적 재산권인데, 지적재산권은 재무상태표에 제공된 무형자산 중 산업재산권과 개발비의 합을 이용하였다. R&D/sales은 매출액 대비 연구개발비이며, 연구개발비는 손익계산서에 제공된 판매 및 일반관리비 중 연구비, 경상연구개발비, 경상개발비, 개발상각의 합을 이용하였다. 재무적 안정성을 나타내는 변수는 다음과 같이 구하였다. 우선 debt/asset은 총자산 대비 부채이고, currentasset/asset은 총자산 대비 유동자산, sales/asset은 총자산 대비 매출액(총자산회전율), EBIT/asset은 총자산 대비 영업이익, retained earnings/asset은 총자산 대비 유보이익(이익잉여금)이다. 끝으로 export/sales은 매출액에서 수출액이 차지하는 비중이다.

2. 실증분석방법

본 연구는 몇 가지 점에서 기존 국내의 연구와 구분된다. 첫째 표본선택편의를 고려한 추정방법을 이용하였다. 둘째 5-6년간의 기업 고용의 성장을 분석하였다. 셋째 지브라의 법칙과 지프의 법칙을 연관시키려 하였다.

가장 먼저 표본선택편의를 고려하여 표본선택모형(sample selction model)을 이용하였다. 보통의 경우 지프라의 법칙을 검정하기 위하여 이용하는 방정식은 다음의 (1)식과 같다.

$$\log(S_{iT}) = \beta_0 + \beta_1 \log(S_{i0}) + \beta_2 X_i + \epsilon_i \quad (1)$$

(1)식에서 $\log(S_{iT})$ 는 i 기업의 T 기 고용의 로그값이며 $\log(S_{i0})$ 는 i 기업의 0기 고용의 로그값이다. 본 연구는 0기와 T 기의 간격을 6년으로 설정하였다. 이 방법은 해외 연구에서 널리 이용하는 방법으로 기업 고용의 1차 자기상관(autocorrelation)에 따른 문제점 등을 피할 수 있도록 해 준다. 한편 X_i 는 기업 고용의 성장을 설명할 수 있는 여러 변수들이다.

지브라의 법칙을 검정한다는 것은 $\log(S_{i0})$ 의 추정계수인 β_1 이 1인가를 검정하는

것이다. β_1 이 1이면 지브라의 법칙이 성립하며, 1보다 작으면 고용 규모가 작은 기업의 고용성장이 더 크다는 뜻이며, 1보다 크면 고용 규모가 큰 기업의 고용성장이 더 크다는 뜻이다.

(1) 식만을 추정한다면 표본선택편의가 발생할 수 있다. 생존자란 T 기에 생존한 기업을 말하는데, (1) 식만의 추정은 생존자만을 이용한다는 뜻이다. 따라서 0기에는 존재하였으나 T 기에는 생존하지 못하고 사망한 기업은 표본에서 제외된다. 이에 따라 표본선택편의가 발생하게 되는데, 그 이유는 명확하다. 사망한 기업을 제외할 경우 생존기업의 고용 증가는 전반적으로 과대평가된다. 게다가 고용 규모가 작은 기업일수록 사망 가능성이 높다면, 사망기업을 제외할 경우 고용 규모가 작은 기업의 고용성장은 과대추정될 수밖에 없다.

이와 같이 특정의 표본이 제외됨으로써 발생하는 편의를 일반적으로 표본선택편의라고 부른다. 또한 생존자만이 표본에 포함된다는 의미에서 생존자편의(survivor bias)라고 부르기로 한다. 표본선택편의를 고려하기 위하여 표본선택모형을 이용한다(Wooldridge, 2002, chap 17). 표본선택모형은 아래와 같다.

$$\begin{aligned}
 z_i^* &= \gamma W_i + u_i, \text{ where } \sigma_u^2 = \text{var}(u_i) \\
 z_i &= 1 \text{ if } z_i^* > 0 \\
 &= 0 \text{ if } z_i^* \leq 0 \\
 \Pr(z_i = 1 | W_i) &= \Phi(\gamma W_i) \\
 \log(S_{iT}) &= \beta_0 + \beta_1 \log(S_{i0}) + \beta_2 X_i + \rho \sigma_\epsilon \lambda(\gamma W_i) + \epsilon_i, \\
 \text{where } \sigma_\epsilon^2 &= \text{var}(\epsilon_i), \quad \rho = \text{corr}(u_i, \epsilon_i), \\
 \lambda(\gamma W_i) &= \phi(\gamma W_i / \sigma_u) / \Phi(\gamma W_i / \sigma_u)
 \end{aligned} \tag{2a}$$

우선 (2a) 식을 선택방정식(Selection equation)이라 부르고 (2c) 식을 결과방정식(Outcome equation)이라 부른다. 선택방정식인 (2a) 식은 Probit 모형으로 기업의 생존 여부를 설명하는 방정식이다. W_i 는 i 기업의 생존에 영향을 미치는 설명변수이고, 이에 따라 z_i^* 가 결정된다. z_i^* 가 0보다 크면 z_i 가 1이 되며, 이는 i 기업이 T 기에 생존한다는 뜻이다. 반면에 z_i^* 가 0과 같거나 작으면 z_i 가 0이 되며, 이는 i 기업이 T 기에 사망한다는 뜻이다. 결국 정규분포를 가정할 경우 기업의 생존여

부는 (2b) 식 $\Phi(\gamma W_i)$ 에 의해 표현할 수 있다. 여기에서 $\Phi(\cdot)$ 는 표준정규분포의 누적분포함수이다. 한편 결과방정식인 (2c) 식은 위의 (1) 식에 대응하는 것이다. 다른 점이 있다면 $\lambda(\gamma W_i)$ 가 설명변수로 포함되어 있다는 것이다. $\lambda(\gamma W_i)$ 은 (2a) 식의 추정결과부터 얻게 되는데, $\phi(\cdot)$ 는 정규분포의 확률밀도함수이다. 흔히 $\lambda(\gamma W_i)$ 를 inverse Mill's ratio라고 부르는데, 이를 설명변수로 포함하는 것이 표본선택모형의 핵심이다. 즉 표본선택편의가 있다면 $\lambda(\gamma W_i)$ 의 추정계수는 유의하게 되며, $\lambda(\gamma W_i)$ 의 추정계수가 유의하다면 표본선택편의가 있다고 판단한다. 이제 (1) 식만을 추정하는 방법의 오류를 쉽게 알 수 있다. 즉 설명변수로 $\lambda(\gamma W_i)$ 을 빠뜨린 것이며 이로 인해 모형설정오류(specification error)를 범하게 된다.

추정방법으로는 Heckman의 2단계 추정(2-step regression)과 완전정보최우추정법(FIML: full information maximum likelihood)이 있다. Heckman의 2단계 추정은 (2a) 식을 추정하여 (2b) 식에서 inverse Mill's ratio를 구한 후 이를 (2c)의 회귀분석에서 설명변수로 이용하는 것이고, FIML은 (2) 식 전체를 추정하는 것이다. Davidson and MacKinnon(2004)은 Heckman의 절차가 inverse Mill's ratio의 추정계수가 유의한가를 검정하는데 유용하지만 FIML이 더 효율적(efficient)임을 보였다. 따라서 본 연구는 두 가지 방법에 의한 추정결과를 모두 제시할 것이지만, 표본선택편의가 있는가를 검정할 때 Heckman의 2단계 추정결과를 이용하고 지브라의 법칙을 검정할 때에는 FIML에 의한 추정결과를 이용한다.

실증분석에서 중요한 내용은 표본선택편의가 있었는가를 검정하는 것과 지브라의 법칙이 성립하는가를 검정하는 것이다. 표본선택편의가 존재한다는 귀무가설은 inverse Mill's ratio의 추정계수가 유의한가에 달려있다. 즉 inverse Mill's ratio의 추정계수가 유의하면 표본선택편의가 존재한다는 귀무가설을 기각할 수 없다. 이하에서는 inverse Mill's ratio를 IMR이라고 약칭한다. 또한 지브라 법칙의 검정과 (2) 식에서 $\log(S_{it})$ 의 추정계수인 β_1 이 1인가의 검정은 같다. 즉 β_1 이 1이면 지브라의 법칙이 성립한다고 판단한다. β_1 이 1이라는 귀무가설을 검정하기 위하여 본 연구는 월드검정(Wald test)을 이용하였다.

기업 고용의 성장에 영향을 미치는 요인으로는 혁신활동(innovating activity), 기업의 연령, 하이테크 산업 등장으로 인한 산업부문 간 차이, 주주의 유한책임(limited legal liability) 정도 등이 알려져 있다(Calvo, 2006). 물론 이 중에서 기업

의 연령은 Jovanovic 가설과 관련하여 중요한 변수이다. 국내 연구에서는 이에 더하여 상장 여부, 관계회사 존재 여부, 재무적 안정성, 수출의존도 등을 설명변수로 이용한 바 있다(김두순, 2016).

결과방정식의 설명변수 X_i 는 다음과 같다. MDIS 자료의 경우 혁신활동의 영향 및 하이테크 산업의 차이를 반영하기 위하여 지적 재산권과 연구개발비, 산업더미 변수를 설명변수로 이용하였다. 기업의 법적 형태가 미치는 영향을 반영하기 위하여 상장 여부, 관계회사 존재 여부를 나타내는 설명변수를 이용하였고, 재무적 안정성이 미치는 영향을 반영하기 위하여 총자산 대비 부채 및 유동자산의 비중을 설명변수로 이용하였다. KIS value 자료의 경우 더 많은 자료를 이용할 수 있었는데, 기업의 연령과 수출의존도를 설명변수로 추가 이용하였다. 다음으로 선택방정식의 설명변수 W_i 는 다음과 같다. 우선 결과방정식에 포함된 설명변수를 모두 이용하였으며 이에 추가하여 총자산 대비 매출액, 영업이익, 유보이윤을 포함하였다. 이 변수들은 기업의 파산확률 추정을 위해 Altman's Z-score를 계산할 때 이용하는 변수들이다.

본 연구는 기업 고용의 성장이 지브라의 법칙을 따르기를 검정한 후 이어서 기업 고용의 분포가 지프의 법칙을 따르는가를 검정하고자 한다.

Gabaix (2009)는 지프의 법칙(Zipf's law)을 따르는 대표적인 변수로 기업의 규모를 꼽았고, 그 이유 중 첫 번째로 지브라의 법칙을 들었다. 즉 기업 규모의 성장이 지브라의 법칙을 따르고 마찰(friction)이 존재한다면 기업 규모의 분포는 지프의 법칙을 따르게 됨을 보였다.

지프의 법칙은 우측 꼬리(right tail)가 두터운 분포에서 성립한다. 우측 꼬리가 두터운 분포로 가장 유명한 것은 소득분배 등에서 널리 이용되는 파레토(Pareto) 분포이다. Gabaix (2009)는 이를 멱법칙분포(power law distribution)라고 부른다.

예를 들면, 기업의 규모인 *size*의 멱법칙분포는 아래와 같이 표현할 수 있다.

$$\Pr[size > x] = \alpha x^{-\beta} \quad (3)$$

즉 *size*가 x 보다 클 확률이 지수(exponent) β 에 의해 결정됨을 알 수 있는데, β 가 작을수록 우측 꼬리는 두텁다. 그리고 β 가 1일 때 지프의 법칙이 성립한다고 말한다.

