

정책연구
2016-04

불평등 지표 개선연구

홍민기

한국노동연구원

목 차

요 약	i
제1장 서 론	1
제2장 개인소득 분포의 비교	5
제1절 머리말	5
제2절 국세통계에서의 개인소득 분포	5
제3절 가계동향조사와 국세통계의 개인소득 분포 비교	9
제3장 지니계수의 보정	13
제1절 지니계수의 계산방법 개요	13
제2절 지니계수 보정방법	14
1. 보정의 어려움	14
2. 보정 방법	15
3. 지니계수 보정방법의 검증	19
4. 지니계수 보정 결과	22
5. 가계금융복지조사를 이용한 지니계수 보정	23
6. 개인소득 지니계수와 가구소득 지니계수	28
제4장 임금 분포의 비교	30
제1절 자료의 특성	30
제2절 임금 분포의 비교	31
1. 일용근로자를 제외한 임금의 분포	31

2. 일용근로자를 포함한 전체 임금의 분포	33
제3절 임금의 평균, 중간값 및 불평등 지표 비교	35
제5장 관리자 소득과 노동소득 분배율	38
제6장 요약 및 결론	43
참고문헌	47

표 목 차

<표 2-1> 소득의 분포(2012년)	8
<표 2-2> 국세통계와 경제활동인구조사의 비교(2012년)	9
<표 2-3> 국세통계와 가계조사의 개인소득 분포 비교(2012년)	10
<표 2-4> 개인소득 불평등 지표	11
<표 3-1> 지니계수 보정의 검증과정	19
<표 3-2> 가구주 소득별 가구원수의 분포	20
<표 3-3> 검증을 위한 지니계수 보정 결과	21
<표 3-4> 지니계수 보정 결과	23
<표 3-5> 국세통계와 가계금융복지조사의 소득분포 비교(2012년) ..	24
<표 3-6> 가계금융복지조사를 이용한 지니계수 보정(2012년)	25
<표 3-7> 가구원수별 가구의 분포	27
<표 3-8> 개인소득과 가구소득 지니계수의 비교	28
<표 4-1> 고용형태별 근로실태조사와 경제활동인구조사 부가조사 비교	30
<표 4-2> 2012년 임금 분포(일용근로자 제외)	32
<표 4-3> 2012년 임금 분포(일용근로자 포함)	34
<표 4-4> 근로소득의 평균과 분위값(천만 원)과 불평등 지수 (2012년)	35

그림목차

[그림 2-1] 개인소득 분포 비교	10
[그림 3-1] 가구주 소득과 가구원수의 관계	18
[그림 3-2] 가구주 소득과 기타 가구원 소득의 관계	18
[그림 3-3] 가구주 소득별 가구원수의 분포	20
[그림 3-4] 가구주 소득과 기타 가구소득의 관계	21
[그림 3-5] 가계금융복지조사와 가계동향조사의 소득 포착률	24
[그림 3-6] 인구총조사 대비 가구 포착률	27
[그림 3-7] 미국 가구, 가족, 개인의 지니계수	29
[그림 4-1] 2012년 임금 분포 비교(일용근로자 제외)	32
[그림 4-2] 2012년 임금 분포 비교(일용근로자 포함)	34

요 약

1. 서 론

가구조사는 일반적으로 고소득을 잘 포착하지 못하는 한계가 있어서 가구조사로 계산한 불평등 지표가 실제의 불평등도를 잘 반영하지 못한다는 지적이 있었다. 본 연구에서는 고소득을 잘 포착하는 국세통계와 가구 조사자료를 결합하여 기존의 불평등 지표를 개선하고자 한다.

조사자료로 소득을 파악하려면 몇 가지 어려움이 있다. 첫째, 표본 추출(sampling)의 문제로, 조사자료에서는 고소득을 과소 포착한다. 둘째, 자료수집의 문제에는 단위 무응답(unit non-response), 항목 무응답(item non-response), 소득 과소보고, 기타 측정오차 등이 있다. 특히 고소득 가구는 소득보고를 회피하기 위해 응답을 기피하거나 소득을 과소 보고하려는 경향이 높다. 셋째, 자료준비의 문제가 있다. 외국에서는 개인이나 가구를 식별할 수 없도록 하기 위해 보고된 고소득을 다양한 방식으로 처리한 후 자료를 공표한다. 고소득을 처리하는 방식은 코딩(coding), 절사(trimming), 중도절단(censoring) 등이 있다.

조사자료의 문제점을 보정하기 위한 기존의 외국 연구들은 단위 무응답이나 코딩의 문제와 같이 조사자료 수집이나 준비의 문제를 다루었지만 고소득을 과소 포착하는 표본추출의 문제를 다루지는 않았다. 본 연구에서는 고소득 과소포착의 문제를 다룬다.

본 연구의 구성은 다음과 같다. 제2장에서는 국세통계와 가계조사의 소득분포를 비교하여 가구조사가 얼마나 고소득을 과소 포착하는지 파악한다. 제3장에서는 가계조사와 국세통계의 개인소득 분포를

비교하여 지니계수를 보정하는 방법을 제시한다. 그리고 제4장에서는 조사자료마다 임금의 분포가 어떻게 다른지를 국세통계를 준거로 비교한다. 제5장에서는 노동소득 분배율에서 관리자 보수가 차지하는 역할에 대해 검토한다. 관리자 보수의 성격에 따라 자본과 노동간의 소득분배를 나타내는 지표인 노동소득 분배율의 수준과 추이가 달라질 수 있다. 이 장에서는 관리자의 보수가 임금총액에서 차지하는 비중을 계산한다.

2. 개인소득 분포의 비교

지니계수를 보정하려면 먼저 소득세 신고 자료와 가계조사의 소득 분포가 어떻게 다른지 알아야 한다. 이 장에서는 국세통계와 가계조사의 소득 분포를 비교한다. 여기서는 2012년 자료를 비교한다.

『국세통계연보』에서 소득에 대한 통계는 소득세 신고 단위별로 여기저기 흩어져 있다. 근로소득 연말정산 자료에는 임금근로자의 임금 분포가 있고, 종합소득세 신고 자료에는 사업소득자의 소득에 대한 정보가 있다. 일용근로자, 보험방문판매자의 개인소득 자료도 따로 있다. 이렇게 여기저기 흩어진 정보를 모아서 소득의 전체 분포를 파악하면 다음과 같다.

국세통계에서 근로소득, 사업소득, 금융소득, 일용소득, 보험 및 방문판매 소득을 보고한 사람을 모두 합하면 2012년 총 24,891,486명이다. 20세 이상 인구의 66.3%를 차지한다.

가계동향조사 분배지표는 가계동향조사와 농가경제조사를 합한 것이다. 본 연구에서 가계동향조사라고 할 때에는 농가경제조사와 합친 조사를 가리킨다.

가계동향조사는 가구를 단위로 하는 조사이다. 국세통계는 개인단위의 조사이다. 국세통계와 비교하기 위해 가계동향조사에 있는 가구원별 임금과 사업소득을 합하여 개인소득을 만들었다. 이상과 같은 방법으로 가계동향조사로부터 개인소득을 만들고, 이를 국세통계

와 비교한 것이 아래 표이다.

전체적으로 보면, 가계동향조사는 국세통계에 비해 중간소득을 많이 포착하고 있다. 이는 가계조사에서 전형적으로 나타나는 현상이다. 가계동향조사로 측정된 불평등 지수는 국세통계에 비해 낮을 것으로 예상된다.

〈국세통계와 가계조사의 개인소득 분포 비교(2012년)〉

(단위: 명)

소득구간	국세통계 (1)	가계동향조사 (2)	비율 (2)/(1)
0 초과 1천만 원 이하	9,867,728	7,801,055	0.791
1천만~2천만 원	5,333,600	5,779,421	1.084
2천만~4천만 원	4,981,571	6,910,355	1.387
4천만~6천만 원	2,297,432	2,976,719	1.296
6천만~8천만 원	1,180,055	1,036,588	0.878
8천만~1억 원	467,234	349,075	0.747
1억~2억 원	471,749	223,164	0.473
2억~3억 원	53,265	13,608	0.255
3억~5억 원	26,926	0	0.000
5억 원 이상	17,866	0	0.000
합 계	24,697,425	25,089,985	1.016

주: 소득이 0을 초과하는 사람만 계산한 것임.

3. 지니계수의 보정

이 장에서는 가계동향조사와 국세통계의 개인소득 분포를 비교하여 지니계수를 보정하고자 한다. 국세통계에서의 소득은 개인단위이고, 지니계수를 계산할 때 소득은 가구단위이다. 국세통계의 개인단위 소득으로부터 가구소득을 추정하여야 하는데, 이를 해결하여야 한다.

지니계수를 보정하는 방법은 가계동향조사와 국세통계의 관계에 따

라 두 가지로 나뉜다. 먼저 소득 포착률이 0인 아닌 구간, 예를 들어 3억 원 미만 구간에 대해서는 가계동향조사와 국세통계자료와의 관계를 이용하여 가계동향조사의 가중치를 조정(reweighting)한다. 그리고 소득 포착률이 0인 구간, 예를 들어 2012년 현재 3억 원 이상인 구간에서는 비반복적인 자료 확대(Non-iterative Data Augmentation) 방법으로 표본을 추가한다.

가중치 보정의 방법은 좀 더 자세히 설명하면 다음과 같다. 먼저 농가경제조사와 가계동향조사에서 각 가구원의 소득을 계산한다. 그리고 국세통계와 비교하여 가구원의 소득이 속한 소득구간에 해당하는 소득 포착률의 역수를 곱하여 가계동향조사의 가중치를 조정한다. 즉, 소득구간 k 에 대해 가계동향조사에서 조사된 인원수를 S_k , 국세통계의 인원수를 T_k , 소득구간별 포착률을 $\pi_k = S_k/T_k$ 이라고 하자. 그리고 가구원 p 의 소득을 y_p 라고 하고 가구 가중치를 w 라고 하자. 가구원의 소득 y_p 가 소득구간 k 에 속하는 경우 이 가구원에 해당하는 새로운 가중치는 $w_p = w/\pi_k$ 가 된다. 그리고 새로운 가구 가중치는 새로운 가구원 가중치의 평균값으로 설정한다.

비반복적인 자료 확대방법에서는 먼저 가계동향조사에서 가구주의 소득 조건부 가구원수 분포를 비모수적 혹은 모수적인 방법으로 추정한다. 그리고 가계동향조사에서 소득과 가구원수 조건부 가구소득 분포를 추정한다. 이렇게 추정한 분포에 기반해서 임의로 많은 수의 자료를 발생시키고 앞에서 가중치를 조정한 자료에 더하여 지니 계수를 계산한다. 이러한 과정을 여러 번 반복하여 평균값을 내서 최종적인 결과로 삼는다.

비반복적 자료 확대방법은 다음의 단계를 거친다.

(1단계) 가구주 소득 조건부 가구원수 h 의 분포 $f_h(h|z)$ 를 가계동향조사로부터 추정한다. 추정은 비모수적(nonparametric)으로 추정할 수도 있고, 예를 들어 프와송 분포(poisson distribution)를 가정하여 모수적으로 추정할 수도 있다. 그리고 소득과 가구원수 조건부 다

른 가구원의 소득 분포 $f_x(x|z, h)$ 를 가계동향조사로부터 추정한다.

(2단계) $\pi_k = 0$ 인 소득구간에 대해 소득구간별로 t_k 개의 소득자료를 생성한다. 개인 가중치는 $w^{(m)} = (T_k - S_k)/t_k$ 와 같다. 이렇게 생성된 소득을 가구주의 소득이라고 하고, 이 벡터를 $z = (z^{(1)}, \dots, z^{(M)})$ 라고 하자.

(3단계) 생성된 가구주 소득 $z^{(m)}$ 에 대해 $f_h(h|z)$, $f_x(x|z, h)$ 로부터 가구원수와 다른 가구원의 소득 ($h^{(m)}$, $x^{(m)}$)을 생성한다.

(4단계) 앞에서 가중치를 조정한 자료에 (3단계)에서 생성한 자료를 합하여 지니계수값을 계산한다.

마지막으로 (1단계)와 (4단계)를 Q 번 반복하고 이를 평균하여 보정 지니계수를 계산한다.

가중치를 조정한 자료와 자료 확대의 방법으로 생성된 자료를 합하여 지니계수를 계산하고 이 과정을 1,000번 반복한 결과가 다음 표에 나와 있다.

〈지니계수 보정 결과〉

	지니계수	표준편차
원자료	0.3379	
$f_h(h z)$ 의 모수적 추정	0.4117	0.0017

자료 확대방법에서 가구원수에 대해 비모수적 추정을 하였을 경우 보정 지니계수는 0.3945이다. 그리고 가구원수에 대해 프외송 추정을 하였을 경우 보정 지니계수는 0.4117이다. 소득이 낮은 구간의 정보를 이용하여 소득이 높은 구간의 가구원수를 예측할 때 모수적 방법이 실제에 더 가깝게 된다는 것을 검증과정에서 확인하였다. 이에 따라, 본 연구에서는 가구원수를 모수적으로 추정하여 지니계수를 보

정한 값인 0.4117을 최종적인 결과로 간주한다. 이는 원래의 지니계수에 비해 21.2% 증가한 값이다.

가계금융복지조사를 이용하여 지니계수를 보정하여 비교하여 본다. 가계금융복지조사에서는 모든 소득구간에서 소득 포착률이 0 이상이므로 가중치를 재조정하는 방법(reweighting)만 적용한다. 새로운 가구를 표본에 포함시키는 자료 확대방법은 사용하지 않는다.

가중치 조정의 방법을 가계금융복지조사에 적용하여 계산한 지니계수값은 0.4016이다. 2012년 가계금융복지조사에서 보정 이전 지니계수 원래 값은 0.3993이다. 보정 지니계수는 원지수에 비해 0.57% 증가하였다. 가계금융복지조사가 가계동향조사에 비해 상대적으로 고소득을 많이 포착하고 있어서 보정의 결과가 원래 값과 크게 차이 나지 않는다.

가구 내에서 가구원들 소득의 상관관계에 따라 가구소득 지니계수와 개인소득 지니계수의 차이가 결정된다. 고소득의 가구원들이 결합하여 가구를 구성하는 경향이 높을수록 가구소득 지니계수와 개인소득 지니계수가 근접한다. 한국에서 가구소득 지니계수와 개인소득 지니계수의 차이가 미국에서보다 더 크다는 것은 한국에서 가구원 간의 소득 상관계수가 미국보다 낮다는 것을 의미한다. 한국에서 여성의 경제활동 참여율이 낮고, 남성과 여성의 임금 및 소득 격차가 크기 때문에 가구주와 배우자의 소득 차이가 미국에 비해 크기 때문이다.

4. 임금 분포의 비교

일용근로자를 포함한 전체 근로자의 임금 평균값의 경우, 국세통계에서는 2,390만 원, 가계동향조사에서는 2,350만 원, 경제활동인구조사에서는 2,520만 원이다. 가계동향조사는 모든 구간에서 국세통계와 임금 분포가 비슷해서 평균값도 비슷한 데 반해 경제활동인구조사는 국세통계에 비해 저임금 포착률이 매우 낮고 중간임금 포착

률이 높아서 평균값이 높다.

