불평등지수 활용 시 유의점 및 개선점

: 국민기초생활보장급여제도의 재분배 효과 추정 예시

김정환
한국조세재정연구원
부연구위원
(jhkim30@kipf.re.kr)

Ⅰ. 서론: 불평등지수의 필요성

일반적으로 '불평등'이라는 단어는 부정적인 가치를 내포하는 경우가 많다. 특히 사회학이나 철학적 연구 중에는 그 기저에 '차별'이 암시되는 불평등에 관한 내용을 쉽게 접할 수 있다. 하지만 본고에서 불평등이라는 단어는 중립적 의미로서 '다름'이라는 뜻을 지닌 용어로 활용됨을 본격적인 논의 이전에 밝힌다.[1] 모든 개개인은 각각 다른 특징을 가지고 있기 때문에 시간이 흐름에 따라 자연스럽게 불평등 상태에 다다르게 된다. 공정한 기회를 부여받고, 공정한 원칙에 의하여 자원을 재분배하는 과정을 거쳐도 부의 불평등과 소득불평등은 피할 수 없는 현상이다. 원하는 것을 가지지 못하면 아쉬워하거나, 낙담하거나, 또는 불행한 감정을 느낄 가능성이 높다.[2] 불평등은 인간의 부정적인 감정을 자연스럽게 야기하는 현상으로 보일 수 있으며 지양하고 싶은 현상이기도 하다.

위에서 설명한 개인적 다양성의 요인 때문이 아니라도 '차별'과 같은, 공정하지 않다고 생각되는 요인들로 야기되는 불평등도 존재한다. 문제 해결을 위한 가장 바람직한 방법은 인문·사회학적으로 심도 있는 논의를 통해 특정 종류의 불평등은 용인하며 다른 종류의 불평등은 지양할 것인지에 대한 사회적 합의에 다다르는 것이다. 하지만 이 과정은 매우 길어질 수도 있고 어쩌면 불가능한 것일지도 모른다. 시간의 흐름에 따라 기존의 많은 가치관의 옳고 그름이 바뀔 수도 있고, 새로운 가치관이 등장할 가능성이 농후하며, 극단적으로 항상 옳은 가치관은 존재하지 않을지도 모르기 때문이다.

1) 최준욱, 『세대간·세대내 불평등과 재분배』, 연구보고서 18-09, 한국조세재정연구원, 2018, p. 22. 기본 개념과 관련 기본 연구에서 가치중립적 불평등을 바라보는 관점에 대한 논의

2) 흔히 질투와 같은 감정이 여기에 포함되며 경제 모형에서 Envy, Jealousy를 효용함수에 포함하여 불평등을 설명하려는 시도들이 존재한다.

이와 같은 다양한 요인으로 검증되지 못한 용인과 지양 사이의 불평등이 미래에는 지양해야 할 불평등의 속성을 가진 것이라면, 불평등의 정도가 악화되는 것을 예방하기 위한 연구들이 필요하다. 불평등 정도가 특정 이상으로 심화된 다음에는 돌이키기 매우 힘들 것이기 때문이다. 위와 같은 배경에서 경제학자들을 중심으로 불평등의 정도를 특정 지수를 활용하여 분석하려는 움직임이 일어났다. 이렇게 개발된 지수를 불평등지수라고 부른다. 경제학자들은 다양한 종류의 불평등지수를 개발하여 부와 소득의 재분배 효과를 측정하는 도구로 활용하고 있다.

이와 같이 중요한 역할을 하는 불평등지수를 활용한 연구에서는 통계적으로 반드시 유의해야 할 점이 있다. 우선 불평등지수라는 통계량은 표본의 극단적인 분위(Extreme Quantile) 값에 민감하게 영향을 받는다는 점을 인지하여야 한다. 어떤 정책의 재분배 효과를 불평등지수로 분석하려고 할 때, 만약 표본의 극단적인 분위 값이 적절하지 않다고 판단된다면 재분배 효과 추정값의 통계적 유효성은 신뢰하기 힘들 수 있다. 본고에서는 이런 점을 보완한 불평등지수 통계량을 소개하며, 분석의 결과가 어떤 방향으로 달라질 수 있는지 관찰하고자 한다.

본고의 구성은 다음과 같다. 제Ⅱ장에서는 몇 가지 불평등지수와 그들의 통계적 성질을 소개하고, 제Ⅲ장에서는 국민기초생활보장급여제도 설명, 제Ⅳ장에서는 앞서 소개한 불평등지수를 국민기초생활보장급여제도에 적용한 예를 들어 불평등지수의 변화를 추정하며, 제Ⅴ장에서는 앞의 유의점을 개선할 방안에 대해 논의하는 것으로 마무리한다.

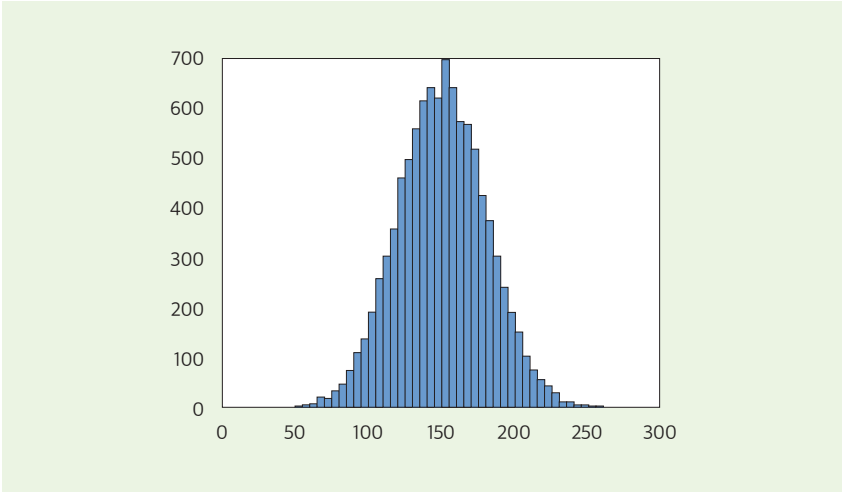
본고에서는 통계적 유의점을 보완한 불평등지수를 통해 분석의 결과가 어떤 방향으로 달라질 수 있는지 관찰하고자 한다.

Ⅱ. 불평등지수 소개 및 통계적 성질

한국 노동자의 평균 임금을 150만원, 표준편차가 30인 정규분포를 따른다고 가정해보자. [그림 1]은 저자가 임의로 가정한 소득분포를 히스토그램으로 나타낸 것이다. 히스토그램을 관찰하면 많은 노동자들이 각기 다른 임금을 받고 있으며 이를 통해 불평등이 존재하는 사회라는 것을 쉽게 확인할 수 있지만, 불평등의 정도가 얼마나 심한지는 쉽게 판단하기 힘들다. 명확하게 비교할 기준이 없으며, 어떤 기준으로 불평등을 정의해야 지양해야 할 개념으로 쉽게 받아들일 수

로렌츠에 따르면 소득에 따라 줄을 세운다고 가정했을 때, 특정 소득 이하의 노동자들이 보유한 소득과 사회에 존재하는 모든 소득의 비율로 불평등도를 추정할 수 있다.

[그림 1] 임의의 소득분포 시뮬레이션 $N(150, 30^2)$



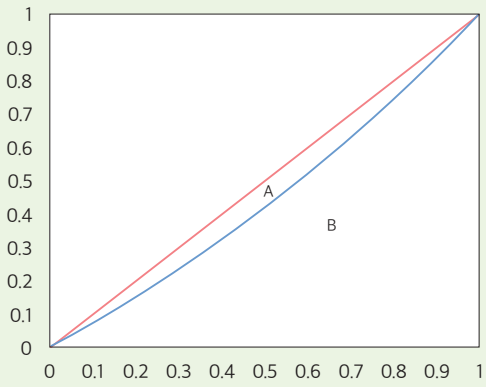
출처: 저자 작성

있을지 고민해야 하기 때문이다. 로렌츠(Lorenz, 1905)는 불평등을 가난한 사람부터 부유한 사람들까지 일렬로 줄을 세운다고 가정했을 때, 특정 소득 이하의 노동자들이 보유한 소득과 사회에 존재하는 모든 소득의 비율로 그 불평등도를 추정할 수 있다고 설명하였다. 위의 아이디어를 활용한 분석의 예를 들면, '상위 20%의 부유한 노동자들이 전체 사회소득의 80%를 가지고 있다' 혹은 '하위 10%의 가난한 노동자들은 전체 사회 소득의 1%도 가지지 못한다'와 같은 분석이 가능해진다는 것을 의미한다. 구체적으로 설명하면 [그림 1]의 내용을 참고하여 다음의 [그림 2]와 같은 그래프를 그릴 수 있다.