Gabaix (2009)는 멱법칙분포의 추정방법을 제시하였다. i 를 기업 규모에 따른 순위(rank)라고 하면 i 의 로그값을 기업 규모를 나타내는 S_i 의 로그값에 대해 회귀분석하여 구한다.

$$\ln(i - 0.5) = a - b \ln(S_i) + e_i \quad (4)$$

Gabaix (2009)는 종속변수로 $\ln(i)$ 가 아니라 $\ln(i - 0.5)$ 를 이용할 것을 제안하였는데, 이는 소표본편의(small sample bias)를 줄이기 위한 것이다. 한편 그는 추정치 \hat{b} 의 점근적 표준오차(asymptotic standard error)가 $\hat{b}(n/2)^{-1/2}$ 임을 보인 바 있다. 끝으로 지프의 법칙을 검정하기 위해서 상위 5%의 기업만을 이용하였다. 이에 따른 차이는 없어 보인다.

Ⅲ. 지브라의 법칙에 대한 실증분석결과

1. 전체 기업을 대상으로 한 검정결과

우선 전체 기업의 대상으로 지브라의 법칙에 대한 실증분석결과를 살펴본다. <Table 2>와 <Table 3>에는 통계청 MDIS를 이용한 결과가 제시되어 있으며, <Table 3>~<Table 8>에는 KIS 데이터베이스를 이용한 결과가 제시되어 있다. <Table 2>와 <Table 3>은 각각 2006~2011년, 2011~2016년의 고용 증가에 대한 회귀분석결과이고, <Table 4>~<Table 8>은 1987~1993년, 1993~1999년, 1999~2005년, 2005~2011년, 2011~2017년의 고용 증가에 대한 회귀분석결과이다.

MDIS 자료는 시계열이 2006년 이후로 짧은 편이고 설명변수로 이용할 만한 정보가 많지 않다는 점에서 단점을 지닌다. 예를 들어 <Table 2>와 <Table 3>에서는 Jovanovic 가설의 검정을 위해 중요한 변수인 기업의 연령(age)을 설명변수로 이용할 수 없었다. 따라서 <Table 2>와 <Table 3>에서 이용한 설명변수는 산업 더미변수, 관계회사 여부(relational), 지적 재산권/자산(IPR/asset), 부채/자산(Debt/asset)이다. 이와 비교할 때 KIS value 자료는 1980년 이후부터 자료를 제공하고 개별 기업의 정보를 상세하게 제공한다. 따라서 <Table 4>~<Table 8>의 결과방정식에는 추가적으로 기업 연령(age), 상장 여부(listed), 연구개발비/매출액(R&D/

sales), 유동자산/자산(currentasset/asset), 수출액/매출액 (export/sales) 등의 설명변수를 이용하였다. 또한 선택방정식에는 결과방정식에 포함된 변수에 추가하여 총자산회전율(sales/asset), 영업이익/자산(EBIT/asset), 유보이익/자산(retained earnings/asset) 을 이용하였다.

모든 표는 회귀분석(regression), 2단계 추정(2-step regression), FIML의 3가지 방법을 이용한 결과를 제시하고 있다. 우선 2열의 회귀분석은 (1) 식만 추정한 결과이다. 3열의 2단계 추정과 4열의 FIML은 (2) 식을 추정하되, 전자는 Heckman의 2단계 절차를 이용한 결과이고 후자는 FIML을 이용한 결과이다. 각 표의 상단에는 선택방정식, 하단에는 결과방정식의 추정결과가 제시되어 있다. 회귀분석에서 선택방정식의 추정결과는 당연히 없다.

지브라의 법칙에 대한 결과를 논의하기 위해 $\log(S_0)$ 의 추정계수를 중심으로 살펴본다. 다음으로 선택방정식과 결과방정식의 나머지 변수에 대한 추정결과를 살펴본다.

〈Table 2〉는 2011년 고용 규모의 로그값을 2006년 고용 규모의 로그값에 대해 추정한 결과이다. 우선 2단계 추정의 결과에서 IMR의 추정계수가 0.542로 유의하였기 때문에 표본선택편의가 있었던 것으로 보인다. 따라서 회귀분석의 결과는 표본선택편의에 의해 영향을 받은 것으로 보아야 한다. 다음으로 $\log(S_0)$ 의 추정계수는 회귀분석의 경우 0.921이고 2단계 추정의 경우 1.031이며 FIML의 경우 0.956이다. 표본선택편의를 고려한다면 $\log(S_0)$ 의 추정계수는 더 큰 것으로 나타났다. 2단계 추정의 경우 $\log(S_0)$ 의 추정계수는 1.031로 1에 매우 가까운 뿐만 아니라 월드검정통계량도 2.21로 추정계수가 1이라는 귀무가설을 기각할 수 없었다. 따라서 2단계 추정결과만을 본다면 지브라의 법칙은 성립하는 것으로 볼 수 있다. 하지만 FIML의 경우 $\log(S_0)$ 의 추정계수는 0.956으로 1과 차이가 있었을 뿐만 아니라 월드검정통계량도 17.36으로 추정계수가 1이라는 귀무가설을 기각하였다. 앞에서 언급한 바와 같이 FIML 추정치의 분산이 더 작은 것으로 알려져 있기 때문에 FIML에 의한 추정결과를 기준으로 판단하는 것이 좋을 것이다. 따라서 〈Table 2〉의 결과를 요약하면 다음과 같다. 표본선택편의는 있는 것으로 보이며 이를 고려하면 $\log(S_0)$ 의 추정계수는 1에 가까워지지만 추정계수가 1이라는 귀무가설은 기각된다. 즉 표본선택편의를 고려하면 지브라의 법칙에 보다 가까운 결과를 얻게 되지만 지브라의 법칙이 성립한다고 말할 수 있는 정도는 아니다.

〈Table 2〉 Estimation results(MDIS) : 2006~2011

	regression	2-step regression	FIML
Selection equation			
constant		-2.172 (-20.94) ***	-2.245 (-26.10) ***
$\log(S_0)$		0.603 (29.19) ***	0.621 (30.01) ***
industry-agriculture		-0.290 (-1.28)	-0.317 (-1.55)
industry-manufacturing		0.091 (2.67) ***	0.067 (2.02) **
industry-construction		-0.230 (-3.87) **	-0.242 (-4.04) ***
industry-wholesale/retail		0.217 (4.61) ***	0.206 (4.50) ***
relational		0.007 (0.25)	0.016 (0.54)
IPR/asset		0.848 (0.26)	0.757 (0.18)
debt/asset		-0.343 (-10.17) ***	-0.345 (-16.15) ***
Outcome equation			
constant	0.493 (12.86) ***	-0.262 (-1.76) *	0.251 (3.78) ***
$\log(S_0)$	0.921 (134.1) ***	1.031 (46.11) ***	0.956 (91.43) ***
industry-agriculture	-0.131 (-1.08)	-0.215 (-1.62)	-0.157 (-0.90)
industry-manufacturing	-0.017 (-1.16)	0.007 (0.40)	-0.008 (-0.55)
industry-construction	-0.001 (-0.04)	-0.064 (-1.84) *	-0.021 (-0.70)
industry-wholesale/retail	-0.060 (-2.97) ***	-0.001 (-0.07)	-0.041 (-2.20) **
relational	-0.050 (-3.83) ***	-0.046 (-3.09) ***	-0.049 (-3.77) **
IPR/asset	-1.201 (-0.97)	-1.025 (-0.70)	-1.175 (-0.41)
debt/asset	-0.024 (-1.38)	-0.128 (-4.65) ***	-0.057 (-2.98) ***
inverse Mill's ratio		0.542 (5.41) ***	
σ_ϵ			0.536 (98.01) ***
ρ			0.325 (6.15) ***
no. of obs			10,692
no. of censored obs.			3,156
Wald statistics		2.01, p-value 0.155	17.36, p-value 0.00

Notes: 1) MDIS data are used. The dependent variable are in 2011 and all the explanatory variables are in 2006.

2) Regression is the estimation results of the model (1). 2-step regression and FIML are the estimation results of the model (2).

3) t-values are in parentheses and *, **, *** are significant at the 10%, 5%, 1% levels respectively.

〈Table 3〉는 2016년 고용 규모의 로그값을 2011년 고용 규모의 로그값에 대해 추정한 결과이다. 2단계 IMR의 추정계수가 0.766로 유의하였기 때문에 표본선택편의가 있었던 것으로 보인다. 다음으로 $\log(S_0)$ 의 추정계수는 회귀분석의 경우 0.929이고 2단계 추정의 경우 1.027이며 FIML의 경우 1.035이다. $\log(S_0)$ 의 추정계수가 1이라는 귀무가설에 대한 월드검정통계량은 2단계 추정의 경우 2.5로 귀무가설을 기각할 수 없었지만 FIML의 경우 23.61로 귀무가설을 기각한다. 하지만 FIML의 경우 귀무가설의 기각은 다른 의미를 지닌다. 즉 $\log(S_0)$ 의 추정계수가 1보다 큼을 의미한다. 이제 〈Table 3〉의 FIML 결과를 중심으로 요약하면 다음과 같다. 표본선택편의는 있는 것으로 보이며 이를 고려하면 $\log(S_0)$ 의 추정계수는 1보다 더 컸다. 이는 다음과 같은 뜻이다. 표본선택편의를 고려할 때 지브라의 법칙이 성립한다고 보기는 어려운데 그 이유는 규모가 큰 기업일수록 고용의 증가율이 더 높았기 때문이다.

〈Table 2〉와 〈Table 3〉의 결과를 비교하면 두 가지로 요약할 수 있다. 첫째 표본선택편의는 중요한 영향을 미치며 이를 고려하면 지브라의 법칙에 가까운 결과를 얻게 된다. 둘째 2006~2011년의 경우 2단계 추정결과와 FIML 추정결과가 달랐지만 2011~2016년의 경우에는 두 결과 모두 고용 규모가 큰 기업의 고용증가율이 더 높았음을 보여준다. 2011~2016년의 결과만을 보면, 기존의 연구와는 매우 다르게 지브라의 법칙을 뛰어넘는 결과를 얻은 셈이다. 다만 기업의 연령 등 설명변수가 누락되었다는 점에서 한계가 있음을 염두에 두어야 한다.

〈Table 4〉~〈Table 8〉는 KIS value 자료를 이용한 결과이다. 앞의 표와 비슷하므로 간략하게 요약한다. 첫째 2단계 추정의 결과에서 IMR의 추정계수는 모두 유의하였기 때문에 표본선택편의가 있었던 것으로 판단된다. 둘째 $\log(S_0)$ 의 추정계수는 회귀분석에 의한 추정치보다 2단계 추정에 의한 추정치가 더 크지만 FIML에 의한 추정치는 약간 더 클 뿐이다. 표본선택편의를 고려한다면 $\log(S_0)$ 의 추정계수는 더 큼을 알 수 있다. 셋째 그림에도 불구하고 $\log(S_0)$ 의 추정계수가 1이라는 귀무가설은 월드검정에 의해 모두 기각되었다. 넷째 $\log(S_0)$ 의 추정계수는 2005년 이전보다 2005년 이후 더 큰 것으로 보이는데, 큰 차이가 있는 것은 아니다.

