

노동정책연구  
2019. 제19권 제1호 pp.157~179  
한국노동연구원

연구논문

## 최상위 소득 대체를 활용한 근로소득 불평등 측정

유혜림\*\*  
민인식\*\*\*

본 연구에서는 조사자료가 소득분포의 최상위를 적절하게 반영하지 못하는 한계를 보완한 수정된 지니계수(top-corrected Gini Coefficient)를 우리나라 근로소득에 대하여 산출하였다. 분석자료는 2013~16년 국세청의 근로소득 1,000분위 자료와 서베이 데이터인 한국노동패널(KLIPS)을 사용하였다. 기초분석 시, 국세청 데이터에 비해 서베이 데이터에서 최상위 근로소득이 차지하는 비율이 낮게 나타나 최상위 소득자에 대하여 두 자료 간 차이가 있음을 확인하였다. 실증분석에서는 국세청 데이터에서 도출한 파레토 Type I 분포와 KLIPS 소득분포를 비교한 후 수정된 지니계수를 산출하였다. 분석결과, 상위 1% 소득수준에 대하여 KLIPS의 결과가 파레토 Type I 분포에서 산출한 결과보다 과소 추계됨을 알 수 있다. 최상위 소득 10%, 5%, 1%에 대하여 수정된 지니계수를 산출한 결과 2016년에 상대적으로 근로소득 불평등이 감소하였으나, 최상위 대체 비율이 높을수록 불평등 정도가 크게 산출된다. 추가적으로 수정된 지니계수, 가계동향조사 지니계수, KLIPS 지니계수를 비교한 결과 수정된 지니계수가 가장 높게 나타났다. 따라서 서베이 데이터에서 최상위 근로소득이 과소추계되는 문제가 있으며 소득 불평등을 정확하게 추정하기 위해서는 소득세 데이터를 활용한 수정된 지니계수를 산출할 필요성이 있다.

핵심용어 : 소득세자료, 소득 불평등, 수정된 지니계수

논문접수일: 2019년 1월 12일, 심사의뢰일: 2019년 1월 17일, 심사완료일: 2019년 2월 12일

\* (제1저자) 경희대학교 경제학과 박사과정(yoohyerim25@naver.com)

\*\* (교신저자) 경희대학교 경제학과 교수(imin@khu.ac.kr)

## I. 서론

소득은 개인의 생활수준 또는 경제적 후생을 나타내는 보편적인 대리변수(proxy)이며, 불평등(inequality)은 자원의 분배가 고르게 분포되어 있지 않은 상태(Litchfield, 1999)를 의미한다. 소득 불평등은 개인에게는 삶의 질과 연관되며 사회적으로는 경제성장에 부정적 영향을 미치는 중요한 요인이므로 이에 대한 현황과 추이를 분석하는 것은 사회적으로 중요하다. 특히 소득을 구성하는 요인 중 근로소득(earning income)은 우리나라 가계소득 불평등의 주된 원인으로(이종철, 2018; 김영귀 외, 2017), 소득 불평등 중 근로소득에 대한 불평등을 분석하는 것은 중요한 의미를 갖는다.

소득 불평등이 경제적 문제가 되면서 국내·외적으로 소득 불평등에 대한 논의가 활발히 전개되고 있다. 그러나 소득 불평등에 대한 연구는 소득분포에 대한 이해와 적절한 자료가 없이는 충분한 분석이 이루어지기 어렵다. 소득 불평등 분석은 불평등 지표(measures)와 자료(data)에 따라 다른 결과가 도출될 수 있다. 관련하여 상세히 설명하면 다음과 같다. 첫째, 불평등 지표의 특성에 따라 소득분포의 구간에서 민감도가 다르다. 소득 불평등은 소득분배 측정에 사용되는 로렌츠 곡선과 지니계수를 기초로 발전해왔다. 그러나 지니계수는 소득분포의 중간의 변화에 민감한 반면 상대적으로 고소득·저소득층의 변화에 둔감한 한계를 갖고 있다(Atkinson, 1970; 이정우, 2010; 장하준, 2014). 둘째, 분석 데이터의 특성에 따라 불평등 결과가 상이하게 나타날 수 있다. 행정자료인 소득세 자료는 최고소득이 상대적으로 다수 관찰되나, 가구 서베이 자료인 조사자료는 최고소득이 명확하게 나타나지 않는다(Bartels & Metzger, 2017). 또한 소득세 자료와 조사자료는 소득(income)의 정의와 관찰단위가 상이하다. 소득의 정의로 가구조사 자료는 소득의 합인 총 소득을 적용하는 데 반해 소득세 자료는 과세대상 소득을 기반으로 한다. 관찰 단위로 조사자료는 가구 수준으로, 소득세 자료는 개인의 과세 대상 단위이다. 불평등 측정 시 조사자료로 분석이 될 경우 상위 소득에 대한 포괄적 조사의 어려움(undercoverage)과 과소보

고(underreporting) 문제가 있어 하향 편의(downward-bias)가 될 수 있다(Bartels & Metzger, 2017).

소득 분포에 대한 연구는 기본적으로 자료에 근거한다. 앞서 살펴본 바와 같이, 조사자료가 소득분포의 최상위를 적절하게 반영하지 못하는 한계를 보완하기 위하여 외국에서는 이전부터 소득세 자료를 이용한 접근이 시도되었다(Bishop, Chiou, & Formby, 1994; Fichtenbaum & Shahidi, 1988; Piketty & Saez, 2003; Schmitt, 2003; Heathcote, Perri, & Violante, 2010; Alvaredo, 2011; Mishel, Bernstein, & Shierholz, 2013)<sup>1)</sup>. 위 선행연구는 조사자료의 최상위 소득을 소득세 자료로 대체하여 산출한 수정된 지니계수를 산출하여 상위 소득집중도에 대해 분석하였다.

수정된 지니계수에 대한 최근 연구는 다음과 같다. Bartels & Metzger(2017)의 연구는 독일과 유럽의 수정된 지니계수(top-corrected Gini coefficient)를 산출하였다.<sup>2)</sup> 먼저, 독일의 수정된 지니계수는 2001~12년 동안 약 0.47~0.50이고 조사자료에서 산출한 지니계수는 약 0.45~0.48의 범위를 나타내었다. 연구 결과 수정된 지니계수가 조사자료 지니계수에 비해 높았으며, 조사자료는 시기별 상위 1% 그룹의 변화를 충분하게 설명하지 못하였다.<sup>3)</sup> 한편, 유럽 11개국에 대한 수정된 지니계수 산출결과는 각 국가별로 상이하였다.<sup>4)</sup> 덴마크, 노르웨이, 네덜란드, 아일랜드에서는 수정된 지니계수와 조사자료에서 산출한 지니계수의 차이가 미미하였다. 수정된 지니계수는 대략적으로 덴마크 0.37(2010), 노르웨이 0.34(2011), 네덜란드 0.35(2012), 아일랜드 0.40(2009) 정도로 나타났다.<sup>5)</sup> 스웨덴과 스위스에서는 두 자료에서 산출한 지니계수의 차이가 컸으며 수

1) Piketty & Saez(2003)의 연구는 경제성장 초기에는 소득 불평등이 증가하나 1인당 소득이 일정 수준에 도달하면 소득 불평등이 완화된다는 Kuznet(1955)의 가설을 비판하였다.

2) 수정된 지니계수에 대한 선행연구의 표기는 adjusting survey-based Gini coefficients 등 다양하나, 본 논문에서는 Bartels & Metzger(2017)의 표기를 따랐다.