[그림 2]의 x축은 가계를 소득이 낮은 순서부터 높은 순서로 일렬로 나열해 둔 상태이다. y축은 각 가계가 가지고 있는 소득의 누적 합을 나타낸다. [그림 2]의 45도 직선[3]인 붉은색 선은 모든 가계가 동일한 소득을 가진 상태로 불평등이 없는 상태를 의미한다. [그림 2]의 45도 직선 아래의 푸른색 곡선이 [그림 1]의 데이터의 분포를 나타낸다. 분포를 어떻게 나타내는지를 살펴보면 다음과 같다. 푸른색 곡선의 x축의 값이 약 0.5인 부분의 y축 값은 0.4 정도이다. 이는 전체 노동자 중 절반 아래에 해당하는 노동자들이 가진 소득의 합은 전체 노동자들의 소득의 합의 약 40%라는 의미이다. 푸른색 곡선은 로렌츠 곡선이라 불리며, 정의를 통해 유추할 수 있듯이 곡선이 45도 붉은색 직선과 멀어질수록 불평등도가 높다는 해석이 가능

3) 완전균등선이라고도 한다.

[그림 2] [그림 1] 소득분포의 로렌츠 곡선



출처: 저자 작성

[그림 2]의 푸른색 곡선(로렌츠 곡선)이 45도 붉은색 직선과 멀어질수록 불평등도가 높다는 해석이 가능하며, 로렌츠 곡선을 활용하여 불평등지수 중 가장 유명한 지니계수를 정의할 수 있다.

하다. 로렌츠 곡선을 활용하여 불평등지수 중 가장 유명한 지니계수를 정의할 수 있다.

[그림 2]에서 두 가지 구분된 영역을 표시하였다. 45도 붉은색 직선과 로렌츠 곡선 사이 부분의 면적을 A라고 정의하자. 그리고 로렌츠 곡선과 x축 사이 부분의 면적을 B라고 정의하자. 경제학자 지니(Gini, 1912)는 불평등지수를 다음과 같이 정의하여 지니계수라 명명하였다. 지니계수는 45도 붉은색 직선과 로렌츠 곡선 사이 부분의 면적인 A와, 45도 붉은색 직선과 x축 사이 면적인 $A+B$의 비율을 의미한다. 식으로는 $\frac{A}{A+B}$로 정의하고, 그 값이 커질수록 불평등도가 심하다는 것을 의미한다.

지니계수는 불평등도 연구의 새로운 지평을 열었다. 그 이후 통계학과 경제학이 발전함에 따라 지니계수 이외에도 다른 불평등지수들이 등장하게 되었다. 이와 같은 불평등지수들은 모두 데이터에서 얻어지는 통계량으로서, 많은 연구자들이 불평등지수들을 통계 이론을 활용하여 분석하는 것에 많은 관심을 가지게 되었다. 후의 논의를 조금 더 수월하게 진행하기 위하여 소득을 확률변수 X라고 정의하고 그 누적확률밀도함수를 $F(\cdot)$라고 가정하자. 로렌츠 곡선을 이와 같은 기호를 활용하여 정의하면 다음과 같다.

식 (1)의 로렌츠 곡선(L)의 수학적 정의를 활용하여 지니계수를 정의하면 식 (2)와 같다. 지니계수 외에 다른 불평등지수로는 Mean Log Deviation, Theil 지수, Coefficient of Variation 등이 있다.

$$L(F(x)) = \frac{1}{\mu}\int_0^x y\,dF(y) \qquad \text{식 (1)}$$

여기서 μ는 소득의 평균을 의미한다. 이와 같이 나타나는 로렌츠 곡선(L)의 수학적 정의를 활용하여 지니계수를 정의하면 다음과 같이 나타낼 수 있다.[4)]

$$G = 1 - 2\int_0^1 L(y)\,dy \qquad \text{식 (2)}$$

지니계수 외에도 다른 불평등지수로는 Mean Log Deviation(=MLD), Theil 지수, 그리고 Coefficient of Variation(=CV) 등이 있다. 위의 세 가지 불평등지수는 '일반화된 엔트로피(Generalized Entropy)'[5)]라는 일반화된 개념으로 쉽게 정의할 수 있다는 장점이 존재한다. 일반화된 엔트로피 지수의 정의는 다음과 같다. α가 0과 1이 아닐 때는

$$GE(\alpha) = \frac{1}{N\alpha(\alpha-1)}\sum_{i=1}^{N}\left[\left(\frac{X_i}{\overline{X}}\right)^{\alpha} - 1\right] \qquad \text{식 (3)}$$

α가 0 또는 1이라면,

$$GE(0) = -\frac{1}{N}\sum_{i=1}^{N}\ln\frac{X_i}{\overline{X}}$$
$$GE(1) = \frac{1}{N}\sum_{i=1}^{N}\frac{X_i}{\overline{X}}\ln\frac{X_i}{\overline{X}} \qquad \text{식 (4)}$$

식 (4)의 $GE(0)$는 MLD를, $GE(1)$은 Theil 지수를 나타내고, 식 (3)을 이용하여 $GE(2)$를 도출하여 이를 CV라고 정의한다. 일반화된 엔트로피를 위에서 소개한 확률변수(X) 및 누적확률함수(F) 기호를 활용하여 다시 표현하면 다음과 같다.

4) 식 (1)과 식 (2)는 Davidson(2009)의 식 (1), (2)와 동일하다.
5) AF Shorrocks(1980)

$$GE(\alpha) = \int \frac{1}{\alpha(\alpha-1)}\left[\left(\frac{y}{\mu}\right)^{\alpha} - 1\right] dF(y), \quad \alpha \neq 0,1$$

$$GE(0) = -\int \log\left(\frac{y}{\mu}\right) dF(y)$$

$$GE(1) = \int \frac{y}{\mu}\log\left(\frac{y}{\mu}\right) dF(y)$$

앞서 소개한 네 가지 불평등지수는 많은 연구에서 활용되고 있는 중요한 불평등지수들이며, 모두 값이 커질수록 불평등도가 심화됨을 의미한다. 계속해서 이들이 가지는 통계적인 성질에 대하여 간단히 소개하고, 이것들이 의미하는 것이 무엇인지 알아보겠다.

불평등지수의 통계적 성질을 잘 분석한 연구 중 하나로 Cowell and Victoria-Feser(1996)를 예로 들 수 있다. 이 연구에서 저자들은 위의 불평등지수들이 Decomposability(분해성)[6]라는 성질을 만족한다고 주장하였다. 뒤에서 밝히겠지만, 이 성질을 만족하는 불평등지수들에 한하여 개선된 불평등지수를 만들어낸 것이 이 연구의 큰 업적이라고 볼 수 있다. 이 성질의 수학적 정의는 다음과 같다.

어떤 불평등지수 $I(F)$가 Decomposability를 만족한다면, 다음과 같이 표현이 가능하다.

$$I(F) = \psi(\nu(F,\mu(F)),\mu(F)) \qquad \text{식 (5)}$$

여기서 $\nu(F,\mu) = \int \phi(x;\mu)dF(x)$ 그리고 ϕ,ψ는 각각 $R^2 \rightarrow R$ 함수이며 ψ는 첫 번째 인수(argument)에 대해서 단조 증가하는 함수이다. 이후의 또 다른 연구인 Cowell and Flachaire(2007)에서 위의 Decomposability 성질을 만족하는 불평등지수들의 극단적인 분위의 데이터 값에 대해 얼마나 민감하게 반응하는지를 이론적으로 보여주고, 시뮬레이션을 통해 검증하였다. 이론적인 논의는 Hampel(1974)의 연구에서 시작된 영향 함수(Influence function)를 활용하였다. 영향 함수를 직관적으로 설명하면, 불평등지수를 주어진 소득분포 데이

앞서 소개한 네 가지 불평등지수는 많은 연구에서 활용되고 있는 중요한 불평등지수들이며, 모두 값이 커질수록 불평등도가 심화됨을 의미한다.

6) 일반화된 엔트로피 외에도 이 성질을 만족하는 지수들에는 Gibrat(1931)의 로그 소득의 분산, Hirschman(1945)의 Hirschman's index, Atkinson(1970)의 Atkinson's index, Kolm(1976)의 Kolm's index 등이 있다. 앞의 지니계수는 이 성질을 만족하지는 않는다.

기존 데이터 부분과 오염된 데이터 부분, 혹은 데이터 값이 변하는 부분의 혼합 분포를 정의하면 식 (6)과 같다.