〈Table 3〉 Estimation results(MDIS) : 2011~2016

	regression	2-step regression	FIML
Selection equation			
constant		-0.929 (-11.85) ***	-1.136 (23.11) ***
$\log(S_0)$		0.391 (24.35) ***	0.510 (54.66) ***
industry-agriculture		0.252 (0.69)	-0.922 (-1.94) **
industry-manufacturing		0.093 (2.88) ***	-0.476 (-19.59) ***
industry-construction		-0.301 (-4.91) **	-0.355 (-9.38) ***
industry-wholesale/retail		0.310 (7.04) ***	0.497 (17.64) ***
relational		0.117 (3.76) ***	0.254 (11.63) ***
listed		0.205 (3.87) ***	0.526 (15.88) ***
IPR/asset		4.81 (0.99)	11.867 (3.98) ***
debt/asset		-0.376 (-12.25) ***	-0.286 (-15.12) ***
Outcome equation			
constant	0.509 (16.29) ***	-0.200 (-1.66) *	-0.264 (-7.39) ***
$\log(S_0)$	0.929 (161.0) ***	1.027 (58.23) ***	1.035 (148.2) ***
industry-agriculture	-0.147 (-1.07)	-0.061 (-0.35)	0.231 (0.67)
industry-manufacturing	-0.068 (-5.13) ***	-0.024 (-1.40)	0.066 (3.97) ***
industry-construction	0.432 (15.26) ***	0.343 (9.43) ***	0.325 (12.61) ***
industry-wholesale/retail	-0.061 (-3.51) ***	0.034 (1.28)	0.032 (1.64) *
relational	0.013 (1.07)	0.045 (2.84) ***	0.039 (2.56) **
listed	-0.023 (-1.35)	0.006 (0.27)	-0.021 (-0.90)
IPR/asset	-2.014 (-0.92)	-1.109 (-0.42)	-2.407 (-1.58)
debt/asset	-0.030 (-1.95) *	-0.183 (-5.98) ***	-0.222 (-17.16) ***
inverse Mill's ratio		0.766 (6.40) ***	
σ_ϵ			0.728 (174.5) ***
ρ			0.987 (6330) ***
no. of obs			11,639
no. of censored obs.			2,437
Wald statistics		2.50, p-value 0.114	23.61, p-value 0.00

Notes: 1) MDIS data are used. The dependent variable are in 2016 and all the explanatory variables are in 2011.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

〈Table 4〉 Estimation results(KIS) : 1987~1993

	regression	2-step regression	FIML
Selection equation			
constant		-0.088 (-0.25)	-0.179 (-0.50)
$\log(S_0)$		0.344 (6.04) ***	0.358 (6.11) ***
industry-agriculture		4.006 (0.02)	4.008 (0.01)
industry-manufacturing		0.025 (0.16)	-0.011 (-0.06)
industry-construction		1.071 (2.54) **	1.074 (2.53) **
industry-wholesale/retail		-0.152 (-0.83)	-0.183 (-0.99)
relational		0.152 (1.36)	0.166 (1.48)
listed		-0.260 (-0.92)	-0.264 (-0.94)
age		0.012 (1.73) *	0.013 (1.74) *
IPR/asset		-208.9 (-3.14) ***	-205.3 (-3.11) ***
R&D/sales		27.47 (0.92)	31.99 (1.00)
debt/asset		-0.453 (-1.38)	-0.412 (-1.26)
currentasset/asset		0.311 (1.18)	0.295 (1.11)
export/sales		0.152 (0.55)	0.131 (0.48)
sales/asset		-0.012 (-0.20)	-0.008 (-0.14)
EBIT/asset		0.829 (0.89)	0.881 (0.95)
retained earnings/asset		-2.142 (-2.99) ***	-2.076 (-2.93) ***
Outcome equation			
constant	1.055 (10.06) ***	0.598 (3.23) ***	1.004 (9.10) ***
$\log(S_0)$	0.821 (57.36) ***	0.875 (37.10) ***	0.827 (55.66) ***
industry-agriculture	-0.308 (-2.00) **	-0.207 (-1.12)	-0.298 (-1.94) *
industry-manufacturing	0.001 (0.02)	0.017 (0.33)	0.003 (0.07)
industry-construction	0.257 (3.78) ***	0.364 (4.26) ***	0.269 (3.93) ***
industry-wholesale/retail	-0.110 (-1.98) **	-0.137 (-2.17) **	-0.113 (-2.04) **
relational	0.191 (5.79) ***	0.223 (5.74) ***	0.195 (5.89) ***
listed	0.226 (4.19) ***	0.184 (2.87) ***	0.222 (4.11) ***
age	-0.007 (-4.42) ***	-0.006 (-3.14) ***	-0.007 (-4.35) ***
IPR/asset	17.618 (0.56)	-27.440 (-0.79)	9.702 (0.31)
R&D/sales	6.603 (1.55)	9.796 (1.93) *	6.967 (1.64)
debt/asset	0.005 (0.06)	-0.026 (-0.26)	0.002 (0.01)
currentasset/asset	-0.088 (-1.14)	-0.036 (-0.40)	-0.081 (-1.05)
export/sales	-0.226 (-3.51) ***	-0.212 (-2.86) ***	-0.225 (-3.49) ***
inverse Mill's ratio		1.022 (3.41) ***	
σ_ϵ			0.629 (57.41) ***
ρ			0.176 (1.40)
no. of obs			1,825
no. of censored obs.			96
Wald statistics		27.84, p-value 0.00	135.5, p-value 0.00

Notes: 1) KIS value data are used. The dependent variable are in 1993 and all the explanatory variables are in 1987.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

〈Table 5〉 Estimation results(KIS) : 1993~1999

	regression	2-step regression	FIML
Selection equation			
constant		0.904 (3.38) ***	0.848 (3.07) ***
$\log(S_0)$		0.179 (4.59) ***	0.187 (4.64) ***
industry-agriculture		0.038 (0.07)	0.048 (0.09)
industry-manufacturing		-0.088 (-0.74)	-0.100 (-0.83)
industry-construction		0.568 (2.37) **	0.555 (2.31) **
industry-wholesale/retail		0.126 (0.79)	0.118 (0.74)
relational		0.238 (2.80) ***	0.241 (2.84) ***
listed		-0.401 (-3.06) ***	-0.394 (-3.00) ***
age		0.007 (1.49)	0.007 (1.53)
IPR/asset		140.16 (1.06)	136.24 (1.04)
R&D/sales		24.49 (1.30)	22.94 (1.25)
debt/asset		-0.404 (-1.71) *	-0.379 (-1.59)
currentasset/asset		-0.154 (-0.75)	-0.149 (-0.73)
export/sales		-0.013 (-0.06)	-0.025 (-0.12)
sales/asset		-0.013 (-0.31)	-0.015 (-0.37)
EBIT/asset		1.084 (1.17)	1.128 (1.22)
retained earnings/asset		-0.395 (-0.61)	-0.340 (-0.52)
Outcome equation			
constant	1.073 (13.15) ***	0.514 (2.30) **	1.055 (12.67) ***
$\log(S_0)$	0.833 (72.25) ***	0.905 (29.71) ***	0.835 (71.21) ***
industry-agriculture	-0.304 (-2.01) **	-0.305 (-1.23)	-0.304 (-2.01) **
industry-manufacturing	0.032 (0.88)	0.016 (0.27)	0.031 (0.87)
industry-construction	0.013 (0.24)	0.185 (1.66)	0.019 (0.33)
industry-wholesale/retail	-0.044 (-0.95)	0.005 (0.06)	-0.042 (-0.91)
relational	0.182 (6.57) ***	0.287 (5.12) ***	0.185 (6.65) ***
listed	0.012 (0.31)	-0.160 (-1.84) *	0.007 (0.17)
age	-0.008 (-6.19) ***	-0.005 (-2.28) **	-0.008 (-6.12) ***
IPR/asset	1.681 (0.79)	3.882 (1.02)	1.749 (0.82)
R&D/sales	5.227 (2.77) ***	8.804 (2.51) **	5.341 (2.83) ***
debt/asset	-0.283 (-3.94) ***	-0.485 (-3.61) ***	-0.290 (-4.02) ***
currentasset/asset	-0.257 (-3.96) ***	-0.284 (-2.70) ***	-0.257 (-3.97) ***
export/sales	0.023 (0.37)	0.016 (0.15)	0.023 (0.36)
inverse Mill's ratio		2.288 (3.25) ***	
σ_ϵ			0.658 (72.27) ***
ρ			0.109 (0.98)
no. of obs			2,753
no. of censored obs.			142
Wald statistics		9.59, p-value 0.00	197.0, p-value 0.00

Notes: 1) KIS value data are used. The dependent variable are in 1999 and all the explanatory variables are in 1993.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

〈Table 6〉 Estimation results(KIS) : 1999~2005

	regression	2-step regression	FIML
Selection equation			
constant		0.277 (2.23) **	0.234 (1.86) *
$\log(S_0)$		0.215 (10.02) ***	0.227 (10.17) ***
industry-agriculture		0.002 (0.00)	-0.013 (-0.05)
industry-manufacturing		0.179 (2.89) ***	0.163 (2.59) ***
industry-construction		-0.168 (-2.02) **	-0.173 (-2.07) *
industry-wholesale/retail		0.146 (1.88) *	0.141 (1.81) *
relational		0.505 (10.56) ***	0.503 (10.52) ***
listed		-0.596 (-6.41) ***	-0.579 (-6.18) ***
age		0.012 (4.37) ***	0.012 (4.38) ***
IPR/asset		1.832 (2.40) **	1.831 (2.42) **
R&D/sales		-0.012 (-0.43)	-0.012 (-0.44)
debt/asset		-0.061 (-0.60)	-0.048 (-0.48)
currentasset/asset		0.040 (0.40)	0.033 (0.32)
export/sales		-0.034 (-0.20)	-0.044 (-0.26)
sales/asset		-0.041 (-2.69) ***	-0.041 (-2.73) ***
EBIT/asset		0.739 (2.31) **	0.722 (2.25) **
retained earnings/asset		-1.622 (-8.17) ***	-1.563 (-7.83) ***
Outcome equation			
constant	1.213 (24.22) ***	0.934 (11.28) ***	1.155 (21.02) ***
$\log(S_0)$	0.773 (97.40) ***	0.802 (74.20) ***	0.779 (94.02) ***
industry-agriculture	-0.353 (-3.56) ***	-0.354 (-3.39) ***	-0.353 (-3.55) ***
industry-manufacturing	0.034 (1.43)	0.058 (2.25) **	0.039 (1.63)
industry-construction	-0.117 (-3.12) ***	-0.153 (3.83) ***	-0.125 (-3.31) ***
industry-wholesale/retail	-0.069 (-2.26) **	-0.059 (-1.83) *	-0.067 (-2.18) **
relational	0.194 (10.70) ***	0.266 (10.52) ***	0.209 (10.94) ***
listed	-0.035 (-1.10)	-0.102 (-2.73) ***	-0.049 (-1.51)
age	-0.009 (-10.34) ***	-0.008 (-8.03) ***	-0.009 (-9.96) ***
IPR/asset	1.185 (5.01) ***	1.437 (5.58) ***	1.238 (5.21) ***
R&D/sales	-0.001 (-0.17)	-0.003 (-0.27)	-0.002 (-0.19)
debt/asset	-0.119 (-2.94) ***	-0.114 (-2.66) ***	-0.118 (-2.91) ***
currentasset/asset	0.140 (3.60) ***	0.124 (3.02) ***	0.137 (3.51) ***
export/sales	-0.025 (-0.43)	-0.029 (-0.47)	-0.026 (-0.44)
inverse Mill's ratio		0.578 (4.44) ***	
σ_ϵ			0.706 (-110.0) ***
ρ			0.171 (2.58) ***
no. of obs			13,957
no. of censored obs.			563
Wald statistics		3.72, p-value 0.054	204.1, p-value 0.00