3) 독일에 대한 분석은 조사자료인 SOEP(German Socio-Economic Panel)와 독일 국세청 데이터(German income tax records)를 사용하였다.

4) 유럽 11개국에 대한 분석은 상위 1%의 가구소득에 대하여 수행되었으며, 조사자료로는 EU-SILC(EU Statistics on Income and Living Conditions)와 최고소득 데이터인 WID(World Wealth and Income Database)를 이용하여 산출하였다.

5) 본문에서 각 국가별 수정된 지니계수는 조사자료와 소득세 자료가 모두 있는 최근 연도에 대하여 대략적인 수치를 제시하였다.

정된 지니계수는 스웨덴 0.35(2012), 스위스 0.34(2010)로 나타났다. 독일과 영국에서는 두 자료에서 산출한 지니계수 차이가 가장 큰 것으로 나타났다. 수정된 지니계수는 독일에서는 조사자료 지니계수보다 4~7% 높았으며, 영국에서는 2~5% 높았다. 수정된 지니계수는 독일 0.40(2012), 영국 0.40(2012)으로 나타났다. Hlasny & Verme(2018)의 연구는 유럽의 31개 국가에 대하여 EU-SILC 자료를 이용하여 종합적으로 분석하였다. 연구결과, 조사자료에서 설문에 대한 응답이 소득과 부정적으로 연관되어 불평등 측정 시 하향편의 가능성을 제시하였으며, 수정된 지니계수로 이 편향을 보정하면 유럽의 지니계수는 3.7%포인트 상향 조정될 수 있음을 밝혔다.

최근 우리나라에서도 관련 연구가 진행되고 있다(김낙년, 2012; 김낙년·김종일, 2013; 김낙년, 2016; 김낙년, 2017; 김낙년, 2018). 위 선행연구에서는 국세통계연보 등의 자료를 이용하여 파레토 보간법(Pareto interpolation)으로 상위 소득 집중도를 산출하였다.<sup>6)</sup> 국세청이 소득세에 대한 자료를 공개한다면 위 방법을 사용하지 않고 상위소득 집중도를 직접 산출할 수 있다. 그러나 자료의 한계로 국세청 자료로 소득 불평등을 분석한 연구는 김낙년(2016)의 연구에 한정되어 있다.<sup>7)</sup>

본 연구에서는 최근 공개된 소득세 자료로 수정된 지니계수(top-corrected Gini coefficients)를 산출하여 우리나라의 근로소득 불평등 현황을 측정하는 것을 목적으로 한다. 분석에서는 최상위 소득구간을 10%, 5%, 1%로 설정하였을 때 수정된 지니계수를 계산한다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서는 분석에서 사용한 국세청의 근로소득 천 분위 자료를 소개하고 국세청 자료와 조사자료인 한국노동패널(Korea Labor & Income Panel Study : KLIPS)로 산출한 근로소득에 대한 최상위 소득구간의 기초통계를 제시한다. 제Ⅲ장에서는 최상위 소득분포를 추정할 시 모수적 분포함수로 선정하는 파레토(Pareto) Type I 분포와 Type II 분포에 대해 설

6) 국세통계연보 외에 가계금융복지조사, 가계동향조사, 가구소비실태조사, 경제활동인구조사, 국민계정 등의 자료를 이용하였다.

7) 국세청 자료는 국회의원의 요청에 따라 제한적으로 공개되며, 홍중학의원실에서 2012년 처음으로 공개하였다. 이 자료는 2007년부터 2012년까지 근로소득과 종합소득의 100분위와 상위 10만 명에 대한 현황자료이다.

명하고 파레토 분포하에서 지니계수를 도출하는 방법을 제시한다. 제IV장에서는 국세청 데이터에서 도출한 파레토 Type I 분포와 KLIPS 소득분포를 비교한 후 수정된 지니계수 산출 결과를 제시한다. 마지막으로 제V장에서 주요 내용을 요약하고 연구의 함의를 제시한다.

## II. 분석 데이터

본 연구에서 기존 소득 불평등 지수를 보완하기 위해 선택한 행정 데이터는 국세청에서 공식발표한 2013~16년 근로소득 천 분위 데이터이다. 해당 데이터는 국세청 소득세과에서 정의당 심상정 의원에게 제출한 임금근로자 천 분위별 총 급여와 결정세액 데이터이다. 해당 데이터는 2018년 9월 2일 정의당 심상정 의원 블로그(<https://blog.naver.com/713sim/221350919960>)에 공개되었다.<sup>8)</sup> [그림 1]은 엑셀파일로 올려진 근로소득 천 분위 2016년 데이터의 일부분을 보여 준다. 2016년 전체 근로소득자는 1,774만 명이고 총 급여는 595조 원 그리고 근로소득세 총액은 30조 8천억 원임을 알 수 있다. 0.1% 구간별로 해당 구간에 속한 근로소득자 수 그리고 총급여를 확인할 수 있다.

[그림 1] 근로소득 천 분위(2016년)

구분	인원	총급여	근로소득금액	소득공제액 합계	과세표준	결정세액
합계	17,740,098	5,959,907	4,399,935	2,751,851	3,208,055	308,539
상위 0.1% 내	17,740	121,433	117,093	9,027	112,405	37,036
상위 0.2% 내	17,740	54,892	51,782	6,373	48,519	13,904
상위 0.3% 내	17,740	43,424	40,486	5,960	37,464	9,984
상위 0.4% 내	17,740	37,408	34,543	5,825	31,583	7,922
상위 0.5% 내	17,740	33,716	30,878	5,811	27,905	6,629
상위 0.6% 내	17,740	31,266	28,451	5,782	25,483	5,797
상위 0.7% 내	17,740	29,435	26,646	5,754	23,682	5,189
상위 0.8% 내	17,740	28,023	25,245	5,786	22,238	4,705
상위 0.9% 내	17,740	26,899	24,137	5,753	21,146	4,335
상위 1.0% 내	17,740	25,992	23,244	5,755	20,237	4,019

자료: 심상정 의원 블로그

8) 당시 심상정 의원 블로그에서는 상위 0.1% 근로소득이 하위 10% 근로소득의 1,000배가 넘는다는 것을 주로 지적하였다.

위 데이터를 이용하면 2016년 구간별 평균 급여액을 쉽게 계산할 수 있다. 상위 0.1% 평균 급여는 6.8억 원이며 상위 1%에 속한 근로자의 평균 급여는 1.4억 원이다. 고액 연봉의 기준인 1억 원의 경우 상위 3.6% 구간에 속한다. <표 1>에서는 2013~16년 국세청 근로소득 천 분위에서 계산할 수 있는 최상위 근로소득자에 대한 기초통계를 보여준다. 상위 0.1% 평균 급여는 2013년 6.37억 원에서 2016년 6.84억 원으로 7.3% 증가하였다. 상위 1% 평균 급여는 1.33억 원에서 1.46억 원으로 9.7% 증가하였다. 연 근로소득 1억 원의 백분위는 2013년 97.2%에서 2016년 96.4%로 낮아졌다. 즉 연 근로소득 1억 이상 비율이 증가하였다는 것을 알 수 있다. 상위 0.1% 그룹의 근로소득이 전체 근로소득에서 차지하는 비율은 2.09%(2013년)에서 2.04%(2016년)로 감소하는 추세를 확인할 수 있다.