전체 근로자 임금의 중간값을 보면, 국세통계에서는 1,550만 원, 가계동향조사에서는 1,790만 원, 경제활동인구조사에서는 2,160만 원이다. 경제활동인구조사는 저임금 포착률이 낮고 중간임금 포착률이 매우 높아서 중간값도 높다. 경제활동인구조사에서 중간 임금값은 평균값과 크게 차이가 나지 않는다.

가계동향조사는 국세통계와 포착률이 비슷하지만 고소득으로 갈수록 포착률이 낮기 때문에 중간값이 국세통계에 비해 높다. 국세통계에서는 고임금이 잘 포착되기 때문에 국세통계에서 나타난 임금의 분포는 조사통계에 비해 고임금쪽 꼬리가 매우 두터운 모양이다. 이 경우 평균값과 중간값의 차이가 크다.

조사통계는 표본의 수가 한정되어 있어서 고소득, 고임금을 포착하는 데는 한계가 있을 수밖에 없다. 그리고 조사통계의 목적이 소득이나 임금의 파악을 일차적인 목표로 하지 않는 경우가 많다. 가계동향조사는 가계의 지출을 정확히 파악하는 것이 일차적인 목표이고 경제활동인구조사는 고용, 비경제활동인구, 실업 등 경제활동을 정확히 파악하는 것이 목적이다. 최근 들어서는 국세자료에서 소득 파악정도가 매우 높아지고 있다. 특히 임금근로자의 임금은 매우 정확하게 파악되고 있다. 따라서 정책판단과 관련하여 임금 분포에 대한 정보가 필요한 경우에는 국세통계를 활용하는 것이 논쟁의 소지를 줄이는 방법이다.

5. 관리자 소득과 노동소득 분배율

노동소득 분배율은 부가가치 중 노동소득이 차지하는 비율이다. 노동소득 분배율을 계산하려면 노동소득의 범위를 정하여야 한다. 그 동안은 사업주의 소득을 자본소득과 노동소득으로 나누는 방법에 대해 많은 논의가 있었다. 그런데 관리자 보수의 성격에 대해서는 논의가 없었다. 그동안은 관리자의 보수를 당연히 임금으로 취급하였

다.

그러나 관리자 보수의 성격을 파악하기 위해서는 보수의 형태보다는 하는 일의 성격에 주목할 필요가 있다. 자본의 규모가 커지면서 자본가의 역할이 경영(관리), 감독, 회계 등으로 분화되었고, 이에 맞추어 경영 위계가 발전하였다. 관리자는 자본가의 역할인 경영과 관리를 하면서 보수를 받는다. 임금의 형태로 보수를 받는다고 해서 관리자의 보수를 임금이라 할 수 없고, 하는 일의 성격에 비추어 보면, 관리자의 보수는 이윤의 일부라고 하는 것이 더 적절하다.

관리자 소득의 성격을 규정하는 문제는 노동소득 분배율 측정과 관련하여 중요하다. 관리자의 소득이 노동소득 분배율의 수준이나 추세를 왜곡할 수 있다. 관리자의 소득을 임금으로 보면, 다른 근로자들의 소득이 정해지거나 감소하더라도 관리자의 소득이 급격히 증가하면 노동소득 분배율에 큰 변화가 없는 것으로 나타난다.

관리자 보수가 노동소득 분배율에 얼마나 영향을 주는가를 판단하기 위해서는 관리자 보수 전체 금액을 계산하여야 한다. 이 절에서는 관리자가 받는 임금을 계산하는 과정에 대해 논의한다.

관리자 보수를 파악하려면 자료에서 직업구분을 이용할 수 있어야 한다. 직업구분은 조사자료에만 있고 국세통계에는 없다. 그런데 조사자료는 국세통계에 비해 고임금을 과소 포착하고 있다는 문제가 있다. 본 연구에서는 조사자료와 최상위 임금비중을 파악한 기존 연구와 결합하여 관리자 보수를 계산하고자 한다. 최상위 임금에 대해서는 조사자료로 직업별 비중을 알 수 있다. 이 직업별 비중을 최상위 임금에 대해 연구한 기존 연구의 통계값과 결합하여 관리자 보수를 계산한다.

본 연구에서는 최상위 임금 5% 경계값을 기준으로 고임금에 속하는 관리자의 보수는 국세통계와 조사자료를 이용하여 계산하고, 나머지 아래쪽 임금이 속하는 관리자의 보수는 조사자료를 이용한다.

고임금이 파레토 분포를 따른다는 가정하에서 계산된 상위 5% 임금집단에 속한 관리자의 1인당 평균임금은 2억 7,726만 원이고, 최상

위 5% 임금집단에 속한 관리자의 전체 임금은 32.4조 원이다. 임금 분포 하위 95% 집단에 있는 관리자의 임금은 조사자료를 이용한다. ‘임금구조기본통계’에서 임금이 8,338만 원 이하인 하위 95%에 속한 관리자의 임금총액은 31.6조 원이다. 임금 상위 5%와 하위 95%에 속한 관리자 임금을 합하면 관리자의 임금총액은 64.1조 원으로 임금총액의 13.6%를 차지한다.

제 1 장 서 론

대표적인 불평등 지표인 지니계수는 가구조사 자료로 계산하여 왔다. 그런데 가구조사는 일반적으로 고소득을 잘 포착하지 못하는 한계가 있어서 가구조사로 계산한 불평등 지표가 실제의 불평등도를 잘 반영하지 못한다는 지적이 있었다. 이러한 문제의식에서 고소득을 잘 포착하는 국세통계를 이용하여 소득 최상위 비중을 분석한 연구가 있었다(예를 들어, 홍민기, 2015b). 본 연구에서는 고소득을 잘 포착하는 국세통계와 가구조사자료를 결합하여 기존의 불평등 지표를 개선하고자 한다.

Cowell and Flachaire(2007)은 고소득이 소득불평등 척도에 크게 영향을 준다는 것을 지적하였다. Burkhauser, Feng, Jenkins, and Larrimore(2012)는 최상위 1% 소득비중의 추세가 전체 소득불평등의 추세를 주도하였다는 것을 보여주었다. 1980년대 이후 영미권을 중심으로 최상위 소득 비중이 늘어나면서 최상위 소득을 포착하기 어려운 조사자료로 소득 불평등을 측정하면 편의(bias)가 발생한다는 문제가 더욱 심각해졌다.

일반적으로 조사자료로 소득을 파악하는 데 발생하는 어려움은 표본추출(sampling)의 문제, 자료수집의 문제, 자료준비의 문제로 요약된다. 첫째, 표본추출(sampling)의 문제는 조사자료에서 최상위 소득을 과소 포착(under-representation)하는 것을 가리킨다. 최상위 소득자는 수가 적고 사전에 미리 누구인지 알 수 없기 때문에 조사자료로 최상위 소득 1%에 속하는 사람들의 소득을 정확히 파악하려면 조사 표본수를 많이 늘려야 하는 문제가 있다.

두 번째, 자료수집의 문제에는 단위 무응답(unit non-response), 항목 무응답(item non-response), 소득 과소보고, 기타 측정오차 등이 있다. 특히 고소득 가구는 소득보고를 회피하기 위해 응답을 기피하거나 소득을 과소 보고하려는 경향이 높다.

세 번째, 자료준비의 문제가 있다. 외국에서는 개인이나 가구를 식별할 수 없도록 하기 위해 보고된 고소득을 다양한 방식으로 처리한 후 자료를 공표한다. 고소득을 처리하는 방식은 코딩(coding), 절사(trimming), 중도 절단(censoring) 등이 있다. 미국 Current Population Survey(CPS)에서는 소득이 특정값을 넘어서는 경우 그 임계점을 넘은 다른 단위의 소득과 뒤섞어 바꾼다(rank proximity swapping). 따라서 임계점을 넘어서는 소득분포는 변하지 않지만 개인소득과 인구특성변수와는 일치하지 않게 된다. 이를 코딩이라 한다. 절사는 소득이 특정값을 넘어서는 경우 표본에서 삭제하는 방법이고, 중도절단은 소득이 특정값을 넘어서는 경우 표본에 유지되지 못하도록 특정값까지만 표시하는 방법이다.

과소포착, 자료수집의 문제, 자료준비의 문제 때문에 조사자료마다 소득 분포가 다르게 된다. Juster and Kuester(1991)는 조사자료를 비교하면서 각각 어떻게 다른지 보여주었다.

조사자료의 문제점을 해결하는 방식은 가중치 조정과 표본추가의 방법이 있다. 가중치 조정(reweighting)은 조사자료에서 주어진 가중치를 다른 자료와 비교하여 바꾸는 방법이다. 미국에서는 지역의 비응답 확률에 대한 정보를 이용하여 가중치를 바꾸는 방법이 사용되었다. 이 방법은 주로 단위 무응답 때문에 발생하는 자료수집의 문제를 해결하기 위해 사용되었다.

표본추가(replacing)의 방법은 이론적 분포를 가정하여 최상위 소득을 추가하는 방법이다. 미국 CPS 최상위 고소득은 코딩이 되고 있는데(top coding) 최상위 소득 분포를 복원하기 위해 표본추가의 방법이 사용되었다(예를 들어, Cowell and Victoria-Feser, 2007; Jenkins et al., 2011; Lakner and Milanovic, 2013).

최근 Hlasny and Verme(2015)의 연구에서도 단위 무응답이나 최상위 소득 코딩의 문제를 다루는 방법에 대해 논의하였다.

조사자료의 문제점을 보정하기 위한 기존의 외국 연구들은 단위 무응답이나 코딩의 문제와 같이 조사자료 수집이나 준비의 문제를 다루었지만 고소득을 과소 포착하는 표본추출의 문제를 다루지는 않았다. 본 연구에서는 고소득 과소포착의 문제를 다룬다.

Alvaredo(2011)는 최상위 소득 비중에 대한 정보를 이용하여 지니계수를 보정하는 방법을 제시하였다. 이 방법을 활용하여 지니계수를 보정하는 것이 한국의 경우 왜 어려운지 간략히 아래에서 설명한다.

최상위 집단의 인구비중을 P , 소득비중을 S 라고 하자. 그리고, 최상위 집단의 지니계수 G^{**} , 하위집단의 지니계수 G^* 라고 하자. 그러면 전체 지니계수는 다음과 같다.

$$G = G^{**}PS + G^*(1-P)(1-S) + S - P$$

참고로, 여기서 $S - P$ 는 집단 간 불평등의 기여분을 계산한 결과이다. 최상위 집단의 인구비중 P 가 매우 작은 경우에 전체 지니계수는

$$\lim_{P \rightarrow 0} G = G^*(1-S) + S$$

와 같다. 최상위 집단이 무한소가 아닐 경우에는 다음 두 단계 과정을 거친다. 먼저, 최상위 소득이 파레토 분포를 따른다고 가정하면,

$$G^{**} = (\beta - 1) / (\beta + 1)$$

이 된다. 여기서 $\beta = \alpha / (\alpha - 1)$ 이고, α 는 파레토 계수이다. 이를 이용하여 전체 지니계수를 계산하면 다음과 같다.

$$G = \frac{\beta - 1}{\beta + 1}PS + G^*(1-P)(1-S) + S - P$$

최상위 소득 비중, 예를 들어 최상위 소득 1% 집단의 소득비중을 알고 있으면 위 식을 이용하여 조사자료의 지니계수를 보정할 수 있다.

위 방법을 한국의 상황에 적용하려면 두 가지 문제점이 있다. 첫째, 최상위 집단의 기준인 P 를 정하여야 한다. 즉, 조사자료의 지니계수값이 99% 소득집단을 대표하는지, 95% 집단을 대표하는지 정하여야 한다. 그

런데 실제 자료를 보면 분명한 단절이 일어나지 않는다. 예를 들어 조사자료가 소득 하위 99%까지는 완벽하게 소득을 포착하다가 최상위 1% 경제값 이상에서는 전혀 소득을 포착하지 못하는 방식으로 소득 조사의 단절이 일어나는 것이 아니라, 최상위 집단으로 갈수록 조사자료의 포착률이 점차적으로 감소한다.

두 번째, 조사자료의 단위와 최상위 집단을 계산한 단위가 다를 경우 위 공식을 적용할 수 없다. 지니계수는 가구단위 조사자료를 이용한다. 최상위 집단을 계산하는 것은 국세통계를 이용하는데, 조세납부 단위가 가구인 나라도 있고 개인인 나라도 있다. 한국을 비롯하여 영국, 호주, 뉴질랜드, 싱가포르, 스웨덴, 캐나다, 일본, 아르헨티나, 스페인, 이탈리아에서는 조세납부 단위가 개인이다. 조세납부 단위가 개인인 나라의 최상위 소득비중은 개인단위이다. 따라서 한국처럼 최상위 소득비중을 개인단위로 계산하는 나라에 대해서는 지니계수를 보정하는데 Alvaredo(2011)가 제안한 방법을 직접 적용할 수 없다.

본 연구의 구성은 다음과 같다. 지니계수를 보정하려면 먼저 소득세 신고 자료와 가구조사의 소득 분포가 어떻게 다른지 알아야 한다. 제2장에서는 국세통계와 가계조사의 소득 분포를 비교하여 가구조사가 얼마나 고소득을 과소 포착하는지 파악한다. 제3장에서는 가계조사와 국세통계의 개인소득 분포를 비교하여 지니계수를 보정하는 방법을 제시한다. 그리고 제4장에서는 조사자료마다 임금의 분포가 어떻게 다른지를 국세통계를 준거로 비교한다. 제5장에서는 노동소득 분배율에서 관리자 보수가 차지하는 역할에 대해 검토한다. 관리자 보수의 성격에 따라 자본과 노동간의 소득분배를 나타내는 지표인 노동소득 분배율의 수준과 추이가 달라질 수 있다. 이 장에서는 관리자의 보수가 임금총액에서 차지하는 비중을 계산한다.

제 2 장 개인소득 분포의 비교

제1절 머리말

지니계수를 보정하려면 먼저 소득세 신고 자료와 가계조사의 소득 분포가 어떻게 다른지를 알아야 한다. 이 장에서는 국세통계와 가계조사의 소득 분포를 비교한다.

여기서는 2012년 자료를 비교한다. 소득세 신고 자료는 『국세통계연보』에 나와 있다. 소득세 신고 자료를 이하에서는 ‘국세통계’ 자료라고 부를 것이다. 여기서 소득이란 임금, 사업소득, 금융소득(배당, 이자)을 합친 것을 말한다. 국세통계에서 부동산임대소득은 사업소득에 포함되어 있다.

제2절 국세통계에서의 개인소득 분포

『국세통계연보』에서 소득에 대한 통계는 소득세 신고 단위별로 여기저기 흩어져 있다. 근로소득 연말정산 자료에는 임금근로자의 임금 분포가 있고, 종합소득세 신고 자료에는 사업소득자의 소득에 대한 정보가 있다. 일용근로자, 보험방문판매자의 개인소득 자료도 따로 있다. 이렇게 여기

저기 흩어진 정보를 모아서 소득의 전체 분포를 파악하는 방법을 아래에서 설명한다.

(1) ‘근로소득 연말정산 신고현황’에서 과세대상자와 과세미달자를 합하여 임금 구간별로 인원과 금액을 정리한다. 연말정산을 한 근로자수는 총 15,768,083명이다.

(2) 근로소득 연말정산자 가운데 다른 소득이 있는 사람은 종합소득을 신고하여야 한다. 이에 대한 정보는 ‘근로소득자의 종합소득금액 신고 현황’에 있다. 총 1,205,501명이 신고를 하였다.