Notes: 1) KIS value data are used. The dependent variable are in 2005 and all the explanatory variables are in 1999.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

〈Table 7〉 Estimation results(KIS) : 2005~2011

	regression	2-step regression	FIML
Selection equation			
constant		0.248 (2.39) **	0.214 (2.06) **
$\log(S_0)$		0.245 (13.76) ***	0.262 (14.21) ***
industry-agriculture		-0.195 (-1.01)	-0.235 (-1.22)
industry-manufacturing		0.021 (0.43)	-0.013 (-0.27)
industry-construction		0.228 (2.68) ***	0.212 (2.49) **
industry-wholesale/retail		0.139 (2.17) **	0.127 (1.97) **
relational		0.579 (14.43) ***	0.576 (14.35) ***
listed		-0.889 (-14.42) ***	-0.890 (-14.40) ***
age		0.014 (6.79) ***	0.014 (6.64) ***
IPR/asset		-0.326 (-1.17)	-0.311 (-1.13)
R&D/sales		-0.004 (-1.40)	-0.004 (-1.36)
debt/asset		-0.205 (-2.44) **	-0.199 (-2.39) **
currentasset/asset		0.067 (0.88)	0.076 (0.99)
export/sales		-0.016 (-0.13)	-0.005 (-0.42)
sales/asset		-0.007 (-0.73)	-0.009 (-0.93)
EBIT/asset		0.613 (2.08) **	0.405 (1.36)
retained earnings/asset		-1.647 (-10.52) ***	-1.614 (-10.36) ***
Outcome equation			
constant	0.606 (15.38) ***	0.165 (2.45) **	0.529 (12.67) ***
$\log(S_0)$	0.862 (137.3) ***	0.916 (96.96) ***	0.871 (134.1) ***
industry-agriculture	0.040 (0.45)	-0.002 (-0.02)	0.033 (0.373)
industry-manufacturing	0.215 (11.22) ***	0.221 (10.19) ***	0.217 (11.26) ***
industry-construction	0.091 (2.92) ***	0.136 (3.87) ***	0.099 (3.16) ***
industry-wholesale/retail	0.128 (5.22) ***	0.159 (5.69) ***	0.133 (5.42) ***
relational	0.102 (6.91) ***	0.228 (10.31) ***	0.124 (8.10) ***
listed	0.000 (0.03)	-0.164 (-4.95) ***	-0.028 (-1.13)
age	-0.009 (-13.51) ***	-0.006 (-7.89) ***	-0.008 (-12.69) ***
IPR/asset	1.306 (9.00) ***	1.217 (7.80) ***	1.290 (8.88) ***
R&D/sales	0.003 (1.50)	0.001 (0.64)	0.003 (1.33)
debt/asset	-0.089 (-2.78) ***	-0.132 (-3.62) ***	-0.096 (-3.01) ***
currentasset/asset	0.159 (5.26) ***	0.156 (4.56) ***	0.159 (5.22) ***
export/sales	-0.136 (-2.74) ***	-0.136 (-2.43) ***	-0.136 (-2.73) ***
inverse Mill's ratio		0.916 (8.96) ***	
σ_ϵ			0.709 (135.9) ***
ρ			0.226 (5.73) ***
no. of obs			11,367
no. of censored obs.			1,133
Wald statistics		78.9, p-value 0.00	389.2, p-value 0.00

Notes: 1) KIS value data are used. The dependent variable are in 2011 and all the explanatory variables are in 2005.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

〈Table 8〉 Estimation results(KIS) : 2011~2017

	regression	2-step regression	FIML
Selection equation			
constant		0.966 (9.10) ***	0.936 (8.82) ***
$\log(S_0)$		0.168 (9.89) ***	0.180 (10.21) ***
industry-agriculture		-0.104 (-0.43)	-0.119 (-0.50)
industry-manufacturing		-0.097 (-1.78) *	-0.128 (-2.28) **
industry-construction		-0.305 (-3.80) ***	-0.318 (-3.95) ***
industry-wholesale/retail		-0.025 (-0.37)	-0.037 (-0.55)
relational		0.421 (9.54) ***	0.424 (9.59) ***
listed		-0.554 (-7.77) ***	-0.559 (-7.83) ***
age		0.008 (4.15) ***	0.008 (4.07) ***
IPR/asset		-1.295 (-3.74) ***	-1.292 (-3.76) ***
R&D/sales		0.020 (0.77)	0.022 (0.85)
debt/asset		0.131 (1.18)	0.153 (1.38)
currentasset/asset		-0.230 (-2.88) ***	-0.241 (-3.04) ***
export/sales		-0.121 (-0.74)	-0.120 (-0.74)
sales/asset		-0.008 (-0.62)	-0.013 (-1.01)
EBIT/asset		0.098 (0.55)	0.111 (0.67)
retained earnings/asset		0.388 (3.13) ***	0.403 (3.29)
Outcome equation			
constant	0.449 (15.30) ***	-0.261 (-2.99) ***	0.426 (14.23) ***
$\log(S_0)$	0.860 (173.6) ***	0.953 (70.70) ***	0.863 (172.0) ***
industry-agriculture	0.129 (1.76) *	0.111 (0.72)	0.128 (1.75) *
industry-manufacturing	0.221 (13.65) ***	0.205 (6.06) ***	0.221 (13.61) ***
industry-construction	0.147 (5.37) ***	-0.022 (-0.39)	0.141 (5.16) ***
industry-wholesale/retail	0.156 (7.70) ***	0.164 (3.86) ***	0.156 (7.70) ***
relational	0.034 (2.67) ***	0.251 (7.62) ***	0.041 (3.16) ***
listed	0.119 (5.42) ***	-0.198 (-3.61) ***	0.109 (4.93) ***
age	-0.005 (-10.36) ***	-0.001 (-0.94)	-0.005 (-10.06) ***
IPR/asset	0.723 (4.53) ***	-0.436 (-1.54)	0.686 (4.29) ***
R&D/sales	-0.000 (0.98)	0.000 (0.08)	-0.000 (-0.05)
debt/asset	-0.061 (-2.31) ***	-0.146 (-2.62) ***	-0.064 (-2.41) ***
currentasset/asset	0.305 (12.36) ***	0.201 (3.86) ***	0.301 (12.21) ***
export/sales	-0.099 (-1.99) **	-0.147 (-1.40)	-0.101 (-2.02) ***
inverse Mill's ratio		3.905 (13.11) ***	
σ_ϵ			0.736 (168.4) ***
ρ			0.167 (3.76)
no. of obs			15,429
no. of censored obs.			617
Wald statistics		12.09, p-value 0.00	736.1, p-value 0.00

Notes: 1) KIS value data are used. The dependent variable are in 2017 and all the explanatory variables are in 2011.

2) the same as Notes 2) and Notes 3) in 〈Table 2〉.

KIS value 자료를 이용한 결과와 MDIS value 자료를 이용한 결과를 비교하면 비슷하다. 첫째 표본선택편의는 있었던 것으로 보이며, 둘째 이를 고려하면 $\log(S_0)$ 의 추정계수는 1에 가까워지지만 추정계수가 1이라는 귀무가설은 기각된다. 셋째 최근의 자료를 이용할 때 $\log(S_0)$ 의 추정계수는 더 컸던 것으로 보이는데, 그 차이가 크다고 보기는 어렵다. 요약하면 다음과 같다. 표본선택편의를 고려하면 지브라의 법칙에 보다 가까운 결과를 얻게 되지만 지브라의 법칙이 성립한다고 말할 수 있는 정도는 아니다. 다만 변화가 조금씩 나타날 뿐이다.

선택방정식과 결과방정식에 포함된 나머지 설명변수에 대한 추정결과를 살펴본다. 선택방정식은 기업의 생존에 영향을 미치는 요인을 추정하고 결과방정식은 기업 고용의 증가에 영향을 미치는 요인을 분석한다. FIML을 이용한 결과를 중심으로 요약하면 다음과 같다.

선택방정식의 추정결과에서 생존가능성에 일관성있게 영향을 미쳤다고 판단되는 변수는 기업의 규모, 관계회사 존재 여부, 기업의 연령이다. 기업 규모의 추정계수는 유의한 (+)이었는데, 규모가 클수록 생존가능성이 높았음을 알 수 있다. 바꾸어 말하면 규모가 작은 기업의 생존가능성이 낮았다. 관계회사 존재 여부의 추정계수는 유의한 (+)인 결과가 많았으므로 관계회사가 있는 기업의 경우 생존 가능성은 높았던 것으로 판단된다. 끝으로 기업의 연령은 KIS value에서만 이용할 수 있는 설명변수였는데, 대부분의 경우 유의한 (+)이었다. 생존 연수가 긴 기업의 생존가능성이 더 높았던 것으로 판단된다. 이 중에서 기업의 규모와 생존 가능성이 (+)의 관계를 갖는다는 결과는 생존자편의와 관련하여 중요하다.

다른 변수의 추정계수는 명확하지 않다. 우선 산업별 더미변수의 추정계수는 유의하지 않은 것이 많았고 시기별 변화도 컸다. 따라서 특정 산업의 생존 가능성이 더 컸다고 말하기는 어려울 것이다. 또한 지적 재산권과 연구개발비의 추정계수도 유의하였다고 보기는 어렵다. 한편 상장 여부가 미치는 영향은 자료에 따라 달랐다. MDIS를 이용한 경우 2011~2016년에는 상장 여부의 추정계수가 유의한 (+)이었지만 KIS value를 이용한 경우 거의 전 시기에 걸쳐 추정계수가 유의한 (-)이었다. 상장기업의 생존 가능성이 낮았다는 결과는 꽤 의외이다. 또한 재무적 건전성을 나타내는 변수들의 추정계수도 유의하지 않은 경우가 많았고 일관성이 있다고 보기도 어렵다. 끝으로 수출비중의 추정계수는 대부분 유의하지 않았다.

다음으로 결과방정식의 추정결과를 살펴본다. 선택방정식에 비해서 일관성있는

결과가 많았는데, 요약하면 다음과 같다. 산업별 더미변수의 경우 농림어업은 유의한 (-)인 결과가 많았고, 제조업은 유의한 (+)인 경우가 많았다. 건설업은 유의한 (+)인 결과가 많았고, 도소매업은 최근의 자료를 이용할 경우 유의한 (+)이 많았다. 벤치마크가 서비스업임을 감안하면 서비스업에 비해 제조업, 건설업, 도소매업의 고용 증가가 컸으며 농림어업의 고용 증가는 적었다고 볼 수 있다. 한편 부채비율의 추정계수는 유의한 (-)이고 유동자산의 추정계수는 유의한 (+)이었다. 재무상태가 안정적인 기업일수록 고용증가는 컸다고 판단된다. 다음으로 관계회사 여부의 추정계수는 (+)이었으나 상장 여부의 추정계수는 시기에 따라 달랐다. 관계회사가 있는 기업의 경우 고용 증가가 컸던 것으로 판단된다. 한편 지적 재산권과 연구개발비의 추정계수는 (+)인 결과가 많았다. 연구개발투자가 활발한 기업의 고용 증가가 컸던 셈이다. 가장 일관성있는 추정결과를 보인 변수는 기업의 연령이다. 즉 기업 연령의 추정계수가 모든 경우 유의한 (-)이었다. 생존 연수가 짧은 기업의 고용증가가 더 컸다고 판단되는데, 이는 Jovanovic 가설을 지지하는 것이다. 끝으로 수출 비중은 유의한 (-)인 결과가 많았다.