<표 1> 최상위 근로소득자 기초통계

(단위: 억 원, %)

	상위 0.1% 평균 급여	상위 1% 평균 급여	연 근로소득 1억 원 위치	상위 0.1% 근로소득이 전체 근로소득에서 차지하는 비율
2013	6.37	1.33	97.2	2.09
2014	6.56	1.37	96.9	2.07
2015	6.55	1.41	96.7	2.01
2016	6.84	1.46	96.4	2.04

본 연구에서 개별 근로자의 근로소득을 조사한 서베이 데이터로 KLIPS를 선택한다. KLIPS의 가구원용 데이터는 15세 이상 가구원의 소득활동 여부, 근로소득자 여부 그리고 근로소득자의 경우 월평균 급여를 조사하고 있다.<sup>9)</sup> 앞선 국세청 데이터와 매칭하기 위해서 KLIPS 16차(2013년)~19차(2016년) 조사차수를 선택한다. 경제활동을 하는 가구원 중 임금근로자(상용직, 임시직, 일용직)만 분석대상에 포함하였다. 따라서 자영업자는 분석에서 제외된다.

9) 월평균 급여에 12를 곱해 연 근로소득으로 환산하였다.

<표 2>에서는 KLIPS 근로소득과 국세청 근로소득의 기초통계량을 비교한 결과를 보여준다. 지면 제약을 고려하여 2016년 근로소득만 비교한다. KLIPS에서 연평균 급여는 가중치 적용여부에 따라 2,930만~3,005만 원이다. 그러나 국세청 데이터에서 연평균 급여는 3,359만 원이다. 서베이 데이터에서 평균 근로소득이 낮은 이유는 고소득 임금근로자가 충분히 반영되지 않았기 때문으로 예상된다. KLIPS 데이터에서 중앙값 소득은 1,923만~1,954만 원으로 추정된다. 그러나 국세청 데이터에서 중앙값은 2,400만 원이다. 가장 근로소득이 높은 임금근로자는 KLIPS에서는 50,400만 원이지만 국세청 데이터에서는 최대값을 확인하기 어렵다. <표 1>에서 확인할 수 있듯이 2016년 상위 0.1% 근로자의 연평균 급여는 68,400만 원이다. 결과적으로 국세청 데이터에서 최대값은 KLIPS 소득 최대값보다 훨씬 높을 것으로 예상된다. <표 2>에서 예상할 수 있듯이 서베이 데이터에는 최상위 근로소득자가 충분히 반영되지 않은 것으로 확인된다. 따라서 최상위 소득자의 제외는 근로소득 불평등 추정에 있어서 편의(bias)를 발생시킬 수 있다.

<표 2> KLIPS와 국세청 근로소득 비교 : 2016년

(단위: 만 원)

통계량	KLIPS (표본)	KLIPS (가중치 고려)	국세청
평균	2,930	3,005	3,359
중앙값	1,923	1,954	2,400
최대값	50,400	50,400	NA
최소값	0	0	0

주: 1) KLIPS에서 가중치는 횡단 가구원 가중치를 사용함.

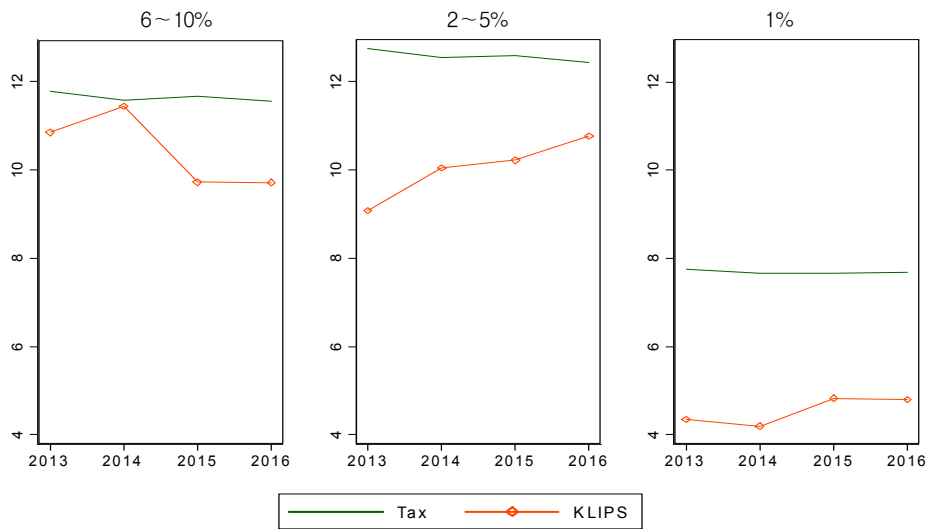
2) KLIPS 임금근로자 수는 5,765명(2016년)임.

최상위 소득그룹이 전체 급여에서 차지하는 비율을 KLIPS와 국세청 데이터에서 비교하고 그 결과를 [그림 2]로 제시한다. 상위 6~10%, 상위 2~5%, 상위 1% 소득그룹이 전체 근로소득에 차지하는 비율에 대한 시계열 그래프이다.<sup>10)</sup> 상위 6~10% 그룹이 전체 근로소득에서 차지하는 비율은 국세청 데이터에서

10) KLIPS 그래프는 횡단 가중치를 고려한 비율을 계산하였다.

11% 중반 정도로 일정하다. KLIPS에서는 2014년 11.4%로 가장 높지만 2016년 9.7%로 다시 감소한다. 상위 2~5% 그룹에서는 국세청 데이터에서 12% 중반으로 거의 변화가 없으나 KLIPS에서는 9.1(2013년)~10.7%(2016년)로 꾸준히 상승하고 있다. 그러나 최상위 1% 근로소득 그룹을 비교하면 그 차이가 분명하며 격차 역시 일정하게 유지된다. 국세청 데이터에서는 7.6~7.7%이지만 KLIPS에서는 4.2~4.8% 비중을 차지한다. [그림 2]를 통해 최상위 1% 근로소득 그룹이 차지하는 비중은 서베이 데이터에서 훨씬 낮게 나타난다는 것을 확인할 수 있다.

[그림 2] 최상위 그룹이 전체 근로소득에서 차지하는 비율



### Ⅲ. 파레토 분포와 지니계수

#### 1. 파레토 Type I과 Type II 분포

소득 또는 자산변수의 분포를 추정하는 모수적 분포함수(parametric distribution function)로 파레토 분포를 사용할 수 있다(Bishop, Chiou, & Formby, 1994;



Fichtenbaum & Shahidi, 1988; Piketty & Saez, 2003; Schmitt, 2003; Heathcote, Perri, & Violante, 2010; Alvaredo, 2011; Mishel, Bernstein, & Shierholz, 2013). 파레토 분포는 오른쪽 꼬리가 매우 두꺼운(heavy-thick) 분포이면서 양의 왜도(positive skewness)를 가진 분포이다. 파레토 분포는 Type I과 Type II 모형으로 구분할 수 있다. Type I 분포는 분포모수가 1개이며 누적분포 함수(cumulative distribution function)는 식 (1)과 같이 정의한다.

$$F(x) = 1 - \left(\frac{x}{x_m}\right)^{-\alpha} \quad (1)$$

식 (1)에서  $x$ 를 개인소득으로 가정하자. 따라서  $x > 0$  이고  $x \geq x_m > 0$  이다. 즉  $x_m$ 은 소득 하한(lower bound)으로 연구자의 일정한 가정에 의해 주어진다. 파레토 Type I 분포는  $x_m$  이상의 소득(일정 이상 상위소득)에 대해서만 분포함수를 정의한다. 분포모수인  $\alpha > 0$ 는 오른쪽 꼬리가 어느 정도 두꺼운지를 측정한다.  $\alpha$ 가 작을수록 두꺼운 꼬리분포(fat-tail distribution)와 매칭된다. 파레토 Type I 분포의  $k$ th 적률은  $k < \alpha$ 인 경우에만 존재한다. 분포모수  $\alpha = 3$ 이면 1차 적률인  $E(X)$ 와 2차 적률인  $E(X^2)$ 만 유한한 값으로 존재하며 3차 적률부터는  $E(X^k) \approx \infty, k \geq 3$  이다.