먼저, 근로소득 연말정산만 한 근로자와 종합소득도 신고한 근로자를 임금구간별로 분리하고자 한다. (1)에서 (2)를 빼야 한다. ‘근로소득자의 종합소득액 신고현황’은 근로소득 금액과 근로소득 외 금액을 합친 종합소득 규모별로 통계가 나와 있다. 이를 근로소득 금액구간별 인원으로 조정하여 빼야 한다. 이를 위해 근로소득 평균액과 임금구간 최소액 간의 관계를 이용하고 삼차 스플라인(cubic spline)의 방법으로 보간(interpolation)하여 종합소득규모별로 되어 있는 통계를 근로소득 규모별 통계로 전환한다. 이후 (1)에서 (2)를 빼면 연말정산만 한 근로자와 종합소득도 신고한 근로자를 임금구간별로 분리할 수 있다.

(3) 종합소득 신고를 한 사람에는 근로소득 연말정산자가 포함되어 있다. ‘종합소득금액 신고현황’에서 (2)를 빼면 종합소득 신고자 가운데에서 이중신고자를 제외하고 종합소득만 신고한 통계를 구할 수 있다. 종합소득만 신고한 사람은 총 3,147,428명이다. 이들이 순수 사업소득자이다.

(4) 일용근로자의 소득 정보는 ‘일용근로 소득자별 지급명세서 제출현황’에 나와 있다. 이 자료에서는 일용소득 보고자 총 7,878,544명을 대상으로 일용소득 금액규모별로 인원과 금액에 대한 정보가 나와 있다. 그런데 이 제출현황 자료의 각주에는 일용소득 보고자 7,878,544명 가운데 5,343,048명이 순수 일용근로자라고 밝히고 있다. 여기서 순수 일용근로자는 다른 근로소득, 사업소득, 금융소득이 없고 일용소득만 있는 사람을 가리킨다. 일용소득 보고자에서 순수 일용근로자가 차지하는 비중은 67.8%

이다. 이 비율을 일용소득 구간별로 적용하여 순수 일용근로자의 인원과 금액을 계산한다.

(5) ‘사업소득 연말정산 신고현황 1’에는 보험모집과 방문판매자의 소득과 인원 에 대한 정보가 있다.

소득세 신고자료에 연금소득이나 기타소득에 대한 정보가 있는데 소득 분포를 계산할 때 포함하지 않았다. 국세통계 ‘거주자의 연금소득 연말정산 신고현황’에 따르면, 연금수령자는 총 3,469,352명이며 이 중 60세 이상이 86.5%(2,999,611명)를 차지한다. 소액 연금소득자가 많아서 연 7백만 원 이하가 연금소득자의 86.6%를 차지한다. 고액 연금수령자는 다른 소득이 있을 가능성이 높아서 다른 소득통계와 중복된다고 생각된다. 이에 따라 개인소득 분포를 계산할 때 연금소득자는 제외하였다.

‘거주자의 기타소득 원천징수 신고현황’을 보면 기타소득자가 총 2,371,473명이다. 이 가운데 97.2%가 1천만 원 이하 소득자이다. 기타소득이 1억 원 이상인 사람은 총 1,541명으로, 연예인, 예술인, 운동선수 등일 것으로 추정된다. 숫자가 적어서 개인소득을 계산할 때 기타소득자는 제외하였다.

이렇게 하여 국세통계로부터 계산한 개인소득의 분포가 <표 2-1>에 나와 있다. 국세통계에서 근로소득, 사업소득, 금융소득, 일용소득, 보험 및 방문판매 소득을 보고한 사람을 모두 합하면 2012년 총 24,891,486명이다. 20세 이상 인구의 66.3%를 차지한다.

소득이 1천만 원 이하인 사람이 전체 소득자의 40.4%를 차지한다. 그리고 1천만 원부터 2천만 원 사이인 소득자가 전체의 21.4%를 차지한다. 그래서 2천만 원 이하 소득자가 전체의 61.9%를 차지한다. 2천만 원 이하 소득자에는 일용근로자가 492만 명, 연말정산신고 근로소득자가 760만 명 있다. 일용근로자가 아닌 근로소득자에도 저소득이 많다는 것이다. 2천만 원부터 4천만 원까지 소득자는 전체의 20.0%이다. 4천만 원 이하 소득자는 전체의 81.9%이다.

1억 원 이상 고소득자가 약 57만 명인데 연말정산 근로소득자가 46.6만 명이고, 종합소득 신고자가 10만 명이다. 연말정산 근로자 중 종합소득 신고도 한 사람은 10만 명이다. 연말정산 신고만 한 1억 원 이상 근로소득

〈표 2-1〉 소득의 분포(2012년)

(단위: 명)

소득구간	근로소득 연말정산	근로소득연 말정산+중 합소득신고	종합소득 신고자	일용근로	보험, 방문판매	합계
0 이하		16,563	177,498			194,061
1천만 원 이하	3,340,364	285,713	1,588,316	4,298,509	354,825	9,867,728
2천만 원 이하	3,717,867	239,343	655,782	626,999	93,609	5,333,600
4천만 원 이하	3,895,747	262,012	383,757	342,652	97,403	4,981,571
6천만 원 이하	1,903,513	166,693	134,737	49,250	43,239	2,297,432
8천만 원 이하	979,809	88,712	67,453	25,638	18,443	1,180,055
1억 원 이하	373,129	49,565	39,184		5,356	467,234
2억 원 이하	336,267	63,973	68,252		3,257	471,749
3억 원 이하	21,883	14,300	16,898		184	53,265
5억 원 이하	7,773	9,438	9,674		41	26,926
5억 원 초과	2,794	9,189	5,877		6	17,866
합 계	14,579,145	1,205,501	3,147,428	5,343,048	616,364	24,891,486

주: 『국세통계연보』를 이용하여 저자가 계산한 것임.

자는 36.9만 명으로 1억 원 이상 소득자의 64.7%를 차지한다. 종합소득자는 1억 원 이상 소득자의 17.6%를 차지한다. 고소득에서 근로소득자의 비중이 높다는 것을 알 수 있다.

취업현황에 대한 공식 통계는 경제활동인구조사를 이용하고 있다. <표 2-2>에서는 국세통계와 경제활동인구조사를 비교하고 있다. 일용근로자를 제외한 임금근로자는 국세통계에서 1,578만 명이고 경제활동인구조사에서는 1,608만 명이다. 경제활동인구조사가 약 2% 정도 더 많다.

사업소득자는 각각 376만 명, 572만 명으로 역시 경제활동인구조사에서 1.52배 정도 많이 파악되었다. 소득을 신고하지 않은 사업소득자가 많아서 이러한 차이가 발생한 것으로 보인다.

일용근로자는 국세통계에서 534만 명, 경제활동인구조사에서는 163만 명으로 파악된다. 국세통계에서는 연간 인원이 파악된 반면, 경제활동인구조사에서는 조사시점(1주) 동안 일용근로를 한 사람을 파악한 것이므

〈표 2-2〉 국세통계와 경제활동인구조사의 비교(2012년)

(단위: 천 명)

	국세통계 (a)	경제활동인구조사 (b)	비율 (b)/(a)
임금근로자(일용 제외)	15,785	16,085	1.02
일용근로자	5,343	1,627	0.30
사업소득자(무급가족 제외)	3,763	5,718	1.52
합 계	24,891	23,430	0.94

로 이러한 차이가 발생한다. 매우 단기간 동안 일용근로를 한 사람이 많아서 조사시점에서는 파악되지 않을 수 있다. 참고로 일용근로자의 90% 이상은 건설직 일용근로자이다. 나머지는 일일 호출근로자이다.

제3절 가계동향조사와 국세통계의 개인소득 분포 비교

가계동향조사는 2006년부터 비농가 1인 이상의 가구를 대상으로 한다. 농가는 ‘농가경제조사’를 통해 조사한다. 가계동향조사 분배지표는 가계동향조사와 농가경제조사를 합한 것이다. 본 연구에서 가계동향조사라고 할 때에는 농가경제조사와 합친 조사를 가리킨다.

가계동향조사는 가구를 단위로 하는 조사이다. 국세통계는 개인단위의 조사이다. 국세통계와 비교하기 위해 가계동향조사에 있는 가구원별 임금과 사업소득을 합하여 개인소득을 만들었다. 기타 가구원의 임금과 사업소득은 가구주나 배우자가 아닌 가구원 한 사람이 얻은 소득이라고 간주하였다.

이상과 같은 방법으로 가계동향조사로부터 개인소득을 만들고, 이를 국세통계와 비교한 것이 <표 2-3>과 [그림 2-1]에 나와 있다. 특징을 정리하면 다음과 같다.

1천만 원 이하 소득에서는 가계동향조사에서 파악된 인원이 국세통계에 비해 적다. 1천만 원 이하 소득자는 가계동향조사에서 803만 명으로 국세

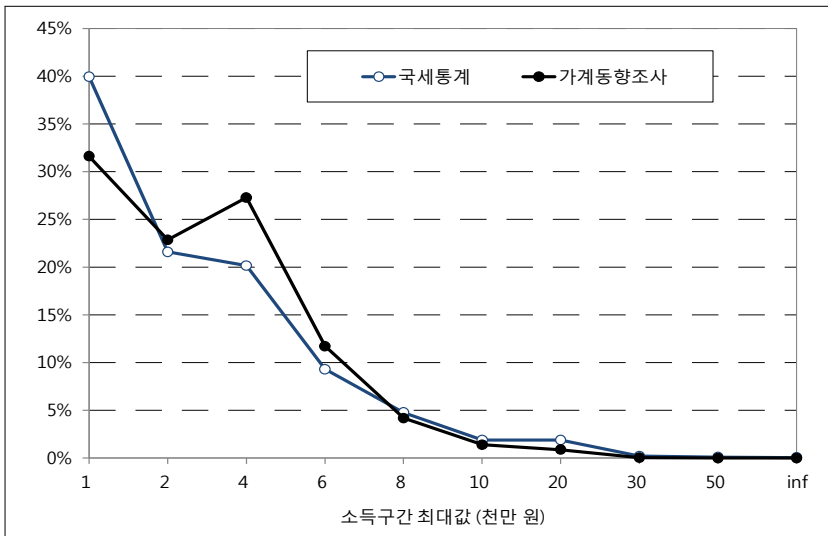
〈표 2-3〉 국세통계와 가계조사의 개인소득 분포 비교(2012년)

(단위: 명)

소득구간	국세통계 (1)	가계동향조사 (2)	비율 (2)/(1)
0 초과 1천만 원 이하	9,867,728	7,801,055	0.791
1천만~2천만 원	5,333,600	5,779,421	1.084
2천만~4천만 원	4,981,571	6,910,355	1.387
4천만~6천만 원	2,297,432	2,976,719	1.296
6천만~8천만 원	1,180,055	1,036,588	0.878
8천만~1억 원	467,234	349,075	0.747
1억~2억 원	471,749	223,164	0.473
2억~3억 원	53,265	13,608	0.255
3억~5억 원	26,926	0	0.000
5억 원 이상	17,866	0	0.000
합 계	24,697,425	25,089,985	1.016

주: 소득이 0을 초과하는 사람만 계산한 것임. 국세통계에서 0 미만의 소득을 얻고 있는 사람 194,061명이 제외된 것임.

(그림 2-1) 개인소득 분포 비교



주: 가로축은 소득구간 최대값을 의미함. 예를 들어 '6'은 소득이 4천만 원에서 6천만 원인 구간을 가리킴.

통계 대비 81.4%이다. 6천만 원 이상의 고소득 구간에서는 가계동향조사에서 파악된 인원이 국세통계에서보다 적다.

소득 1천만 원에서 6천만 원까지 중간소득 구간에서는 가계동향조사에서 파악된 인원이 국세통계보다 많다. 특히 중간소득에 해당하는 2천만~4천만 원에서는 가계동향조사가 국세통계보다 39.1% 더 많다.

전체적으로 보면, 가계동향조사는 국세통계에 비해 중간소득을 많이 포착하고 있다. 이는 가계조사에서 전형적으로 나타나는 현상이다. 가계동향조사로 측정된 불평등 지수는 국세통계에 비해 낮을 것으로 예상된다.

국세통계는 소득구간별로 인원과 금액이 나와 있는 자료이다. 이러한 자료를 그룹화 자료(grouped data)라고 한다. 그룹화 자료로부터 지니계수를 구하는 방법은 여러 가지가 있다. 이에 대해서는 부록에서 설명하였다.

<표 2-4>에서는 개인소득 불평등 지표를 보여주고 있다. 가계동향조사에서 10분위 소득이 매우 적다. 10분위 소득이 국세통계에서는 250만 원이고 가계동향조사에서는 120만 원이다. 50분위 소득은 국세통계에서 1,470만 원이고 가계동향조사에서 1,800만 원이다. 90분위 소득은 각각 5,780만 원, 5,150만 원이다.

p_{90}/p_{50} 비율은 국세통계에서 3.94로, 가계동향조사의 2.86에 비해 높다. 가계동향조사가 상대적으로 중간소득을 많이 포착하고 고소득을 적게

<표 2-4> 개인소득 불평등 지표

(단위: 천만 원)

	국세통계	가계동향조사	가계금융복지조사
p10	0.25	0.12	0.36
p50	1.47	1.80	2.04
p90	5.78	5.16	6.00
p50/p10	5.85	14.62	5.67
p90/p50	3.94	2.86	2.94
p90/p10	23.09	41.85	16.67
지니계수	0.584	0.486	0.470

주: 소득이 0 이상만 포함된 것임.

12 불평등 지표 개선연구

포착한 결과이다. 그런데 $p50/p10$ 의 비율은 가계동향조사가 국세통계보다 높다. 가계동향조사에서 10분위 소득이 매우 낮은 까닭이다.

국세통계의 지니계수는 0.584이고, 가계동향조사의 지니계수는 0.486이다. 가계동향조사가 중간소득을 많이 포착하고 있어서 지니계수가 낮다.

제 3 장

지니계수의 보정

이 장에서는 가계동향조사와 국세통계의 개인소득 분포를 비교하여 지니계수를 보정하고자 한다. 지니계수 보정은 국세통계에 비해 가계동향조사에서 과소 추정된 가구소득을 추정하는 작업이다.

제1절 지니계수의 계산방법 개요

$y \in (a, b)$, $a \geq 0$, $b \leq \infty$ 의 밀도함수(density function)를 $f(y)$, 분포를 $F(y)$ 라고 하면 지니계수는 다음과 같다.

$$G = \frac{1}{2\mu} \iint_a^b |x - y| f(x) f(y) dx dy$$

위 지니계수는 다음과 같이 공분산 형식으로 표현될 수 있다(Lerman and Yitzhaki, 1984; Shalit, 1985).

$$G = \frac{2}{\mu} cov[y, F(y)] \quad \text{where } \mu = \bar{y}$$

가중치가 있는 경우 지니계수를 계산하는 방법은 다음과 같다(Shalit, 1985 참조). 가중치를 w_i 라고 하자. 즉, $\sum_{i=1}^n w_i = 1$. 소득 y_i 를 작은 것에

서 큰 것의 순서대로 정렬한다. 즉 $(j < k \Leftrightarrow y_j < y_k)$. 이 경우 $F(y)$ 의 추정량(estimator)은

$$\hat{F}_i(y) = \sum_{j=0}^{i-1} w_j + \frac{w_i}{2}, \quad \text{where } w_0 = 0 \quad (1)$$

y 와 $F(y)$ 의 가중 공분산(weighted covariance)을 이용하여 지니계수를 계산할 수 있다. 가중치가 있는 경우 지니계수는 다음과 같이 계산된다.

$$G = \frac{2}{y} \sum_{i=1}^n w_i (y_i - \bar{y})(\hat{F}_i - \bar{F}), \quad (2)$$

여기서 \bar{y} , \bar{F} 는 가중평균값이다.