이제 선택방정식과 결과방정식의 주요결과를 설명변수별로 정리하면 다음과 같다. 첫째 기업의 규모와 생존 가능성은 (+)의 관계를 갖는다. 둘째 기업의 연령은 생존 가능성과 (+)의 관계를 갖지만, 고용 증가와는 (-)의 관계를 갖는다. 셋째 연구개발투자는 생존 가능성과 유의한 관계를 갖는다고 보기 어렵지만, 고용 증가와는 (+)의 관계를 갖는다. 넷째 재무적 안정성은 생존 가능성과 일관성있는 관계를 갖는다고 보기 어렵지만 고용 증가와는 (+)의 관계를 갖는다. 다섯째 관계회사의 존재 여부는 생존 가능성과 및 고용 증가와 모두 (+)의 관계를 갖는다. 끝으로 산업별 차이, 상장 여부, 수출 비중 등의 효과는 일관성이 있다고 보기 어렵다.

2. 분위별 검정결과

다음으로 과거 고용규모의 분위별로 표본을 나눈 후 수행한 검정결과를 살펴본다. <Table 9>는 과거 기업고용의 분위별로 추정한 결과를 요약하고 있다. KIS value 자료를 이용하였고 FIML에 의해 추정한 결과이며 표에는 지브라의 법칙 검정에 초점을 맞추기 위하여 $\log(S_0)$ 의 추정계수만이 제시되어 있다. 앞에서 수행한 것과 같이 $\log(S_0)$ 는 6년 전의 값을 이용하였다. 따라서 2004년의 행에 제시된 결

과에서 종속변수인 기업 고용의 로그값은 2004년의 값이며 $\log(S_0)$ 를 포함한 설명 변수는 6년 전인 1998년의 값을 이용하였다. 2004년 이전의 추정결과도 비슷한 편이지만 FIML의 추정치가 수렴하지 않는 경우가 있어서 <Table 9>에는 제시하지 않았다. 끝으로 괄호 안은 $\log(S_0)$ 가 1이라는 귀무가설에 대한 윌드검정통계량이다.

여기에서는 종업원수가 10명 이하인 기업과 3천명 이상인 대기업을 제외하였다. 이에 따라 고용 규모의 연도별 분위값에는 큰 차이가 없었다. 예를 들어 고용 규모의 20%, 40%, 60%, 80% 분위값은 2017년 28, 48, 81, 161명이었고, 2011년 28, 49, 83, 164명, 2006년 27, 47, 80, 160명, 2001년 27, 48, 85, 177명이었다. 따라서 40% 분위값-60% 분위값에 속하는 기업의 종업원수는 평균 48-82명, 60% 분위값-80% 분위값에 속하는 기업의 종업원수는 평균 83-163명, 80% 분위값 이상에 속하는 기업의 종업원수는 평균 164명 이상이다. 80% 분위값 이상에 속하는 기업이라고 해도 대부분 종업원 500명 이하의 중소 규모 기업이다.

분위별 추정결과는 매우 두드러진 특징을 보여준다. 첫째 상위 기업으로 갈수록 $\log(S_0)$ 의 추정계수가 커진다는 점이다. 예를 들어 2017년 경우 하위 20% 기업의 경우 0.59이었지만 20-40% 기업의 경우 0.74, 40-60% 기업의 경우 0.94, 60-80% 기업의 경우 0.86이었으며 80% 이상인 기업의 경우 1.02에 이른다. 80% 이상인 기업의 경우 2008년 이후 $\log(S_0)$ 의 추정계수는 1을 넘는 편이었다. 둘째 하위 20%와 20-40% 그리고 60-80%인 기업의 경우 $\log(S_0)$ 가 1이라는 귀무가설은 모두 기각되었지만 40-60%와 80% 이상인 기업의 경우 귀무가설이 기각되지 않는 경우가 많았다는 점이다. 40-60%에 속하는 기업의 경우 2011년 이후 귀무가설을 기각할 수 없었으며 80% 이상에 속하는 기업의 경우 2006년 이후 귀무가설을 기각할 수 없는 경우가 많았다. 게다가 80% 이상인 기업의 경우 2011, 2013, 2014년 귀무가설이 기각되었지만 추정계수가 1.04이었으므로 귀무가설의 기각은 고용 규모가 큰 기업일수록 고용 증가가 훨씬 더 컸음을 의미한다. 즉 지브라의 법칙을 뛰어넘은 결과인 셈이다.

이 결과는 고용 증가와 기존의 고용규모 간에 비선형관계가 있었음을 시사한다. 즉 기존의 고용규모가 일정 수준(예를 들어 48명) 이하로 영세한 기업의 경우 기존 고용 규모가 작을수록 고용 증가는 큰 편으로 지브라의 법칙은 성립하지 않는다. 하지만 기존의 고용 규모가 이 수준을 넘은 기업의 경우 고용 증가는 기존의 고용 규모와 관계없이 무작위적으로 이루어지므로 이 기업들의 경우 지브라의 법칙이 성

립하는 것으로 보인다. 게다가 기존의 고용 규모가 더 큰 수준(예를 들어 164명)을 넘는 기업의 경우에는 기존의 고용 규모가 클수록 고용 증가가 큰 편이다. 즉 이 기업들의 경우에는 지브라의 법칙을 뛰어넘는 현상이 나타나기도 한다.

〈Table 9〉 Estimation Results by Quantiles

	~20%Q	20%Q~40%Q	40%Q~60%Q	60%Q~80%Q	80%Q~
2004	0.42 (159.3) ***	0.84 (3.29) *	0.72 (7.35) **	0.75 (10.4) ***	0.81 (44.1) ***
2005	0.46 (131.1) ***	0.78 (7.56) *	0.80 (4.37) **	0.84 (3.22) *	0.95 (2.96) *
2006	0.52 (138.1) ***	0.66 (17.62) ***	0.74 (6.75) **	0.77 (5.46) **	0.96 (2.37)
2007	0.55 (104.6) ***	0.69 (16.66) ***	0.67 (8.94) **	0.74 (6.55) **	0.99 (0.09)
2008	0.48 (160.9) ***	0.67 (18.62) ***	0.74 (6.74) **	0.84 (3.21) *	1.02 (0.97)
2009	0.52 (150.0) ***	0.71 (17.44) ***	0.79 (6.71) *	0.81 (3.20) *	1.03 (2.37)
2010	0.53 (163.1) ***	0.73 (11.69) ***	0.79 (7.94) ***	0.83 (6.13) *	1.02 (0.99)
2011	0.57 (148.6) ***	0.76 (9.76) ***	0.92 (0.01)	0.88 (2.91) *	1.04 (4.00) *
2012	0.57 (148.3) ***	0.71 (18.87) ***	0.90 (1.34)	0.85 (4.76) **	1.02 (1.80)
2013	0.54 (180.0) ***	0.76 (12.43) ***	0.95 (0.35)	0.88 (3.46) *	1.04 (5.48) *
2014	0.54 (235.2) ***	0.74 (17.32) ***	0.90 (1.93)	0.82 (9.25) *	1.04 (7.22) ***
2015	0.56 (213.1) ***	0.84 (6.26) ***	0.89 (1.95)	0.88 (3.45) *	1.02 (2.17)
2016	0.55 (262.2) ***	0.76 (13.94) ***	1.03 (0.28)	0.85 (4.55) **	1.02 (1.99)
2017	0.59 (251.1) ***	0.74 (20.07) ***	0.94 (0.68)	0.86 (5.34) **	1.02 (2.00)

Notes: 1) KIS value data are used. All the explanatory variables are in 6 years ago.

2) All the estimation results of the model (2) are by FIML.

3) Wald statistics are in parentheses and *, **, *** are significant at the 10%, 5%, 1% levels respectively.

언급한 바와 같이 고용 규모의 연도별 분위값에는 큰 차이가 없었으며 40% 분위 값~60% 분위 값에 속하는 기업은 작은 중소기업, 80% 분위 값 이상에 속하는 기업은 큰 중소기업, 중견기업, 그리고 대기업이다. 이는 본 연구의 표본이 대부분 중소기업임을 의미하는 것이다. 따라서 앞에서 수행한 바 있는 전체 기업을 대상으로 한 검정결과에서 지브라의 법칙이 기각되었던 이유는 고용 규모가 대략 48명 이하인 영세 기업이 다수 포함되었기 때문일지도 모른다. 즉 지브라의 법칙이 성립하지 않는다는 실증분석결과는 영세기업효과에 불과할 수도 있다.

IV. 이전 국내 연구결과와의 비교

앞에서 제시한 분위별 결과는 과거 고용규모의 분위별로 표본을 나눈 후 수행한 추정결과이다. 여기에서는 분위회귀분석(quantile regression)을 이용한 결과를 제시한다. 이를 통하여 분위회귀분석을 수행하였던 이전의 국내 연구결과와 비교한다. 방법론의 주요 차이점은 후자가 분위회귀분석을 이용하였고, 기업 고용의 성장률을 t 기와 $t-1$ 동안 측정하였다는 점이다.

첫째 분위회귀분석에 대해 논의한다. 표한형·홍성철(2014)이 추정한 방정식을 간략하게 제시하면 아래의 (5)식과 같다. 그들은 이 방정식을 분위회귀분석에 의해 추정하였다.

$$\log(S_{i,t}) - \log(S_{i,t-1}) = \alpha_0 + \alpha_1 \log(S_{i,t-1}) + \alpha_2 X_i + v_i \quad (5)$$

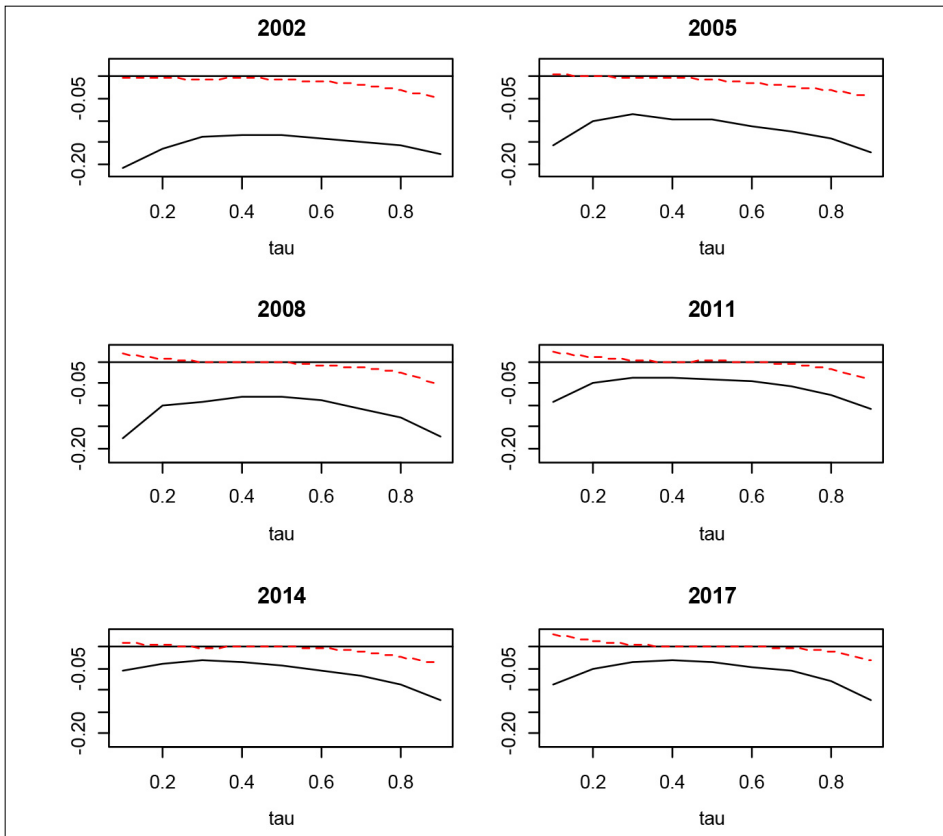
(5)식에서 종속변수는 기업 고용의 성장률이다. 따라서 지브라의 법칙이 성립한다면 α_1 은 0이 된다. 물론 표본선택편의를 고려하지 않았으므로 이를 조정하는 항은 없다. 표본선택이 있음을 고려한다면 (5)식은 결과방정식이 되며 이에 추가하여 선택방정식이 필요하다. 또한 표본선택편의를 제거하기 위하여 결과방정식도 수정되어야 한다.