파레토 Type I 분포에서는 분포모수가 1개인 데 비해 파레토 Type II 분포에서 분포 모수는 식 (2)와 같이 2개를 가진다.  $\sigma$ 와  $\theta$ 는 각각 척도 모수(scale parameter)와 형상 모수(shape parameter)에 해당하며 모두 양의 값이다.  $\theta > 0$  조건을 만족하면 오른쪽 꼬리가 두꺼운 분포가 된다. 누적분포함수는 식 2에서 제시한다.<sup>11)</sup>

$$F(x) = 1 - \left[1 + \theta \frac{x - x_m}{\sigma}\right]^{-\frac{1}{\theta}}, \quad x > x_m > 0 \quad (2)$$

파레토 Type II 분포의  $k$ th 적률은 역시  $k < 1/\theta$ 인 경우에만 존재한다. 파레토 Type I 분포와 Type II 분포의 관계는 다음과 같다는 것을 쉽게 증명할 수

11) 파레토 Type II 분포는 Lomax 분포라고도 부른다.

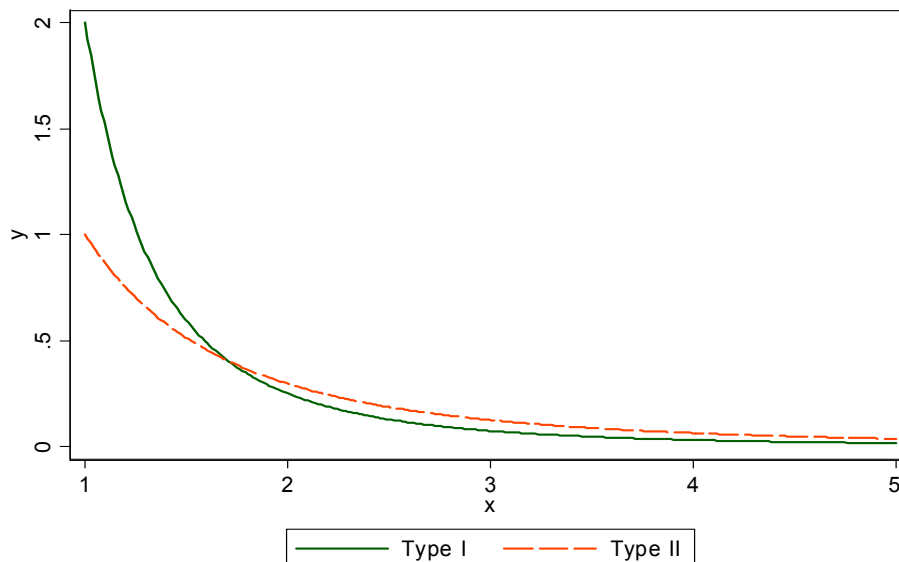
있다. 파레토 Type II 분포에서  $\theta = \frac{1}{\alpha}$  그리고  $\sigma = \frac{x_m}{\alpha}$ 로 제약하면 파레토 Type I 분포가 된다. 2개의 분포모수를 모두  $\alpha$ 의 함수로 제약하면 파레토 Type I 분포와 같아진다.

$$P \text{ Type II } (\theta = 1/\alpha, \sigma = x_m/\alpha) \Rightarrow P \text{ Type I } (\alpha)$$

파레토 Type II 분포는 분포 모수를 1개 더 가짐으로써 다양한 소득분포 함수에 유연하게 대응할 수 있다. 그러나 적합도(goodness of fit) 측면에서 Type I 분포와 Type II 분포에서 큰 차이가 없다고 알려져 있다(Jenkins, 2017).

[그림 3]에서는 파레토 Type I과 Type II 분포의 확률밀도함수를 그래프로 표현하여 비교한다. Type I 분포에서  $\alpha = 2$ 로 가정하였고 Type II에서는  $\theta = \frac{1}{\alpha} = 0.5$  그리고 척도모수  $\sigma = 1$ 로 가정하여 그린 결과이다.  $x$ 의 최소값  $x_m = 1$ 로 가정한다. 따라서 PDF는  $x > 1$  범위에서만 나타난다. Type II 분포에서 오른쪽 꼬리가 더 두꺼운 형상임을 알 수 있다. 극단적으로 높은 소득에 대한 밀도(density)가 더 크게 나타난다.

[그림 3] 파레토 Type I과 Type II 분포의 PDF



## 2. 지니계수 도출

본 소절에서는 근로소득 천 분위 데이터를 이용하여 파레토 Type I 분포모수  $\alpha$ 를 추정한 후 해당 분포로부터 지니계수를 얻는 과정을 설명한다. Atkinson (2007)에서는 식 (1)의 파레토 Type I 분포에서 모수  $\alpha$ 를 식 3과 같이 추정한다.<sup>12)</sup>

$$\hat{\alpha} = \frac{1}{1 - \left( \frac{\log(S_j/S_i)}{\log(P_j/P_i)} \right)} \quad (3)$$

위 식에서  $P$ 와  $S$ 에 대해 다음과 같이 정의한다.

$P_j$ : 그룹  $j$ 의 인구비율

$P_i$ : 그룹  $i$ 의 인구비율

$S_j$ : 그룹  $j$ 가 전체 소득에서 차지하는 비율

$S_i$ : 그룹  $i$ 가 전체 소득에서 차지하는 비율

가령  $i = 0.1\%$ 이고  $j = 1\%$ 라면  $S_{i=0.1\%}$ 와  $S_{j=1\%}$ 는 각각 상위 0.1% 근로소득이 전체 근로소득에서 차지하는 비율 그리고 상위 1% 근로소득이 전체 근로소득에서 차지하는 비율로 이해할 수 있다.  $\log\left(\frac{P_{j=1\%}}{P_{i=0.1\%}}\right) = \log(10) = 2.302$  이고  $\log\left(\frac{S_{j=1\%}}{S_{i=0.1\%}}\right)$ 는 제II장에서 언급한 근로소득 천 분위 국세청 데이터에서 구할 수 있다. 최상위 0.1% 그룹의 근로소득에서 차지하는 비중( $S_{i=0.1\%}$ )이 높아질수록  $\hat{\alpha}$ 은 작아지는 것을 확인할 수 있다. 파레토 Type I 분포에서 분포모수  $\alpha$ 가 작아질수록 불평등(지니계수)이 커지기 때문에 직관과 일치하는 결과이다.

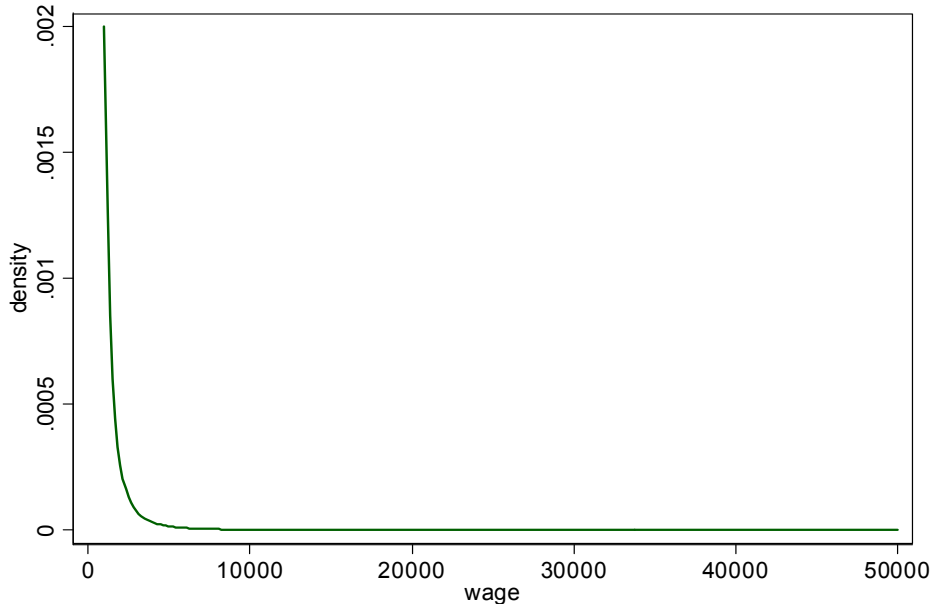
파레토 Type I 분포의 하한(lower bound)  $x_m$ 은 분포모수  $\alpha$ 가 주어지면 식 (3)을 이용하여 계산할 수 있다.

$$\hat{x}_m = (1 - F(x))^{1/\hat{\alpha}} \times x \quad (4)$$

12) 식 3의 도출 과정은 김낙년(2012, 2016)의 연구를 참고할 수 있다.

서베이 데이터에서 상위 1% 근로소득을 파레토 Type I 분포로 대신하고자 하는 경우  $F(x) = 0.99$ 이고  $x$ 는 상위 1%에 해당하는 서베이 데이터 소득이 된다. 예를 들어,  $\hat{\alpha} = 2$  그리고 서베이 데이터에서 상위 1% 근로소득을 1억 원으로 가정하면  $\hat{x}_m = 1000$  만 원이 된다.  $\hat{x}_m = 1000$ (만 원)이고  $\hat{\alpha} = 2$ 인 파레토 Type I 분포의 PDF는 [그림 4]에서 제시한다.<sup>13)</sup>

(그림 4) 적합된 근로소득 분포 : 파레토 Type I 분포



본 연구에서는 하위 99% 지니계수는 KLIPS 표본을 이용하고 상위 1% 그룹의 지니계수는 파레토 Type I 분포로부터 도출한다. 두 지니계수를 병합하여 전체 그룹에 대한 수정된 지니계수를 계산하는 것이 목적이다. Atkinson(2007)

13) [그림 4]에서 오른쪽 꼬리가 평평하게 보이는 것은 확률분포 값의 척도(scale) 때문이다. 따라서 과도하게 두터운 꼬리(fat tail)처럼 보일 수 있다. 그러나 본문에서 제시한  $x_m = 1,000$ 만 원,  $\alpha = 2$ 로 가정한 후 (식 1)의 CDF 공식을 이용하여 오른쪽 꼬리에 해당하는 확률을 구하면 아래와 같이 소득이 높아질수록 구간 확률이 감소한다는 것을 확인할 수 있다.

구간	$\Pr(1 \times 10^4 \leq X < 2 \times 10^4)$	$\Pr(2 \times 10^4 \leq X < 3 \times 10^4)$	$\Pr(3 \times 10^4 \leq X < 4 \times 10^4)$	$\Pr(4 \times 10^4 \leq X < 5 \times 10^4)$
확률	0.7500%	0.1388%	0.0486%	0.0225%

과 Alvaredo(2011)에서 전체 그룹의 지니계수가 서로 겹치지 않는(non-overlapping) 서브 그룹의 지니계수로 분해되는 방식을 설명하고 있다. 전체 표본을 상위 1%와 하위 99%의 두 그룹으로 나누었을 때 전체 그룹의 불평등 정도는 다음과 같이 계산할 수 있다.

$$\begin{aligned} \text{전체 그룹의 불평등} &= \text{상위 1\% 그룹 내 불평등} + \\ &\quad \text{하위 99\% 그룹 내 불평등} + \\ &\quad \text{두 그룹 간 불평등} \end{aligned} \quad (5)$$

식 (5)에서 두 그룹 간(between groups) 불평등은 각 근로소득자가 자신이 속한 그룹의 평균적인 근로소득이라고 가정할 때 계산되는 불평등으로 정의한다. 식 (5)의 불평등 분해(inequality decomposition)는 지니계수뿐 아니라 여타 불평등 지수에도 모두 적용될 수 있다.

Alvaredo(2011)에 따르면 수정된 지니계수는 다음과 같이 그룹 간 불평등과 그룹 내 불평등으로 계산한다.

$$G^{correct} = G^{**}PS + G^*(1-P)(1-S) + S - P \quad (6)$$

$P$ 는 대체하고자 하는 최상위 임금근로자 비율이고  $S$ 는 최상위 근로소득이 전체 근로소득에서 차지하는 비율이다.  $S$ 는 국세청 데이터에서 구할 수 있다.  $G^{**}$ 는 국세청 데이터에 근거한 파레토 Type I 분포로부터 계산한 지니계수이다. 파레토 Type I 분포의 모수 추정치  $\hat{\alpha}$ 이 주어지면 지니계수  $G^{**} = \frac{1}{2\hat{\alpha} - 1}$ 로 구할 수 있다(Cowell 1989).  $G^*$ 는 최상위 소득 그룹을 제외한 나머지 그룹에서 계산한 KLIPS 지니계수이다. 하위 99% 그룹의 지니계수  $G^*$ 는 식 (7)과 같이 계산한다.

$$G^* = \frac{\sum_{i=1} \sum_{j=1} |x_i - x_j|}{2 \sum_{i=1} \sum_{j=1} x_j} \quad (7)$$

위 식에서  $x_i$ 와  $x_j$ 는 서베이 데이터(KLIPS)에서 하위 99% 그룹에 속한 근로소득에 해당한다.

## IV. 실증 분석결과

### 1. 파레토 분포의 적합

본 소절에서는 국세청 근로소득 천 분위 데이터에서 도출한 파레토 Type I 분포와 KLIPS 소득분포를 비교한다. KLIPS의 상위 1% 근로소득 그룹에 대한 경험적 분포(empirical distribution)와 모수적 분포인 파레토 Type I 분포 적합 결과를 비교한다.

식 (1)을 다시 쓰면  $\log(1 - F(y))$  와  $\log(x)$  는 서로 선형관계임을 확인할 수 있다. 식 (8)에서  $\alpha \log(x_m)$  은 양의 상수(positive constant)가 되기 때문에  $\log(1 - F(y))$  와  $\log(x)$  의 관계는 기울기가  $-\alpha < 0$  인 우하향 직선이 된다는 것을 알 수 있다.

$$\begin{aligned} F(x) &= 1 - \left(\frac{x}{x_m}\right)^{-\alpha} \\ \Rightarrow 1 - F(x) &= \left(\frac{x}{x_m}\right)^{-\alpha} \\ \Rightarrow \log(1 - F(x)) &= -\alpha \log\left(\frac{x}{x_m}\right) \\ \Rightarrow \log(1 - F(x)) &= -\alpha \log(x) + \alpha \log(x_m) \end{aligned} \quad (8)$$

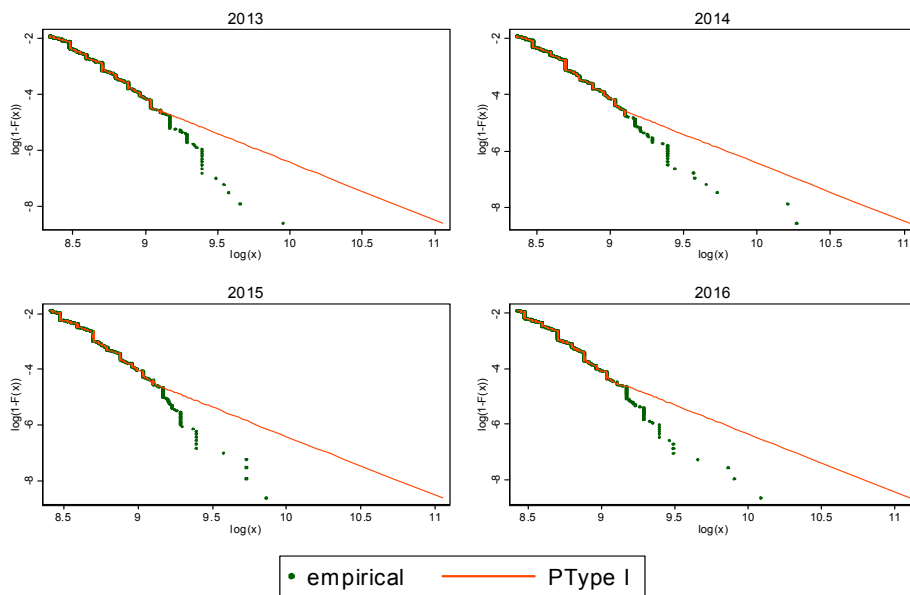
KLIPS 근로소득( $x$ )을 이용하여  $\log(1 - F(y))$  와  $\log(x)$  변수 관계를 표현한 결과는 [그림 5]에서 제시한다. [그림 5]는 Zipf plot이라고도 부른다.<sup>14)</sup> [그

14) Zipf plot은 미국의 언어학자 조지 킹즐리 지프(George Kingsley Zipf)가 소개한 Zipf's law에서 유래되었다. 이 법칙은 요소 개체의 규모는 요소 개체의 순위(rank)에 반비례한다는 것이 주요 내용이다. 이러한 분포는 언어뿐 아니라 성씨 분포, 도시인구 분포 등 사회현상에서 자주 발견되는 경험적 분포로 알려져 있다. [그림 5]에서 요소 개체 규모는 근로소득( $x$ )이고 요소 개체 순위는  $1 - F(x)$ 가 된다.  $x$ 가 최대값인 경우 100백분위 값( $F(x)=1$ )이므로 rank=1이 된다. 따라서 [그림 5]는 Zipf's law에서 일반적으로 정의하는 X축과 Y축 변수에 대해 로그를 취한 것과 같다고 해석할 수 있다.

림 5]에서 산포도(scatter plot)는 KLIPS 조사연도(wave)별 상위 1% 근로소득에 대한 경험적 분포이다. 2016년 상위 1% 소득그룹 최소값을  $x_{99}$ 로 표현하면  $x$  축 값은  $\log(x_{99})$  이고  $y$  축 값은  $\log(1-0.99)=\log(0.01)$  이 된다. 실선 부분은 99~100% 누적확률에 대응하는 소득을 파레토 Type I 분포에서 도출하여 로그 취한 값이다.  $0.99 \leq F(x) \leq 1$  로 주어졌을 때 파레토 Type I 분포에서 도출한 소득( $x^*$ )은 식 (9)와 같이 구할 수 있다.

$$x^* = \frac{x_m}{[(1-F(x))]^{\frac{1}{\alpha}}} \tag{9}$$

[그림 5] Zipf plot



<표 3>에서는 [그림 5]의 Zipf Plot을 작성하는 데 필요한 분포 모수인  $\alpha$ 와  $x_m$ 의 연도별 추정치를 정리한다.  $\hat{\alpha}$ 은 식 (3)에서  $j=1\%$ 와  $i=0.1\%$ 로 가정하였고  $S_j$ 와  $S_i$ 는 국세청 천 분위 데이터를 이용하였다.  $x_m$ 은 식 (4)를 이용하여 계산하였고  $x$ 는 KLIPS 데이터에서 구한 상위 1%에 속한 근로자의 근로소득을 선택한다.

〈표 3〉 Zipf plot 작성을 위한  $\alpha$  와  $x_m$ 

(단위: 만 원)

	$\hat{\alpha}$	$\hat{x}_m$
2013	2.067	983.5
2014	2.070	983.6
2015	2.118	1,064.7
2016	2.104	1,075.5

[그림 5]와 <표 3>에 대한 해석은 다음과 같다. 첫째, KLIPS 표본에 있는 상위 1% 소득수준은 파레토 Type I 분포에서 구한 상위 1% 임금수준보다 훨씬 낮다. KLIPS 서베이에서 상위 1% 소득수준이 과소추계되고 있다는 것을 알 수 있다. 따라서 근로소득 불평등을 나타내는 지니계수 역시 과소추정될 가능성이 있다. 둘째, KLIPS 서베이 데이터로 판단하였을 때 최상위 소득분포는 파레토 Type I 분포로 적합하는 것이 적절하다. [그림 5]의 경험적 분포에서 도출한 Zipf plot은 거의 우하향하는 직선에 가까운 형태임을 알 수 있다. 셋째, [그림 5]에서 실선의 기울기 ( $\hat{\alpha}$ )가 완만할수록(작을수록) 상위 1% 그룹이 전체 근로소득에서 차지하는 비중이 커진다. 즉 근로소득 불평등이 커지는 것을 알 수 있다.<sup>15)</sup> <표 3>에서 정리한  $\hat{\alpha}$ 를 통해 2013~14년에 비해 2015~16년에 근로소득 불평등이 줄어들었다는 것을 알 수 있다.

## 2. 수정된 지니계수 계산

하위 99%와 상위 1%의 지니계수를 각각 KLIPS 서베이 데이터와 파레토 Type I 분포에서 도출한 후 식 (6)에 따라 수정된 지니계수를 계산한다. 파레토 Type I 분포에서 최상위 근로소득을 어느 정도 대체할지에 따라 수정된 지니계수는 달라질 수 있다. 분석결과의 강건성(robustness)을 확보하기 위해 본 연구에서는 식 (3)의  $j$ 와  $i$  조합을 <표 4>와 같이 다양하게 정의한다(Bartels & Metzger, 2017).

15) 파레토 Type I 분포에서 지니계수는  $\frac{1}{2\hat{\alpha}-1}$  이기 때문이다.



〈표 4〉  $j$ 와  $i$ 에 따른 분포모수  $\alpha$  추정치

	$j = 10\%$ $i = 1\%$	$j = 5\%$ $i = 1\%$	$j = 1\%$ $i = 0.1\%$
2013	3.075	3.070	2.067
2014	3.054	3.049	2.070
2015	3.074	3.065	2.118
2016	3.033	3.013	2.104

수정된 지니계수를 얻기 위해 하위 그룹의 지니계수는 서베이 데이터에서 근로소득을 이용하여 경험적으로 계산하고 최상위 그룹의 지니계수는 파레토 Type I 분포에서 계산한다. 두 그룹(하위 그룹과 최상위 그룹)의 지니계수를 통합하기 위한 가중치는 식 (6)을 따라 정한다. 국세청 데이터를 이용하여 대체하고자 하는 최상위 그룹( $P$ )은 상위 10%, 상위 5%, 상위 1%로 다양하게 정의한다. 따라서 식 (6)에서  $s$ 는 국세청 데이터에서 상위 10%, 상위 5% 그리고 상위 1%가 전체 근로소득에서 차지하는 비율로 이해할 수 있다.  $G^{**}$ 는 <표 4>에 주어진  $\hat{\alpha}$ 를 이용하여 상위 10%, 5% 그리고 1% 그룹의 지니계수를 얻을 수 있다. <표 5>에서는 KLIPS 서베이 데이터에서 모든 임금근로자를 이용한 지니계수와 본 연구의 실증분석 결과 계산된 지니계수를 비교한다.