제2절 지니계수 보정방법

1. 보정의 어려움

국세통계에서의 소득은 개인단위이고, 지니계수를 계산할 때 소득은 가구단위이다. 국세통계의 개인단위 소득으로부터 가구소득을 추정하여야 하는데, 이를 해결하는 과정이 쉽지 않다.

해결해야 할 상황을 좀 더 구체적으로 살펴보면 다음과 같다. 가구소득을 y , 가구원수를 h , 가구 가중치를 w_h 라고 하자. 균등화된 가구소득은 $y_e = y/\sqrt{h}$ 이며, 가구원수 가중치는 $w = w_h h$ 이다. 지니계수를 계산할 때에는 균등화된 가구소득과 가구원수 가중치를 이용한다.

국세통계에서는 소득이 개인화되어 있다. 가구의 소득을 y_i , 가구의 재산소득 및 사적이전소득을 y_h 라고 하면 가구소득 y 는 다음과 같다.

$$y = \sum_{i=1}^h y_i + y_h$$

ϕ 는 가구공통소득 y_h 가 각 가구원에게 할당되는 비율이라고 하면, 국세통계에서 관측되는 소득은 $x_i = y_i + \phi_i y_h$ 이다. 지니계수를 계산할 때 필요한 y, h 는 관측되지 않는다. 관측되지 않는 (y, h) 를 일종의 결측자료(missing data)라고 볼 수 있다.

2. 보정방법

지니계수의 보정을 구체적으로 설명하면 다음과 같다. 가계동향조사에서 조사된 소득자료를 y_{obs} 라고 하고, 이 자료로 계산한 현재 공식 지니계수를 $G(y_{obs})$ 라고 하자. 그리고 가계동향자료의 표본에 포함되지 않은 소득자료를 z 라고 하고, 가계동향조사의 표본에 포함된 자료와 포함되지 않은 자료를 완전한 자료 $y_c = (y_{obs}, z)$, 완전한 자료로 추정한 지니계수를 $G(y_c)$ 라고 하자.

지니계수의 보정이란 y_c 를 추정하여 $G(y_c)$ 를 계산하는 것을 말한다. 구체적으로는 국세통계에서 관측된 개인화된 소득 x_i 로부터 결측자료인 (y, h) 를 추정하고 이를 이용하여 지니계수를 계산하는 것이다.

지니계수를 보정하는 방법은 가계동향조사와 국세통계의 관계에 따라 두 가지로 나뉜다. 먼저 소득 포착률이 0인 아닌 구간, 예를 들어 3억 원 미만 구간에 대해서는 가계동향조사와 국세통계자료와의 관계를 이용하여 가계동향조사의 가중치를 조정(reweighting)한다. 그리고 소득 포착률이 0인 구간, 예를 들어 2012년 현재 3억 원 이상인 구간에서는 비반복적인 자료 확대(Non-iterative Data Augumentation) 방법(McLachlan and Krishnan, 2008, 6.4 참조)으로 표본을 추가한다.

가중치 보정의 방법을 좀 더 자세히 설명하면 다음과 같다. 먼저 농가경제조사와 가계동향조사에서 각 가구원의 소득을 계산한다. 농가경제조사에서는 가구의 소득만 있고 가구원의 소득이 없어서, 가구의 근로소득과 사업소득을 합한 값을 가구원의 소득으로 간주한다. 가구의 소득을 한 가구원에게 몰아주는 방법이다. 가계동향조사에서는 가구원의 소득을 알 수 있다. 이 정보를 이용하여 가구원별로 근로소득과 사업소득을 합하여

가구원의 소득을 만든다. 나머지 가구원에 대한 소득은 가구에서 세 번째 가구원에게 몰아주는 방식으로 계산한다.

그리고 국세통계와 비교하여 가구원의 소득이 속한 소득구간에 해당하는 소득 포착률의 역수를 곱하여 가계동향조사의 가중치를 조정한다. 즉, 소득구간 k 에 대해 가계동향조사에서 조사된 인원수를 S_k , 국세통계의 인원수를 T_k , 소득구간별 포착률을 $\pi_k = S_k/T_k$ 라고 하자. 그리고 가구원 p 의 소득을 y_p 라고 하고 가구 가중치를 w 라고 하자. 가구원의 소득 y_p 가 소득구간 k 에 속하는 경우 이 가구원에 해당하는 새로운 가중치는 $w_p = w/\pi_k$ 가 된다. 그리고 새로운 가구 가중치는 새로운 가구원 가중치의 평균값으로 설정한다. 즉, 새로운 가구 가중치 w^n 은

$$w_n = \sum_{p=1}^3 w_p/3$$

비반복적인 자료 확대방법에서는 먼저 가계동향조사에서 가구주의 소득 조건부 가구원수 분포를 비모수적 혹은 모수적인 방법으로 추정한다. 그리고 가계동향조사에서 소득과 가구원수 조건부 가구소득 분포를 추정한다. 이렇게 추정한 분포에 기반해서 임의로 많은 수의 자료를 발생시키고 앞에서 가중치를 조정한 자료에 더하여 지니계수를 계산한다. 이러한 과정을 여러 번 반복하여 평균값을 내서 최종적인 결과로 삼는다.

가계동향조사에서 과소 조사된 소득구간을 $k = 1, \dots, K$ 라고 하자. 가계동향조사에서 조사된 소득구간 k 에 속하는 인원수를 S_k 라고 하고, 소득구간 k 에서 국세통계의 인원수를 T_k 라고 하면 소득구간별 포착률은 $\pi_k = S_k/T_k$ 과 같다. 비반복적 자료 확대방법은 다음의 단계를 거친다.

(1단계) 가구주 소득 조건부 가구원수 h 의 분포 $f_h(h|z)$ 를 가계동향조사로부터 추정한다. 추정은 비모수적(nonparametric)으로 추정할 수도 있고, 예를 들어 프와송 분포(poisson distribution)를 가정하여 모수적으로 추정할 수도 있다. 그리고 소득과 가구원수 조건부 다른 가구원의 소득 분포 $f_x(x|z, h)$ 를 가계동향조사로부터 추정한다.

(2단계) $\pi_k = 0$ 인 소득구간에 대해 각 소득구간별로 t_k 개의 소득자료를 생성하여 총 M 개의 소득을 생성한다. 즉, $M = \sum_{k=1}^K t_k$. 자료 생성은 각 구간에서 일양분포(uniform distribution)을 가정하여 생성한다. 따라서 개인 가중치는 $w^{(m)} = (T_k - S_k)/t_k$ 와 같다. 이렇게 생성된 소득을 가구주의 소득이라고 하고, 이 벡터를 $z = (z^{(1)}, \dots, z^{(M)})$ 라고 하자.

(3단계) 생성된 가구주 소득 $z^{(m)}$ 에 대해 $f_h(h|z)$, $f_x(x|z, h)$ 로부터 가구원수와 다른 가구원의 소득 ($h^{(m)}$, $x^{(m)}$)을 생성한다. 이로부터 균등화 소득 $y_e^{(m)}(z^{(m)}) = (z^{(m)} + x^{(m)})/\sqrt{h^{(m)}}$ 과 가중치 $w^{(m)}$ 를 생성한다.

(4단계) 앞에서 가중치를 조정한 자료에 (3단계)에서 생성한 자료를 합하여 지니계수값을 계산한다.

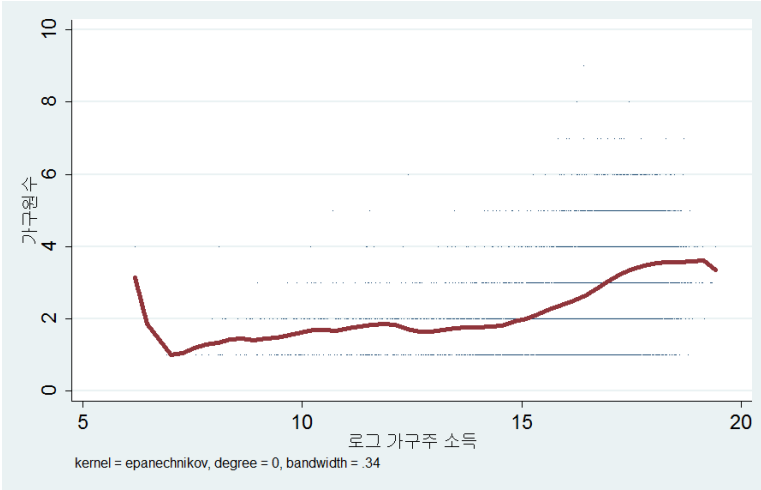
마지막으로 (1단계)와 (4단계)를 Q 번 반복한다. 즉, q 번째 반복에서 생성된 자료를 $z^q = (z^{(1)}, \dots, z^{(M)})'$ 라고 하면, 최종적으로 보정 지니계수는 다음과 같다.

$$\hat{G} = \frac{1}{Q} \sum_{q=1}^Q G(y_{obs}, y_e^q(z^q))$$

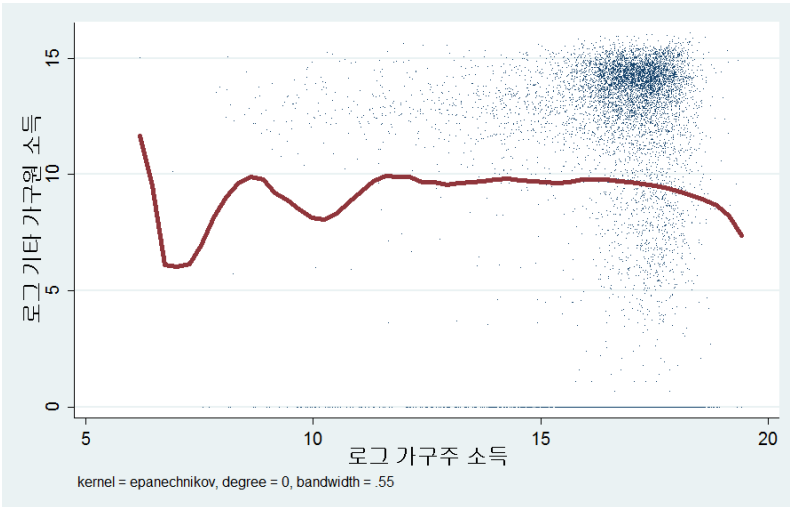
참고로, 가구주의 소득과 가구원수의 관계를 추정할 때 어떤 방식이 좋은지를 직관적으로 판단하는 데 도움을 얻고자 가구주의 소득과 가구원수의 관계를 비모수적으로 추정한 것이 [그림 3-1]에 나와 있다. 이 그림에서 가로축은 가구주 소득의 로그값이고 세로축은 가구원수이다. 그림에서 나타나듯이, 가구주 소득이 증가하면 가구원수도 증가하는 경향이 있는데 그 증가율은 매우 완만하다. 가구주 소득의 로그값과 가구원수의 관계가 선형에 가까워서 모수적(parametric) 추정을 해도 무방한 것으로 보인다.

가구주 소득의 로그값과 가구주를 제외한 다른 가구원 소득의 로그값의 관계를 비모수적으로 추정한 것이 [그림 3-2]이다. 가구주 소득의 로그값이 11, 즉 월 6만 원 수준까지는 가구주 소득과 다른 가구원 소득의 관계가 매우 불규칙적이지만 대부분의 관측치가 있는 구간에서는 가구주

[그림 3-1] 가구주 소득과 가구원수의 관계



[그림 3-2] 가구주 소득과 기타 가구원 소득의 관계



소득이 늘어나도 다른 가구원수의 소득은 거의 늘어나지 않고 일정하다. 다만 가구주 소득이 매우 높으면 기타 가구원의 소득이 약간 감소한다. 지니계수 보정을 위해서는 가구주 소득이 아주 높은 구간에 대해 외삽(extrapolation)을 해야 하는데 소득이 아주 높은 구간에서의 관련성이 분

명치 않을 수 있어서 가구주 소득과 기타 가구원 소득의 관계를 추정할 때 비모수적 방법을 사용할 것이다.

3. 지니계수 보정방법의 검증

여기서는 위에서 제시한 방법이 타당한지를 검증한다. 이를 위해, 가계동향조사 원자료에서 일부분을 누락시킨 후 남은 자료를 대상으로 위에서 제시한 방법을 사용하여 지니계수를 보정하고 이를 원자료의 지니계수값과 비교한다.

먼저, 2012년 가계동향조사 자료에서 연 소득 6백만 원 이상이면 모두 제외하고, 그 이하 소득에 대해서는 임의로 누락을 시켰다. 이렇게 임의로 누락한 자료와 원자료를 비교한 것이 <표 3-1>이다. 임의로 누락한 결과 원자료에서 73.0%의 가구가 남았다. 임의로 누락시킨 자료의 지니계수는 0.290으로 원자료의 지니계수 0.329에 비해 11.9% 감소하였다.

먼저 임의누락자료의 소득 포착률이 0보다 크고 1보다 작은 8천만 원 이하의 구간에서는 가계동향조사 원자료와 비교하여 임의누락자료에서

<표 3-1> 지니계수 보정의 검증과정

소득구간 번호	구간최소 (천만 원)	인원(명) 가계동향조사	인원(명) 임의누락자료	비율
1	1천만 원 이하	7,801,055	5,954,848	0.763
2	1천만~2천만 원	5,779,421	4,926,841	0.852
3	2천만~4천만 원	6,910,355	5,410,105	0.783
4	4천만~6천만 원	2,976,719	1,690,135	0.568
5	6천만~8천만 원	1,036,588	250,450	0.242
6	8천만~1억 원	349,075	0	0
7	1억~2억 원	223,164	0	0
8	2억 원 초과	13,608	0	0
합계		25,089,985	18,232,379	0.730
지니계수		0.3291	0.2900	

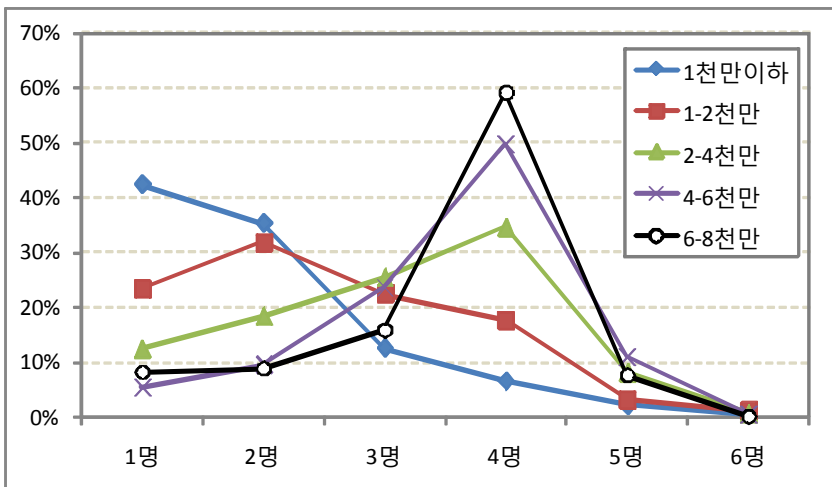
의 가중치를 조정하였다. 그리고 소득 포착률이 0인 8천만 원 이상의 구간에서는 자료 확대방법에 따라 자료를 생성하였다. 가구주 소득 조건부 가구원수 $f_h(h|z)$ 의 분포가 <표 3-2>와 [그림 3-3]에 나와 있다. 이 정보를 활용하면 비모수적 추정이 된다. 이 분포에는 8천만 원 이하에 대해서만 추정이 가능한데 8천만 원 이상의 구간에 대해서는 외삽을 하여야 한다. 본 연구에서는 8천만 원 소득자의 가구원 분포가 6천만~8천만 원 소득자와 같다고 가정한다.

한편, 가구주 소득 조건부 가구원수 분포를 모수적인 방법으로 추정할

<표 3-2> 가구주 소득별 가구원수의 분포

가구원수 \ 소득범주	1명	2명	3명	4명	5명	6명	합계
1천만 원 이하	0.425	0.354	0.126	0.067	0.023	0.005	1.0
1천만~2천만 원	0.234	0.319	0.225	0.177	0.033	0.012	1.0
2천만~4천만 원	0.125	0.184	0.255	0.346	0.082	0.008	1.0
4천만~6천만 원	0.055	0.096	0.236	0.498	0.110	0.005	1.0
6천만~8천만 원	0.082	0.089	0.159	0.592	0.077	0.002	1.0

[그림 3-3] 가구주 소득별 가구원수의 분포

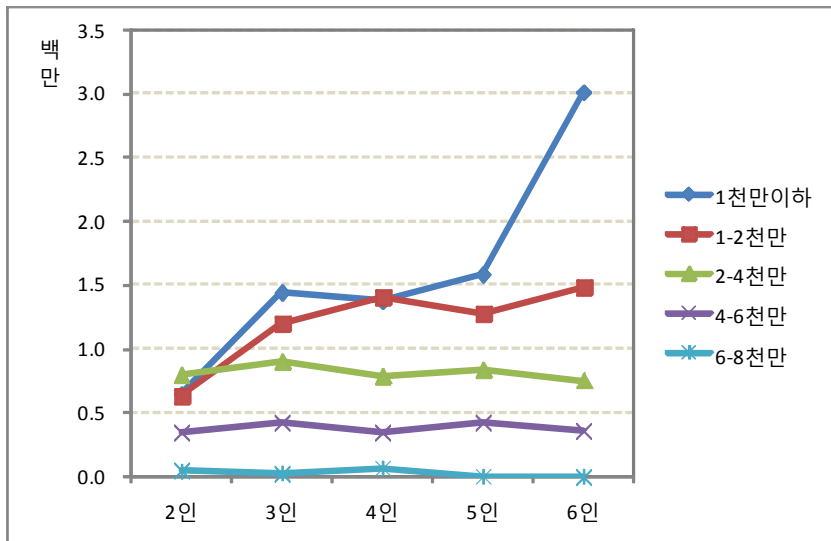


수도 있다. 이 연구에서는 (가구원수-1)을 종속변수로 하고 가구주 소득의 로그값은 독립변수로 하여 프와송 추정을 하였다. 생성된 자료에서의 가중치가 임의누락자료의 평균 가중치와 근접하도록 생성자료의 총수를 결정하였다. 이렇게 하여 새로 생성된 자료는 총 14,533개이며, 새로 생성된 자료의 평균 가중치는 500이다.

그리고 소득과 가구원수 조건부 다른 가구원의 소득 분포 $f_x(x|z, h)$ 를 비모수적으로 추정된 결과가 [그림 3-4]에 나와 있다.

이렇게 하여 생성된 자료를 기반으로 하여 1,000번을 반복하여 지니계수를 계산한 결과가 <표 3-3>에 나와 있다. 임의누락자료에서 지니계수값이 0.290이었고, 가계동향조사 원자료에서 지니계수값은 0.3291이었다.

[그림 3-4] 가구주소득과 기타 가구소득의 관계



<표 3-3> 검증을 위한 지니계수 보정 결과

	평균	표준편차
지니계수 참값	0.3291	
$f_h(h z)$ 의 비모수적 추정	0.3241	0.00159
$f_h(h z)$ 의 모수적 추정	0.3265	0.0017

8천만 원 이하의 구간에서는 가중치를 조정하고 8천만 원 이상인 구간에서 가구주 소득 조건부 가구원수 소득을 비모수적으로 추정하고 자료를 확대하여 계산하면 지니계수값이 0.3241이다. 그리고 8천만 원 이하 구간에서 가중치를 조정하고 8천만 원 이상인 구간에서 가구주 소득 조건부 가구원수 소득을 프와송 분포를 가정하여 모수적으로 추정하고 자료를 확대하여 계산하면 지니계수값이 0.3265이다.

두 값이 모두 지니계수 참값에 약간 작지만 매우 근접하였다. 가구주 소득 조건부 가구원수 소득을 비모수적으로 추정하였을 때보다 모수적으로 추정하였을 때 지니계수 참값에 더 근접하였다. 이는 외삽의 문제가 중요함을 의미한다. 비모수적 추정에서는 가구주 소득이 8천만 원 이상인 가구원수를 실제보다 과소 추정한다. 반면, 모수적 추정을 하면 8천만 원 이상의 가구원수를 비모수적 추정에 비해 상대적으로 정확히 추정을 한다. 따라서 소득이 낮은 구간의 정보를 이용하여 소득이 높은 구간에 대해 외삽을 할 때에는 모수적인 평탄화(smoothing)의 방법이 적절하다는 것을 발견할 수 있다.

4. 지니계수 보정 결과

여기서는 본 연구에서 제시한 보정방법에 따라 지니계수를 보정한 결과를 보고한다. 먼저 소득 파악률이 0보다 크고 1보다 작은 3억 원 이하 구간에 대해서는 국세통계와 비교하여 가계동향조사의 가중치를 조정하였다. 그리고 소득 파악률이 0인 3억 원 이상의 구간에서는 자료 확대의 방법으로 자료를 추가하였다.

마지막 구간 5억 원 이상에서는 상한값이 없어서 자료 확대의 방법으로 자료를 생성할 때 상한값을 정하여야 한다. 마지막 구간의 상한값은 국세통계자료에서의 평균소득을 참고하여 결정한다. 2012년 국세통계자료에서 5억 원 이상인 사람은 총 17,866명이고 이들의 평균소득은 12억 3,300만 원이다. 평균소득이 이렇게 되도록 마지막 구간 최대값을 19억 6,594만 원으로 정하였다.

가중치를 조정한 자료와 자료 확대의 방법으로 생성된 자료를 합하여

〈표 3-4〉 지니계수 보정 결과

	지니계수	표준편차
원자료	0.3379	
$f_h(h z)$ 의 비모수적 추정	0.3945	0.0016
$f_h(h z)$ 의 모수적 추정	0.4117	0.0017

지니계수를 계산하고 이 과정을 1,000번 반복한 결과가 <표 3-4>에 나와 있다.

자료 확대방법에서 가구원수에 대해 비모수적 추정을 하였을 경우 보정 지니계수는 0.3945이다. 그리고 가구원수에 대해 프와송 추정을 하였을 경우 보정 지니계수는 0.4117이다. 소득이 낮은 구간의 정보를 이용하여 소득이 높은 구간의 가구원수를 예측할 때 모수적 방법이 실제에 더 가깝게 된다는 것을 앞의 검증과정에서 확인하였다. 이에 따라, 본 연구에서는 가구원수를 모수적으로 추정하여 지니계수를 보정한 값인 0.4117을 최종적인 결과로 간주한다. 이는 원래의 지니계수에 비해 21.2% 증가한 값이다.

5. 가계금융복지조사를 이용한 지니계수 보정

자산, 부채, 소득 등을 파악하기 위해 2010년 가계금융조사가 실시되었고 2012년 표본이 개편되고 복지부문이 추가되면서 가계금융복지조사로 명칭이 변경되었다. 이 조사는 전국 읍면동에 거주하는 1인 이상의 가구를 조사대상으로 하여 약 20,000가구를 면접조사한다.

가계금융복지조사에서 가구원의 소득 분포와 국세통계를 비교한 것이 <표 3-5>이다. 소득구간별로 국세통계 인원 대비 가계금융복지조사의 인원 비율, 즉 소득 포착률은 1천만~6천만 원 사이 중간소득 구간에서 120~150%이다. 가계금융복지조사가 중간소득 구간을 과대 포착하고 있다.

이러한 경향은 가계동향조사에서도 나타난다. [그림 3-5]에서는 가계금융복지조사와 가계금융조사의 소득 포착률을 비교하여 보여주고 있다. 여기서 소득 포착률은 국세통계에 대비한 것이다. 1천만~6천만 원 중간

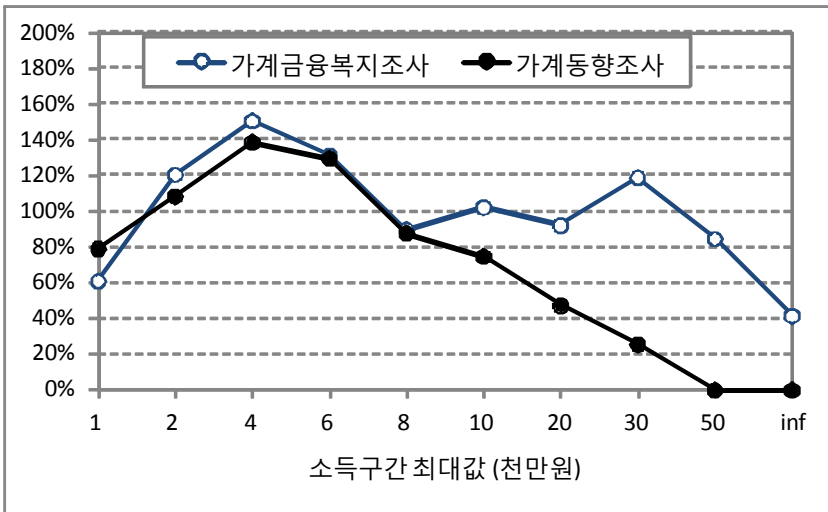
〈표 3-5〉 국세통계와 가계금융복지조사의 소득분포 비교(2012년)

(단위: 명)

소득구간	국세통계 (1)	가계금융복지조사 (2)	비율 (2)/(1)
0초과 1천만 원 이하	9,867,728	6,016,198	0.610
1천만~2천만 원	5,333,600	6,427,847	1.205
2천만~4천만 원	4,981,571	7,496,588	1.505
4천만~6천만 원	2,297,432	3,014,596	1.312
6천만~8천만 원	1,180,055	1,060,656	0.899
8천만~1억 원	467,234	477,955	1.023
1억~2억 원	471,749	434,933	0.922
2억~3억 원	53,265	63,313	1.189
3억~5억 원	26,926	22,794	0.847
5억 원 이상	17,866	7,423	0.415
합 계	24,697,425	25,022,303	1.013

주: 소득이 0을 초과하는 사람만 계산한 것임. 국세통계에서 0 미만의 소득을 얻고 있는 사람 194,061명이 제외된 것임.

〔그림 3-5〕 가계금융복지조사와 가계동향조사의 소득 포착률



소득 구간에서 가계동향조사의 소득 포착률은 110~140%이다. 1천만~6천만 원 중간소득 구간에서 가계금융복지조사의 소득 포착률이 가계동향조사보다 약간 높지만 거의 비슷하다고 할 수 있다.

8천만 원 이상 고소득 구간에서는 가계금융복지조사와 가계동향조사의 소득 포착률이 크게 차이가 난다. 8천만 원부터 3억 원 구간에서 가계금융복지조사의 소득 포착률은 92~129%로 국세통계와 차이가 많이 나지 않는다. 반면 이 구간에서 가계동향조사의 소득 포착률은 급격히 감소하여 30% 미만으로 감소한다.

전체적으로 가계금융복지조사와 가계동향조사의 소득 포착률은 8천만 원 이상 고소득에서 크게 차이가 난다. 가계금융복지조사의 소득 포착률이 고소득에서 상대적으로 매우 높기 때문에 가계금융복지조사의 지니계수가 크게 나타난다.

앞에서는 가계동향조사를 이용하여 지니계수를 보정하는 작업을 하였는데, 여기서는 가계금융복지조사를 이용하여 지니계수를 보정하여 비교하여 본다. 가계금융복지조사에서는 모든 소득구간에서 소득 포착률이 0 이상이므로 가중치를 재조정하는 방법만 적용한다. 새로운 가구를 표본에 포함시키는 자료 확대방법은 사용하지 않는다.

가중치를 조정하는 방법은 앞서와 동일하다. 소득구간 k 에 대해 가계금융복지조사에서 조사된 인원수를 S_k 라고 하고, 국세통계의 인원수를 T_k 이라고 하면 소득구간별 포착률은 $\pi_k = S_k / T_k$ 라고 하자. 가구원 p 의 소득을 y^p 라고 하고 가구 가중치를 w 라고 하자. 가구원의 소득 y_p 가 소득구간 k 에 속하는 경우 이 가구원에 해당하는 새로운 가중치는 $w_p = w / \pi_k$ 가 된다. 그리고 새로운 가구 가중치는 새로운 가구원 가중치의 평균값으로 설정한다.

보정결과가 <표 3-6>에 나와 있다. 가중치 조정의 방법을 가계금융복지조사에 적용하여 계산한 지니계수값은 0.4016이다. 2012년 가계금융복

<표 3-6> 가계금융복지조사를 이용한 지니계수 보정(2012년)

원래값	보정값
0.3993	0.4016

지조사에서 보정이전 지니계수 원래 값은 0.3993이다. 보정 지니계수는 원지수에 비해 0.57% 증가하였다. 가계금융복지조사가 가계동향조사에 비해 상대적으로 고소득을 많이 포착하고 있어서 보정의 결과가 원래 값과 크게 차이하지 않는다.

가계동향조사와 가계금융복지조사를 이용하여 지니계수를 보정할 때 그 기준을 동일하게 국세통계로 설정하였기 때문에 어떤 조사자료를 이용하더라도 보정한 결과는 원칙적으로 동일하여야 한다. 그런데, 가계동향조사를 이용하여 보정한 지니계수는 0.4117이고 가계금융복지조사를 이용해 보정한 지니계수는 0.4016으로 약간 차이가 난다. 이러한 차이가 발생할 수 있는 이유는 몇 가지가 있다. 먼저, 가구 내 가구원 간의 소득 상관관계가 두 조사에서 다를 수 있다. 가구원의 개인소득 분포가 동일하더라도 두 조사에서 가구원 간의 소득 상관관계가 높으면 가구소득으로 측정하는 지니계수가 높을 수 있다. 둘째, 가구소득과 가구원수의 상관관계가 다를 수 있다. 지니계수는 균등화 가구소득으로 계산하는데 균등화 가구소득은 가구소득을 가구원수 제곱근으로 나눈 값이다. 가구소득이 동일하더라도 가구원수가 적으면, 즉 가구소득과 가구원수 간의 상관관계가 작으면 가구소득이 동일하더라도 지니계수가 높을 수 있다. 이상과 같은 두 가지 이유는 모두 표본설계, 표본추출, 가중치 부여방식과 관련이 있다. 가계동향조사와 가계금융복지조사 모두 가구의 대표성을 중심으로 표본을 추출하고 가중치를 부여한다. 표본조사와 관련된 과정이 다르기 때문에 동일하게 국세통계를 기준으로 지니계수를 보정해도 다른 결과가 나올 수 있다.

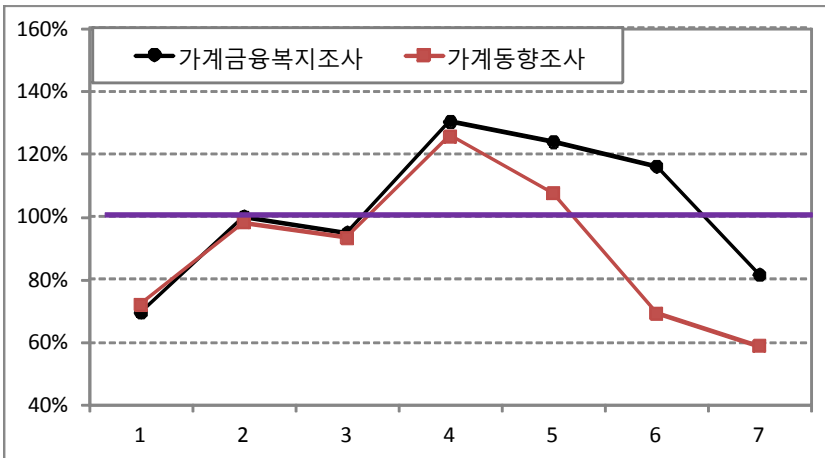
가계금융조사와 가계동향조사에서의 가구 분포와 인구총조사에서의 가구 분포를 비교한 것이 <표 3-7>에 나와 있다. 인구총조사에서 조사된 가구수의 분포가 참이라고 할 수 있다. 인구총조사와 비교하여 가구 포착률이 얼마나 되는지를 비교하여 보여준 것이 [그림 3-6]이다.

두 조사자료는 공통적으로 1인 가구를 과소 포착하며 4인 가구를 과대 포착한다. 두 조사가 다른 점은 5인 이상 가구 포착률이다. 가계동향조사는 5인 가구 포착률이 107%로 인구총조사와 근접하지만 6인, 7인 가구의 포착률은 각각 69.3%, 58.9%로 과소 포착하고 있다. 반면, 가계금융복지

〈표 3-7〉 가구원수별 가구의 분포

가구원수	가구수			비율	
	인구총조사 (1)	가계금융복지조사 (2)	가계동향조사 (3)	가계금융복지조사 (2)/(1)	가계동향조사 (3)/(1)
1	4,609,541	3,204,305	3,329,289	0.695	0.722
2	4,514,773	4,523,734	4,435,620	1.002	0.982
3	3,855,325	3,662,909	3,603,214	0.950	0.935
4	3,751,810	4,891,597	4,716,509	1.304	1.257
5	1,025,818	1,271,678	1,103,765	1.240	1.076
6	231,432	269,011	160,479	1.162	0.693
7	73,980	60,416	43,590	0.817	0.589
합계	18,062,679	17,883,651	17,392,465	0.990	0.963

〔그림 3-6〕 인구총조사 대비 가구 포착률



주: 가로축은 가구원수.

조사는 5인, 6인, 7인 가구 포착률은 각각 124.0%, 116.2%, 81.7%로, 5인과 6인 가구를 과대 포착하고 있다. 7인 가구의 경우 가계금융복지조사가 과소 포착하지만 그 정도는 가계동향조사에 비해 많이 포착하고 있다. 가계금융복지조사에서는 가구원수가 많은 가구를 많이 포착하고 있는데 가

구원수가 많은 가구에 고소득자가 있을 가능성이 더 높기 때문에 가계금융복지조사에서 지니계수가 높게 측정된다.

6. 개인소득 지니계수와 가구소득 지니계수

<표 3-8>에서는 개인소득 지니계수와 가구소득 지니계수를 비교하여 보여주고 있다. 가계동향조사와 가계금융복지조사 같은 조사자료에서 개인소득 지니계수는 각각 0.486, 0.470인데, 가구소득 지니계수는 0.338과 0.399이다. 가계동향조사로 보면, 가구소득과 개인소득 지니계수의 차이는 0.148이다. 국세통계로 측정한 개인소득 지니계수는 0.584이고 가계동향조사로 측정된 보정 지니계수는 0.412로서 그 차이는 0.172이다.

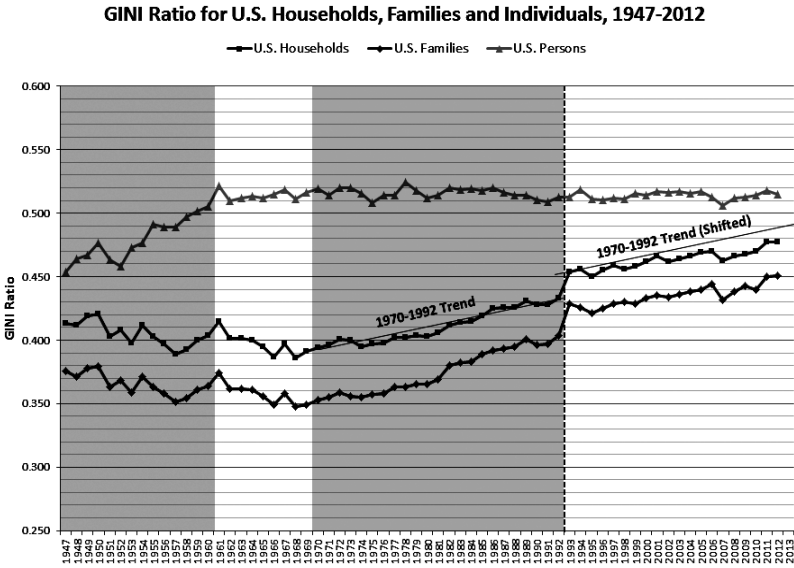
[그림 3-7]에서는 Kitov and Kitov(2013)가 연구한 미국 가구와 개인소득 지니계수를 보여주고 있다. 미국에서 1970년대부터 1980년대까지 개인소득 지니계수는 0.52 정도이고 가구소득 지니계수는 0.41 정도였다. 둘의 차이는 0.11 정도이다. 미국에서는 개인소득과 가구소득의 지니계수 차이가 점점 줄어드는 추세로 2010년대에는 그 차이가 0.3까지 줄어들었다.

가구 내에서 가구원들 소득의 상관관계에 따라 가구소득 지니계수와 개인소득 지니계수의 차이가 결정된다. 고소득의 가구원들이 결합하여 가구를 구성하는 경향이 높을수록 가구소득 지니계수와 개인소득 지니계수가 근접한다. 한국에서 가구소득 지니계수와 개인소득 지니계수의 차이가 미국에서보다 더 크다는 것은 한국에서 가구원 간의 소득 상관관계가 미국보다 낮다는 것을 의미한다. 한국에서 여성의 경제활동 참여율이 낮고, 남성과 여성의 임금 및 소득 격차가 크기 때문에 가구주와 배우자의 소득 차이가 미국에 비해 크기 때문이다.

<표 3-8> 개인소득과 가구소득 지니계수의 비교

	가계동향조사	가계금융복지조사	국세통계
개인소득 지니	0.486	0.470	0.584
가구소득 지니	0.338	0.399	
가구소득 보정 지니계수	0.412	0.402	

[그림 3-7] 미국 가구, 가족, 개인의 지니계수



Data Sources: U.S. Census Bureau, Kitov (2013), Author's Calculations

© Political Calculations 2013

주목할 만한 것은 미국에서의 추세이다. 대규모 가구의 비중이 줄어들고 1인이나 2인 가구가 늘어나면, 개인소득 분포가 변하지 않더라도, 가구원 간의 소득 상관관계가 낮아진다. 그러면 개인소득 지니계수는 변하지 않더라도 가구소득 지니계수는 증가한다. 한국에서도 가구의 분포가 소규모 가구 쪽으로 변화하면 미국에서와 마찬가지로 가구소득 지니계수가 증가할 수 있다.

제 4 장

임금 분포의 비교

제1절 자료의 특성

이 절에서는 국세통계와 여러 조사자료의 임금 분포를 비교한다. 여기서 비교되는 조사자료는 고용형태별 근로실태조사¹⁾, 경제활동인구조사이다.

<표 4-1>에서는 두 조사를 비교하여 보여주고 있다. 고용형태별 근로실태조사는 임금근로자 1인 이상 전 산업(공공행정, 국방, 가사서비스업,

<표 4-1> 고용형태별 근로실태조사와 경제활동인구조사 부가조사 비교

	고용형태별 근로실태조사	경제활동인구조사 부가조사
조사기관	고용노동부	통계청
조사단위	사업체	가구
조사범위	자영자, 고용주, 무급가족종사자 등 비임금근로자를 제외한 임금근로자	매월 15일 현재 표본가구 내에 상주하는 자 중 만 15세 이상인 자(임금 및 비임금근로자)
조사주기	1회	2회(3월, 8월)
임금기준	6월 급여 계산기간 기준	최근 3개월간의 평균임금
시작년도	2000년	2001년 8월

주 : 임금구조부문은 1968년도에 시작하여, 2008년 고용형태별 근로실태조사와 통합됨.

1) 2002년 비정규근로자실태조사에서 시작. 2003년 정규직과 비정규직으로 확대. 2008년 임금구조기본통계조사를 고용형태별 근로실태조사로 통합.

국세제외)을 조사범위로 하며, 조사대상은 1인 이상 임금근로자를 고용하고 있는 사업체에 종사하고 있는 정규, 비정규 근로자이다. 자영자, 고용주, 무급가족종사자는 제외된다.

경제활동인구조사는 가구를 단위로 한다. 조사범위는 매월 15일에 만 15세 이상인 임금 및 비임금근로자이다. 경제활동인구조사에서는 매년 3월과 8월 부가조사에서 최근 3개월간의 평균임금을 묻고 있다. 본 연구에서는 8월 조사의 임금을 이용하여 분석한다.

제2절 임금 분포의 비교

1. 일용근로자를 제외한 임금의 분포

국세통계와 조사자료의 임금분포를 비교한 것이 <표 4-2>와 [그림 4-1]에 제시되어 있다. 이 표에서는 임금근로자 가운데 일용근로자는 제외한 것이다. 국세통계에서 근로소득 연말정산자는 총 1,577만 명이다. 근로소득 연말정산 통계에는 일용자가 포함되지 않는다. 고용형태별 근로실태조사에서는 일용근로자를 제외한 임금근로자가 총 1,311만 명 포착된다. 국세통계 대비 포착률은 83.1%이다. 경제활동인구조사에서는 일용근로자를 제외한 근로자가 총 1,614만 명 포착된다. 국세통계 대비 포착률은 102%이다.

소득구간별로 파악된 인원을 비교하여 보면 다음과 같은 특징이 나타난다. 첫째, 임금 0 이상 1천만 원 이하 구간에서 국세통계에 비해 조사자료의 포착률이 낮다. 소득이 매우 낮은 1천만 원 이하 구간에서 국세통계 대비 고용형태별 근로실태조사의 포착비율은 35.0%이고, 경제활동인구조사의 포착비율은 41.0%이다.

둘째, 1천만 원 초과 4천만 원 이하의 소득구간에서는 국세통계에 비해 조사자료의 포착률이 더 높다. 특히 중간소득자가 많이 포함된 2천만 원부터 4천만 원까지에서 조사자료의 포착률이 국세통계에 비해 높다. 2천만

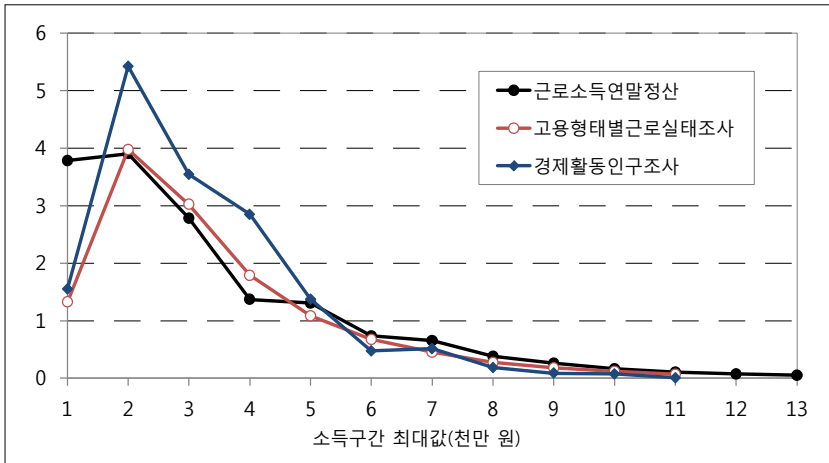
〈표 4-2〉 2012년 임금 분포(일용근로자 제외)

(단위: 명)

소득구간	근로소득 연말정산	고용형태별 근로실태조사		경제활동 인구조사	
	인원 (1)	인원 (2)	포착률 (2)/(1)	인원 (3)	포착률 (3)/(1)
1천만 원 이하	3,781,952	1,325,432	0.350	1,550,252	0.410
2천만 원 이하	3,906,324	3,974,536	1.017	5,419,072	1.387
4천만 원 이하	4,155,562	4,811,494	1.158	6,393,398	1.539
6천만 원 이하	2,043,759	1,761,691	0.862	1,853,785	0.907
8천만 원 이하	1,038,172	728,620	0.702	697,984	0.672
1억 원 이하	426,839	304,258	0.713	162,811	0.381
2억 원 이하	369,819	185,600	0.502	59,333	0.160
3억 원 이하	27,748	13,233	0.477	1,759	0.063
5억 원 이하	11,812	2,785	0.236	0	0
5억 원 초과	6,096	681	0.112	0	0
합 계	15,768,083	13,108,329	0.831	16,138,395	1.023

〈그림 4-1〉 2012년 임금 분포 비교(일용근로자 제외)

(세로축: 백만 명)



주: 가로축은 소득구간 최대값을 뜻함. 예를 들어 2는 1천만 원부터 2천만 원까지의 소득구간을 의미함. 경제활동인구조사에서 임금은 8월 부가조사에서 계산한 것임.

원에서 4천만 원까지 소득구간에서 국세통계 대비 고용형태별 근로실태 조사의 포착률은 115.8%이고, 경제활동인구조사의 포착률은 153.9%이다.

셋째, 4천만 원 이상의 소득구간에서는 국세통계에 비해 조사자료의 포착률이 낮다. 특히 소득이 증가할수록 조사자료의 포착률이 감소한다.

2. 일용근로자를 포함한 전체 임금의 분포

일용근로자를 포함하여 임금근로자의 임금 분포를 보여주고 있다. 고용형태별 근로실태조사는 사업체 단위의 조사로 일용근로자가 잘 포착되지 않았을 가능성이 높기 때문에 여기서는 가계동향조사와 경제활동인구조사를 국세통계 자료와 비교한다.

일용자를 포함한 전체 근로자 임금 1천만 원 이하 구간에서 경제활동인구조사와 가계동향조사의 포착률은 29.3%와 72.7%이다. 상대적으로 가계동향조사의 포착률이 높은 편이다. 임금 1천만 원부터 4천만 원까지 구간에서 두 조사의 포착률은 100%를 넘는다. 이 구간에서 경제활동인구조사의 포착률은 130~148%로, 가계동향조사의 102~113%에 비해 매우 높다. 가계동향조사가 이 중간임금 구간에서 상대적으로 국세통계와 비슷한 인원을 포착하고 있다.

4천만 원 이상 임금구간에서는 임금이 높을수록 두 조사의 포착률이 감소한다. 두 조사자료를 비교하면, 고임금 구간에서 가계동향조사의 포착률이 경제활동인구조사에 비해 높은 편이다. 가계동향조사의 포착률이 경제활동인구조사에 비해 모든 임금구간에서 국세통계에 가깝다. 최근 국세통계가 적어도 임금에 대해서는 정확히 보고한다고 보고 이를 준거로 삼는다면, 가계동향조사가 임금의 분포에 대해서는 좀 더 정확한 정보를 준다고 할 수 있다.

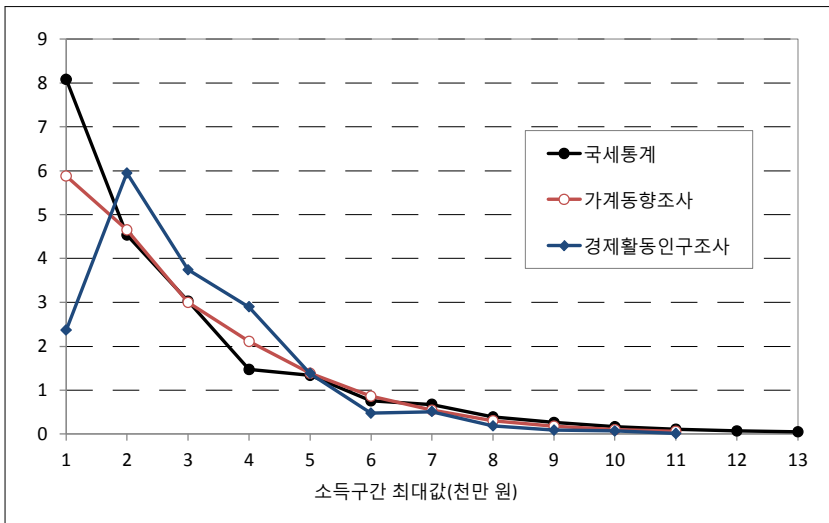
〈표 4-3〉 2012년 임금 분포(일용근로자 포함)

(단위: 명)

소득구간	국세통계	경제활동인구조사		가계동향조사	
	인원 (1)	인원 (3)	포착률 (3)/(1)	인원 (2)	포착률 (2)/(1)
1천만 원 이하	8,080,461	2,367,319	0.293	5,878,046	0.727
2천만 원 이하	4,533,323	5,944,535	1.311	4,649,381	1.026
4천만 원 이하	4,498,214	6,638,636	1.476	5,114,339	1.137
6천만 원 이하	2,093,009	1,861,436	0.889	2,243,661	1.072
8천만 원 이하	1,063,810	697,984	0.656	853,838	0.803
1억 원 이하	426,839	162,811	0.381	278,585	0.653
2억 원 이하	369,819	59,333	0.160	171,118	0.463
3억 원 이하	27,748	1,759	0.063	8,987	0.324
5억 원 이하	11,812	0	0	0	0
5억 원 초과	6,096	0	0	0	0
합 계	21,111,131	17,733,814	0.840	19,197,956	0.909

〈그림 4-2〉 2012년 임금 분포 비교(일용근로자 포함)

(세로축: 백만 명)



주: 가로축은 소득구간 최대값을 뜻함. 예를 들어 2는 1천만 원부터 2천만 원까지의 소득구간을 의미함.

제3절 임금의 평균, 중간값 및 불평등 지표 비교

일자리나 복지정책에서 임금의 평균값이나 중간값이 기준이 되는 경우가 많다. 그런데 정책마다 사용하는 자료가 달라서 어떤 것을 평균이나 중간값으로 삼아야 하는 것 자체도 논쟁의 대상이 된다. 여기서는 여러 자료로 평균값과 중간값을 비교하여 어떤 자료를 이용할 경우 논란의 여지가 줄어들 수 있는지 살펴보고자 한다. 아울러 여러 가지 불평등 지표도 비교해 보고자 한다.

<표 4-4>에서는 여러 가지 자료로 계산한 평균값, 분위값, 불평등 지수를 비교하여 보여주고 있다. 이 표의 값들은 모두 0을 초과한 임금만을 대상으로 계산한 것이다. 일용근로자를 포함한 전체 근로자의 임금 평균값의 경우, 국세통계에서는 2,390만 원, 가계동향조사에서는 2,350만 원, 경제활동인구조사에서는 2,520만 원이다. 앞에서 보았듯이 가계동향조사는 모든 구간에서 국세통계와 임금 분포가 비슷해서 평균값도 비슷한 데 반해 경제활동인구조사는 국세통계에 비해 저임금 포착률이 매우 낮고

<표 4-4> 근로소득의 평균과 분위값(천만 원)과 불평등 지수(2012년)

	일용포함 전체근로자			일용 제외		
	국세 통계	가계동향조사	경제활동인구조사	국세 통계	고용형태근로조사	경제활동인구조사
평균값	2.39	2.35	2.52	2.99	3.03	2.66
p10	0.26	0.15	0.84	0.42	1.13	1.02
p50	1.55	1.79	2.16	2.09	2.37	2.24
p90	5.80	5.23	4.80	6.59	5.87	4.80
p50/p10	5.92	11.72	2.57	5.02	2.10	2.20
p90/p50	3.75	2.93	2.22	3.14	2.48	2.14
p90/p10	22.22	34.28	5.71	15.79	5.19	4.71
지니계수	0.533	0.483	0.347	0.466	0.365	0.330

주: 임금이 0을 초과한 근로자를 대상으로 계산한 것임.

중간임금 포착률이 높아서 평균값이 높다.

국민계정을 이용해서도 임금의 평균값을 구할 수 있다. 국민계정의 임금 및 급여는 현물을 포함하고 있어서 이를 제외하면 현금으로 지급된 임금총액이 된다. 2012년 국민계정에서 임금 및 급여에서 현물을 제외한 임금은 약 478.3조 원이다. 경제활동인구조사에서 파악된 임금근로자는 1,771.2만 명이다. 따라서 국민계정으로부터 임금 평균값을 계산하면 2,700만 원이다. 국민계정으로 계산한 평균값은 국세통계에서 연말정산 자료와 일용소득을 합쳐서 계산한 임금 평균값보다 높다. 국민계정에서 임금 및 급여를 계산할 때에는 여러 가지 조사통계를 활용하고 있는데 현재로서는 구체적인 산출방법을 파악하지 못해서 통계의 정확성을 논할 수 없다.

각 자료별로 전체 근로자 임금의 중간값을 보면, 국세통계에서는 1,550만 원, 가계동향조사에서는 1,790만 원, 경제활동인구조사에서는 2,160만 원이다. 경제활동인구조사는 저임금 포착률이 낮고 중간임금 포착률이 매우 높아서 중간값도 높다. 경제활동인구조사에서 중간임금값은 평균값과 크게 차이가 나지 않는다.

가계동향조사는 국세통계와 포착률이 비슷하지만 고소득으로 갈수록 포착률이 낮기 때문에 중간값이 국세통계에 비해 높다. 국세통계에서는 고임금이 잘 포착되기 때문에 국세통계에서 나타난 임금의 분포는 조사통계에 비해 고임금 쪽 꼬리가 매우 두터운 모양이다. 이 경우 평균값과 중간값의 차이가 크다.

경제활동인구조사의 임금 10분위값은 840만 원으로 국세통계의 260만 원, 가계동향조사의 150만 원에 비해 매우 높다. 경제활동인구조사의 90분위값은 4,800만 원으로 국세통계의 5,800만 원, 가계동향조사의 5,230만 원에 비해 낮다. 가계동향조사는 국세통계에 비해 고임금이 적게 포착되어 90분위값도 국세통계에서보다 작다. 그런데 가계동향조사에서는 10분위값이 매우 적게 나타난다. 이는 조사방식과 관련이 있는 것으로 보인다.

임금의 지니계수를 비교하여 보면, 국세통계에서는 0.533이고 가계동향조사에서는 0.483이고 경제활동인구조사에서는 0.347이다. 경제활동인구조사에서의 임금 분포는 국세통계에서의 임금 분포에 비해 저임금 쪽과 고임금 쪽이 모두 중간임금 쪽으로 줄어든 형태이다. 따라서 경제활동인

구조사에서의 지니계수가 국세통계에 비해 매우 낮다. 경제활동인구조사는 3개월간의 평균소득을 조사하기 때문에 실제 소득에 비해 평탄화된 소득이 조사될 가능성이 높다. 경제활동인구조사는 소득 파악을 주된 목적으로 하는 조사가 아니라 고용형태를 파악하기 위한 조사이기 때문에 소득분포의 파악에는 한계가 있다.

일용임금을 제외한 통계값이 표의 오른쪽에 나와 있다. 전체적인 경향은 일용을 포함하였을 때와 비슷하다. 경제활동인구조사의 임금 분포는 국세통계에 비해 저임금과 고임금이 중간 쪽으로 몰린 형태여서 10분위 값은 국세통계에 비해 높고, 90분위는 국세통계에 비해 낮다. 따라서 경제활동인구조사로 계산된 지니계수값은 국세통계에 비해 낮다. 고용형태별 근로실태조사에서의 임금분포는 경제활동인구조사와 마찬가지로 국세통계에 비해 안쪽으로 몰린 형태이다. 지니계수를 비교하면 고용형태별 근로조사에서는 경제활동인구조사에 비해 고임금이 많이 포착되어서 지니계수값이 약간 높다(각각 0.365와 0.330).

지금까지 국세통계와 여러 가지 조사통계에서의 임금 분포를 비교하여 보았다. 조사통계는 표본의 수가 한정되어 있어서 고소득, 고임금을 포착하는 데는 한계가 있을 수밖에 없다. 그리고 조사통계의 목적이 소득이나 임금의 파악을 일차적인 목표로 하지 않는 경우가 많다. 가계동향조사는 가계의 지출을 정확히 파악하는 것이 일차적인 목표이고 경제활동인구조사는 고용, 비경제활동인구, 실업 등 경제활동을 정확히 파악하는 것이 목적이다. 조사자료를 이용하여 임금의 정확한 분포를 파악하는 것은 어렵다. 국세통계는 조세의 획득을 목적으로 한다. 최근 들어서는 국세자료에서 소득 파악정도가 매우 높아지고 있다. 특히 임금근로자의 임금은 매우 정확하게 파악되고 있다. 따라서 정책판단과 관련하여 임금 분포에 대한 정보가 필요한 경우에는 국세통계를 활용하는 것이 논쟁의 소지를 줄이는 방법이다.

제 5 장

관리자 소득과 노동소득 분배율

노동과 자본 간의 분배지표로서 노동소득 분배율이 사용되고 있다. 노동소득 분배율은 부가가치 중 노동소득이 차지하는 비율이다. 노동소득 분배율을 계산하려면 노동소득의 범위를 정하여야 한다. 노동소득의 범위가 분명치 않아 논쟁거리가 될 만한 것들이 있다.²⁾ 첫째, 임금노동자의 범위가 분명치 않다. 관리자나 사업주의 소득 중 얼마가 자본소득이고 얼마가 노동소득인가가 분명치 않다. 둘째, 기업 경영진이 가진 스톡옵션은 노동소득인가 자본소득인가? 셋째, 일할 때 사회보험료를 내고 은퇴 뒤에 연금이나 건강보험 혜택을 받는다. 그렇다면 연금이나 건강보험 혜택도 노동소득에 포함되는가?

이상과 같은 문제에서 그동안은 사업주의 소득을 자본소득과 노동소득으로 나누는 방법에 대해 많은 논의가 있었다. 그런데 관리자 보수의 성격에 대해서는 논의가 없었다. 그동안은 관리자의 보수를 당연히 임금으로 취급하였다.

주주는 배당의 형식으로 이윤의 일부를 보수로 받고 개인사업주는 이윤을 소득으로 하는 데 반해, 관리자는 임금의 형태로 보수를 받기 때문에 이들의 소득원천은 이윤과 임금으로 서로 다른 것처럼 보인다. 그러나 관리자 보수의 성격을 파악하기 위해서는 보수의 형태보다는 하는 일의

2) 노동소득 분배율의 측정과 관련된 문제는 Krueger(1999), Boggio et al(2009)에서 상세히 다루고 있다.

성격에 주목할 필요가 있다. 자본의 규모가 커지면서 자본가의 역할이 경영(관리), 감독, 회계 등으로 분화되었고, 이에 맞추어 경영 위계가 발전하였다. 관리자는 자본가의 역할인 경영과 관리를 하면서 보수를 받는다. 임금의 형태로 보수를 받는다고 해서 관리자의 보수를 임금이라 할 수 없고, 하는 일의 성격에 비추어 보면, 관리자의 보수는 이윤의 일부라고 하는 것이 더 적절하다. 주주, 관리자, 사업주의 보수의 원천은 이윤이라 할 수 있다. 관리자의 보수를 이윤으로 볼 것인가 임금으로 볼 것인가는 결국 관리자가 보수를 받는 형식에 주목할 것인가 아니면 관리자가 어떤 일을 해서 보수를 받는가 하는 내용에 주목할 것인가에 달려 있다. 형식에 주목하면 임금이 되고 내용에 주목하면 이윤이 된다.

관리자의 보수가 소득 불평등을 주도한 핵심 요인이라는 것은 외국에서 이미 많이 밝혀진 사실이다. 자본주의 발전국에서 관리자의 보수가 증가한 이유에 대해서는 주주자본주의 발전의 결과로 설명한다. 자본주의가 발전하면서 자본의 규모가 커지면서 상대적으로 개별 주주의 지분이 감소하고, 이에 따라 적은 몫으로 큰 자본을 통제하기에 어려움이 발생한다. 영국과 미국에서는 주주와 경영의 분리가 발생하였고, 1970년대 이후 주주와 경영권을 일치시키기 위해 노력하였다. 이러한 노력 중 하나가 관리자의 보수를 성과와 연계(스톡옵션)시킨 것이다. 관리자에 대한 외부 노동시장이 넓어서 관리자의 보수가 급격히 증가하였다. 한편으로 경영자가 이사회에 영향력을 행사하여 자신의 보수를 높이기도 하였다.

영국이나 미국에서는 관리자의 보수에서 스톡옵션과 같이 주식으로 성과를 보상하는 방식의 비율이 늘어나고 있다. 관리자는 스톡옵션을 행사하거나 주식을 팔아서 소득을 얻는데 이를 온전히 임금이라고 간주하는 것도 논쟁거리가 될 수 있다.

관리자 소득의 성격을 규정하는 문제는 노동소득 분배율 측정과 관련하여 중요하다. 관리자의 소득이 노동소득 분배율의 수준이나 추세를 왜곡할 수 있다. 미국에서는 관리자의 소득이 다른 근로자의 소득에 비해 급격하게 증가하였다. 관리자의 소득을 임금으로 보면, 다른 근로자들의 소득이 정체하거나 감소하더라도 관리자의 소득이 급격히 증가하면 노동소득 분배율에 큰 변화가 없는 것으로 나타난다.

관리자 보수가 노동소득 분배율에 얼마나 영향을 주는가를 판단하기 위해서는 관리자 보수 전체 금액을 계산하여야 한다. 이 절에서는 관리자가 받는 임금을 계산하는 과정에 대해 논의한다.

관리자 보수를 파악하려면 자료에서 직업구분을 이용할 수 있어야 한다. 직업구분은 조사자료에만 있고 국세통계에는 없다. 그런데 조사자료는 국세통계에 비해 고임금을 과소 포착하고 있다는 문제가 있다. 본 연구에서는 조사자료와 최상위 임금비중을 파악한 기존 연구와 결합하여 관리자 보수를 계산하고자 한다. 최상위 임금에 대해서는 조사자료로 직업별 비중을 알 수 있다. 이 직업별 비중을 최상위 임금에 대해 연구한 기존 연구의 통계값과 결합하여 관리자 보수를 계산한다.

관리자 보수를 파악하기 전에 먼저 임금근로자의 임금총액을 계산하여야 한다. 여기서는 2014년 자료로 예를 든다. 먼저, 국민계정의 임금 및 급여 가운데 현물을 제외한 임금을 구한다. 국민계정의 임금 및 급여는 약 573조 원인데, 현금을 제외한 임금은 531조 원이다.³⁾ 국세통계에서 일용소득은 58.6조 원이다. 일용소득을 제외한 임금근로자의 임금총액은 약 473조 원이다. 일용소득을 제외하는 이유는 관리자 보수 계산을 위해 사용하는 자료는 노동부의 조사자료인 ‘임금구조기본통계’인데 이 자료에는 일용소득이 빠져 있기 때문이다.

본 연구에서는 최상위 임금 5% 경계값을 기준으로 고임금에 속하는 관리자의 보수는 국세통계와 조사자료를 이용하여 계산하고, 나머지 아래쪽 임금에 속하는 관리자의 보수는 조사자료를 이용한다.

홍민기(2015)에 따르면, 2014년 최상위 임금 5%의 경계값은 8,338.5만 원이다. 그리고 최상위 임금 5% 집단이 전체 임금에서 차지하는 임금비중은 22.0%이다. 임금구조기본통계⁴⁾ 조사로 계산하면 최상위 5% 임금집단에서 관리자가 차지하는 인원비중은 12.5%이고 최상위 1% 임금집단에서 차지하는 비중은 25.3%이다.

3) 현물을 계산하는 방법에 대해서는 홍민기(2015) 참조.

4) ‘고용형태별 근로실태조사’에서는 관리자의 비중이 1.4% 정도에 불과하다. ‘고용형태별 근로실태조사’는 일용, 비정규직 등 다양한 고용형태의 근로자를 많이 포착하는 대신 관리자는 매우 적게 포착하고 있다. 관리자의 비중이 매우 적게 포착되어서 이 자료를 이용하면 근로자의 총임금이 매우 적은 것으로 나타난다.

다음으로 최상위 5% 임금집단에 속하는 관리자의 평균임금을 구하여야 한다. 제한된 정보를 이용하여 평균임금을 계산하여야 하기 때문에 관리자의 임금 분포에 대한 가정이 필요하다. 최상위 분포가 파레토 분포를 따른다고 가정하고 있는 기존 연구들(Atkinson et al., 2011)을 감안하여 본 연구에서도 관리자의 고임금 부분은 파레토 분포를 따른다고 가정한다. 파레토 분포의 가정하에서 평균임금을 구하는 과정은 다음과 같다 (홍민기, 2016 참조).

임금 y 의 밀도함수 $f(y)$ 와 임금이 y_j 이상인 사람의 비중 $H_j(y)$ 은 다음과 같다.

$$f(y) = \alpha \frac{m^\alpha}{y^{1+\alpha}}$$

$$H_j(y) = \int_{y_j}^{\infty} f(y)dy = (m/y_j)^\alpha, \quad (\text{파라미터 : } \alpha, m) \quad (1)$$

임금 1%, 5% 집단의 경계값을 y_j, y_k 라고 이 집단에 속한 직업 c 의 인원수를 각각 N_j^c, N_k^c 라고 하자. 최상위 임금집단에서 각 직업이 차지하는 비중을 알 수 있기 때문에 최상위 소득집단에 속한 직업의 인원수 N_j^c, N_k^c 는 알 수 있다. 그런데 조사자료는 표본을 통해 조사를 한 것이기 때문에 직업 c 의 전체 인원수 N^c 는 정확히 알 수 없다. 전체 인원수 N^c 를 알지 못하는 상황에서 평균임금을 계산하는 방법은 다음과 같다. 식 (1)에 따라 다음의 두 식이 모두 성립한다.

$$N_j^c/N^c = (m/y_j)^\alpha, \quad N_k^c/N^c = (m/y_k)^\alpha \quad (2)$$

이 두 식의 비례관계를 이용하면 전체 인원수 N^c 가 없어도 다음과 같이 파라미터 α 값을 구할 수 있다.

$$\alpha = \frac{\ln(N_k^c/N_j^c)}{\ln(y_j/y_k)} \quad (3)$$

다음으로, y_j 이상인 임금의 평균값 y^* 는 다음과 같다.

$$y^*(y) = \left[\int_y^\infty yf(y)dy \right] / \left[\int_y^\infty f(y)dy \right] = \frac{\alpha}{\alpha-1}y \quad (4)$$

위에서 구한 α 를 대입하면 y_j 이상인 소득의 평균값 y^* 를 구할 수 있다. α 값을 구할 때에는 임금 1%와 5% 경계값과 비중을 이용하여 계산한다.

이렇게 계산된 상위 5% 임금집단에 속한 관리자의 1인당 평균임금은 2억 7,726만 원이고, 최상위 5% 임금집단에 속한 관리자의 전체 임금은 32.4조 원이다.

임금 분포 하위 95% 집단에 있는 관리자의 임금은 조사자료를 이용한다. '임금구조기본통계'에서 임금이 8,338만 원 이하인 하위 95%에 속한 관리자의 임금총액은 31.6조 원이다. 임금 상위 5%와 하위 95%에 속한 관리자 임금을 합하면 관리자의 임금총액은 64.1조 원으로 임금총액의 13.6%를 차지한다.

제 6 장

요약 및 결론

가구조사는 일반적으로 고소득을 잘 포착하지 못하는 한계가 있어서 가구조사로 계산한 불평등 지표가 실제의 불평등도를 잘 반영하지 못한다는 지적이 있었다. 본 연구에서는 고소득을 잘 포착하는 국세통계와 가구조사자료를 결합하여 기존의 불평등 지표를 개선하고자 하였다. 연구 결과를 요약하면 다음과 같다.

지니계수를 보정하려면 먼저 소득세 신고 자료와 가계조사의 소득 분포가 어떻게 다른지를 알아야 한다. 이 연구에서는 2012년 국세통계와 가계조사의 소득 분포를 비교하였다. 전체적으로 보면, 가계동향조사는 국세통계에 비해 중간소득을 많이 포착하고 있다. 이는 가계조사에서 전형적으로 나타나는 현상이다.

지니계수를 보정하는 방법은 가계동향조사와 국세통계의 관계에 따라 두 가지로 나뉜다. 먼저 소득 포착률이 0인 아닌 구간, 예를 들어 3억 원 미만 구간에 대해서는 가계동향조사와 국세통계자료의 관계를 이용하여 가계동향조사의 가중치를 조정한다. 그리고 소득 포착률이 0인 구간, 예를 들어 2012년 현재 3억 원 이상인 구간에서는 비반복적인 자료 확대(Non-iterative Data Augmentation) 방법으로 표본을 추가한다.

가중치를 조정한 자료와 자료 확대의 방법으로 생성된 자료를 합하여 지니계수를 계산한 결과 0.4117로 계산되었다. 이는 원래의 지니계수에 비해 21.2% 증가한 값이다. 가중치 조정의 방법을 가계금융복지조사에 적용하여 계산한 지니계수값은 0.4016이다. 2012년 가계금융복지조사에

서 보정 이전 지니계수 원래 값은 0.3993이다. 보정 지니계수는 원지수에 비해 0.57% 증가하였다. 가계금융복지조사가 가계동향조사에 비해 상대적으로 고소득을 많이 포착하고 있어서 보정의 결과가 원래 값과 크게 차이나지 않는다.

그리고 이 연구에서는 여러 가지 자료에서의 임금 분포를 비교하였다. 조사통계는 표본의 수가 한정되어 있어서 고소득, 고임금을 포착하는 데 한계가 있을 수밖에 없다. 그리고 조사통계의 목적이 소득이나 임금의 파악을 일차적인 목표로 하지 않는 경우가 많다. 최근 들어서는 국세자료에서 소득 파악정도가 매우 높아지고 있다. 특히 임금근로자의 임금은 매우 정확하게 파악되고 있다. 따라서 여러 가지 정책에 기초자료로 사용되는 임금 평균값이나 중간값을 얻고자 할 때에는 국세통계를 활용하는 것이 논쟁의 소지를 줄이는 방법이다.

관리자 소득의 성격을 규정하는 문제는 노동소득 분배율 측정과 관련하여 중요하다. 관리자의 소득이 노동소득 분배율의 수준이나 추세를 왜곡할 수 있다. 관리자의 소득을 임금으로 보면, 다른 근로자들의 소득이 정체가거나 감소하더라도 관리자의 소득이 급격히 증가하면 노동소득 분배율에 큰 변화가 없는 것으로 나타난다.

고임금이 파레토 분포를 따른다는 가정하에서 계산된 상위 5% 임금 집단에 속한 관리자의 1인당 평균임금과 전체 임금을 계산하고, 임금분포 하위 95% 집단에 있는 관리자의 임금은 '임금구조기본통계'로 계산한 결과 2014년 관리자의 임금총액은 64.1조 원으로 임금총액의 13.6%를 차지한다.

본 연구에서는 2012년 한 해에 대해서만 지니계수 보정작업을 하였다. 앞으로 여러 기간에 대해 보정작업을 하여 온전한 시계열 자료를 만드는 것이 다음의 연구과제이다.

[보론 1] 그룹화된 자료로 지니계수를 계산하는 방법

구간별 자료(grouped data)로 지니계수를 계산하는 방법은 여러 가지이다. 이하에서는 Abounoori and McCloughan(2003)가 제안한 방법을 요약한다.

총 n 명, 소득구간의 수 K , 구간 k 에 있는 인원수 n_k 이라고 하자. 전체 소득평균을 \bar{y} , 구간별 소득평균을 \bar{y}_k 라고 하자. 그러면 지니계수는 다음과 같이 계산된다.

$$G = C \sum_{k=1}^K w_k \left(1 - \frac{\bar{y}_k}{\bar{y}} \right)$$

여기서 $C = 2/(n(n+1))$,

$$w_k = \frac{1}{2} \left\{ \sum_{k=k}^K n_k \left(\sum_{k=k}^K n_k + 1 \right) - \sum_{k=k+1}^K n_k \left(\sum_{k=k+1}^K n_k + 1 \right) \right\}$$

[보론 2] 가계동향조사 분배지표 구성과 관련된 문제

가계동향 분배지표는 가계동향조사와 농가경제조사를 합친 것이다. 가계동향 분배지표를 구성할 때에는 몇 가지 과정을 거친다. 주요한 과정을 요약한다.

(1) 2010년부터 근로, 사업, 재산소득이 음인 농가는 0으로 치환하였다. 참고로, 가계동향조사에서는 사업소득이 음이 없다.

(2) 농가경제조사로 분배지표를 만들 때 다음과 같이 한다.

(a) 근로소득=근로수입(계)-농외취업비용

(b) 사업소득=농업소득+겸업소득

(단, 더한 값이 음수이면 0으로 처리한다)

(3) 재산소득=자본수입(계)-자본수입관련비용=재산소득

(단, 뺀 값이 음수이면 0으로 처리한다)

(4) 사적이전소득=사적이전금

시장소득은 근로소득, 사업소득, 재산소득, 사적이전소득의 합이다.

각 항목별로 음수를 0으로 처리한 후 합친다는 것은 OECD의 지침이다.

참고문헌

- 홍민기(2015a), 「최상위 임금 비중의 장기 추세(1958~2013)」, 『산업노동 연구』 21(1), pp.191~220.
- _____ (2015b), 「최상위 소득 비중의 장기 추세(1958~2013)」, 『경제발전 연구』 21(4), pp.1~34.
- _____ (2016), 「최상위 소득 집단의 직업 구성과 직업별 소득 분배율」, 『사회경제평론』.

Abounoori and McCloughan(2003), “Simple way to calculate the Gini Coefficient for grouped as well as ungrouped data,” *Applied Economics Letters*, pp.505~509.

Alvaredo(2011), “A note on the relationship between top income shares and the Gini coefficient,” *Economic Letters* 110, pp.274~277.

Atkinson, A., T. Piketty, and E. Saez(2011), “Top Income in the Long Run of History,” *Journal of Economic Literature* 49(1), pp.3~71.

Boggio, L., V. Dall’Aglia, and M. Magnani(2009), “On the Labor Shares in Recent Decades: A Survey,” DISCE Working Paper.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore(2012), “Recent trends in top income shares in the United States: Reconciling estimates from March CPS and IRS tax return data,” *Review of Economics and Statistics* 94(2), pp.371~388.

Cowell, F. A. and M.-P. Victoria-Feser(2007), “Robust Lorenz curves: a semiparametric approach,” *Journal of Economic Inequality* 5, pp.21~35.

Cowell, F.A. and E. Flachaire(2007), “Income distribution and inequality measurement: The problem of extreme values,” *Journal of*

Econometrics 141(2), pp.1044~1072.

Hlasny, Vladimir and Paolo Verme(2015), *Top Incomes and the Measurement of Inequality: A Comparative Analysis of Correction Methods using Egyptian, EU and US Survey Data.*

Jenkins, S. P., R. V. Burkhauser, S. Feng, and J. Larrimore(2011), “Measuring inequality using censored data: a multiple-imputation approach to estimation and inference,” *Journal of the Royal Statistical Society* 174(1), pp.63~81.

Juster, F. T., and K. A. Kuester(1991), “Differences in the measurement of wealth, wealth inequality and wealth composition obtained from alternative US wealth surveys,” *Review of Income and Wealth* 37(1) pp.33~62.

Kitov, I. and O. Kitov(2013), “The Dynamics of Personal Income Distribution and Inequality in the United States,” MPRA working paper, Fifth Meeting of the Society for the Study of Economic Inequality(ECINEQ).

Krueger, A.(1999), “Measuring Labor’s Share,” *American Economic Review Papers and Proceedings* 89(2), pp.45~51.

Lakner, C. and B. Milanovic(2013), “Global income distribution from the fall of the Berlin Wall to the great recession,” World Bank Policy Research working paper series #6719.

McLachlan, G. and T. Krishnam(2008), *The EM Algorithm and Extensions*, 2nd Edition, John Wiley & Sons, Inc.

◆ 執筆者

- 홍민기(한국노동연구원 연구위원)

불평등 지표 개선연구

- 발행연월일 | 2016년 12월 26일 인쇄
2016년 12월 30일 발행
- 발 행 인 | 방 하 남
- 발 행 처 | **한국노동연구원**
☎ 대표 (044) 287-6080 Fax (044) 287-6089
300147 세종특별자치시 시청대로 370
세종국책연구단지 경제정책동
- 조판·인쇄 | 거목정보산업(주) (044) 863-6566
- 등록 일자 | 1988년 9월 13일
- 등록 번호 | 제13-155호

© 한국노동연구원 2016 정가 4,000원

ISBN 979-11-260-0101-9