본 연구도 이 장에서 분위회귀분석결과를 제시한다. 다만 본 연구는 표본선택이 있을 때 분위회귀분석방법을 이용한다는 점에서 다르다. 이를 위하여 Arellano and Bonhomme(2017)의 방법을 이용하였으며 방법론에 관한 자세한 내용은 부록에서 설명한다.

둘째 표한형·홍성철(2014), 김두순(2016)은 기업 고용의 성장률을 t 기와 $t-1$ 간격으로 측정하였다. 즉 (5)식의 좌변에서 확인할 수 있듯이 1년 동안 기업 고용의 성장률을 종속변수로 한다. 이 경우 에러의 자기상관이 있을 수 있으므로 (5)식과 같은 추정방정식을 얻게 되었고 $\log(S_{i,t-1})$ 가 설명변수가 되었다. 하지만 본 연구는 6년의 간격을 두고 기업 고용의 성장을 분석한다. 이에 따라 에러의 자기상관은 문제가 되지 않을 수 있지만 그 대신 표본선택이 중요한 문제가 된다. 왜냐하면 1년만에 파산하는 기업은 많지 않겠지만 6년 동안 파산한 기업은 적지 않을 것이기 때문이다. 따라서 분석기간이 1년이면 표본선택편의는 크게 중요한 문제가 아닐 수

있겠지만 분석기간이 6년이면 표본선택편의는 중요한 문제가 된다. 바꾸어 말하면 표본선택편의를 분석하기 위해서는 꽤 긴 기간 동안 기업 고용이 어떻게 변화하였는가를 분석하는 것이 좋다는 뜻이다. 이런 의미에서 보면 분위회귀분석을 수행한 기존 연구와 달리 본 연구가 갖는 가장 큰 차이는 기업고용의 성장을 분석하는 기간이 6년으로 길다는 점이다.

〈Figure 1〉 Comparison with the previous study



〈Figure 1〉은 표본선택이 있을 때 분위회귀분석을 이용하여 (5)식을 추정한 결과를 보여준다. 두 가지 결과가 제시되었는데 하나는 분석기간을 1년으로 설정한 결과이고 다른 하나는 6년으로 설정한 결과이다. 여기에서 분석기간이란 기업고용의 성장을 측정하는 기간을 말한다. 즉 분석기간이 1년(또는 6년)이면 종속변수가 t 기의 변수일 때 설명변수는 $t-1$ 기(또는 $t-6$ 기)의 변수이다. 지면 절약을 위해

2002, 2005, 2008, 2011, 2014, 2017년의 결과 중에서 α_1 의 추정치만을 제시하였다. 각 그림에서 실선은 분석기간이 6년, 점선은 분석기간이 1년이다.

결과를 요약하면 다음과 같다. 첫째 분석기간이 1년일 때 α_1 의 추정치는 분위값이 낮을 때 크고 분위값이 높을 때 작다. 즉 α_1 의 추정치는 분위값 τ 에 대해 우하향하는데 이는 표한형·홍성철(2014)과 비슷하다. 반면에 분석기간이 6년일 때 α_1 의 추정치는 분위값이 낮을 때 작고 분위값이 중간일 때 크지만 분위값이 높을 때 작다. 즉 α_1 의 추정치는 분위값 τ 에 대해 역U자형인데, 김두순(2016)과 비슷하다. 따라서 분석기간이 6년인 경우 특징은 α_1 의 추정치가 분위값이 낮을 때와 분위값이 높을 때 더 작았다는 점이다. 분위값이 낮을 때 α_1 의 추정치가 더 작았다는 결과는 3장과 비슷하지만 분위값이 높을 때 α_1 의 추정치가 더 작았다는 결과는 3장과 다르다. 이 중에서 전자는 표본선택의 효과 때문인 것으로 설명할 수 있다.

둘째 분석기간이 1년이면 α_1 의 추정치는 분위값 τ 가 낮을 때 (+)가 되기도 한다. 이는 지브라의 법칙을 지지하는 결과에 해당한다. 반면에 분석기간이 6년이면, α_1 의 추정치는 모든 분위값 τ 에서 0보다 작으므로 지브라의 법칙과는 거리가 멀다. 셋째 그림에도 불구하고 분석기간이 6년인 경우 α_1 의 추정치는 시간이 흐를수록 모든 분위값에서 0에 근접한다. 즉 지브라의 법칙이 성립하는 것은 아니지만 그 방향으로 변화가 나타나고 있음을 알 수 있다.

3장에서 얻은 결과와 이 장에서 얻은 결과의 차이점을 살펴보자. 표본선택을 고려하고 분석기간을 6년으로 설정한 경우 3장에서 얻은 결과에 따르면 과거 고용 규모가 40~60% 분위에 속하는 기업과 80% 이상 분위에 속하는 기업의 경우 지브라의 법칙이 성립하는 결과를 얻었다. 하지만 이 장에서 얻은 결과에 따르면 지브라의 법칙이 성립하는 경우는 없었으며 다만 중간 정도의 분위값 τ 에서 지브라의 법칙에 더 가까운 결과를 얻었을 뿐이다. 그 중에서도 가장 두드러진 차이점은 분위값이 높을 때 지브라의 법칙과 상당한 거리를 보인다는 점이다.

물론 공통점도 있다. 하나는 3장의 결과에서 40% 이하인 기업의 경우 지브라의 법칙과 상당한 거리를 보였는데 이 장의 결과에서도 분위값이 낮을 때 지브라의 법칙과는 상당한 거리를 보였다는 점이다. 다른 하나는 최근에 나타나는 변화가 지브라의 법칙을 향하고 있다는 점이다. 즉 3장에서도 $\log(S_0)$ 의 추정계수는 최근 들어 증가하고 있다. 또한 〈Figure 1〉에서도 모든 분위값 τ 에서 α_1 의 추정치는 증가하

여 최근에는 0에 접근하고 있다. 이 두 가지 결과는 지브라의 법칙이 성립하는 방향으로 변화가 나타나고 있음을 시사하는 것이다.

이전의 국내 연구가 수행한 바 있는 분위회귀분석을 표본선택 하에 수행하고 이를 3장에서 얻은 결과와 종합하면 다음과 같이 요약할 수 있다. 즉 지브라의 법칙이 성립한다는 근거는 일부 분위 또는 일부 기간에서 나타나고 있을 뿐이므로 여전히 논쟁적이다. 하지만 두 가지 특징을 확인할 수 있었다. 첫째 표본선택을 고려한다면 고용규모가 작은 기업의 고용증가가 더 컸다는 견해는 근거가 약해 보인다. 둘째 최근에 나타나고 있는 전반적인 변화는 지브라의 법칙을 향하여 진행되고 있는 것으로 판단된다.

이제 본 연구결과와 이전 국내 연구결과가 다른 이유에 대해 생각해본다. 가장 큰 차이는 분석기간에 있다고 판단된다. 즉 분석기간이 1년인 경우 표본선택편의가 크지 않고 분석기간이 6년인 경우 표본선택편의가 크기 때문이다. 물론 그 이유는 1년 만에 파산하는 기업은 적은 반면에 6년 동안에 파산하는 기업은 많다는 데 있다. 따라서 분석기간이 1년인 경우 분위회귀분석결과는 표본선택을 고려하거나 또는 고려하지 않거나 크게 다르지 않았다. 하지만 분석기간이 6년인 경우 분위회귀분석결과는 표본선택에 의해 큰 영향을 받는다.

본 연구는 지브라의 법칙을 검증하기 위해 꽤 긴 기간 동안 기업고용이 어떻게 변화했는가를 분석하는 것이 좋다고 판단했다. 예를 들어 Audretsch et al. (2004)는 1987년과 1991년을 비교함으로써 분석기간을 4년으로 설정하였고, Calvo (2006)는 1990년과 2000년을 비교함으로써 분석기간을 10년으로 설정하였다. 그런데 분석기간이 길수록 파산하는 기업이 많아지게 되고 이로 인해 표본선택이 중요해지므로 이를 고려한 방법론을 이용했던 것이다.

V. 지브라의 법칙과 지프의 법칙

지금까지의 분석결과를 종합하면 우리나라 기업고용의 경우 지브라의 법칙이 성립했다고 보기는 어렵다. 일부 규모의 기업 또는 일부 기간의 경우 지브라의 법칙이 성립하거나 또는 가까운 결과를 얻기는 했지만 상이한 결과도 적지 않았다. 다만 3장과 4장의 결과에서 공통적으로 확인할 수 있는 결과 중 하나는 최근 변화가 지브라의 법칙을 향하고 있다는 점이다. 여기에서는 이러한 변화를 지프의 법칙과

관련하여 생각해 보고자 한다.

지브라의 법칙이 성립하면 - 그리고 마찰이 존재하다는 가정을 자명한 것으로 받아들이는다면 - 기업 규모의 분포는 지프의 법칙을 따르게 된다. 여기에서는 지프의 법칙에 대한 추정결과를 제시한다. 〈Table 10〉에는 MDIS 자료와 KIS value 자료를 이용하여 추정한 결과가 제시되어 있다. MDIS 자료는 2006~2016년 자료를 이용하였으며 KIS value 자료는 1987~2017년 자료를 이용하였다. 연도별로 방정식을 추정하였고 각 연도의 표본수와 지프 승수의 추정치가 제시되어 있다. 이 분포는 우측 꼬리를 표현하기 위한 것으로 상위 5%의 기업만을 이용한다.

〈Table 10〉 Estimation Results of Zipf's Law

year	MDIS data		KIS data	
	no. of samples	Zipf exponent	no. of samples	Zipf exponent
1987			102	1.43 (0.20)
1988			112	1.45 (0.19)
1989			120	1.43 (0.19)
1990			131	1.43 (0.18)
1991			148	1.39 (0.16)
1992			154	1.40 (0.16)
1993			154	1.38 (0.16)
1994			176	1.33 (0.14)
1995			236	1.29 (0.12)
1996			274	1.28 (0.11)
1997			340	1.24 (0.10)
1998			367	1.24 (0.09)
1999			410	1.24 (0.09)
2000			440	1.26 (0.08)
2001			473	1.28 (0.08)
2002			542	1.31 (0.08)
2003			597	1.30 (0.08)
2004			607	1.28 (0.07)
2005			617	1.26 (0.07)
2006	534	1.33 (0.08)	661	1.24 (0.07)
2007	533	1.31 (0.08)	710	1.26 (0.07)
2008	542	1.31 (0.08)	757	1.24 (0.06)
2009	540	1.28 (0.08)	792	1.21 (0.06)
2010	548	1.29 (0.08)	813	1.22 (0.06)
2011	581	1.28 (0.07)	845	1.21 (0.06)
2012	596	1.27 (0.07)	911	1.18 (0.06)
2013	607	1.26 (0.07)	957	1.15 (0.05)
2014	616	1.24 (0.07)	1,030	1.16 (0.05)
2015	619	1.25 (0.07)	1,066	1.17 (0.05)
2016	619	1.22 (0.07)	1,154	1.17 (0.05)
2017			1,272	1.18 (0.05)

Notes: 1) Regression is the estimation results of the model (4).