터에서 계산한다고 가정할 때, 데이터 일부분의 값이 변한다면, 불평등지수의 값이 얼마나 민감하게 반응하는지를 계산하여 보여줄 수 있는 이론적 분석방법을 의미한다. 다른 표현으로는 통계량에 데이터 오염(Data contamination)이 일어났을 경우의 강건성을 분석한다고도 한다. 기존 데이터 부분과 오염된 데이터 부분, 혹은 데이터 값이 변하는 부분의 혼합 분포(Mixture distribution)를 식 (6)과 같이 정의한다.

$$G_\epsilon = (1-\epsilon)F + \epsilon H \qquad \text{식 (6)}$$

여기서 $H(y) = 1(y \geq z)$으로 정의하고, z를 데이터 오염이 일어난 값이라고 해석한다. 이때 불평등지수 $I(F)$의 영향 함수를 다음과 같이 정의한다.

$$IF(x) = \lim_{\epsilon \to 0} \frac{I((1-\epsilon)F + \epsilon H) - I(F)}{\epsilon} \qquad \text{식 (7)}$$

위의 식 (7)은 Gateaux 미분(differentiability)이라고 불리며, 연쇄 법칙(Chain rule)을 만족한다. 자세히 살펴보면, 분자는 데이터의 분포 중 ϵ만큼의 비율이 기존 소득분포인 F에서 오지 않고 H라는 새로운 분포에서 온다고 가정할 때의 변화량을 의미하며, 분모를 함께 고려하여 불평등지수의 순간변화율을 계산하고자 하는 아이디어에서 비롯된 결과이다. 식 (5)에서 정의된 ψ를 두 함수 ν, μ에 대하여 Gateaux 미분(differentiability) 연쇄 법칙을 활용하여 영향 함수를 도출하면 다음과 같다.

$$IF(z;I,F) = \frac{\partial \psi}{\partial \nu_{G_\epsilon}} \frac{\partial \nu_{G_\epsilon}}{\partial \epsilon}\bigg|_{\epsilon=0} + \frac{\partial \psi}{\partial \mu_{G_\epsilon}} \frac{\partial \mu_{G_\epsilon}}{\partial \epsilon}\bigg|_{\epsilon=0} \qquad \text{식 (8)}$$

식 (5)의 μ, ν의 정의를 활용하여 계산하면 다음과 같은 식이 도출된다.

$$IF(z;I,F) = \frac{\partial \psi}{\partial \nu_F}(\phi(z) - \nu_F) + \frac{\partial \psi}{\partial \mu_F}(z - \mu_F) \qquad \text{식 (9)}$$

이해를 돕기 위하여 여기까지의 내용을 간략히 정리하자면, 위에서 네 가지의 불평등지수를 소개하였다. 그리고 일반적인 엔트로피에 해당하는 불평등지수들의 공통적인 성질로 Decomposability라는 것을 소개하였으며, 그 성질을 만족하는 불평등지수들은 위의 식 (9)를 만족한다고 볼 수 있다. 식 (9)는 불평등지수의 영향 함수를 나타낸 것으로 z라는 특정 데이터 오염 값, 혹은 특정 데이터 값 변화에 대한 민감도를 나타낸 것으로 이해 가능하다.

좋은 통계량이 가지는 성질 중에 일치성(Consistency)이라는 성질이 있다. 이 성질을 만족하려면 수학적으로 Glivinko-Cantelli 법칙(Theorem)을 만족하여야 하는데, 그 가정 중에 중요한 부분이 영향 함수가 유계(Bounded) 함수이어야 한다는 것이다. 그러나 식 (9)에서 관측할 수 있듯이 영향 함수가 데이터 오염 값인 z에 직접적으로 의존하기 때문에 만약 z가 충분히 크거나, 작은 값을 가진다면, 불평등지수가 일치성을 가진다고 단정하기 힘들다는 해석이 가능해진다.[7] 즉 데이터의 극단 분포 부분에서의 변화에 더 민감하다고 볼 수 있다. 지니계수는 Decomposability 성질을 만족하지 않지만 영향 함수를 분석하면 비슷한 결론을 이끌어낼 수 있기 때문에 위와 유사한 문제를 공유한다고 이해할 수 있다.

영향 함수가 데이터 오염 값에 직접적으로 의존하기 때문에 데이터의 극단 분포 부분에서의 변화에 더 민감하다고 볼 수 있다.

Ⅲ. 불평등지수 개선 대안

앞서 불평등지수가 왜 데이터 변화에 민감할 수 있으며, 특히 극단 값에 더 크게 반응할 수 있다는 점을 소개하였다. 위의 사실과 더불어 흥미로운 것은 경제학 연구에서 소득과 재산의 분포를 적절한 가정들을 활용하면 파레토 분포(Pareto distribution)를 가진다는 이론이 다수 존재한다는 것이다.[8] 소득이라는 확률 변수 X가 파레토 분포를 가진다는 수학적인 정의는 다음과 같은 우측 꼬리(Right tail) 분포를 가진다는 의미이다.

$$P(X > x) \sim x^{-\theta} \qquad \text{식 (10)}$$

이 경우 확률누적밀도함수는 다음과 같다고 가정할 수 있다.

7) 자세한 내용은 Cowell and Flachaire(2007)의 부록 부분을 참고

8) Wold and Whittle(1957), Gabaix(2009), Benhabib and Bisin(2018) 등을 참고

Cowell and Flachaire(2007)는 기존의 불평등지수의 아이디어를 유지하되 경제학 이론에서 도출된 파레토 분포의 성질 또한 반영할 수 있는 준모수적 불평등지수 추정량 활용 방안을 제시하였다.

$$F(x) = 1-(x/x_0)^{-\theta} \text{ 이때 } x \geq x_0$$

부가적으로 설명하면 파레토 분포는 정규분포보다 극단 값이 자주 나타날 수 있는 분포 형태를 가진다. 따라서 경제학 모형에서 도출한 결론인, 소득이나 자산과 같은 확률 변수는 극단 값 부분의 출몰 빈도가 상대적으로 높다는 점과 통계량으로서 불평등지수의 수학적 성질인 극단 값의 변화에 민감하다는 두 가지 사실이 모두 작용하여 기존의 불평등지수가 잠재적으로 가질 수 있는 문제점을 확인하였다. Cowell and Flachaire(2007)의 연구에서는 위와 같은 문제점을 개선할 새로운 방법을 제시하였다. 기존 불평등지수의 아이디어를 유지하되 경제학 이론에서 도출된 파레토 분포의 성질 또한 반영할 수 있는 준모수적(Semiparametric) 불평등지수 추정량을 활용하는 방안이다. 이러한 방법을 활용한 모의실험의 결과 어느 정도 개선되는 모습을 보여주었다. 그러나 위의 방법을 활용할 때는 파레토 분포가 잘 정의될 수 있는 조건을 만족해야 하는 제약이 존재하기 때문에 활용 시 유의하여야 한다.

기존의 방법론은 불평등지수를 비모수적으로(Nonparametric) 추정하는 방식으로 극단적인 값에 민감하다. 새로 제시된 방법은 분포의 우측 꼬리 부분을 파레토 분포를 활용하여 모수화한 다음 준모수적 추정법을 활용하는 방식으로 추정하는 것이다. 우선 주어진 소득 데이터가 파레토 분포를 따를 것이라는 경제학 이론 가정하에 파레토 분포의 모수를 추정하기 위해 표본을 크기 순서로 나열한 다음, 전체 소득자 N명 중에 적절한 고소득자인 $k \leq N$를 정한다. k번째 고소득자보다 높은 소득을 수령하는 소득분포의 우측 꼬리 부분의 표본만을 활용하여 최대우도추정법을 통해 파레토 분포의 모수를 추정한다. 실제 수학 공식을 통해 추정치를 도출할 때는 다음과 같은 식을 활용한다.

$$H_{k,N} = \frac{1}{k}\sum_{i=0}^{k-1}\log X_{(N-i)} - \log X_{(N-k+1)}, \text{ 추정치는 } \hat{\theta} = \frac{1}{H_{k,N}}$$

다음 순서는 파레토 분포 부분의 모수를 추정하여 비모수적 추정을 활용한 기존의 불평등지수와 함께 활용하기 위해 가중 평균을 활용한다. 분포 우측 꼬리를 제외한 부분은 기존의 비모수적 방법론을 활용하여 불평등지수를 구한 다음

$(1-p_{tail})$의 가중치를 곱하고, 위의 모수적 방법론을 활용한 불평등지수는 파레토 분포를 활용하여 불평등지수를 구하는 공식에 대입한 다음 p_{tail}의 가중치를 곱하여 최종적으로 불평등지수를 보정한다. 보정을 거치면 각 분포의 주요 성질을 나타내는 Moment 값들이 뒤에 따로 더해진 파레토 분포에서 비롯된 p_{tail} 부분과 함께 결합되어 값들이 보정된 것을 확인할 수 있다.

$$\mu^* = \frac{1}{N}\sum_{i=1}^{N^*} X_{(i)} + p_{tail}\frac{\hat{\theta}x_0}{\hat{\theta}-1}, \quad \nu_\alpha^* = \frac{1}{N}\sum_{i=1}^{N^*} X_{(i)}^\alpha + p_{tail}\frac{\hat{\theta}x_0^\alpha}{\hat{\theta}-\alpha}$$

여기서 $N^* = (1-p_{tail})N$. 이를 활용하여 일반적인 엔트로피, $GE(\alpha)$를 도출하면 다음과 같다.

$$GE^*(\alpha) = [\nu_\alpha^*/(\mu^*)^\alpha - 1]/(\alpha^2-\alpha)$$

단 $\hat{\theta} \geq \alpha$일 때 파레토 분포가 잘 정의되기 때문에 이 조건을 만족하여야 한다. 위의 p_{tail}값, k값을 적절히 조절하여 이 조건을 만족하게 만드는 것이 중요하다. 위의 일반적인 엔트로피 지수와 유사하게 Theil 지수와 MLD를 정의하면 다음과 같이 구할 수 있다. 이 경우 $\hat{\theta} \geq 1$를 만족하여야 한다.

$$GE^*(1) = \nu_1^*/\mu^* - \log\mu^*$$

$$GE^*(0) = \log\mu^* - \nu_0^*$$

$$\nu_1^* = \frac{1}{N}\sum_{i=1}^{N^*} X_{(i)}\log X_{(i)} + p_{tail}\frac{\hat{\theta}x_0}{\hat{\theta}-1}\left[\log x_0 + \frac{1}{\hat{\theta}-1}\right]$$

$$\nu_0^* = \frac{1}{N}\sum_{i=1}^{N^*}\log X_{(i)} + p_{tail}\left[\log x_0 + \frac{1}{\hat{\theta}}\right]$$

앞에서 소개한 보정된 불평등지수를 활용하기 위해서는 k와 p_{tail}을 결정해야 하는데 실증연구자들을 위해 Davidson and Flachaire(2007)은 시뮬레이션 결과를 활용하여 $k = N/10$, x_0는 $N(1-p_{tail})$번째 표본을 선택,

파레토 분포 부분의 모수를 추정하여 비모수적 추정을 활용한 기존의 불평등지수와 가중 평균을 함께 활용하여 불평등지수를 보정하였다.

본 장에서는 기초생활보장제도를 자세히 알아보고, 간단한 경제모형을 통해 불평등지수가 어떻게 변하는지 살펴본다.

$p_{tail} = 0.04N^{-1/2}$ 가이드라인으로 권장한다.

앞서 불평등지수가 어떤 잠재적 문제점을 가지고 있으며 어떻게 보정될 수 있는지 살펴보았다. 다음 장에서는 기초생활보장제도를 소개하고 보정된 불평등지수를 어떻게 활용 가능한지 아주 단순한 경제 모형을 활용한 예시를 통해 기초생활보장제도의 재분배 효과 추정에 대하여 간단하게 서술해 보고자 한다.

Ⅳ. 기초생활보장제도 설명

기초생활보장제도의 목적은 생활이 어려운 사람에게 필요한 급여를 실시하여 이들의 최저생활을 보장하고 자활을 돕는 것이다. 저소득 계층을 정책의 목표로 설정하여 소득재분배 효과를 불평등지수를 통해 관찰할 수 있다. 사실 불평등지수의 극적인 변화, 즉 재분배 효과를 관찰하기에는 저소득층보다는 고소득층을 정책 목표로 하는 상황이 효과가 쉽게 나타난다고 볼 수 있다. 그렇다면 저소득층을 기준으로 분석할 경우에는 기존 불평등지수를 활용해도 보정된 불평등지수와 별 차이가 없이 강건하게 활용할 수 있는 것인지, 아니면 또 다른 유

<표 1> 2020년도 급여종류별 수급자 선정기준

(단위: 원)

구분	1인 가구	2인 가구	3인 가구	4인 가구	5인 가구	6인 가구	7인 가구
생계급여 (중위소득 30%)	527,158	897,594	1,161,173	1,424,752	1,688,331	1,951,910	2,216,915
의료급여 (중위소득 40%)	702,878	1,196,792	1,548,231	1,899,670	2,251,108	2,602,547	2,955,886
주거급여 (중위소득 45%)[1]	790,737	1,346,391	1,741,760	2,137,128	2,532,497	2,927,866	3,325,372
교육급여 (중위소득 50%)	878,597	1,495,990	1,935,289	2,374,587	2,813,886	3,253,184	3,694,858

주: 1) 2022년도는 46%로 변동

출처: 보건복지부, 「2020년 국민기초생활보장사업안내」, 2020.

<표 2> 2020년도 기준 중위소득

(단위: 원)

구분	1인 가구	2인 가구	3인 가구	4인 가구	5인 가구	6인 가구	7인 가구
기준 중위소득	1,757,194	2,991,980	3,870,577	4,749,174	5,627,771	6,506,368	7,389,715

출처: 보건복지부, 「2020년 국민기초생활보장사업안내」, 2020.

의할 점들이 있는 것인지 궁금증을 가질 수 있다. 지금부터는 기초생활보장제도에 대해 자세히 알아보고, 간단한 경제모형을 통해 불평등지수가 어떻게 변하는지 살펴보자.

우선 기초생활보장제도 수급자 선정기준에 대해 알아보자. <표 1>은 2020년도 급여종류별 수급자 선정기준이 되는 소득인정액을 나타낸 것이다. 예를 들어, 1인 가구 기준으로 만약 가구원의 한 달 소득인정액이 50만원이라면 이는 생계급여 기준인 52만 7,158원 이하이므로 생계급여 수급 대상으로 선정된다. 소득인정액의 산출법에 대해서는 아래에서 더 자세히 알아보도록 한다. 수급 기준은 중위소득 개념을 기준으로 선정되며, 중위소득은 국민 전체를 소득 기준으로 정렬했을 때 절반에 해당하는 소득 수준을 의미한다. 매년 물가상승률 등 경제 전반적인 변수와 지표를 분석하여 다음 연도에 중위소득을 미리 예상하여 발표하며, 각 급여의 수급 기준은 중위소득에 연동되어 결정되고 있다. <표 2>는 2020년 기준 중위소득을 의미한다. 가구원의 수에 따라 중위소득이 크게 영향을 받기 때문에 중위소득은 가구의 크기에 따라 따로 관리한다.

8인 이상 가구는 가구 숫자가 적기 때문에 별도의 분석보다는 7인 가구 기준 중위소득에서 6인 가구 기준 중위소득의 차액을 7인 가구 기준 중위소득에 더하여 산정하는 방식으로 구한다. 생계급여를 제외하면 위의 수급자 선정기준은 최소로 만족해야 할 선정기준으로 이해해야 한다. 의료, 주거, 교육 급여는 다양한 각도에서 불평등을 해소하기 위해 만들어진 제도로 이해하여야 하며 각 분야에서 더 필요로 하는 요건을 만족하는 대상에게 우선적으로 제공된다. 조금 더 상세한 제도의 이해를 위해 주요 용어 및 산식들을 알아보자.

우선 소득인정액이라는 개념을 알아보자. 소득인정액은 가구가 현재 가진 소득과 잠재적으로 활용할 수 있는 재원에 대한 조정치를 의미하는 것으로 소득평가액과 재산소득환산액이라는 두 가지 개념의 합으로 나타난다.

소득인정액 계산식

- 소득인정액 = 소득평가액 + 재산소득환산액

소득인정액을 이루고 있는 소득평가액과 재산소득환산액 중 소득평가액은 실제 가구소득에서 가구의 특성에 따른 지출비용 부분과 근로소득공제 부분을 제

소득인정액은 가구가 현재 가진 소득과 잠재적으로 활용할 수 있는 재원에 대한 조정치를 의미하는 것으로, 소득평가액과 재산소득환산액의 합으로 나타난다.

한 부분을 의미한다. 여기서 근로소득공제 부분은 평소에 세금납부 시 적용되는 근로소득공제와는 다른 것으로 기초생활보장제도 수급 기준 소득평가액을 계산할 때만 적용되는 개념으로서 근로소득이나 사업소득의 30%를 공제해주는 것을 의미한다. 이는 소득평가액을 산출함에 있어서 저소득층의 노동의욕 감소를 막기 위한 것으로 보인다. 소득평가액을 구성하는 각 항목에 대해 자세히 살펴보자. 실제소득은 근로소득, 사업소득, 재산소득, 이전소득의 합으로 구성되어 있으며 이는 아주 다양한 종류의 소득을 모두 포함하는 개념으로 보인다. 특히 이전소득에 부양비 등의 항목 등이 포함되는 것으로 보아 기초생활보장제도의 목적은 가구원들이 스스로 부양할 능력이 없음과 동시에 정부를 제외하고는 도움받을 곳이 없는 가구를 정책의 목표로 하고 있음을 보여준다. 소득평가액에서 제하는 두 가지 중 하나인 가구 특성에 따른 지출비용에 해당하는 부분을 알아보면 장애, 질병, 양육, 국가유공 등의 특수한 경우에 발생하는 지출비용은 소득으로 산정되지 않게 되어 있다. 다음의 계산식은 한눈에 알아보기 쉽게 정리한 것이다.

소득평가액 계산식

- 소득평가액 = 실제소득 - 가구 특성에 따른 지출비용 - 근로소득공제[9)]
- 실제소득 = 근로소득[10)] + 사업소득[11)] + 재산소득[12)] + 이전소득[13)]
- 가구 특성에 따른 지출비용 = 장애요인으로 인한 금품[14)] + 질병요인으로 인한 금품[15)] + 양육요인으로 인한 금품[16)] + 국가유공요인으로 인한 금품[17)]

앞서 소개한 소득평가액의 명확한 이해를 돕기 위해 간단한 예시를 하나 소개한다.

예시 1) 소득평가액 계산

- 어떤 가구의 근로소득이 2천만원, 사업소득이 500만원, 재산소득이 50만원, 이전소득이 100만원이라고 하자. 만약 이 가구가 가구 특성에 따른 지출이 없다고 가정한다면, 이때 이 가구의 소득평가액은 실제소득 2,650만원(2,000+500+50+100)에서 근로소득공제((2,000+500)*(0.3))를 제외한 1,900만원이다.

앞에서 소득인정액의 첫 번째 부분인 소득평가액의 개념과 산출방법을 알아

9) 근로소득과 사업소득의 30% 공제
10) 상시근로자 소득 + 일용근로자 소득 + 자활근로소득 + 공공일자리 소득
11) 농업소득 + 임업소득 + 어업소득 + 기타사업소득
12) 임대소득 + 이자소득 + 연금소득
13) 사적이전소득 + 부양비 + 공적이전소득
14) 장애수당 및 장애아동 수당 및 보호수당 + 장애인연금법 기초급여액 및 부가 급여액 + 고엽제후유의증 등 환자지원 + 자동차손해배상 보장법 재활보조금 + 대한장애인체육회에 등록된 사람으로서 국제경기대회에서 입상한 사람이 국민체육공단으로부터 받은 경기력향상연구연금 월정금
15) 만성질환 등의 치료·요양·재활로 인하여 3개월 이상 지속적으로 지출하는 의료비 + 희귀난치성질환자 의료비 지원사업에 따른 의료비 중 호흡보조기 대여료, 기침유발기 대여료 및 간병비, 특수식이구입비 + 한센인 피해자 위로지원금
16) 한부모가족지원 아동양육비 및 추가아동양육비 + 한부모가족지원사업 안내에 따른 청소년 한부모 자립지원촉진수당 + 소년소녀가정 부가급여 + 자동차손해배상 보장법 피부양보조금 + 양육보조금 + 농어민 가구의 보육시설 이용 자부담
17) 국가유공자, 독립유공자, 보훈대상자, 체육유공자에 대한 생활조정수당 + 참전명예수당 중 1인 가구 기준 중위소득의 20% 이하에 해당하는 금액 + 독립유공자 (손)자녀 생활지원금

보았다. 이제는 재산소득환산액의 개념과 산출방법에 대해 알아보자. 재산소득환산액은 크게 세 부분으로 나누어 생각해 볼 수 있다. 첫 번째로 주거재산, 일반재산, 금융재산 각각의 일정 비율을 소득으로 환산하여 더한 부분, 두 번째로 가계부채 총차감액 부분과 마지막으로 자동차 재산가액 부분으로 나누어진다. 재산소득환산 부분에 부채를 차감하고 자동차 재산가액은 그 자체를 소득으로 환산하여 더하게 된다. 계산식은 다음과 같다.

재산소득환산액 계산식

- 재산소득환산액 = 주거재산*주거재산소득환산율 + 일반재산*일반재산소득환산율 + 금융재산*금융재산소득환산율 - 부채 총차감액 + 자동차 재산가액

재산소득환산액의 첫 부분에 해당하는 주거재산, 일반재산, 금융재산 각각의 환산액을 어떻게 도출하는지 자세히 알아보자. 재산 산출에 있어 본 제도에서 가장 중요한 것은 특정 재산에 대한 공제가 어떤 형태로 이루어져 있는지를 파악하는 것이다. 주거재산을 산정함에 있어 공제가 이루어질 때 우선적으로 고려되는 점은 주택을 소유하고 있는지의 유무이다. 만약 자기 소유의 부동산이라면 전체를 고려하고, 임차보증금의 경우에는 95%의 금액만 고려된다. 주거재산은 그 성격이 기본적인 생활에 반드시 필요한 것으로서, 주거재산 소득환산율은 타 소득환산율에 비해 낮게 산정된다. 또한 주거지역별로 주거재산에 포함될 수 있는 금액의 상한선을 정하여 고가의 사치스러운 주택에 대하여 높은 재산 값을 산출하게 하였다. 만약 고가의 주택에 거주한다면 특정 상한선까지의 금액은 주거재산으로 산정되지만, 그 외의 금액은 일반재산으로 산정된다. <표 3>은 대도시, 중소도시, 농어촌의 주거재산 한도액을 보여준다.

재산 중 일부는 기본적인 생활유지에 필요하다고 인정하여 보건복지부에서

재산소득환산액에서 가장 중요한 것은 특정 재산에 대한 공제가 어떤 형태로 이루어져 있는지를 파악하는 것이다.

<표 3> 수급자의 주거용 재산 한도액

(단위: 만원)

구분	대도시	중소도시	농어촌
생계·주거·교육급여	12,000	9,000	5,200
의료급여	10,000	6,800	3,800

출처: 보건복지부, 「2020년 국민기초생활보장사업안내」, 2020.

재산 중 일부는 기본적인 생활유지에 필요하다고 인정하여 보건복지부에서 거주지별로 정한 공제금액을 기본재산 공제제도라고 한다.

<표 4> 거주지별 기본재산(공제)액

(단위: 만원)

구분	대도시	중소도시	농어촌
생계·주거·교육급여	6,900	4,200	3,500
의료급여	5,400	3,400	2,900

출처: 보건복지부, 「2020년 국민기초생활보장사업안내」, 2020.

거주지별로 정한 공제금액을 기본재산 공제제도라고 한다. <표 4>는 대도시, 중소도시, 농어촌별로 재산에서 공제되는 액수를 의미한다.

또한 금융재산 중 일부는 생활준비금 명목과 장기금융저축 명목으로 공제가 적용된다. 생활준비금이란 의료비, 관혼상제비 등을 의미하며 가구당 500만원을 공제하며, 장기금융저축은 최대 연간 500만원, 총한도는 1,500만원까지 공제가 가능하다.

위의 재산 공제제도는 순차적으로 주거용 재산, 일반재산, 금융재산에 적용된다. 이는 부채가 존재할 경우 부채 차감 순서로도 적용된다. 그리고 자동차 재산 가액에서는 부채를 차감하지 않는다고 명시되어 있다.

주거, 일반, 금융, 자동차 재산을 조사하여 공제제도와 부채 차감을 적용한 후, 최종적으로 소득환산율을 각각 곱하고 더하여 최종 재산소득환산액을 산출한다. 소득환산율은 2020년 기준 주거용 재산은 1.04%, 일반재산은 4.17%, 금융재산은 6.26%, 자동차 재산은 100% 적용한다. 복잡한 산출 과정의 이해를 돕기 위해 아래의 예시를 소개한다.

예시 2) 재산소득환산액

- 어떤 가구가 대도시에 거주하며 전세보증금이 1억 3천만원, 일반재산이 2천만원, 금융재산이 5천만원, 장기금융저축은 없으며, 부채가 1천만원, 자동차 차량가액이 500만원이라고 가정하자. 이 가구의 재산소득환산액을 산출하면 다음과 같다.
- 임대보증금 1억 3천만원 중 95%인 12,350만원이 주거용 재산한도액 1억 2천만원을 초과하였기 때문에 초과한 350만원은 일반재산에 더해진다. 따라서 기존의 일반재산 2천만원에 주거용 재산한도액 초과분 350만원을 더한 2,350만원으로 산출된다. 기본재산 공제제도를 활용하여 공제 우선순위에 따라 주거재산 한도액인 1억 2천만원에서 기본재산(공제)액 6,900만원을 제외한 5,100만원이 주거용 재산으로 재산출된다. 금융재산의 경우 5천만원에서 생활준비금 명목 500만원을 공제한 4,500만원으로 계산된다. 부채가 1천만원이기 때문에 차감 1순위인 주거용 재산에서 1천만원이 차감되어 주거용 재산이 4,100만원으로 계산된다. 자동차 차량가액 500만원은 그대로 더해진다.
  그러므로 최종적인 재산소득환산액은 {(4,100만원*0.0104) + (2,350만원*0.0417) + (4,500만원*0.0626) + 500만원 = 9,223,350원}이다.

<표 5> 급여의 종류 및 내용

급여종류	내용	비고
생계급여	기준 중위소득 30%에 해당하는 금액과 가구의 소득인정액과의 차액을 지급	
의료급여	근로능력 유무에 따라 1종, 2종으로 구분하여 지급	근로능력: 2020년 기준 18세 이상(2002년생)부터 64세(1956년생)까지 1종[1]
주거급여	국토교통부 장관이 정하는 기준에 따라 지급	
교육급여	교육부 장관이 정하는 기준에 따라 입학금, 수업료, 교과서대, 부교재비, 학용품비 지급	
해산급여	수급자가 출산 시 1인당 70만원 지급	단, 교육급여만 받는 수급자는 제외
장제급여	수급자 사망 시 80만원을 장제를 실제 행하는 자에게 지급	단, 교육급여만 받는 수급자는 제외

주: 1) 근로능력이 없다고 판단되는 경우는 장애의 정도가 심한 장애인, 질병·부상 후유증으로 근로능력을 상실한 사람, 20세 미만 중·고교 재학생, 국가유공자 3급 이상 등에 해당된다.
출처: 보건복지부, 「2020년 국민기초생활보장사업안내」, 2020.

생계급여는 빈곤층의 노동 의욕 저하를 불러일으킬 수 있다는 잠재적 문제점이 존재하여 사회의 사중손실을 증가시킬 가능성이 있다.

기초생활보장제도의 수급 기준과 소득인정액을 어떻게 산출하는지에 대하여 알아보았다. 산출된 소득인정액을 기준으로 기초생활보장제도에서 의료급여 다음으로 많은 재정지출 비율을 차지하는 생계급여가 지급된다. 생계급여는 앞서 살펴본 것처럼 기준 중위소득 30%에 해당하는 금액과 가구의 소득인정액과의 차이를 지급한다. 이와 같은 제도는 빈곤층의 노동의욕 저하를 불러일으킬 수 있다는 잠재적 문제점이 존재하여 사회의 사중손실을 증가시킬 가능성이 있다. 따라서 추후 제도 개편에서 중대한 사안이 될 것으로 보인다. 생계급여를 제외한 타 급여는 소득이 아닌 다각도의 빈곤 형태의 문제점을 가진 가구를 돕기 위해 활용된다. 해당 급여에는 의료급여, 주거급여, 교육급여, 해산급여, 장제급여 등이 있으며 특히 의료, 주거, 교육 급여는 앞에서 소개한 중위소득에 연관된 수급 조건이 별도로 존재한다. <표 5>에 급여 종류와 내용을 정리하였다.

Ⅴ. 기초생활보장제도와 불평등지수 적용 사례

1. 데이터, 가정 및 모형

앞서 서술한 기초생활보장제도 외에도 수많은 수급 특례가 존재하지만 현재

본고의 분석 내용은 불평등지수의 민감도와 보정에 대한 실증적 응용에 목적이 있기에 기초생활보장제도의 재분배 효과에 대한 추정에 다소 한계가 많을 가능성이 있음을 밝힌다.

활용하는 데이터로 관찰이 불가능하기에 본 연구에서는 데이터 셋에서 관찰 가능한 부분들만 최대한 반영하여 분석에 활용하였다. 본 분석에 앞서 다음의 분석 내용은 불평등지수의 민감도와 보정에 대한 실증적 응용에 목적이 있기에 기초생활보장제도의 재분배 효과에 대한 추정에 다소 한계가 많을 가능성이 있음을 밝힌다. 분석에 활용한 데이터는 통계청 가계금융복지조사패널의 2020년도 데이터로 관찰하기 힘든 부분에 대하여 다음과 같은 가정을 추가하여 분석하였다.

모든 가구는 대도시에 거주한다고 가정하였다. 또한 차량의 경우 일반형으로 소유하여 생계형 소유는 없다고 가정하였는데 이는 생계형으로 소유할 경우에 자동차 재산의 공제가 다르게 적용되기 때문이다. 또한 주거급여를 산정할 때, 임대의 경우에는 1급지(서울)로 적용하였으며, 자가의 경우에는 3년에 1회 지급하는 경보수 금액인 457만원의 3분의 1을 적용하였다.

위의 가정들은 최소한의 미시적인 데이터 정보가 주어진 생계급여와 주거급여에 대한 가정이다. 의료급여의 경우에는 어떤 가구가 수급하였는지에 대한 수급 자료가 부족하여 「2020년도 생계급여 및 의료급여경상보조 사업 예산안 현황」(국회예산정책처, 2019) 자료를 참고하여 생계급여(4조 3천억원)와 의료급여(7조원)의 비율을 표본의 생계급여비의 총합에 곱한 다음, 표본의 의료급여 대상의 수로 나눈 금액을 급여 대상에게 지급한다고 가정하였다. 부가적으로 설명하면 데이터의 생계급여를 받는 대상과 전체 인구에서 생계급여를 받는 대상의 비율이 데이터에서 의료급여 수급 소득조건을 만족하는 대상과 전체 인구에서 의료급여를 받는 대상의 비율이 유사할 것이라는 가정을 추가했다고 볼 수 있다. 또한 각 의료급여 수급자는 비슷한 수준의 급여를 받는다는 추가적인 가정을 적용하였다. 교육급여의 경우에도 미시적인 데이터로 수급 대상을 확인할 수 없었기 때문에 의료급여와 유사한 가정을 적용하였으며, 그 결과 「2021년도 예산안 위원회별 분석」(국회예산정책처, 2020) 자료를 참고하여 생계급여(4조 3천억원)와 교육급여(1,016억원)의 비율을 표본의 생계급여비의 총합에 곱한 다음, 표본 중 교육지출이 있는 동시에 교육급여 대상을 만족하는 표본의 수로 나눈 금액을 급여 대상에게 지급한다고 가정하였다. 해산급여와 장제급여는 본고에서는 우선 배제하였으며, 금리$(=r)$는 2020년 기준금리 0.5%를 가정하였다. 분석에 활용할 소득세율을 추정하기 위하여 식 (11)의 Benabou(2002), Heathcote et al.(2017) 모형을 가정한다. 모수(λ, τ)는 실제소득에 누진세율을 가정하고 최소자승법으

로 추정하도록 한다.

$$\log D(y_i) = \log\lambda + (1-\tau)\log y_i \qquad \text{식 (11)}$$

식 (11)에서 y_i는 소득, $D(y_i)$는 가처분소득, λ는 평균적인 세금의 크기, τ는 누진세를 반영한다고 가정한다.

여기서 y_i는 소득, $D(y_i)$는 가처분소득, λ는 평균적인 세금의 크기, τ는 누진세를 반영한다고 가정한다. 소득세율을 구한 후에는 간단한 2기간 모형을 상정하여 각 가계 i는 다음과 같은 최적화문제를 푼다고 가정한다.

$$\max_{C_{i,1}, C_{i,2}} \log C_{i,1} + \beta \log C_{i,2}$$
$$\text{subject to } C_{i,2} = (1+r)(y_i - t(\phi_i) - C_{i,1} + \phi_i) + (y_i - t(\phi_i))$$

여기서 ϕ_i는 기초수급지원액, y_i는 경상소득, $t(\phi_i) = \pi\lambda y_i^{(1-\tau)}$, π는 총 기초수급지원금액과 총 세금추정액의 비율로서 총 기초수급지원액과 총 세금추정액을 같게 만들어준다. 시간에 대한 할인율은 $\beta = 0.95$를 적용하였다. 만약 기초수급정책이 없다고 가정하면 각 가계 i는 다음과 같은 최적화문제를 풀게 된다.

$$\max_{C_{i,1}, C_{i,2}} \log C_{i,1} + \beta \log C_{i,2}$$
$$\text{subject to } C_{i,2} = (1+r)(y_i - C_{i,1}) + y_i$$

기초수급정책 시행하의 1기 소비와 기초수급정책을 시행하지 않을 시 1기 소비의 불평등지수 변화를 분석 시도하였다. 또한 불평등지수가 표본에 민감하게 반응하는 정도를 보기 위해 데이터 오염 값을 추가하여 불평등지수의 변화를 측정하였다. 아래에서 분석 결과에 대해 표를 활용하여 소개하겠다.

2. 분석 결과

가. 데이터 기초통계량 vs 2020년 실제 현황

<표 6>은 2020년 가계금융복지패널데이터의 기초통계량과 위의 가정을 활용하여 표본 중 잠재적으로 급여를 수급했을 가구의 통계량을 나타낸 것이다. 대략 전체 표본 중 약 3.6%가 생계급여를 수급했음을 알 수 있다. 표본 중 잠재

불평등지수는 재분배 효과를 매우 단순한 방법으로 수치화한 지수로, 정책을 시행할 때나 정책의 효과를 정확하게 할 때는 그 한계가 명확함을 인지하여야 한다.

<표 6> 데이터 기초통계량

표본수(N)	평균 가구실제소득	평균 가구금융재산	생계급여 수급가구	의료급여 수급가구
18,064가구	5,494만원	7,284만원	646가구(3.58%)	1,691가구(9.36%)
주거급여 수급가구	교육급여 수급가구	생계급여 총금액	평균 가처분소득	추정세수
2,333가구(12.92%)	2,863가구(15.85%)	13억 4,710만원	4,482만원	8,091억원

주: 생계급여 총금액과 추정세수의 2020년 가구수(2,148만가구)비율로 보정하면 생계급여 총 금액보정액은 약 1조 6천억원, 추정세수 보정액은 약 965조원임

출처: 저자 작성

<표 7> 2020년 실제 현황

총가구수	생계급여 수급가구	의료급여 수급가구
약 2,148만가구	약 100만가구(4.1%)	약 109만가구
주거급여 수급가구	교육급여 수급가구	2020년 세수
약 137만가구	약 21만가구	약 285조원

출처: 보건복지부, 『국민기초생활보장 수급자 현황』, 2020.

적으로 지급된 생계급여 총금액은 약 13억 5천만원으로 도출되었다.

<표 7>은 전국의 가구를 대상으로 데이터를 정리한 것이다. 생계급여 기준에 대한 자료 이외에는 추정된 부분이 많기 때문에 실제 데이터와 통계량의 차이가 많이 나는 것을 알 수 있다. 추후 소비, 세금의 실제 데이터를 활용해 통계량의 차이를 감소시킬 수 있는지 분석이 필요할 것으로 보인다. 본고에서 기초생활보장제도에 대한 분석은 이와 같은 한계점을 가지기 때문에 큰 의미를 부여하기는 힘들다고 보여진다. 하지만 불평등지수에 대한 논의와 실제 제도에 적용하기 위하여 위의 가정들을 추가하였으며 아래에서 불평등지수의 변화에 대한 논의를 진행하고자 한다.

불평등지수를 활용한 분석에 앞서 해석의 유의점에 대해 논의한다. 우선 불평등지수는 재분배 효과를 매우 단순한 방법으로 수치화한 지수로 정책을 시행할 때나, 정책의 효과를 정확하게 측정하고자 할 때는 그 한계가 명확함을 인지하여야 한다. 예를 들어, 두 가지 정책 A와 B가 존재한다고 가정하고, 두 정책의 시행으로 지니계수가 각각 0.02씩 변했다고 가정하자. A를 시행하였을 경우 지니계수는 0.3에서 0.28로 변하였고, B를 시행하였을 경우 지니계수는 0.28에서

0.26으로 변했다고 가정하자. 연구자에 따라서는 B정책의 효과가 더 큰 것 같다고 판단하는 경우도 충분히 있을 수 있을 것이다. 그러나 지니계수는 각 경제주체에게 무슨 일이 발생하였는지는 관심이 없다. 다만 부와 재산, 소득 등이 어떻게 분포되어 있는지만 나타낼 뿐이다. 정책 A와 B의 지니계수의 차이가 0.02로 같으므로 두 정책의 재분배 효과가 동일하다고 생각하기도 어렵다. 또한 지니계수의 정의가 수학적으로 복잡하기에 단순히 지니계수의 비율이 얼마나 변했는지를 재분배 효과로 보는 것도 명확하지 않다. 만약 완벽하게 동일한 두 세계에 다른 정책을 시행하여 불평등지수의 차이를 본다면, 재분배 효과의 크기 비교 정도는 가능할 것으로 보이지만, 실제로 그 정도를 분석, 비교하기가 매우 어렵다는 한계점이 있다. 다음 분석을 해석할 때에는 위의 유의점을 염두에 두되 통계량이 어떤 변화를 가진다는 것을 살펴보는 것을 주요 분석목적으로 하고자 한다.

기존의 불평등지수의 정의를 활용하여 제도의 재분배 효과를 확인한 결과, 제도를 시행하였을 경우의 불평등지수가 제도가 없을 때보다는 낮아진 것을 확인할 수 있다.

나. 불평등지수별 변화: 제도를 시행할 경우와 하지 않을 경우

<표 8>은 기존의 불평등지수의 정의를 활용하여 제도의 재분배 효과를 확인한 내용이다. 제도를 시행하였을 경우의 불평등지수가 제도가 없을 때보다는 낮아진 것을 확인할 수 있다. 참고로 OECD에 보고된 한국의 시장소득 지니계수는 2011년 0.418에서 2020년 0.405로 낮아진 것으로 나타나서 상대적으로 유사한 지니계수를 가진다고 볼 수 있다. MLD, Theil 지표, CV의 경우에는 위에서 소개한 통계적 성질인 Decomposability를 만족하는 불평등지수로서 보정된 불평등지수를 따로 확인할 수 있으므로, 뒤에서 데이터 오염에 따른 차이에 얼마나 민감하게 반응하는지 알아볼 수 있다. 지니계수는 본 연구자가 아는 한 보정된 통계량이 개발되지 않았다. 이는 앞으로 경제학자들이 풀어야 할 흥미로운 연구과제로 사료된다.

<표 8> 1기의 추정 소비 값의 불평등지수

불평등지수	지니계수	MLD(=GE(0))	Theil(=GE(1))	CV(=GE(2))
제도 시행 X	0.4445	0.3807	0.3450	0.5126
제도 시행 O	0.4394	0.3562	0.3366	0.5054

출처: 저자 작성

<표 9>에서 가장 낮은 소비를 한 표본을 제거하였을 경우 지니계수가 아주 미세하게 변하고, 가장 많은 소비를 한 표본을 제거하였을 때는 상대적으로 지니계수가 많이 바뀌는 것을 확인할 수 있다.

다. 불평등지수별 변화: 제도 시행 유무와 극단 표본 1개를 제거할 때

다음으로는 데이터 오염에 대한 민감도를 살펴보자. <표 9>의 표본 중에 가장 낮은 소비를 한 표본을 제거하였을 경우 지니계수가 아주 미세하게 변함을 알 수 있다. 그와 비교하여 소비가 가장 많은 표본을 제거하였을 때는 상대적으로 지니계수가 많이 바뀌는 것을 확인할 수 있다. 대체로 일반적인 엔트로피에 해당하는 불평등지수인 MLD, Theil, CV 모두 지니계수와 유사한 변화를 보였다. 조금 다른 각도에서 데이터 오염에 대한 실험을 설계하여 데이터 오염이 하나의 표본에서만 일어나되 그 정도가 심하게 나타나 오염된 데이터의 값이 크게 상승하는 경우에는 어떻게 될 것인지 실증데이터를 통해 다음과 같이 실험해 보았다.

<표 9> 표본의 최솟값과 최댓값을 제거할 때 1기의 추정 소비 값의 불평등지수

불평등지수		지니계수	MLD(=GE(0))	Theil(=GE(1))	CV(=GE(2))
최솟값 제거	제도 X	0.4445	0.3790	0.3450	0.5125
	제도 O	0.4394	0.3558	0.3366	0.5054
최댓값 제거	제도 X	0.4433	0.3787	0.3396	0.4705
	제도 O	0.4382	0.3542	0.3312	0.4636

출처: 저자 작성

라. 불평등지수별 변화: 제도를 시행할 경우 데이터 오염 값이 분포의 우측 꼬리에 1개 존재하는 경우

<표 10>은 데이터의 최댓값을 표본에 추가한 경우와 최댓값에 표준편차만큼 더 큰 숫자를 표본에 추가한 경우 등을 실험한 결과이다. 오염된 데이터가 현재

<표 10> 표본의 최댓값과 표본 표준편차를 더한 데이터 오염 값이 존재할 때 불평등지수

불평등지수	지니계수	MLD(=GE(0))	Theil(=GE(1))	CV(=GE(2))
max	0.4406	0.3581	0.3420	0.5469
max+σ	0.4406	0.3582	0.3422	0.5491
max+2σ	0.4407	0.3582	0.3424	0.5513
max+3σ	0.4407	0.3583	0.3426	0.5536

출처: 저자 작성

주어진 표본의 최댓값과 큰 차이가 나지 않으며 오염된 데이터의 개수가 적다면 불평등지수는 대체로 강건함을 보여주었다. 다음은 보정된 불평등지수에 대한 실험 결과를 관찰해보자.

<표 11>은 지니계수를 제외한 다른 불평등지수들을 보정된 불평등지수 값으로 대체한 결과로, <표 9>와 비교하면 최댓값을 제거했을 때 상대적으로 불평등지수의 값이 유사하게 추정되었음을 알 수 있다.

마. 준모수 방법론을 활용한 불평등지수

1) 불평등지수별 변화: 제도 시행 유무와 극단 표본 1개를 제거할 때

<표 11>은 지니계수를 제외한 다른 불평등지수들을 보정된 불평등지수 값으로 대체한 결과이다. <표 9>와 비교하면 최솟값을 제거했을 때 상대적으로 불평등지수의 값들이 더 많이 상승한 사실과, 최댓값을 제거했을 때 상대적으로 불평등지수의 값이 유사하게 추정되었음을 알 수 있다. 최솟값을 제거할 경우에 기존 불평등지수와 보정된 불평등지수를 산출하는 과정에 어떤 영향을 주는지를 살펴보자. 기존 불평등지수에서는 오른쪽 꼬리 부분의 파레토 분포 부분을 고려하지 않는다. 오른쪽 꼬리 부분을 포함한 모든 표본이 비슷한 영향력을 행사하여 불평등지수를 구성하게 된다. 그러나 보정된 불평등지수의 경우에는 우측 꼬리 표본 부분은 따로 분리되어 파레토 분포의 모수를 추정하는 데 활용되고 나머지 남은 표본이 보정된 불평등지수에 부분적인 영향을 주기 때문에 기존 불평등지수에 미치는 영향력보다 더 큰 영향을 주며, 우측 꼬리 표본 부분의 파레토 분포 모수 추정 영향으로 더욱 불평등도가 큰 상황으로 인식하게 된다.

<표 11> 표본의 최솟값과 최댓값을 제거할 때 1기의 추정 소비 값의 준모수적 불평등지수

불평등지수		MLD(=GE(0))	Theil(=GE(1))	CV(=GE(2))
최솟값 제거	제도 X	0.3775	0.3407	0.4858
	제도 O	0.3542	0.3323	0.4788
최댓값 제거	제도 X	0.3785	0.3394	0.4784
	제도 O	0.3539	0.3310	0.4715

출처: 저자 작성

바. 불평등지수별 변화: 제도를 시행할 경우 데이터 오염 값이 분포의 우측 꼬리에 1개 존재하는 경우

<표 10>의 내용과 유사하게 <표 12>는 데이터 오염 값이 한 개 존재할 때, 그

불평등지수 활용에 있어 연구자가 특히 주의할 사항은 불평등지수가 통계량이라는 것을 인지하고 있어야 한다는 것이다.

<표 12> 표본의 최댓값과 표본 표준편차를 더한 데이터 오염 값이 존재할 때 준모수적 불평등지수

불평등지수	MLD(=GE(0))	Theil(=GE(1))	CV(=GE(2))
max	0.3561	0.3363	0.5091
max+σ	0.3561	0.3363	0.5091
max+2σ	0.3561	0.3363	0.5091
max+3σ	0.3561	0.3363	0.5091

출처: 저자 작성

크기에 어떤 영향을 받는지 정리하였다. 보정된 불평등지수를 활용할 때, 불평등도가 더 크게 추정되었음을 알 수 있다. <표 12>의 내용을 해석하면 고소득층의 데이터 오염에 대하여 보정된 불평등지수는 상대적으로 덜 민감하게 반응하여 강건성을 가지는 것으로 보인다. 그 이유는 파레토 분포의 극단 값에 의해 고소득층이 불평등도에 영향을 주는 것을 경제학 이론을 반영한 통계적 논리를 활용하여 선제적으로 반영한 결과로 볼 수 있다. 참고로 다음은 Cowell and Flachaire(2007)에서 발췌한 몇 가지 이론적 결과이며 실증적으로 보정된 불평등지수를 활용할 경우 참고하면 좋을 것으로 사료된다.

- $GE(\alpha)$ 만약 $\alpha > 1$이라면 분포의 우측 꼬리 데이터 오염 값에 민감
- $GE(\alpha)$ 만약 $\alpha < 0$이라면 분포의 좌측 꼬리 데이터 오염 값에 민감
- 지니계수는 GE지수들보다 분포 우측 꼬리 데이터 오염 값에 덜 민감
- $GE(\alpha)$의 α가 작을수록 데이터 오염 값에 덜 민감

VI. 결론: 유의점 및 권장사항

본고에서는 재분배 효과를 추정할 때 중요한 불평등지수의 잠재적인 한계점 및 개선점과 그 응용에 대하여 알아보았다. 연구자가 특히 주의할 사항은 불평등지수가 통계량이라는 것을 인지하고 있어야 한다는 것이다. 신뢰구간 등에 대한 논의 없이 통계량만 제시하는 경우 독자들에게 편향성 있는 정보가 전달될

가능성이 있으므로, 이 경우에는 더욱 편향이 최소화된 양질의 불평등지수를 제시할 수 있도록 노력해야 한다. 본고에서는 불평등지수가 극단적인 분위 값이나 데이터 오염 값에 민감하게 반응할 수 있음을 인지하여 분석하는 경우와 보정된 불평등지수를 활용하도록 노력할 필요가 있다는 점을 강조한다. 불평등지수를 활용할 때의 권장사항으로는 첫째, 가장 좋은 방법은 전수 데이터를 활용하는 것이며, 이 경우 극단 분위 값의 민감성을 신경쓸 필요가 없다는 점이다. 우측 꼬리 극단에서 발생하는 표본이 경제학 이론을 바탕으로 발생한 파레토 분포라 할지라도 모든 표본을 관찰할 수 있다면 연구자는 전체 분포를 알기 때문에 이러한 한계점에서 자유로울 수 있다. 둘째, 표본을 활용한 연구 시에는 가능하면 극단 분위의 표본이 충분한 데이터를 활용할 수 있으면 더 나을 것으로 판단된다. 기존 불평등지수뿐만 아니라 보정된 불평등지수를 활용할 때에도 만약 극단 분위의 표본이 충분하다면 데이터 오염에 대하여 강건한 결과를 보여줄 수 있기 때문이다. 끝으로, 앞의 경우들을 현실적으로 달성하기 힘들다면 위에서 소개된 준모수적 방법론을 활용하여 데이터 오염 값에 덜 민감한 보정된 불평등지수를 활용하는 것도 좋은 방안이라고 생각된다. KIPF

불평등지수를 활용할 때 가장 좋은 방법은 전수 데이터를 활용하는 것이다. 모든 표본을 관찰할 수 있다면 연구자는 전체 분포를 알기 때문에 이러한 한계점에서 자유로울 수 있다.

<참고문헌>

국회예산정책처, 『2020년도 예산안 위원회별 분석(보건복지위원회·여성가족위원회 소관)』, 2019.

_____, 『2021년도 예산안 위원회별 분석(교육위원회 소관)』, 2020.

최준욱, 『세대간·세대내 불평등과 재분배』, 연구보고서 18-09, 한국조세재정연구원, 2018.

보건복지부, 「국민기초생활보장사업안내」, 2020.

_____, 「국민기초생활보장 수급자 현황」, 2020.

통계청, 「가계금융복지조사」, 2000.

Atkinson, