통계연구(2018), 제23권 제2호, 74-95

소득집중도와 소득분포함수[1)]

윤종인[2)]

요약

Piketty and Saez(2003), 김낙년(2012)은 파레토 보간을 이용하여 소득집중도를 추정한다. 본 연구는 이에 더하여 자연대수 정규분포 보간을 이용하여 소득집중도를 추정하였다. 또한 두 보간법의 성과를 비교하기 위하여 다음과 같은 방법을 이용하였다. 우선 소득집중도가 알려진 소득계급에 대하여 두 보간법을 이용하여 소득집중도를 추정하였다. 그리고 두 보간법 중 어느 것이 알려진 소득집중도에 더 잘 부합하는가를 비교하였다. 주요 결과는 다음과 같다. 첫째, 파레토 보간과 자연대수 정규분포 보간을 이용하여 추정한 소득집중도의 추이는 크게 다르지 않았으며, 두 추정치 모두 최근까지 상승하였던 것으로 보인다. 둘째, 파레토 보간을 이용한 소득집중도 추정치는 자연대수 정규분포를 이용한 추정치보다 크다. 셋째 자연대수 정규분포 보간을 이용한 소득집중도 추정치가 파레토 보간을 이용한 것보다 실제 소득집중도에 더 잘 부합하는 것으로 나타났다. 따라서 소득집중도를 추정하는 방법으로는 자연대수 정규분포 보간이 더 좋은 방법이라고 판단된다. 그럼에도 불구하고 파레토 보간을 이용한 소득집중도가 크게 과대 추정되었다고 보기는 어렵다.

주요용어 : 소득집중도, 파레토 보간, 자연대수 정규분포 보간

1. 서론

Piketty and Saez(2003)는 Simon Kuznets의 가설, 즉 경제성장이 진행되는 초기 단계에는 소득불균등이 증가하지만 1인당 소득이 일정 수준에 도달하면 소득불균등이 완화된다는 가설을 비판하였다. 이후 그들의 연구는 여러 나라를 대상으로 확대되었는데, 1980년대 이후에는 소득불균등의 심화가 일반적인 경향이었다고 지적하였다. 국내연구에서도 김낙년(2012)은 비슷한 결과를 얻었다. 1990년대 말 이후 소득불균등이 심화되었는데, 상위 0.1%의 소득집중도는 1999년 2.1%에서 2010년 4.24%로, 상위 1%의 소득집중도는 1999년 7.55%에서 2010년 11.93%로 확대되었다.

Piketty and Saez(2003)는 소득불균등을 측정하는 새로운 방법을 제시하였다. 즉 조세당국의 자료를 이용하였고 소득집중도(top income share)라는 지표를 이용하였다. 물론 Kuznets(1955)도 상위 5%의 소득집중도를 인용했고, Feenberg and Poterba (1993)도 파레토분포(Pareto distribution)를 이용하여 소득집중도를 추정했지만, 그럼에도 불구하고 Piketty and Saez(2003)의 방법론적 기여는 적지 않다. 조세당국의 자료는 각 소득계급에 속한 소득자들의 수와 소득 합계를 제공한다. 따라서 이 자료를 이용하여 상위 소득자들의 소득집중도를 추정하려면 소득분포를 가정해야 한다. Feenberg and Poterba(1993), Piketty and Saez(2003), 김낙년(2012) 모두는 파레토분포를 이용하였는데, 이 분포는 두터운 우측 꼬리(right tail)를 잘 설명하는 것으로 알려져 있다. 하지만 소득의 분포함수로 파레토분포가 가장 적합한가 여부는 여전히 실증적인 문제이다. 원종학·성명재(2007)는 통계청의 가계동향 조사자료를 이용하여 우리나라의 경우 자연대수 정규분포(lognormal distribution)가 적합하다는 연구

---

1) 이 논문은 2018년도 백석대학교 대학연구비에 의하여 수행된 것이다.
2) 교신저자. 충남 천안시 동남구 문암로 76, 백석대학교 경상학부, 교수. E-mail: jiyoon@bu.ac.kr

결과를 제시하였다. 박명호·전병목(2014)도 상위 5%를 넘어서면 우리나라의 소득분포는 파레토분포를 따르지 않는다는 결과를 보고한 바 있다. 소득분포를 표현하기 위하여 어떤 분포함수가 적합한가는 응용연구에서 중요한 의의를 지닌다. 예를 들어 우리나라의 소득분포가 자연대수 정규분포를 따른다는 기존 문헌에 기초하여 한종석 외(2015)는 노동생산성 충격이 자연대수 정규분포를 따른다고 가정한 후 근로소득분포를 생성할 수 있음을 언급한 바 있다.

물론 상위 소득자들의 소득집중도를 추정하기 위하여 전 계층의 소득분포가 파레토분포를 따라야 하는 것은 아니다. 토론자들의 지적에 대해 김낙년(2016)이 답변한 바와 같이, 상위 소득자들의 소득집중도를 추정하는 데에는 우측 꼬리에서 파레토분포가 적합한가 여부가 중요하다. 따라서 중요한 것은 소득분포의 우측 꼬리에서 보간법(interpolation)을 이용하여 소득집중도를 추정할 때 파레토분포가 적합한가이다. 상위 소득자들의 소득집중도를 추정하기 위하여 이용할 수 있는 조세자료는 상위 소득구간의 누적자료뿐이다. 즉 개인의 소득자료를 이용할 수는 없으므로 소득분포함수를 직접 추정하는 일은 불가능하다. 그러므로 우측 꼬리의 소득분포에만 관심을 갖는다고 하더라도 어떤 함수가 적합한가를 판단하는 일은 쉬운 일이 아니다.

본 연구는 첫째 상위 소득구간의 누적자료만을 이용할 수 있고, 둘째 상위 소득자의 소득집중도를 추정하기 위하여 보간법을 이용하려 할 때 어떤 분포함수를 이용하는 것이 더 좋은가에 대한 실증적인 판단을 시도한다. 즉 소득집중도를 추정하기 위한 방법으로 파레토분포를 이용한 보간법과 자연대수 정규분포를 이용한 보간법 중 어느 것이 더 좋은가에 관한 실증적 판단을 시도하려는 것이다.

연구방법을 간략히 소개하면 다음과 같다. 우선 김낙년(2012)과 거의 같은 방법으로 우리나라의 소득집중도를 추정한다. 다만 파레토분포와 함께 자연대수 정규분포를 이용함으로써 두 가지 보간법에 의해 소득집중도를 추정한다. 문제는 어느 분포함수가 우리나라의 국세청 자료와 잘 부합되는가를 판단하는 일이다. 본 연구는 두 분포를 가정한 후 종합소득자료와 근로소득자료를 이용하여 특정 소득계급의 소득집중도를 추정하고 이를 실제 값과 비교하였다. 이에 따라 소득집중도의 실제 값과 가까운 추정치를 제시하는 분포가 실제의 소득분포를 잘 설명한다고 판단한다.

본 연구의 결과는 다음과 같다. 첫째 파레토분포를 가정하는 경우보다 자연대수 정규분포를 가정하는 경우 소득집중도는 더 낮았다. 둘째 종합소득자료를 이용하였을 때 0.1% 근방의 최상위 소득계급에 대해서만 파레토분포의 성과가 더 우수했던 것으로 나타났다. 반면에 종합소득자료를 이용한 1%, 5%, 10% 그리고 근로소득자료를 이용한 모든 경우에 자연대수 정규분포의 성과가 더 우수했다.

본 논문은 다음과 같이 진행된다. 우선 2절에서는 자료와 추정절차에 대해 설명한다. 3절에서는 파레토 보간과 자연대수 정규분포 보간에 대해 설명하고, 두 분포의 성과를 비교하는 방법에 대해 설명한다. 4절에서는 파레토 보간과 자연대수 정규분포 보간을 이용한 소득집중도 추정결과를 제시한다. 5절에서 두 분포의 성과를 비교하고 6절에서 결론을 맺는다.

2. 자료와 추정절차

본 연구는 Piketty and Saez(2003), 김낙년(2012)을 따라 국세청 및 국민계정 자료를 이용하여 소득집중도를 추정한다. 표본기간은 2005~2015년이다. 그 이전 자료의 경우 많은 가정이 필요할 뿐만 아니라 본 연구의 목적이 파레토분포와 자연대수 정규분포의 성과를 비교하는 것이기 때문에 표본기간을 2005년 이후로 정하였다. 따라서 2005~2010년의 경우에는 김낙년(2012)과 본 연구를 직접 비교할 수 있다.

파레토분포를 가정한 추정방법은 Piketty and Saez(2001)에 자세하게 설명되어 있으며, Piketty and Saez(2003) 이후에도 여러 나라의 자료에 맞추어 추정방법이 개선되어 왔다. 우리나라 자료에 대해서는 김낙년(2012)이 상세하게 설명하고 있다. 이하에서 설명하듯이 그는 많은 가정을 도입하였는데, 자료제약이 많다는 점을 감안하면 현실적인 선택이다. 이에 본 연구는 자료에 관한 한 그의 방법을 따르기로 한다. 중복을 피하되 본 연구의 방법을 도입하기 위한 수준에서 자료의 이용방법을 설명한다.

〈표2.1〉에 표본 통계의 요약이 제시되어 있다. 가장 먼저 통제총수(control total)에 해당하는 20세 이상 인구와 개인소득(personal income)이 제시되어 있다. 통제총수의 의미는 다음과 같다. 즉 이하에서 이용하는 인원수의 비율은 모두 20세 이상 인구로 나누어 준 것이며, 여러 소득의 비율도 모두 개인소득으로 나누어 준 것이다.

20세 이상 인구는 통계청 자료를 이용하였으며, 개인소득은 김낙년(2012)을 따라 국민계정과 산업연관표를 이용하여 구하였다. 국민계정 제도부문별 소득계정에서 개인부문의 소득, 즉 피용자보수, 영업잉여, 재산소득에서 고용주의 사회부담금, 자가주택의 귀속임료, 금융중개서비스를 차감하여 구하였다. 이렇게 구한 20세 이상 인구, 개인소득, 1인당 개인소득이 〈표2.1〉의 2~4열에 제시되어 있다.

한편 국세청 자료의 연도별 총계는 〈표2.1〉의 5~6열에 종합소득 자료, 7~8열에 근로소득 자료가 제시되어 있다. 종합소득 자료는 인원수와 종합소득금액, 근로소득 자료의 경우 인원수와 급여총계이다. 이로부터 몇 가지 사실을 알 수 있다. 첫째로 (종합소득금액-종합소득금액에 포함된 근로소득금액+근로소득 급여총계)/개인소득의 비율을 보면 2005년 52.2%에서 2015년 70.0%로 꾸준히 상승하여 왔다. 이는 국민소득 중 과세 대상이 꾸준히 확대되어 왔음을 의미한다. 종합소득금액에 포함된 근로소득금액은 별도로 추정해야 하는 것으로 이하에서 추정방법을 설명하게 될 것이다. 둘째 (근로소득 급여총계/종합소득금액)의 비율은 2005년 4.13배에서 서서히 하락하여 2015년 2.76배에 이른다. 많이 하락하기는 했지만 여전히 근로소득이 큰 비중을 차지한다. 이는 근로소득의 분포함수가 다른 소득의 분포함수보다 더 지배적임을 의미한다.

〈표 2.1〉 표본통계의 요약

연도	20세 이상 인구 (만 명)	개인소득 (십억 원)	1인당 개인소득 (만 원)	종합소득		근로소득	
				인원 (만 명)	종합소득금액 (십억 원)	인원 (만 명)	급여총계 (십억 원)
2005	3,588.12	512,404.4	1,428.06	227.95	54,103.3	610.67	223,690.1
2006	3,635.54	545,497.5	1,500.46	273.65	65,001.1	662.06	267,961.5
2007	3,678.53	576,107.3	1,566.14	307.44	77,124.1	774.87	294,936.7
2008	3,729.39	599,019.9	1,606.22	358.44	85,082.5	798.10	312,464.1
2009	3,770.98	618,591.3	1,640.40	357.08	90,225.7	854.12	322,430.4
2010	3,814.59	658,137.6	1,725.32	378.52	100,266.8	924.44	356,351.4
2011	3,869.66	696,680.3	1,800.36	395.67	111,446.4	993.50	392,163.7
2012	3,918.70	724,626.5	1,849.15	435.29	126,023.2	1,061.23	425,888.6
2013	3,966.36	755,527.7	1,904.84	456.47	134,370.0	1,123.86	458,819.0
2014	4,024.32	778,748.8	1,935.10	505.26	144,782.7	866.32	418,694.6
2015	4,076.28	814,481.0	1,998.10	548.27	164,098.6	922.92	452,614.8

주1) 개인소득은 한국은행이 발표하는 국민계정 제도부문별 소득계정에서 개인부문의 소득, 즉 피용자보수, 영업잉여, 재산소득에서 고용주의 사회부담금, 자가주택의 귀속임료, 금융중개서비스를 차감하여 구하였음.

주2) 종합소득 및 근로소득 자료는 국세청 국세통계연보로부터 구하였음. 근로소득 자료의 경우 인원수의 변동이 큰 편인데, 이는 결정세액이 없는 자를 포함하는 방식이 달라지기 때문임

셋째 (종합소득금액 중 근로소득금액/종합소득금액)은 2005년 18.7%에서 꾸준히 상승하여 2015년 28.5%에 이른다. 이것 역시 근로소득의 분포함수가 다른 소득의 분포함수보다 더 지배적임을 의미한다. 끝으로 2012~2013년의 경우 근로소득 인원의 변동이 큰 편인데, 국세청 자료가 하위 소득자의 자료를 조금씩 다른 방식으로 포함하거나 또는 제외하기 때문이다. 하지만 이하에서는 통제총수를 이용하여 상위 소득자의 소득집중도를 구할 것이기 때문에 하위 소득자의 자료는 결과에 영향을 미치지 않는다.

소득계급별 자료의 경우 2015년 예가 〈표2.2〉에 제시되어 있는데, 이를 중심으로 국세청 자료의 이용절차를 설명한다. 이 절차를 따르면 몇 차례 보간법을 이용하게 된다. Piketty and Saez(2003), 김낙년(2012)은 파레토 보간(Pareto interpolation)을 이용하였는데, 이에 더하여 본 연구는 자연대수 정규분포 보간을 함께 이용한다. 따라서 이하에서 보간법은 두 분포를 이용한 방법 모두를 의미한다. 보간법을 제외한 내용은 김낙년(2012)을 따랐으므로 그에 준하여 간략하게 설명한다.

전체 과정을 (A)~(E)로 구분하여 설명한다. (A)~(D)는 종합소득과 근로소득을 합산하여 단일한 소득계급을 가진 소득 자료를 구하는 절차이며, (E)는 이렇게 구한 자료를 이용하여 소득집중도를 추정하는 절차이다. (A)~(D)의 복잡한 절차가 필요한 이유는 두 가지이다. 첫째 종합소득 자료에는 근로소득이 포함되어 있으므로 종합소득금액과 근로소득 급여총계를 그대로 더하면 중복금액이 있다. 따라서 중복되는 근로소득을 배제하는 절차가 필요하다. 둘째 모든 소득금액을 종합소득금액 기준(또는 급여총액 기준)으로 전환

한 후 종합소득자료와 근로소득자료를 합산해야 한다.

(A) 종합소득금액에 포함된 근로소득은 중복되는 금액이므로 이를 근로소득자료에서 차감한다. 이를 위하여 종합소득 계급구간별로 근로소득의 비중이 일정하다고 가정한다. 예를 들어 〈표2.2〉 8천만 원~1억 원 소득계급의 경우 (종합소득금액에 포함된 근로소득/종합소득금액)은 40.4%이었는데, 이 비율을 이용한다. 이에 따라 소득계급구간이 달라지는데, 8천만 원~1억 원에서 3,232.8만 원~4,041.1만 원으로 바뀌게 된다. 즉 바뀐 소득계급구간에 따라 종합소득금액에 포함된 근로소득금액을 구하게 된다. 하지만 〈표2.2〉의 하단에 제시된 근로소득 자료를 보면 계급구간이 이와 다르다. 따라서 (종합소득 자료로부터 구한) 바뀐 소득계급구간 자료를 이용하여 (근로소득 자료의 계급구간인) 4천만 원 이상의 근로소득금액을 추정해야 한다. 이 때 보간법을 이용한다. 이와 같이 계급구간이 바뀔 때마다 보간법을 이용하여 계급구간을 일치시킨 후 합산 또는 차감해야 하는데, 이하 (B)~(D)에서 보간법을 이용한다는 것은 이런 의미이다. 결과적으로 종합소득금액에 포함된 근로소득금액과 근로소득자료의 근로소득금액의 소득계급구간이 같아졌으므로 후자에서 전자를 차감하면 중복금액이 배제된 근로소득금액을 구하게 된다.

(B) (A)에서 구한 근로소득 통계는 근로소득금액 기준이므로 이를 급여총액 기준으로 전환한다. 이 때 소득계급별 전환율은 (급여총액/근로소득금액)의 비율을 이용하되 소득계급구간이 달라지므로 이를 일치시키도록 보간법을 이용한다.

(C) 종합소득 통계에 포함된 근로소득을 급여총액 기준으로 전환해야 한다. 이 때 (B)에서 이용한 바 있는 전환율을 이용하는데, 이에 따라 소득계급구간도 달라진다. 따라서 이를 일치시키도록 보간법을 이용하여 추정한 후 합산한다.

(D) 종합소득 통계와 근로소득 통계의 중복부분이 배제되고 모든 금액이 급여총액 기준으로 전환되었으므로 이를 합산하면 된다. 다만 두 자료의 소득계급구간이 다르기 때문에 이를 일치시키도록 보간법을 이용하여 추정한 후 합산한다.

(E) 이제 소득계급별 인원수와 소득을 구하였으므로 보간법을 이용하여 상위 0.1%, 1%, 5%, 10%의 경계소득, 평균소득, 소득집중도를 추정한다.

(A)~(D)에서 이용하는 보간법과 (E)에서 이용하는 보간법을 구분해야 한다. 즉 파레토 보간의 경우 (A)~(D)에서 이용하는 방법과 (E)에서 이용하는 방법이 다르다. 그 이유는 전자의 경우 소득계급 구간의 상단과 하단을 모두 이용하지만 후자의 경우 소득계급구간의 상단과 하단 중 하나만을 이용하기 때문이다. 따라서 김낙년(2012)은 전자를 ‘파레토 보간 2’ 그리고 후자를 ‘파레토 보간’이라고 불렀다. 하지만 자연대수 정규분포 보간의 경우 소득계급구간의 상단과 하단을 모두 이용할 수밖에 없다. 따라서 자연대수 정규분포를 가정하는 경우 (A)~(E)에서 이용하는 보간법은 모두 같다. 이에 대한 자세한 내용은 이하 연구방법에서 설명한다.

〈표 2.2〉 2015년 자료

소득계급 (백만 원)	누적인원 (만 명)	누적소득 (십억 원)	$H(y)$ (%)	$G(y)$ (백만 원)	소득집중도 (%)
종합소득금액					
0 ~ 10	548.27	164,098.6	13.45	4.03	20.15
10 ~ 20	288.58	151,681.2	7.08	3.72	18.62
20 ~ 40	173.73	134,989.5	4.26	3.31	16.57
40 ~ 60	89.89	111,201.4	2.21	2.73	13.65
60 ~ 80	57.34	95,063.6	1.41	2.33	11.67
80 ~ 100	38.94	82,209.6	0.96	2.02	10.09
100 ~ 200	27.62	72,001.2	0.68	1.77	8.84
200 ~ 300	9.12	46,707.3	0.22	1.15	5.73
300 ~ 500	4.78	36,110.8	0.12	0.89	4.43
500 ~	2.05	25,678.9	0.05	0.63	3.15
근로소득 급여총계					
0 ~ 10	1,726.32	566,729.0	42.35	13.90	69.58
10 ~ 20	1,050.97	503,238.9	25.78	12.35	61.79
20 ~ 40	683.61	417,267.4	16.77	10.24	51.23
40 ~ 60	320.50	272,884.8	7.86	6.69	33.50
60 ~ 80	151.45	167,574.2	3.72	4.11	20.57
80 ~ 100	71.56	101,495.9	1.76	2.49	12.46
100 ~ 200	36.17	64,577.3	0.89	1.58	7.93
200 ~ 300	5.24	20,735.1	0.13	0.51	2.55
300 ~ 500	2.22	12,983.3	0.05	0.32	1.59
500 ~	0.73	7,123.4	0.02	0.17	0.87

주) 근로소득 급여총계의 경우 결정세액이 없는 자를 포함하였으므로 소득계급 1천만 원 이하인 구간의 인원 및 근로소득 급여총계가 〈표2.1〉과 다르다

3. 파레토 보간과 자연대수 정규분포 보간의 방법

조세자료는 특정 소득계급에 속한 소득자의 수와 소득의 합계를 제공한다. 그러므로 소득집중도를 추정하려면 소득분포함수를 가정하고 보간법(interpolation)을 이용해야 한다. 본 연구는 파레토분포와 자연대수 정규분포를 가정하여 소득집중도를 추정하고 두 추정 결과를 비교할 것이다. 이하에서는 이를 파레토분포 보간과 자연대수 정규분포 보간으로 부른다. 본 절에서는 우선 소득집중도를 추정하는 방법에 대해 설명하고, 이어서 두 분포함수 보간의 성과를 비교하는 방법에 대해 설명한다.

먼저 실제 자료를 이용하여 소득집중도를 구하는 절차를 개관하기로 하자. 〈표2.1〉에는 2015년의 경우 소득계급별로 종합소득금액과 근로소득 급여총계의 자료가 제시되어 있

다. 통제총수에 해당하는 20세 이상 인구를 N 그리고 전체 개인소득을 Y라고 하면 2015년에는 각각 4,076.3만 명과 814.4조 원이었다.

근로소득자료에서 근로소득금액이 1억 원 이상인 소득계급의 예를 대상으로 설명한다. 인원과 금액은 누적된 값으로 근로소득금액이 1억 원 이상인 소득자는 총 36.2만 명이며 이들의 소득은 총 64조 5,773억 원이다. $H(y)$와 $G(y)$의 개념에 대해서는 이하에서 설명하기로 하고, 여기에서는 계산방법만 간단히 소개한다. $H(y)$는 소득이 y 이상인 소득자가 N에서 차지하는 비중으로 36.2만 명/4,076.3만 명=0.89%가 된다. 한편 $G(y)$는 64조 5,773억 원/4,076.3만 명=158.4만 원이 된다. 그렇다면 소득집중도는 $NG(y)/Y$ =7.93%가 된다. 이를 정리하면 다음과 같이 말할 수 있다. 상위 0.89%에 속하는 소득자는 소득이 1억 원 이상이며 소득집중도는 7.93%이다. 여기에서 1억 원을 경계소득(threshold income)이라 부른다.

물론 상위 0.1%, 1%, 5%, 10%의 소득집중도는 자료에 없다. 따라서 우리는 소득분포함수를 가정하고 이에 근거하여 보간해야 한다. 예를 들어 보자. 상위 0.89%에 속하는 소득자는 소득이 1억 원 이상이며, 소득집중도는 7.93%이다. 또한 상위 1.76%에 속하는 소득자는 소득이 8천만 원 이상이며, 소득집중도는 12.46%이다. 그런데 상위 1%는 0.89%~1.76% 범위에 속하므로 이 소득자들의 경계소득도 8천만 원~1억 원의 구간에 있을 것이며, 소득집중도도 7.93%~12.46%의 구간에 있을 것이다. 이를 찾는 방법이 보간법이고 이를 위해 소득분포함수의 가정이 필요하다.

3.1 소득집중도 추정방법

본 연구는 누적자료를 이용하여 소득집중도를 추정할 때 어느 분포함수를 이용하는 것이 좋은가에 관심을 갖는다. 바꾸어 말하면, 어느 분포함수를 가정한 보간법을 이용하는 것이 좋은가를 다루고자 한다. 비교하려는 분포함수는 파레토분포와 자연대수 정규분포이다. 물론 소득분포함수로 제안된 것은 매우 많다[Chotikapanich(2008)]. 하지만 이 두 분포함수가 가장 널리 이용되고 있으며, 국내에서의 토론도 이 두 분포함수를 대상으로 전개되었기 때문에 파레토분포와 자연대수 정규분포만을 이용한다.

물론 파레토분포와 자연대수 정규분포는 모두 우측의 두터운 꼬리를 표현할 수 있는 분포함수이다. 본 연구의 목적에 비추어 기존 연구로부터 알려진 두 분포함수의 차이를 언급한다면 다음과 같다. 첫째 전 계층의 소득분포를 표현하기에 더 적합한 분포함수는 자연대수 정규분포이지만 우측꼬리를 표현하는 데에는 파레토분포가 더 많이 이용되고 있다. 둘째 파레토분포는 매우 간단한 형태를 지니고 있지만 자연대수 정규분포는 상대적으로 복잡한 형태를 지니고 있다. 파레토분포가 간단한 형태를 갖고 있다는 장점은 – 소득구간의 누적자료와 같이 – 제한된 자료만을 가지고 있을 때 특히 유용하다. 이 두 가지 이유 때문에 소득집중도를 추정하기 위한 기존 연구는 자연대수 정규분포가 아니라 파레토분포를 이용하였던 셈이다.

이제 파레토분포를 이용하여 소득집중도를 추정하는 방법, 즉 '파레토 보간'에 대해 설명한다. 여기에도 몇 가지 방법이 있지만 Piketty and Saez(2004), 김낙년(2012)가 개선했던 방법으로 2절의 추정절차 (E)에서 이용하는 방법이다.

파레토분포가 널리 이용되어 왔던 이유는 두터운 우측 꼬리(right tail)를 잘 반영할 뿐만 아니라 그 형태가 단순하여 쉽게 이용할 수 있기 때문이다. 우선 파레토분포의 누적확률분포함수와 확률밀도함수는 식 (3.1) 및 식 (3.2)과 같다.

$$F(y) = 1-\left(\frac{k}{y}\right)^{\alpha}, \text{ where } k > 0,\ \alpha > 1 \tag{3.1}$$

$$f(y) = \frac{\alpha k^{\alpha}}{y^{1+\alpha}} \tag{3.2}$$

따라서 소득이 y 이상인 소득자가 차지하는 확률은 아래의 $H(y)$ 이다.

$$H(y) = \int_{y}^{\infty} f(x)dx = 1-F(y) = \left(\frac{k}{y}\right)^{\alpha} \tag{3.3}$$

이제 (3.2)식을 이용하면 부분기댓값(partial expectation)에 해당되는 $G(y)$ 를 아래와 같이 구할 수 있다.

$$G(y) = \int_{y}^{\infty} x f(x)dx = \frac{\alpha}{\alpha-1}\frac{k^{\alpha}}{y^{1-\alpha}} \tag{3.4}$$

$G(y)$ 는 소득이 y 이상인 구간의 기댓값이다.

한편 소득이 y 이상인 소득자들의 평균소득은 $G(y)/H(y)$가 되는데, 이는 (3.3)식과 (3.4)식으로부터 다음과 같이 도출할 수 있다.

$$G(y)/H(y) = \frac{\alpha}{\alpha-1}y \tag{3.5}$$

파레토분포를 이용하여 소득집중도를 추정하는 절차는 다음과 같다.

첫째 모수인 α와 k를 알아야 하는데, 이를 위해 첫째 식 (3.3)과 식 (3.5)을 이용하여 α와 k를 구한다. 즉 상위 1%의 소득집중도를 구하려면 1%와 가장 가까운 $H(y)$ 를 찾고 이로부터 y를 구한다. 〈표2.2〉의 예를 보면 근로소득의 경우 1%와 가장 가까운 $H(y)$ 는 0.89%이므로 y는 1억 원이 된다. 이렇게 구한 y를 식 (3.5)에 대입하면 α를 구할 수 있고 이어서 식 (3.3)에 y와 α를 대입하면 k를 구할 수 있다.

둘째 α와 k를 알고 있으므로 식 (3.3)에서 $H(y)$ =p%로 놓으면, 상위 p%를 정하는 경계소득 y를 구할 수 있다. 즉 소득이 y 이상이면 상위 p%에 해당되는 소득자들이다. 이제 경계소득 y를 식 (3.4)에 대입하면 $G(y)$ 를 구할 수 있다.

셋째 $NG(y)$는 부분합(partial sum)으로 소득이 y 이상인 소득자들의 소득의 합이다. 따라서 소득집중도는 $NG(y)/Y$이 된다.

자연대수 정규분포도 두터운 꼬리를 잘 반영하지만 파레토분포에 비해서는 다루기 조금 복잡한 형태이다. 자연대수 정규분포함수는 아래의 식 (3.6)과 같다.

$$F(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]dx \tag{3.6}$$

물론 자연대수 정규분포의 경우에도 소득이 y 이상인 소득자가 차지하는 확률은$H(y)=1-F(y)$이다. 한편 부분기댓값 $G(y)$는 다음과 같다.

$$G(y) = \int_{y}^{\infty} x f(x)\,dx = \exp\left(\mu + \frac{1}{2}\sigma^2\right)\Phi\left[\frac{\mu+\sigma^2-\ln y}{\sigma}\right],$$
$$\text{where } \Phi[\ \ ] \text{ is standard normal distribution function} \tag{3.7}$$

자연대수 정규분포를 이용하여 소득집중도를 추정하는 절차는 다음과 같다. 첫째 모수인 μ와 σ를 알아야 하는데, 이 절차가 파레토분포의 경우보다 복잡하다. 즉 파레토분포의 경우 하나의 소득계급 자료만으로 모수를 구할 수 있지만 자연대수분포의 경우에는 두 개의 소득계급 자료가 필요하다. 즉 상위 1%의 소득집중도를 구하려면 1%와 가장 가까운 하위의 $H(y_i)$와 상위의 $H(y_{i+1})$를 찾아야 한다. 〈표2.2〉의 예를 보면 근로소득의 경우 1%와 가장 가까운 하위의 $H(y_i)$는 1.76%이고 상위의 $H(y_{i+1})$는 0.89%이다. 그리고 각각에 해당하는 소득 y_i은 8천만 원이고 y_{i+1}은 1억 원이 된다. 이제 소득이 y_i 이상 또는, y_{i+1} 이상인 소득자의 누적확률은 아래의 식 (3.8)과 같다.

$$H(y_i) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_i} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]dx \tag{3.8a}$$

$$H(y_{i+1}) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_{i+1}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]dx \tag{3.8b}$$

또한 소득이 y_i과 y_{i+1}이상인 구간의 부분기댓값은 아래의 식 (3.9)과 같다.

$$G(y_i) = \int_{y_i}^{\infty} x f(x)\,dx = \exp\left(\mu + \frac{1}{2}\sigma^2\right)\Phi\left[\frac{\mu+\sigma^2-\ln y_i}{\sigma}\right] \tag{3.9a}$$

$$G(y_{i+1}) = \int_{y_{i+1}}^{\infty} x f(x)\,dx = \exp\left(\mu + \frac{1}{2}\sigma^2\right)\Phi\left[\frac{\mu+\sigma^2-\ln y_{i+1}}{\sigma}\right] \tag{3.9b}$$

이제 y_i, y_{i+1}와 $H(y_i)$, $H(y_{i+1})$을 알고 있으므로 식 (3.8)을 이용하면 μ와 σ를 구할 수 있다. 또한 y_i, y_{i+1}와 $G(y_i)$, $G(y_{i+1})$를 알고 있으므로 식 (3.9)을 이용하면 μ와 σ를 구할 수 있다. 다만 파레토분포의 경우 닫힌 해(closed form solution)를 구할 수 있었

던 것과 달리 자연대수 정규분포의 경우에는 수치해(numerical solution)를 구할 수밖에 없다. 이를 위하여 본 연구는 뉴턴-랩슨 기법(Newton Rhapson method)을 이용하였다[Hamilton(1994)]. 이 기법에 대해서는 부록에서 간략하게 설명한다. 또한 식 (3.8) 또는 식 (3.9)을 이용할 수 있는데, 본 연구는 파레토분포의 경우와 맞추기 위하여 식 (3.9)을 이용하여 μ와 σ를 추정하였다.

둘째 μ와 σ를 알고 있으므로 식 (3.6)과 식 (3.7)을 이용하여 $G(y)$를 추정한다. 우선 식 (3.6)을 이용하여 $H(y) = 1 - F(y)$ = p%로 놓으면, 상위 p%를 정하는 경계소득 y를 구할 수 있다. 이제 경계소득 y를 식 (3.7)에 대입하면 $G(y)$를 구할 수 있다.

셋째 파레토분포의 경우와 마찬가지로 자연대수 정규분포의 경우에도 $NG(y)$는 소득이 y 이상인 소득자들의 소득의 합이며, 소득집중도는 $NG(y)/Y$이 된다.

3.2 파레토 보간과 자연대수 정규분포 보간의 비교방법

파레토분포 보간의 경우에는 모수인 α와 k를 구하기 위하여 하나의 소득계급 자료만을 이용해도 된다. 이것이 Piketty and Saez(2004)가 방법론에서 보여준 공헌 중 하나이다. 하지만 자연대수 정규분포의 경우 모수인 μ와 σ를 구하려면 두 개의 소득계급 자료가 필요하다. 따라서 파레토분포 보간과 자연대수 정규분포 보간의 성과를 비교하려면 두 개의 소득계급 자료를 이용하는 방법을 써야 한다.

파레토분포 하에서 두 개의 소득계급 자료를 이용하는 방법은 이미 Feenberg and Poterba(1993)에서 시도된 바 있는데, Piketty and Saez(2002)는 이 방법을 개선하였다. 그들은 소득집중도를 추정할 때 하나의 소득계급 자료만 이용하는 방법을 이용하였지만 자료의 중간추계과정에서 소득계급 구간이 바뀔 때에는 두 개의 소득계급자료를 이용하는 방법을 썼다. 2절의 설명을 보완하면, (E)에서는 하나의 소득계급 자료만 이용하는 방법을 이용하였지만, (A)~(D)에서는 두 개의 소득계급자료를 이용하는 방법을 썼다. 김낙년(2012)은 전자를 ‘파레토 보간’, 후자를 ‘파레토보간 2’라고 불렀다. 여기에서는 파레토분포 보간 2를 설명한다.

우선 식 (3.3)에서 y를 구해서 식 (3.4)에 대입하면 다음의 식 (3.10)이 도출된다.

$$G(y)^{\alpha} = cH(y)^{\alpha-1}, \text{ where } c = \left(\frac{\alpha k}{\alpha-1}\right)^{\alpha} \tag{3.10}$$

두 개의 소득계급을 y_i, y_{i+2}라 하면 식 (3.11)이 된다.

$$G(y_i)^{\alpha} = cH(y_i)^{\alpha-1} \tag{3.11a}$$

$$G(y_{i+2})^{\alpha} = cH(y_{i+2})^{\alpha-1} \tag{3.11b}$$

〈표2.2〉에서 예를 들면 y_i는 8천만 원이고 y_{i+2}는 2억 원이다. 따라서 아래의 식 (3.12)을 도출할 수 있다.

$$[G(y_i)/G(y_{i+2})]^{\alpha} = [H(y_i)/H(y_{i+2})]^{\alpha-1} \tag{3.12}$$

식 (3.12)에서 다른 변수의 값을 모두 알고 있으므로 α를 구할 수 있다.

다음으로 식 (3.3)과 식 (3.4)을 각각 이용하면 아래의 식 (3.13)을 도출할 수 있다.

$$H(y_{i+1})/H(y_i) = [y_i/y_{i+1}]^{\alpha} \tag{3.13a}$$

$$G(y_{i+1})/G(y_i) = [y_i/y_{i+1}]^{\alpha-1} \tag{3.13b}$$

〈표2.2〉에서 예를 들면 y_i는 8천만 원이고 y_{i+1}은 1억 원이다. 다른 변수의 값을 알고 있으므로 식 (3.13a)과 식 (3.13b)을 이용하면 $H(y_{i+1})$와 $G(y_{i+1})$을 구할 수 있다. 따라서 우리는 경계소득이 y_{i+1}일 때 소득집중도를 추정할 수 있다.

쉽게 말하면 소득계급 y_i, y_{i+2}의 알려진 자료를 이용하여 파레토분포 하에서 소득계급 y_{i+1}의 소득집중도 추정치를 구한 셈이다. 그런데 우리는 이미 소득계급 y_{i+1}의 소득집중도 실제 값을 알고 있으므로 파레토분포 보간에 의해 구한 추정치와 실제 값을 비교하면 파레토분포가 얼마나 타당한가를 판단할 수 있다.

파레토분포 보간 2에 대응하는 자연대수 정규분포 보간은 앞의 소득집중도 추정방법에서 설명한 것과 같다. 다만 μ와 σ를 구하는 과정에서 이용하는 소득계급이 y_i, y_{i+2}이고 그 이후에는 소득계급 y_{i+1}을 이용한다는 점만 다르다. 결국 우리는 소득계급 y_i, y_{i+2}의 알려진 자료를 이용하여 자연대수 정규분포 하에서 소득계급 y_{i+1}의 소득집중도 추정치를 구한다. 그 다음에는 이 추정치를 소득계급 y_{i+1}의 소득집중도 실제 값과 비교하면 된다. 추정치와 실제 값의 차이가 작다면 자연대수 정규분포는 타당하다고 볼 수 있다.

〈표2.2〉의 예를 보면 y_i=8천만 원, y_{i+2}=2억 원의 자료를 이용하여 y_{i+1}=1억 원의 소득집중도 추정치를 구할 수 있다. 이 추정치를 y_{i+1}=1억 원의 소득집중도 실제 값인 7.93%와 비교하면 된다. 물론 이와 같은 방법으로 소득집중도 추정치를 구할 수 있는 소득계급 y_{i+1}은 1천만 원~3억 원일 때이다.

본 연구는 파레토분포 보간과 자연대수 정규분포 보간을 이용하여 어떤 보간법에 의한 추정치가 실제 값과 가까운가를 비교한다. 실제 값과 더 가까운 추정치를 산출하는 보간법이 더 우수하다고 볼 수 있을 것이다. 물론 이 방법이 명시적인 검정결과를 제시하는 것은 아니다. 다만 제한된 자료만을 이용하여 소득집중도를 추정할 때 그 결과를 비교하기 위하여 시도할 수 있는 방법일 뿐이다.

4. 소득집중도 추정결과

파레토분포를 가정한 경우의 추정치와 자연대수 정규분포를 가정한 경우의 추정치를 제시하고 이를 비교한다. 이 결과는 〈표4.1〉에 제시되어 있다. 파레토분포를 가정한 경우의 추정치는 2~5열, 자연대수 정규분포를 가정한 경우의 추정치는 6~9열에 제시되어 있다. 각각 상위 0.1%, 1%, 5%, 10%에 대한 추정치를 제시한다. 한편 상단에는 경계소득, 중단에는 평균소득, 하단에는 소득집중도의 추정치가 제시되어 있다.

파레토 보간을 이용한 결과는 김낙년(2012)과 비교할 수 있다. 그는 파레토 보간을 이용하여 2010년까지의 추정결과를 제시한 바 있다. 따라서 2010년 자료를 이용한 추정치를 비교하기로 한다. 경계소득은 김낙년(2012)에 따르면 0.1%, 1%, 5%, 10%의 경우 2억 8,388.6만 원, 1억 494.8만 원, 5,706.3만 원, 3,620.1만 원이었으나 본 연구결과에서는 3억 59.9만 원, 1억 636.5만 원, 5,875.2만 원, 4,082.3만원으로 본 연구의 추정치가 더 크다. 평균소득은 김낙년(2012)에 따르면 0.1%, 1%, 5%, 10%의 경우 6억 9,384.8만 원, 1억 9,552.9만 원, 9,713.4만 원, 7,192.9만 원이었으나 본 연구결과에서는 7억 1,226.8만 원, 2억 664.0만 원, 1억 532.1만 원, 8,197.0만원으로 본 연구의 추정치가 더 크다. 소득집중도는 김낙년(2012)에 따르면 0.1%, 1%, 5%, 10%의 경우 4.24%, 11.93%, 29.64%, 43.90%이었으나 본 연구의 결과에서는 4.13%, 11.97%, 30.52%, 47.51%이었으며 0.1%의 결과를 제외하면 본 연구의 추정치가 더 크다. 요약하면 김낙년(2012)와 비교할 때 큰 차이는 없지만 본 연구의 경계소득, 평균소득, 소득집중도가 더 큰 편이라고 볼 수 있다.

다음으로 파레토 보간을 이용하여 추정한 소득집중도의 2005~2015년 추이를 살펴보자. 0.1%의 소득집중도는 2005년 3.03%에서 2006년 3.57%로 크게 상승한 후 2015년 4.49%까지 지속적으로 상승하고 있다. 1%의 소득집중도는 2005년 11.94% 이후 2011년 12%를 넘었고 2015년에는 12.97%에 이르고 있다. 5%의 소득집중도는 2005년 42.08%로 높았으나 이후 2009년 29.59%까지 낮아졌고 서서히 상승하여 2015년 33.18%에 이른다. 10%의 소득집중도는 2005년 51.60%에서 2010년 47.51%까지 낮아졌으나 이후 상승하여 2015년 51.08%까지 높아졌다. 요약하면 파레토 보간을 이용할 경우 소득집중도는 2010년 이후에도 전반적으로 상승하였던 것으로 보인다.

이제 자연대수 정규분포 보간을 이용한 추정결과를 살펴보자. 파레토 보간을 이용한 결과와 비교하고자 하는데, 간결한 설명을 위해 2015년 결과를 대상으로 경계소득과 평균소득의 추정치부터 살펴본다. 경계소득은 파레토 보간을 가정할 경우 0.1%, 1%, 5%, 10%의 경우 4억 1,326.4만 원, 1억 1,924.0만 원, 7,223.2만 원, 4,883.1만 원이었으나 자연대수 정규분포를 가정할 경우 4억 3,530.2만 원, 1억 4,690.6만 원, 7,621.0만 원, 5,428.2만원으로 후자의 추정치가 더 크다. 평균소득은 파레토 분포를 가정할 경우 0.1%, 1%, 5%, 10%의 경우 8억 9,801.6만 원, 2억 5,910.8만 원, 1억 3,257.5만 원, 1억 206.9만 원이었으나 자연대수 정규분포를 가정할 경우 7억 369.0만 원, 2억 7,019.8만 원, 1억 2,142.0만 원, 9,335.6만원으로 전자의 추정치가 더 크다. 요약하면 경계소득은 자연대수 정규분포 보간을 이용할 경우 더 크고, 평균소득은 파레토 보간을 이용할 경우 더 크다.

〈표 4.1〉 소득집중도 추정결과

구분	파레토분포 보간				자연대수 정규분포 보간			
	0.1%	1%	5%	10%	0.1%	1%	5%	10%
경계소득(천 원)								
2005	20,597.77	6,468.31	3,031.88	1,212.93	20,540.80	8,035.77	3,908.48	1,193.08
2006	23,592.12	9,088.90	5,671.23	4,074.61	27,909.78	9,964.68	5,899.33	4,235.27
2007	24,983.40	8,877.62	5,200.82	3,754.20	26,984.50	10,026.08	5,730.92	4,221.25
2008	26,455.06	9,382.15	5,485.55	3,896.07	28,668.79	10,642.04	5,935.42	3,731.01
2009	27,055.85	9,451.65	5,490.50	3,899.92	29,186.60	10,687.58	5,941.19	4,347.80
2010	30,059.86	10,636.53	5,875.22	4,082.27	32,868.55	11,928.40	6,274.98	4,439.40
2011	33,837.89	11,649.88	6,195.17	4,253.09	36,634.29	12,976.16	6,620.62	4,701.02
2012	34,607.82	12,194.56	6,499.08	4,510.10	37,717.14	13,517.57	6,914.09	5,102.90
2013	36,271.67	12,823.84	6,742.99	4,646.59	39,419.74	14,106.69	7,152.63	5,306.89
2014	38,819.21	11,356.12	6,898.62	4,716.76	41,345.86	13,756.60	7,299.01	5,194.58
2015	41,326.35	11,924.03	7,223.24	4,883.18	43,530.15	14,690.55	7,621.02	5,428.21
평균소득(천 원)								
2005	43,253.61	17,057.44	12,019.80	7,369.07	43,463.84	11,039.27	8,082.78	7,343.74
2006	53,622.95	16,472.62	9,036.64	7,374.74	48,167.98	16,933.09	8,624.81	6,850.67
2007	61,987.20	17,877.33	9,267.22	7,329.23	62,361.59	18,025.40	8,691.83	6,776.10
2008	64,036.09	18,594.90	9,674.32	7,637.42	64,198.15	18,757.20	9,025.34	7,003.47
2009	64,490.05	18,750.96	9,707.69	7,643.90	64,692.53	18,920.83	9,059.06	7,065.54
2010	71,226.77	20,653.94	10,532.06	8,196.99	70,694.20	20,501.93	9,759.86	7,568.27
2011	78,574.74	22,469.58	11,323.85	8,776.45	77,986.05	22,070.00	10,419.22	8,031.14
2012	79,078.16	22,969.44	11,706.54	9,119.28	78,030.02	22,517.48	10,766.04	8,268.41
2013	80,629.49	23,751.92	12,146.02	9,492.15	79,348.90	23,217.23	11,167.94	8,530.30
2014	83,270.23	24,359.77	12,569.46	9,682.09	67,356.88	25,560.50	11,563.05	8,891.53
2015	89,801.59	25,910.76	13,257.52	10,206.95	70,368.97	27,019.75	12,141.95	9,335.58
소득집중도(%)								
2005	3.03	11.94	42.08	51.60	3.04	7.73	28.30	51.42
2006	3.57	10.98	30.11	49.15	3.21	11.29	28.74	45.66
2007	3.96	11.41	29.59	46.80	3.98	11.51	27.75	43.27
2008	3.99	11.58	30.12	47.55	4.00	11.68	28.10	43.60
2009	3.93	11.43	29.59	46.60	3.94	11.53	27.61	43.07
2010	4.13	11.97	30.52	47.51	4.10	11.88	28.28	43.87
2011	4.36	12.48	31.45	48.75	4.33	12.26	28.94	44.61
2012	4.28	12.42	31.65	49.32	4.22	12.18	29.11	44.71
2013	4.23	12.47	31.88	49.83	4.17	12.19	29.31	44.78
2014	4.30	12.59	32.48	50.03	3.48	13.21	29.88	45.95
2015	4.49	12.97	33.18	51.08	3.52	13.52	30.38	46.72

다음으로 자연대수 정규분포를 가정한 경우 소득집중도의 추이는 다음과 같다. 0.1%의 소득집중도는 2005년 3.04%에서 2011년 4.33%까지 상승하였으나 이후 하락하여 2015년 3.52%에 머물고 있다. 1%의 소득집중도는 2005년 7.73%에서 지속적으로 상승하여 2011년 12%를 넘었고 2015년에는 13.52%에 이르고 있다. 5%의 소득집중도는 2005년 28.30% 이후 큰 변동이 없었으나 서서히 상승하여 2015년 30.38%에 이른다. 10%의 소득집중도는 2005년 51.42%에서 2009년 43.07%까지 낮아졌으나 이후 서서히 상승하여 2015년 46.72%까지 높아졌다. 요약하면 0.1%의 소득집중도는 최근 낮아지면서 안정된 편이지만 1%, 5%, 10%의 소득집중도는 상승하고 있다.

따라서 소득집중도의 추이를 보면 자연대수 정규분포를 가정한 경우 0.1%의 소득집중도만이 최근 하락하고 있다. 그 이외의 추정치는 파레토 분포를 가정하거나 자연대수 정규분포를 가정하거나 최근까지 꾸준히 상승하여 왔던 것으로 보인다.

끝으로 가장 중요한 일은 파레토 분포를 가정한 경우와 자연대수 정규분포를 가정한 경우 소득집중도의 수준을 비교하는 것이다. 2010~2015년 소득집중도는 자연대수 정규분포를 가정한 경우 더 낮은 편이었다. 예를 들어 2015년의 경우 0.1%, 5%, 10%의 소득집중도는 자연대수 정규분포를 가정한 경우 0.97%p, 2.8%p, 4.36%p 더 낮았다. 반면에 1%의 소득집중도는 자연대수 정규분포를 가정한 경우 0.55%p 더 높았다. 요약하면 파레토 분포를 가정한 경우의 소득집중도가 전반적으로 더 큰 편이라고 볼 수 있다.

두 분포를 이용한 차이가 가장 두드러진 경우는 0.1% 소득집중도이다. 첫째 이미 언급한 바와 같이 0.1% 소득집중도는 파레토분포를 가정한 경우 계속해서 상승하고 있지만 자연대수 정규분포를 가정한 경우 최근 하락하면서 안정되고 있는 것으로 보인다. 둘째 두 분포를 가정한 경우 0.1% 소득집중도의 차이가 가장 크다. 2015년을 예로 들면 파레토분포를 가정한 경우 4.49%, 자연대수 정규분포를 가정한 경우 3.52%로 그 차이는 0.97%p이다. 하지만 3.52%보다 4.49%는 27.5%나 더 크다. 반면에 10% 소득집중도의 경우 파레토분포를 가정한 경우 51.08%, 자연대수 정규분포를 가정한 경우 46.72%로 그 차이는 4.36%p이지만 51.08%는 46.72%보다 9.3% 더 클 뿐이다. 따라서 두 분포를 가정할 경우 소득집중도의 추이와 수준에서 가장 두드러진 차이를 보이는 집단은 상위 0.1%일 때이다. 가장 우측 꼬리에서 차이가 가장 크다.

요약하면 다음과 같다. 첫째 파레토 보간을 이용하거나 자연대수 정규분포 보간을 이용해도 최근 소득집중도는 전반적으로 상승하여 왔던 것으로 보인다. 다만 자연대수 정규분포를 이용할 때 0.1%의 소득집중도가 최근 하락하고 있다는 점이 특징적일 뿐이이다. 둘째 파레토 보간을 이용한 경우보다 자연대수 정규분포 보간을 이용한 경우 소득집중도의 추이는 더 낮은 편이었다. 그 차이가 가장 큰 경우는 상위 0.1%, 즉 소득분포의 매우 우측 꼬리에서 나타났다.

파레토분포 대신에 자연대수 정규분포를 가정하더라도 소득집중도의 차이가 크다고 볼 수 있을지는 의문이다. 다만 상위 0.1%의 소득집중도만이 다른 움직임을 보였다는 결과는 주목할 만한 것이다. 이제 남은 과제는 두 분포 중 어느 것의 성과가 더 좋은가를 판단하는 일이다. 다음 절에서 살펴보기로 한다.

5. 파레토 보간와 자연대수 정규분포 보간의 비교

이 절에서는 소득집중도에 초점을 맞추어 파레토 보간을 이용한 경우와 자연대수 정규분포 보간을 이용한 경우의 성과를 비교한다. 이 결과는 〈표5.1〉의 종합소득을 이용한 결과와 〈표5.2〉의 근로소득을 이용한 결과에서 제시되고 있다.

표에서 $H(y)$ 와 소득집중도의 열은 주어진 종합소득 및 근로소득 자료로부터 직접 계산된 실제 값이다. 파레토와 자연대수 정규분포라는 열에는 각각 파레토 보간을 이용한 경우와 자연대수 정규분포 보간을 이용한 경우 소득집중도 추정치가 제시되어 있다. 따라서 실제 값인 소득집중도와 파레토 및 자연대수 정규분포의 열에 제시된 추정치를 비교하면 파레토와 자연대수 정규분포의 성과를 비교할 수 있다. 자연대수 정규분포의 추정치에는 * 표시가 되어 있는데, 이는 자연대수 정규분포를 이용한 추정치와 실제 값의 차이가 파레토분포를 이용한 추정치와 실제 값의 차이보다 적음을 나타낸다. 즉 자연대수 정규분포의 성과가 더 좋다는 뜻이다.

우선 〈표5.1〉 종합소득 자료를 이용한 추정치를 보면 $H(y)$ 가 클 때는 자연대수 정규분포 보간의 성과가 더 좋고, $H(y)$ 가 작을 때는 파레토 보간의 성과가 더 좋음을 알 수 있다. 즉 상위 10%, 5%, 1%의 경우에는 자연대수 정규분포 보간의 성과가 더 좋고, 상위 0.1%의 경우에는 파레토 보간의 성과가 더 좋다. 이는 파레토분포의 성과가 소득분포의 매우 우측 끝 꼬리에서 좋음을 의미한다. 다만 2015년의 경우 전반적으로 파레토 보간의 성과가 더 좋다. 하지만 〈표5.2〉 근로소득 자료를 이용한 추정치를 보면 거의 모든 경우에 자연대수 정규분포의 성과가 더 좋다. 특히 상위 0.1% 정도의 가장 우측 꼬리에서도 그러하다는 점은 주목할 만하다.

종합소득 자료와 근로소득 자료를 이용한 경우 차이가 있는데, 그 이유는 다음과 같이 설명할 수 있다. 첫째 $H(y)$ 를 계산할 때 통제총수는 같지만 종합소득금액과 근로소득금액 중에는 후자가 더 크다. 따라서 종합소득 자료를 이용한 경우에 $H(y)$ 가 더 적은 편이다. 즉 〈표5.1〉의 경우 소득분포에서 조금 더 우측 꼬리에 있는 자료를 이용하게 된다. 따라서 종합소득 자료를 이용한 경우 조금 더 우측 꼬리를 대상으로 하기 때문에 파레토분포의 성과가 더 좋을 수 있다. 둘째 종합소득의 분포가 근로소득의 분포보다 더 두터운 우측 꼬리를 가질 수 있다. 따라서 우측 꼬리로 갈수록 파레토분포의 성과가 더 좋다면 종합소득 자료를 이용할 때에는 파레토 보간의 성과가 더 좋고, 근로소득 자료를 이용할 때에는 자연대수 정규분포 보간의 성과가 더 좋을 것이다. 첫째 이유보다는 둘째 이유가 더 설득력 있는 것으로 보이는데, 〈표5.2〉의 결과에서도 $H(y)$ 가 0.1% 근방인 경우가 많았기 때문이다. 즉 근로소득 자료를 이용한 경우에도 대단히 우측에 있는 꼬리를 대상으로 하였으며 그 경우에도 자연대수 정규분포의 성과가 더 좋다.

이상의 결과를 요약하면 파레토 보간의 성과가 자연대수 정규분포 보간의 성과보다 더 좋다고 말하기는 어렵다. 오히려 자연대수 정규분포 보간의 성과가 더 좋다고 판단된다. 다만 예외적 결과도 있었는데, 종합소득을 이용하여 상위 0.1%의 소득집중도를 추정할 경우이다. 이 때에는 파레토 보간의 성과가 더 좋았는데, 이는 종합소득의 소득집중도가 상대적으로 더 높았기 때문인 것으로 판단된다.

〈표 5.1〉 파레토분포와 자연대수 정규분포 보간의 비교(종합소득)

연도	계급	$H(y)$	소득집중도	파레토	자연대수 정규분포	연도	계급	$H(y)$	소득집중도	파레토	자연대수 정규분포
2006	20	1.92	9.14	8.94	9.05 *	2011	20	2.96	12.93	12.69	12.82 *
	40	0.97	7.36	7.05	7.17 *		40	1.56	10.76	10.35	10.60 *
	60	0.57	6.05	5.98	6.03 *		60	0.99	9.20	9.06	9.08 *
	80	0.38	5.16	5.13	5.15 *		80	0.66	7.93	7.89	7.92 *
	100	0.26	4.50	4.48	4.52		100	0.47	6.97	6.99	7.06
	200	0.08	2.84	2.88	2.93		200	0.16	4.65	4.61	4.74
	300	0.04	2.19	2.20	2.26		300	0.08	3.64	3.63	3.69
2007	20	2.21	10.47	10.25	10.38 *	2012	20	3.55	14.27	14.00	14.03 *
	40	1.15	8.57	8.22	8.36 *		40	1.90	11.74	11.25	11.40 *
	60	0.69	7.13	7.04	7.10 *		60	1.13	9.70	9.58	9.63 *
	80	0.46	6.11	6.08	6.10 *		80	0.73	8.22	8.19	8.21 *
	100	0.32	5.35	5.34	5.38		100	0.50	7.13	7.14	7.26
	200	0.10	3.42	3.46	3.52		200	0.17	4.66	4.64	4.76
	300	0.05	2.65	2.66	2.72		300	0.09	3.62	3.62	3.68
2008	20	2.42	11.00	10.74	10.90 *	2013	20	3.67	14.63	14.35	14.42 *
	40	1.24	8.96	8.58	8.76 *		40	1.97	12.12	11.61	11.86 *
	60	0.75	7.45	7.35	7.41 *		60	1.22	10.20	10.03	10.09 *
	80	0.49	6.36	6.33	6.35 *		80	0.81	8.72	8.66	8.68 *
	100	0.35	5.55	5.54	5.59		100	0.57	7.57	7.58	7.67
	200	0.11	3.54	3.58	3.64		200	0.18	4.86	4.83	4.96
	300	0.06	2.74	2.75	2.81		300	0.10	3.72	3.72	3.78
2009	20	2.61	11.46	11.21	11.31 *	2014	20	3.85	15.22	14.94	14.99 *
	40	1.33	9.26	8.88	9.06 *		40	1.98	12.51	12.06	12.39 *
	60	0.79	7.66	7.56	7.61 *		60	1.25	10.67	10.52	10.58 *
	80	0.52	6.51	6.48	6.51 *		80	0.86	9.25	9.19	9.20 *
	100	0.36	5.67	5.67	5.72		100	0.61	8.11	8.15	8.18
	200	0.12	3.61	3.63	3.70		200	0.20	5.26	5.21	5.35
	300	0.06	2.77	2.78	2.83		300	0.11	4.03	4.03	4.07
2010	20	2.79	12.14	11.90	12.03 *	2015	20	4.26	16.57	16.36	16.33
	40	1.47	10.00	9.58	9.76 *		40	2.21	13.65	13.23	13.53 *
	60	0.89	8.35	8.24	8.29 *		60	1.41	11.67	11.56	11.55
	80	0.58	7.14	7.11	7.14 *		80	0.96	10.09	10.09	10.04
	100	0.41	6.26	6.25	6.31		100	0.68	8.84	8.91	8.93
	200	0.13	4.05	4.07	4.14		200	0.22	5.73	5.73	5.87
	300	0.07	3.13	3.14	3.20		300	0.12	4.43	4.45	4.48

주1) 계급의 단위는 백만 원이며, 나머지 수치의 단위는 %임.
주2) $H(y)$ 와 소득집중도의 열은 주어진 종합소득자료로부터 직접 계산된 실제 값이며, 파레토와 자연대수 정규분포라는 열은 각각 파레토 보간을 이용한 경우와 자연대수 정규분포 보간을 이용한 경우 소득집중도 추정치임. 자연대수 정규분포의 열에 표시된 *은 자연대수 정규분포를 이용한 추정치와 실제 값의 차이가 파레토분포를 이용한 추정치와 실제 값의 차이보다 적음을 나타냄.

〈표 5.2〉 파레토분포와 자연대수 정규분포 보간의 비교(근로소득)

연도	계급	$H(y)$	소득 집중도	파레토	자연대수 정규분포	연도	계급	$H(y)$	소득 집중도	파레토	자연대수 정규분포
2006	20	9.90	36.59	76.65	32.70 *	2011	20	12.49	42.11	73.31	38.50 *
	40	3.78	19.55	32.05	16.76 *		40	5.68	26.43	38.83	28.52 *
	60	1.22	8.88	15.11	9.22 *		60	2.39	14.92	22.83	14.82 *
	80	0.49	4.82	7.11	4.96 *		80	1.12	8.98	12.73	8.79 *
	100	0.23	2.98	3.77	3.51 *		100	0.52	5.49	7.03	6.53 *
	200	0.04	1.10	1.41	1.32 *		200	0.08	1.99	2.51	2.35 *
	300	0.02	0.76	0.83	0.78 *		300	0.04	1.31	1.43	1.37 *
2007	20	9.61	35.74	78.28	32.24 *	2012	20	13.63	44.84	76.53	41.47 *
	40	3.68	19.27	33.47	16.86 *		40	6.22	28.31	41.23	25.69 *
	60	1.24	9.15	16.00	9.48 *		60	2.68	16.22	24.70	16.04 *
	80	0.51	5.16	7.75	5.37 *		80	1.26	9.78	13.93	9.55 *
	100	0.25	3.37	4.23	3.89 *		100	0.59	5.96	7.70	7.07 *
	200	0.04	1.40	1.79	1.64 *		200	0.09	2.09	2.63	2.45 *
	300	0.02	1.01	1.10	1.05 *		300	0.04	1.32	1.46	1.39 *
2008	20	10.21	37.55	78.07	34.20 *	2013	20	14.42	46.58	78.72	44.34 *
	40	4.20	21.40	35.78	18.72 *		40	6.72	29.88	43.19	27.21 *
	60	1.51	10.62	18.00	10.62 *		60	2.96	17.42	26.28	17.18 *
	80	0.60	5.80	8.83	6.05 *		80	1.41	10.59	15.05	10.38 *
	100	0.29	3.68	4.74	4.33 *		100	0.67	6.53	8.39	7.64 *
	200	0.05	1.50	1.89	1.74 *		200	0.10	2.24	2.83	2.64 *
	300	0.02	1.05	1.15	1.09 *		300	0.04	1.40	1.56	1.47 *
2009	20	10.10	36.36	75.60	33.11 *	2014	20	16.08	49.96	76.56	46.23 *
	40	4.15	20.73	34.65	0.00		40	7.32	31.91	43.78	34.42 *
	60	1.49	10.29	17.43	10.28 *		60	3.28	18.88	27.60	18.59 *
	80	0.60	5.61	8.55	5.85 *		80	1.58	11.58	16.15	11.41 *
	100	0.28	3.57	4.59	4.19 *		100	0.78	7.28	9.10	8.35 *
	200	0.05	1.45	1.83	1.69 *		200	0.12	2.43	3.05	2.92 *
	300	0.02	1.01	1.11	1.05 *		300	0.05	1.54	1.66	1.60 *
2010	20	11.58	39.96	71.95	36.35 *	2015	20	16.77	51.23	77.26	47.65 *
	40	5.09	24.31	36.51	21.78 *		40	7.86	33.50	45.36	31.19 *
	60	2.07	13.29	20.47	12.98 *		60	3.72	20.57	29.48	19.80 *
	80	0.88	7.54	10.97	7.67 *		80	1.76	12.46	17.46	12.45 *
	100	0.42	4.72	5.98	5.54 *		100	0.89	7.93	9.90	8.95 *
	200	0.07	1.79	2.22	2.07 *		200	0.13	2.55	3.22	3.09 *
	300	0.03	1.17	1.28	1.22 *		300	0.05	1.59	1.71	1.64 *

주1) 계급의 단위는 백만 원이며, 나머지 수치의 단위는 %임.

주2) $H(y)$와 소득집중도의 열은 주어진 근로소득자료로부터 직접 계산된 실제 값이며, 파레토와 자연대수 정규분포라는 열은 각각 파레토 보간을 이용한 경우와 자연대수 정규분포 보간을 이용한 경우 소득집중도 추정치임. 자연대수 정규분포의 열에 표시된 *은 자연대수 정규분포를 이용한 추정치와 실제 값의 차이가 파레토분포를 이용한 추정치와 실제 값의 차이보다 적음을 나타냄.

결국 우리는 종합소득과 근로소득을 합산한 전체 소득의 집중도에 관심을 갖는다. 그런데 종합소득을 이용한 상위 0.1%의 경우에만 파레토 보간의 성과가 좋고, 그 이외의 경우에는 종합소득이나 근로소득 중 어느 것을 이용하더라도 그리고 상위 몇 %이더라도 자연대수 정규분포의 성과가 좋다. 그러므로 종합소득과 근로소득의 상대적인 비중은 중요한 문제이다. 종합소득과 근로소득 중 어느 것이 더 큰가에 따라 전체 소득을 이용할 때 두 보간법의 성과가 달라질 것이기 때문이다. 즉 종합소득이 많을수록 파레토 보간의 성과가 조금 더 좋을 수 있고, 반면에 근로소득이 많을수록 자연대수 정규분포의 성과가 더 좋을 수 있다.

앞의 〈표2.1〉을 보면 (근로소득 급여총계/종합소득금액)의 비율은 2005년 4.13배에서 서서히 하락하여 2015년 2.76배에 이르는 것으로 나타났다. 또한 (종합소득금액 중 근로소득금액/종합소득금액)은 2005년 18.7%에서 꾸준히 상승하여 2015년 28.5%에 이르는데, 이는 근로소득자 중 종합소득과세 대상자가 증가하였음을 의미한다. 이를 이용하여 (근로소득 급여총계/종합소득금액 중 근로소득금액을 제외한 금액)의 비율을 계산해 보면 2005년 5.09배에서 2015년 3.86배로 하락하였다. 요약하면 2005년 이후 근로소득의 상대적 비중은 감소하여 왔던 것으로 보인다. 하지만 그럼에도 불구하고 여전히 근로소득의 상대적 비중은 압도적이다.

결론적으로 다음과 같이 말할 수 있다. 첫째 근로소득의 상대적 비중이 압도적이라는 사실과 둘째 종합소득을 이용한 상위 0.1%의 경우에만 파레토 보간의 성과가 좋고 그 이외의 경우에는 자연대수 정규분포 보간의 성과가 더 좋다는 결과는 다음을 시사한다. 종합소득과 근로소득을 합산한 전체 소득의 집중도를 추정할 때에는 자연대수 정규분포 보간의 성과가 파레토 보간의 성과보다 더 좋을 것이다.

6. 결론 및 시사점

Piketty and Saez(2003), 김낙년(2012)은 조세자료를 이용하여 상위 소득자의 소득집중도를 추정하였다. 조세자료는 각 소득계급에 속한 소득자의 수와 소득 합계를 제공하며, 이를 이용하여 소득집중도를 추정하기 위하여 그들은 파레토분포를 가정하였다. 하지만 이에 대한 반론도 제기되었다. 원종학·성명재(2007)는 우리나라의 경우 자연대수 정규분포(lognormal distribution)가 적합하다는 연구결과를 제시한 바 있다. 이에 본 연구는 어떤 분포가 타당한가에 대한 검증을 시도하였다.

우선 본 연구는 두 분포를 가정하여 소득집중도를 추정하였다. 또한 어느 분포의 추정치가 실제 소득집중도에 잘 부합하는가를 검증하기 위해 다음과 같은 방법을 이용하였다. 즉 파레토분포와 자연대수 정규분포를 가정한 후 (소득집중도가 이미 알려진) 특정 소득계급의 소득집중도를 추정하였고 이렇게 구한 추정치를 알려진 실제 값과 비교하는 방식을 이용하였다. 따라서 소득집중도의 실제 값과 추정치가 가까운 분포함수가 실제의 소득분포를 잘 설명한다고 판단한다.

주요 실증분석결과는 다음과 같다. 첫째 파레토 보간을 이용하거나 자연대수 정규분포 보간을 이용해도 최근 소득집중도는 전반적으로 상승하여 왔던 것으로 보인다. 다만 자연

대수 정규분포를 이용할 때 0.1%의 소득집중도가 최근 하락하고 있다는 점이 특징적이다. 둘째 파레토 분포를 가정한 경우보다 자연대수 정규분포를 가정한 경우 소득집중도의 수준은 더 낮은 편이었다. 예를 들어 2015년의 경우 0.1%, 5%, 10%의 소득집중도는 자연대수 정규분포를 가정한 경우 0.97%p, 2.8%p, 4.36%p 더 낮았다. 두 분포를 이용한 결과의 차이가 가장 큰 경우는 상위 0.1%, 즉 소득분포의 매우 우측 꼬리에서 나타났다. 셋째 종합소득을 이용한 상위 0.1%의 경우에만 파레토 보간의 성과가 좋고, 그 이외의 경우에는 – 종합소득이나 근로소득 중 어느 것을 이용하더라도 그리고 상위 몇 %이더라도 – 자연대수 정규분포의 성과가 좋다.

이에 따라 소득집중도를 추정할 때 두 분포 중 어느 것을 이용해야 하는가는 몇 가지 실증적 사실에 근거해야 한다. 첫째 근로소득의 상대적 비중이 압도적이라는 사실과 둘째 종합소득을 이용한 상위 0.1%의 경우에만 파레토 보간의 성과가 좋고 그 이외의 경우에는 자연대수 정규분포 보간의 성과가 더 좋다는 결과는 다음을 시사한다. 종합소득과 근로소득을 합산한 전체 소득의 집중도를 추정할 때에는 자연대수 정규분포 보간의 성과가 파레토 보간의 성과보다 더 좋을 것이다.

물론 그렇다고 해서 두 분포를 이용한 소득집중도의 차이가 크다고 볼 수 있을지에 대해서는 논의의 여지가 남아 있다. 자연대수 정규분포를 가정하더라도 소득집중도는 낮은 편이 아니며, 최근에도 상승하였던 것으로 보이기 때문이다. 다만 상위 0.1%의 소득집중도가 최근 하락 또는 안정되었다는 결과만이 특이하다. 그럼에도 불구하고 파레토분포를 이용한 소득집중도 추정치가 과장될 수 있으므로 향후에는 자연대수 정규분포를 이용하는 방법도 고려해야 할 것이다.

끝으로 기존 국내 연구에서 제시된 바와 같이 전 계층의 소득분포함수로 자연대수 정규분포함수가 적합하다면 본 연구의 결과는 자연대수 정규분포의 활용가능성을 더 높여준다고 말할 수 있다. 자연대수 정규분포는 전 계층의 소득분포를 표현하기에 적합할 뿐만 아니라 상위 계층의 소득분포도 잘 표현할 수 있기 때문에 소득분포함수를 가정하는 응용연구의 경우 자연대수 정규분포의 가정은 큰 무리가 없는 선택이다.

(2018년 4월 5일 접수, 2018년 5월 10일 수정, 2018년 6월 1일 채택)

감사의 글

본 논문을 심사해 주신 익명의 심사자께 감사드립니다.

부록

자연대수정규분포의 모수를 추정하기 위한 수치최적화기법(Numerical optimization)으로 뉴턴-랩슨 기법에 대하여 설명한다. 자세한 내용은 Hamilton(1994, pp.133~141)을 참조할 수 있으며, 여기에서는 기본 원리를 설명하는 것으로 국한한다.

어떤 비선형함수 $f(x)$의 해가 r이라고 하자. 닫힌 해를 구할 수 없을 때 다음과 같은 최적화기법을 이용하여 이 방정식의 수치해를 구한다. 우선 x_0가 r에 대한 좋은 추측치(guess)라고 하자. 그렇다면 r=x_0+h과 같을 것이고 h는 작은 값이 된다. 이제 이 함수에 대한 선형근사치는 다음과 같다

$$0 = f(r) = f(x_0 + h) = f(x_0) + hf'(x_0)$$

따라서 이를 정리하면 다음과 같다.

$$h \approx -\frac{f(x_0)}{f'(x_0)}$$

이를 r=x_0+h에 대입하면 다음과 같다.

$$r \approx x_0 - \frac{f(x_0)}{f'(x_0)}$$

이 원리를 이용하여 다음과 같은 반복과정(iteration)을 수행하면 된다. 즉 추측치 x_0에서 시작하여 x_1을 구한다.

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

이어서 다음을 이용하여 x_2, x_3, x_4,...를 구한다.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

h가 매우 작은 값을 갖게 되면 x_{n+1}을 수치해로 볼 수 있다. 물론 자연대수 정규분포의 경우 모수가 2개이므로 반복과정은 더 복잡하지만 기본원리는 위와 같다.

참고문헌

김낙년(2012), 한국의 소득집중도 추이와 국제비교, 1976-2010: 소득세자료에 의한 접근, 〈경제분석〉 18(3), 75-114

김낙년(2016), 한국의 개인소득 분포: 소득세자료에 의한 접근, 1976-2010: 소득세자료에 의한 접근, 〈한국경제의 분석〉 22(3), 147-208

박명호, 전병목(2014), 소득분배 변화와 정책과제: 소득집중도와 소득이동성 분석을 중심으로, 한국조세재정연구원

원종학, 성명재(2007), 소득분배 격차 확대의 원인과 정책 대응방향, 한국조세연구원

한종석, 윤성주, 최승문(2015), 근로소득 불평등 변화에 대한 실증분석과 정책적 함의, 한국조세재정연구원

Chotikapanich, Duangkamon(2008), *Modeling Income Distributions and Lorenz Curves*, Springer

Feenberg, Daniel R. and James M. Poterba(1993), Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns, *Tax Policy and the Economy*, 145-177

Hamilton James, D.(1994), *Time Series Analysis*, Princeton University Press

Kuznets, Simon(1955), Economic Growth and Economic Inequality, *American Economic Review*, 1–28

Piketty, Thomas, and Emmanuel Saez(2001), Income Inequality in the United States, 1913-1998, Working Paper

Piketty, Thomas, and Emmanuel Saez(2003), Income Inequality in the United States, 1913–1998, *Quarterly Journal of Economics*, 1-39

Top Income Share and Income Distribution Function

Jong In Yoon[1)]

Abstract

Piketty and Saez(2003), Kim(2012) estimate the top income shares by Pareto interpolation. We estimate the top income shares by Pareto and Lognormal interpolation. Also, to evaluate the performances of two interpolations, we estimate the top income shares of percentiles which is known, and then compare the estimated top income shares to the known top income share. If the estimated top income shares by Lognormal interpolation fits the known top income share better than those by Pareto interpolation, we can conclude that Lognormal intepolation is better than Pareto interpolation.

Main results are as follows. First, estimates of the top income shares by Pareto and Lognormal interpolation show similar trends and have risen recently. Second, estimates of the top income shares by Lognormal interpolation are lower than those by Pareto interpolation. Third, the estimated top income shares by Lognormal interpolation fit the known top income shares better than those by Pareto interpolation. We conclude that estimates by Lognormal interpolation is better than those by Pareto interpolation but differences between them is not so large.

Key words : Top Income Share, Pareto Interpolation, Lognormal Interpolation

1) (Corresponding author) Professor, Division of Business and Commerce, Baekseok University, 76 Munam-Ro, Dongnam-Gu, Cheonan, Chungnam 31065, Korea. E-mail: jiyoon@bu.ac.kr