〈표 5〉 수정된 지니계수

	KLIPS 지니계수	수정된 지니계수		
		최상위 10% 대체 $P = 0.1$	최상위 5% 대체 $P = 0.05$	최상위 1% 대체 $P = 0.01$
2013	0.329	0.400	0.380	0.355
2014	0.330	0.396	0.379	0.355
2015	0.327	0.397	0.377	0.354
2016	0.323	0.391	0.371	0.348

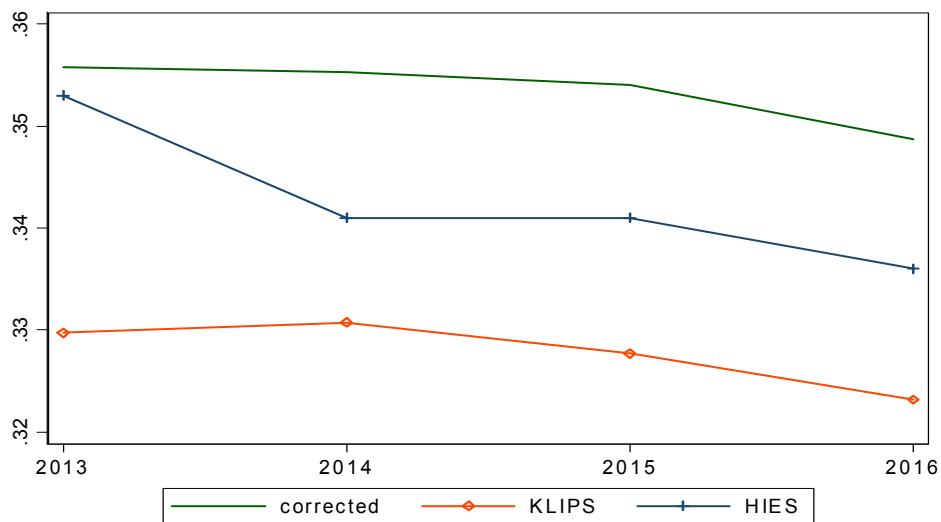
주: KLIPS 지니계수는 가중치를 고려하지 않은 지니계수이다.

최상위 소득을 대체하지 않고 KLIPS 임금표본을 그대로 이용하여 계산한 지니계수는 0.323~0.330이고 2013~15년에 비해 2016년에 0.323으로 근로소득 불

평등이 줄어들고 있음을 알 수 있다. 그러나 수정된 지니계수는 KLIPS 지니계수에 비해 훨씬 큰 값이라는 것을 알 수 있다. 최상위 소득 1% 그룹을 대체하면 0.355~0.348로 KLIPS 지니계수에 비해 약 7.5% 더 높다. 대체하고자 하는 상위 그룹 비율이 더 높을수록 수정된 지니계수는 KLIPS 지니계수에 비해 더 큰 값을 갖게 된다. 상위 10% 소득그룹을 대체하면 수정된 지니계수는 0.391~0.400으로 KLIPS 지니계수에 비해 20%나 높은 수준이다. 수정된 지니계수에서도 최근 4년 중 2016년에 근로소득 불평등이 가장 낮은 수준으로 나타난다.

[그림 6]에서는 KLIPS 표본 데이터를 이용한 지니계수와 수정된 지니계수 추세를 비교하고 있다. 추가적으로 가계동향조사(Household Income and Expenditure Survey : HIES) 서베이에서 조사된 시장소득을 이용하여 계산된 지니계수와 비교한다.<sup>16)</sup> 시장소득은 임금근로자뿐 아니라 자영업자의 소득까지 포함한 불평등이라는 점에서 본 연구의 지니계수와 차이가 있다. KLIPS 지니계수는 가중치를 고려하지 않은 경우이고 수정된 지니계수는 최상위 1%를 대체하여 계산된 결과이다. 수정된 지니계수는 0.348~0.355로 가장 높은 수준이고 근로소득

[그림 6] 지니계수 추세 비교



16) 가계동향조사 지니계수는 통계청 국가통계포털 kosis.kr에서 얻을 수 있다.

을 포함한 전체 소득에 대한 가계동향조사(HIES) 지니계수는 수정된 지니계수보다 낮은 0.336~0.353이다. 또한 2016년은 2013~15년에 비해 근로소득 불평등이 줄어들었다는 것을 세 종류의 지니계수에서 모두 확인할 수 있다.

<표 5>와 [그림 6] 결과는 서베이 데이터에 나타난 근로자의 소득을 이용하여 지니계수를 계산하는 것에 비해 국세청 데이터에 나타난 최상위 근로소득을 결합한 수정된 지니계수에서 근로소득 불평등이 훨씬 높게 나타난다는 것을 보여주고 있다. 제II장에서 서베이 데이터에서 최상위 근로소득이 과소추계 되는 문제가 있다는 점을 이미 지적하였다. 이러한 문제를 보완하고 근로소득 불평등에서 더 나아가 전체 소득 불평등을 정확하게 추정하기 위해서 조세 데이터에서 도출한 상위 소득분포를 활용할 필요가 있다는 것을 실증분석에서 보여주고 있다.

## V. 결론 및 시사점

본 연구에서는 조사자료가 소득분포의 최상위를 적절하게 반영하지 못하는 한계를 보완하기 위한 방법으로 제시된 수정된 지니계수(top-corrected Gini coefficient)를 우리나라 근로소득에 대하여 산출하였다. 분석자료로는 심상정의원이 공개한 2013~16년 근로소득 천 분위 자료를 이용하였으며 조사자료는 KLIPS를 사용하였다. 최상위 소득분포를 모수적 분포인 파레토 Type I 분포로 가정하고 분포모수를 추정한 후 수정된 지니계수를 산출하였다.

분석결과, KLIPS 근로소득과 국세청 근로소득은 기초통계에서도 차이가 있었다. 첫째, 국세청 자료의 근로소득 평균과 중앙값은 표본 조사자료인 KLIPS에 비해 높았다. 따라서 최상위 소득자 제외는 근로소득 불평등 추정에 있어서 편의를 발생시킬 수 있다. 둘째, 전체 근로소득에서 상위 6~10%, 2~5%, 1% 그룹이 차지하는 비중을 살펴본 결과 근로소득에 대한 두 자료의 차이가 나타났다. 먼저 국세청 데이터에서는 각 그룹별 비중이 일정하게 유지되는 반면, 서베이 데이터에서는 연도별 변동이 있다. 또한 최상위 1% 근로소득 그룹이 차지하는 비중은 서베이 데이터에서 현저하게 낮게 나타나는 것을 확인할 수 있다.

국세청 최상위 근로소득 데이터에 기초하여 파레토 Type I 분포를 적합하여 분석한 결과는 다음과 같다. KLIPS 표본에 있는 상위 1%의 소득수준은 파레토 Type I 분포에서 구한 상위 1% 소득수준보다 훨씬 낮게 나타나 서베이 데이터에서 상위 1% 소득수준이 과소추계됨을 알 수 있다. KLIPS 서베이 데이터는 경험적 분포에서 도출한 Zipf plot의 우하향 직선과 가까운 형태이므로 파레토 Type I 분포를 적합하여 분석하는 것이 적절하였다. 근로소득 불평등은 2013/14년에 비해 2015/16년에 감소함을 알 수 있었다. 다음으로, 수정된 지니계수 산출결과를 상위 10%, 5%, 1%로 구분하여 KLIPS 지니계수와 비교하였다. 최근 년도일수록 근로소득 불평등이 줄어들고 있음을 알 수 있으나, KLIPS에서 산출한 지니계수에 비해 수정된 지니계수가 훨씬 큰 값을 알 수 있다. 또한 대체하는 상위그룹 비율이 증가할수록 근로소득 불평등은 더 높게 나타났다. 마지막으로, 수정된 지니계수, 가계동향조사, KLIPS 지니계수로 추세를 비교한 결과, 세 종류의 지니계수 모두 감소하는 추세를 보였으나, 수정된 지니계수, 가계동향조사 지니계수, KLIPS 지니계수 순으로 불평등이 높게 나타났다. 따라서 서베이 데이터에서 최상위 근로소득이 과소추계되는 문제가 있으며 이러한 문제를 보완하여 근로소득 불평등에서 더 나아가 전체 소득 불평등을 정확하게 추정하기 위해서는 조세 데이터에서 도출한 최상위 소득분포를 활용할 필요성이 있다.

본 연구는 다음과 같은 한계를 지니고 있다. 첫째, 연구에서는 자료의 한계로 근로소득에 대해서만 분석하였다. 향후 추가적으로 자료가 공개된다면 다른 소득원에 대해서도 실증분석을 할 수 있을 것이다. 둘째, 본 연구는 소득 불평등 산출이 목적이기 때문에 불평등 요인분해에 대해서는 다룰 수 없었다. 셋째, 자료의 한계로 2013~16년에 해당하는 불평등만 산출하였다. 장기적인 기간에 대한 국세청 데이터가 공개되면 소득 불평등에 대한 보다 장기적인 시계열 연구가 될 것이다. 추가적으로 본고에서 수행한 파레토 Type I 분포 적합뿐만 아니라 파레토 Type II 분포를 이용하여 소득 불평등을 추정한다면 연구의 범위를 확장할 수 있을 것이다.

## 참고문헌

- 김낙년(2012). 「한국의 소득집중도 추이와 국제비교, 1976~2010: 소득세 자료에 의한 접근」. 『경제분석』 18 (3): 75~114.
- \_\_\_\_\_(2016). 「한국의 개인소득 분포: 소득세 자료에 의한 접근」. 『한국경제의 분석』 22 (3): 147~208.
- \_\_\_\_\_(2017). 「한국에서의 부와 상속, 1970~2014」. 『경제사학』 64: 127~160.
- \_\_\_\_\_(2018). 「한국의 소득집중도: 1933~2016」. 『한국경제포럼』 11 (1): 1~32.
- 김낙년·김종일(2013). 「한국 소득분배 지표의 재검토」. 『한국경제의 분석』 19 (2): 1~50.
- 김영귀·남시훈·금혜윤·김낙년(2017). 『대외개방이 국내 분배구조에 미치는 영향』. 한국조세재정연구원.
- 이정우(2010). 『불평등의 경제학』. 서울: 후마니타스.
- 이종철(2018). 「한국의 가계소득 불평등과 근로소득의 역할」. 『경제발전연구』 24: 35~68.
- 장하준(2014). 『장하준의 경제학 강의』. 서울: 도서출판 부키.
- 국회의원 심상정 의원 블로그. <https://blog.naver.com/713sim/221350919960>.
- Alvaredo, F.(2011). “A Note on the Relationship between Top Income Shares and the Gini Coecient.” *Economics Letters* 110: 274~277.
- Atkinson, A. B.(1970). “On the measurement of inequality.” *Journal of economic theory* 2 (3): 244~263.
- Atkinson, A. B.(2007). *Measuring Top Incomes: Methodological Issues*. In A. B. Atkinson and T. Piketty (eds.). *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, Chapter 2: 18~42. Oxford: Oxford University Press.
- Bartels C. and M. Metzging(2017). “An Integrated Approach for Top-Corrected

- Ginis.” IZA Institution of Labor Economics.
- Bishop, J. A., J. R. Chiou, and J. P. Formby(1994). “Truncation Bias and the Ordinal Evaluation of Income Inequality.” *Journal of Business and Economic Statistics* 12 : 123~127.
- Cowell, F. A.(1989). “Sampling variance and decomposable inequality measures.” *Journal of Econometrics* 42 : 27~41.
- Fichtenbaum, R. and H. Shahidi(1988). “Truncation Bias and the Measurement of Income Inequality.” *Journal of Business and Economic Statistics* 6 : 335~337.
- Heathcote, J., F. Perri, and G. L. Violante(2010). “Unequal We Stand : An Empirical Analysis of Economic Inequality in the United States : 1967~2006.” *Review of Economic Dynamics* 13 (1) : 15~51.
- Hlasny, V. and P. Verme(2018). “Top Incomes and Inequality Measurement : A Comparative Analysis of Correction Methods Using the EU SILC Data.” *Econometrics* 6 (2) : 1~30.
- Jenkins, S. P.(2017). “Pareto distributions, top incomes, and recent trends in UK income inequality.” *Economica* 84 : 261~289.
- Kuznet, S.(1955). “Economic Growth and Economic Inequality.” *American Economic Review* 45 : 1~28.
- Litchfield, J. A.(1999). “Inequality : Methods and Tools.” World Bank PovertNet website : <http://www.worldbank.org/poverty>.
- Mishel, L., J. Bernstein, and H. Shierholz(2013). “The State of Working America.” 12th ed. Ithaca, NY : Cornell University Press.
- Piketty, T. and E. Saez(2003). “Income Inequality in the United States, 1913~1998.” *Quarterly Journal of Economics* 118 (1) : 1~39.
- Schmitt, J.(2003). *Creating a Consistent Hourly Wage Series from the Current Population Survey’s Outgoing Rotation Group, 1979~2002*. Washington, DC : Center for Economic and Policy Research.

abstract

---

## A New Approach to Measuring Earning Inequality by Replacing Top Incomes

Yoo Hyerim · Min Insik

In this study, we employ top-corrected Gini coefficients to supplement the limits on which the survey data did not adequately reflect the top of the income distribution. We use both the National Tax Service of the 1000th-percentile recent income data and the Korea Labor and Income Panel Study(KLIPS) in 2013-2016. In the descriptive analysis, the ratio of the highest income in survey data was lower than that of the National Tax Service, implying that excluding the highest income earners could be biased in estimating earning inequality. Empirical analysis shows the comparison of Pareto Type I distribution using the National Tax Service data and the KLIPS income distribution. And we derive the result of top-corrected Gini coefficients.

The analysis indicates that for the top 1% earning income level, the results of KLIPS are underestimated than those obtained from the Pareto Type I distribution. In detail, top-corrected Gini coefficients for 10%, 5%, and 1% of the highest income in the National Tax Service data shows lower inequality in 2016. As the replacement rate of the highest income higher from 1% to 10%, the degree of inequality is growing greater. In addition, comparing top-corrected, Household Income and Expenditure Survey and KLIPS Gini coefficients, the top-corrected Gini coefficient measures the worst income inequality. Therefore, this study highlights that the highest income may be understated in the survey data, and in order to accurately estimate earning inequality, it is necessary to utilize a revised Gini coefficient incorporating income tax data.

Keywords : tax data, earning inequality, top-corrected gini coefficient