2) Standard errors are in parentheses.

MDIS 자료를 이용한 결과에 따르면 지프 승수는 2006년 1.33이었으나 이후 점차 낮아졌던 것으로 보이며 2016년에는 1.22이었다. 한편 KIS value 자료를 이용한 결과에 따르면 1990년대 초반까지 1.4를 넘었으나 이후 계속 낮아져서 2017년에는 1.18이었다. 지프 승수가 1보다 크기 때문에 고용 규모의 경우 지프의 법칙이 성립한다고 보기는 어렵다 - 이 값은 해외 연구에서 알려진 값에 비해서도 큰 편이므로 우리나라에서는 아직 지프의 법칙이 성립한다고 말할 수 없다. 하지만 지프 승수가 지속적으로 낮아져 왔음에 주목할 필요가 있다.

지프의 법칙이 성립한다고 보기는 어렵지만 시간이 흐를수록 지프 승수가 작아져 지프의 법칙에 가까운 방향으로 변화가 나타나고 있음은 흥미롭다. 기업 고용의 증가가 지브라의 법칙을 따르고 마찰이 존재한다면 고용 규모의 분포는 지프의 법칙을 따르게 될 것이기 때문이다. 즉 지브라의 법칙을 향한 변화는 지프의 법칙을 향한 변화를 설명할 수 있다.

물론 지프의 법칙이 성립한다고 해서 지브라의 법칙이 성립한다고 말할 수는 없다. 마찰이 있을 때 지브라의 법칙은 지프의 법칙을 위한 충분조건이지 필요조건은 아니기 때문이다. 다만 3장과 4장에서 얻은 결과(지브라의 법칙을 향한 변화)와 5장에서 얻은 결과(지프의 법칙을 향한 변화)를 연관시키는 것은 충분히 생각해 볼 수 있다. 즉 기업고용의 성장에서 지브라의 법칙을 향한 변화가 나타나고 있다면 기업 고용의 분포에서 지프의 법칙을 향한 변화가 나타나날 수 있다는 뜻이다. 따라서 우리나라 기업고용의 분포가 서서히 (선진국에서 관찰되고 있듯이) 지프의 법칙이 성립하는 방향으로 변화하고 있다면 그 이유를 찾기 위해 지브라의 법칙을 향한 변화가 나타나고 있는가를 분석하는 것은 중요한 이슈가 된다.

VI. 결론 및 시사점

본 연구는 기업의 성장이 규모와 독립적이며 무작위적이라는 지브라의 법칙을 검정하였다. 기업의 규모를 종업원수로 정의하면 지브라의 법칙은 고용 규모의 증가가 기존의 고용 규모와 무관하다는 뜻이다. 표본선택모형을 이용하여 다음과 같은 결과를 얻을 수 있었다.

첫째 기업의 고용 규모와 생존 가능성은 (+)의 관계를 갖는 것으로 나타났다. 이는 소기업의 생존 가능성이 낮음을 의미하는 것으로 생존기업만을 이용한 추정에서

소기업의 고용 증가가 과대하게 추정되는 표본선택편의를 초래한다. 이는 표본선택 모형을 이용한 검정결과에서도 확인되었다. 즉 생존기업만을 이용한 검정결과는 모형설정오류를 범하게 된다.

둘째 지브라의 법칙과 관련된 결과를 요약하면 다음과 같다. 기존의 고용규모와 고용의 증가의 관계는 여전히 논쟁적이다. 일부 기업군의 경우 지브라의 법칙이 성립하거나 그것에 가까운 결과를 얻기도 했지만 이에 반하는 결과도 적지 않았기 때문이다. 즉 표본선택을 고려하더라도 지브라의 법칙이 성립한다고 말할 수 있는 정도는 아니다. 하지만 이전의 연구와 같이 분위회귀분석을 수행하는 등 여러 방법을 이용할 때 표본선택을 고려한다면 최소한 다음의 두 가지 결과를 말할 수 있다. 하나는 기존의 고용 규모가 작은 영세기업의 고용 증가가 더 많았다고 보기는 어렵다. 고용 증가가 가장 컸다고 판단되는 기업은 기존의 고용 규모가 중위에 속하는 기업들이었다. 이는 고용의 증가와 기존의 고용 규모 간에 비선형관계가 존재함을 시사한다. 또 다른 하나는 최근 지브라의 법칙에 가까운 방향으로 전반적인 변화가 관찰되었다는 점이다. 이는 분위회귀분석결과에서도 확인되었는데, 모든 분위에 속하는 기업들에 대해 고용의 증가와 기존의 고용 규모 간의 관계는 더 강화되어 왔던 것으로 보인다.

이외의 주요 변수에 대한 결과는 다음과 같다. 셋째 기업의 연령은 생존 가능성과 (+)의 관계를 갖고 고용 증가와는 (-)의 관계를 갖는 것으로 나타났다. 기업의 연령과 고용 증가가 갖는 (-)의 관계는 Jovanovic 가설을 지지하는 것이다. 넷째 연구개발투자는 생존 가능성과 유의한 관계를 갖지 않지만, 고용 증가와는 (+)의 관계를 갖는다. 다섯째 재무적 안정성은 생존 가능성과 일관성있는 관계를 갖는다고 보기 어렵지만 고용 증가와는 (+)의 관계를 갖는다. 다섯째 관계회사의 존재 여부는 생존 가능성과 및 고용 증가와 모두 (+)의 관계를 갖는다.

끝으로 지브라의 법칙에 가까운 방향으로 최근의 변화가 나타나고 있다는 결과는 최근 우리나라 고용 규모 분포가 보이는 변화와 관련이 있을 것으로 보인다. 즉 지프의 법칙에 가까운 방향으로 변화가 나타나고 있는데, 이는 지브라의 법칙과 관련된 변화 때문일 수도 있다.

본 연구의 결과는 일자리 창출과 관련된 정책에 시사점을 제공한다. 그 중에서 지브라의 법칙 성립 여부는 일자리창출과 관련된 중소기업정책을 논의할 때 중요한 고려사항이 되어야 한다. 물론 일자리 창출이 기존의 고용규모에만 의존하는 것은

아니다. 예를 들어 연령이 낮은 기업의 고용 증가는 큰 편인데, 스타트업 기업을 육성하는 정책은 일자리 창출에 도움이 됨을 시사한다. 또한 일자리 창출을 위해서는 연구개발투자가 많은 혁신형 기업을 육성해야 하며, 이 기업들의 재무적 안정성이 중요함도 확인할 수 있었다. 끝으로 관계회사의 존재가 일자리 창출에 도움이 된다는 결과도 흥미해볼 만하다. 보다 광범위한 차원에서 일자리 창출을 위해 어떠한 정책과 노력이 필요한가에 대한 향후 연구를 기대해본다.

■ 참 고 문 헌

1. 김낙년, “한국의 소득집중도 추이와 국제비교, 1976-2010: 소득세 자료에 의한 접근,” 『經濟分析』, 제18권 제3호, 한국은행, 2012, pp.75-114.
(Translated in English) Kim, Nak Nyeon, “Income Concentration in Korea, 1976-2010: Evidence from Income Tax Statistics,” *Economic Analysis*, Vol. 18, No. 3, Bank of Korea, 2012, pp.75-114.
2. 김두순, 『기업의 생존과 고용성장에 관한 실증분석』, 한국고용정보원, 2016.
(Translated in English) Kim, Doo Soon, *An Empirical Study on the Survival of the firm and Employment Growth*, Korea Employment Information Service, 2016.
3. 박현숙 · 김주한, “한국제조업에서의 Gibrat 법칙과 Jovanovic가설에 대한 연구,” 『상업교육연구』, 제27권 제5호, 2013, pp.121-134.
(Translated in English) Park, Hyun Sook and Joo Han Kim, “Tests of Gibrat’s Law and Jovanovic Hypothesis for Korean Manufacturing Industry,” *Journal of Business Education*, Vol. 27, No. 5, 2013, pp.121-134.
4. 성낙일 · 김민창, “우리나라 광 · 제조업 부문 중소 사업체의 생존 및 성장: 실증분석,” 『산업조직연구』, 제21권 제4호, 2013, pp.19-50.
(Translated in English) Sung, Nakil and Min Chang Kim, “Survival and Growth of Small and Medium-sized Plants in the Korean Mining and Manufacturing Sector: Empirical Analysis,” *Korean Journal of Industrial Organization*, Vol. 21, No. 4, 2013, pp.19-50.
5. 윤충한 · 손준철, “경기지역 IT산업의 대 · 중소기업간 성장성 분석,” 『한국산학기술학회논문지』, 제15권 제4호, 2014, pp.2376-2381.
(Translated in English) Yoon, Choong-Han and Jong Chil Son, “The Comparative Analysis about the Firm Growth Between Large Enterprises and SMEs in the IT Companies located in Gyeonggi-do,” *Journal of the Korea Academia-Industrial Cooperation Society*, Vol. 15,

- No. 4, 2014, pp.2376-2381.
6. 이인권, 『한국기업의 성장동학에 관한 연구』, 한국경제연구원, 2002.
(Translated in English) Lee, In Kwon, *A Study on Korean Enterprises' Growth Dynamics*, Korean Economic Research Institute, 2002.
7. 표한형 · 홍성철, “분위회귀분석을 이용한 한국 기업의 성장결정요인 분석,” 『산업조직연구』, 제22권 제1호, 2014, pp.61-94.
(Translated in English) Pyo, Hannhyung and Sungcheol Hong, “The Determinants of Growth for Korean Firms using Quantile Fixed Regression,” *Korean Journal of Industrial Organization*, Vol. 22, No. 1, 2014, pp.61-94.
8. 최희갑, “Zipf 분포에 기초한 한국의 기업 규모 분포 분석,” 『통계연구』, 제11권 제2호, 2006, pp.73-95.
(Translated in English) Choi, Heegab, “Zipf Plots and the Size Distribution of Korean Firms,” *Journal of The Korean Official Statistics*, Vol. 11, No. 2, 2006, pp.73-95.
9. Arellano, M., and S. Bonhomme, “Quantile Selection Models with an Application to Understanding Changes in Wage Inequality,” *Econometrica*, Vol. 85, No. 1, 2017, pp.1-28.
10. _____, “Sample Selection in Quantile Regression: A Survey,” in *Handbook of Quantile Regression*, ed. by Koenker, R., V. Chernozhukov, X. He, and L. Peng, Chapman & Hall/CRC, 2018, pp.209-223.
11. Audretsch, D. B., L. Klomp, E. Santarelli, and A. R. Thurik, “Gibrat’s Law: Are the Services Different?” *Review of Industrial Organization*, Vol. 24, No. 3, 2004, pp.301-324.
12. Bentzen, J., E. S. Madsen, and V. Smith, “Do Firms’ Growth Rates Depend on Firm Size?” *Small Business Economics*, Vol. 39, No. 4, 2012, pp.937-947.
13. Buchinsky, M., “Quantile Regression with Sample Selection: Estimating Women Returns to Education in the U.S.,” *Empirical Economics*, Vol. 26, No. 1, 2001, pp.87-113.
14. Calvo, J. L., “Testing Gibrat’s Law for Small, Young and Innovating Firms,” *Small Business Economics*, Vol. 26, 2006, pp.117-123.
15. Daunfeldt, S., and D. Halvarsson, “Are High-growth Firms One-hit Wonders? Evidence from Sweden,” *Small Business Economics*, Vol. 44, No. 2, 2015, pp 361-383.
16. Davidson, R., and J. G. MacKinnon, *Econometric Theory and Methods*, Oxford University Press, 2004.
17. Evans, D. S., “Tests of Alternative Theories of Firm Growth,” *Journal of Political Economy*, Vol. 95, No. 4, 1987, pp.657-674.
18. Hall, B. H., “The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector,” *Journal of Industrial Economics*, Vol. 35, No. 4, 1987, pp.583-606.
19. Haltiwanger, J., R. S. Jarmin, and J. Miranda, “Who Creates Jobs? Small Versus Large Versus Young,” *Review of Economics and Statistics*, Vol. 95, No. 2, 2013, pp.347-361.
20. Koenker, R., “Quantile Selection Models: An R Vignette,” manuscript, 2017.

21. Lawless, M. "Age or Size? Contributions to Job Creation," *Small Business Economics*, Vol. 42, No. 4, 2014, pp.815-830.
22. Mansfield, E., "Entry, Gibrat's Law, Innovation, and the Growth of Firms," *American Economic Review*, Vol. 52, No. 5, 1962, pp.1023-1110.
23. Neumark, D., B. Wall, and J. Zhang, "Do Small Business Create More Jobs?: New Evidence for the United States from the National Establishment Time Series," *Review of Economics and Statistics*, Vol. 93 No. 1, 2011, pp.16-29.
24. Stam, E., "Growth Beyond Gibrat: Firm Growth Processes and Strategies," *Small Business Economics*, Vol. 35, No. 2, 2010, pp 129-135.
25. Wooldridge, J. W., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

부록 : 표본선택 하에서 분위회귀분석

표본선택이 있을 때 분위회귀분석을 수행하는 방법에 대하여 설명한다. Buchinsky (2001)가 제시한 방법이 있지만 일치성 (consistency) 을 보장하지 못한다는 문제가 지적되어 왔다. 최근에 이에 대한 대안으로 Arellano and Bonhomme (2017) 이 제시되었으므로 본 연구는 이 방법을 이용하였다. 이하에서는 Arellano and Bonhomme (2018)의 설명을 간략히 소개한다.

우선 Heckman의 표본선택모형을 요약하면 다음과 같다.

Y^* 의 일부만이 표본으로 선택되고 선택된 표본을 Y 라고 하자. 만약 Y^* 를 모두 표본으로 갖고 있다면 아래의 (A. 1)식을 추정하면 된다.

$$Y^* = X'\beta + \epsilon \quad (\text{A. 1})$$

하지만 Y^* 의 일부만을 표본으로 갖게 되기 때문에 아래와 같이 선택방정식과 결과방정식을 모두 이용해야 한다. (A. 2)식이 선택방정식이고 (A. 3)식이 결과방정식이다.

$$D = 1[\eta \leq W'\gamma] \quad (\text{A. 2})$$

$$E(Y^* | D = 1, W = w) = X'\beta + E(\epsilon | D = 1, W = w) = X'\beta + \Lambda(W),$$

$$\text{where } \Lambda(W) = E(\epsilon | \eta \leq W'\gamma, W = w) \quad (\text{A. 3})$$

probit모형에 의해 (A. 2)식을 추정하며, 이로부터 inverse Mill's ratio를 구하여 결과방정식인 (A. 3)식에 이용하는 것이 Heckman의 표본선택모형이다. 이 모형에서 $\Lambda(W)$ 이 중요한 역할을 하는데, inverse Mill's ratio에 의한 것으로 표본선택편의를 제거해준다. 끝으로 우리가 관찰할 수 있는 변수는 (Y_i, W_i, D_i) , $i = 1, 2, \dots, N$ 이며, $W_i = (X_i, M_i)$ 임을 확인해 둔다.

이제 표본선택이 있을 때 분위회귀분석모형을 살펴보자. (A. 1)에 해당하는 식이

(A. 4) 이다.

$$Y^* = X' \beta(U), \text{ where } U \sim \text{Uniform}(0,1) \quad (\text{A. 4})$$

위 식을 익숙한 분위회귀분석의 형태로 정리하면 다음의 (A. 5) 식과 같다.

$$Q(\tau, X) = X' \beta(\tau) \quad (\text{A. 5})$$

역시 Y^* 를 모두 표본으로 갖고 있다면 (A. 5) 식을 추정하는 것으로 충분하다. 하지만 Y^* 의 일부만을 표본으로 갖게 되기 때문에 분위회귀분석모형에서도 선택방정식과 결과방정식을 이용해야 한다. 아래의 (A. 6) 식이 선택방정식이고 (A. 8) 이 결과방정식이다. 물론 선택방정식인 (A. 6) 식은 probit모형이고 결과방정식인 (A. 8) 식은 분위회귀분석모형이다.

$$D = 1[V \leq P(W)],$$

where $P(W) = \Phi(W'\gamma)$, $V = \Phi(\eta)$, $\Phi()$ is standard normal cdf. (A. 6)

$$\Pr(Y^* \leq x' \beta | D = 1, W = w) = \Pr(U \leq \tau | V \leq \Phi(z'\gamma), W = w)$$

$$= G(\tau, \Phi(z'\gamma); \rho) \quad (\text{A. 7})$$

$$Q^s(\tau, W) = X' \beta(\tau^*(W)) \quad (\text{A. 8})$$

복잡해 보이지만, (A. 6) 식은 (A. 2) 식에 대응하는 선택방정식이고 (A. 8) 식은 (A. 3) 식에 대응하는 결과방정식이다. 여기에서 중요한 것은 (A. 7) 식인데, $G(\tau, \Phi(z'\gamma); \rho)$ 은 모수가 ρ 인 (U, V) 의 조건부 copula이다. 이것이 조건부 copula인 이유는 $V \leq \Phi(w'\gamma)$ 가 주어졌을 때 $U \leq \tau$ 의 copula이기 때문이다. 그런데 이 조건부 copula는 (A. 6) 식으로부터 초래되는 표본선택을 (A. 8) 식의 결과방정식에 반영하여 표본선택편의를 제거하는 역할을 한다. 즉 앞에서 설명한 Heckman의 모형에서 inverse Mill's ratio가 하는 역할을 수행한다.

따라서 (A. 6) 식을 추정한 후 이로부터 조건부 copula의 모수를 구한 후 이를 (A. 8)에 이용하면 된다. 이를 추정하기 위한 Arellano and Bonhomme(2017)의

절차는 3단계로 이루어진다. (Y_i, W_i, D_i) , $i = 1, 2, \dots, N$ 은 관찰할 수 있는 변수이며 $W_i = (X_i, M_i)$ 임을 환기하자.

$$\hat{\gamma} = \underset{a}{\operatorname{argmin}} \sum_{i=1}^N D_i \ln \Phi(W_i' a) + (1 - D_i) \ln \Phi(-W_i' a), \quad (\text{A. 9})$$

$$\hat{\rho} = \underset{c}{\operatorname{argmin}} \left\| \sum_{i=1}^N \sum_{l=1}^L D_i \varphi(\tau_l, W_i) [1[Y_i \leq X_i' \hat{\beta}(\tau_l, c)] - G(\tau_l, \Phi(W_i' \gamma; c))] \right\|, \\ \text{where } \|\cdot\| \text{ is Euclidean norm.} \quad (\text{A. 10})$$

$$\hat{\beta}(\tau) = \underset{b(\tau)}{\operatorname{argmin}} \sum_{i=1}^N D_i [\hat{G}_{\tau_i}(Y_i - X_i' b(\tau))^+ + (1 - \hat{G}_{\tau_i})(Y_i - X_i' b(\tau))^-], \\ \text{where } a^+ = \max(a, 0), \quad a^- = \max(-a, 0). \quad (\text{A. 11})$$

1단계에서 (A. 9) 식을 추정한 후 2단계에서 조건부 copula의 모수 ρ 를 추정하되 (A. 10) 식을 이용한다. ρ 를 구하였으므로 3단계에서 (A. 10) 식을 이용하여 $\beta(\tau)$ 를 추정한다.

(A. 9) 식과 (A. 10) 식은 각각 probit모형과 GMM을 이용하여 추정하고, (A. 11) 식은 분위회귀분석모형이므로 선형계획법(linear programming)을 이용하여 추정한다. (A. 11) 식을 보면 표본선택 하의 분위회귀분석모형이 어떻게 달라지는가를 확인할 수 있다. 즉 보통의 분위회귀분석모형에서 이용하는 분위값 τ 대신에 조건부 copula인 \hat{G}_{τ_i} 를 이용한다. 쉽게 말하면 분위값 $\tau=0.1, 0.2, \dots, 0.9$ 대신에 회전된(rotated) 분위값 \hat{G}_{τ_i} 를 이용한다고 보면 된다. 따라서 Arellano and Bonhomme (2017)은 (A. 11) 식을 rotated quantile regression이라고 부른다.

M. Arellano와 S. Bonhomme도 Matlab로 작성된 source code를 제공한다. 본 연구는 Koenker (2017)가 통계프로그램 R로 작성한 source code를 이용하였다. 그는 Arellano and Bonhomme (2017)를 그대로 수행하였다. 예를 들어 copula로는 Frank copula를 이용하였고 $\varphi(\tau_l, W_i)$ 도 Arellano and Bonhomme (2017)와 같은 함수를 이용하였다.

The Growth of Employment and Gibrat's Law*

Jong In Yoon**

Abstract

This article analyzes the growth of employment and tests Gibrat's law in the employment of Korean firms. It is reasonable to suppose that sample selection bias plays an important role in the estimated results. Therefore, we utilize the procedure in which a probit survival equation is first estimated to correct for sample selection bias. To test the validity of Gibrat's law, we use Heckman's two-step estimator and Arellano and Bonhomme(2017)'s quantile selection regression.

Most of the results by Heckman's procedure reject Gibrat's law on an aggregate level, which means that small firms grow faster than large firms. But Gibrat's law is confirmed more often than it is rejected for firms over a certain size(firms larger than 40% quantile). And for firms larger than 80% quantile, the results show that large firms grow faster than small firms. Results by quantile selection model show that the employment of medium-size firms growth faster than small and large firms. We think the results might highlight the validity of Zipf's law in the employment of Korean firms.

Key Words: Gibrat's law, sample selection bias(survivor bias), sample selection model, quantile selection regression, Zipf's law

JEL Classification: D21, L22, L25, L26, M13

Received: Oct. 10, 2018. Revised: Feb. 19, 2019. Accepted: March 22, 2019.

* I would like to thank anonymous referees for their helpful comments but all the remaining errors are mine. This research was supported by Baekseok University in 2019.

** Associate Professor, Division of Business and Commerce, Baekseok University, Munam-ro 76, Dongnam-gu, Cheonan 31065, Korea, Phone: +82-41-550-0525, e-mail: jiyoon@bu.ac.kr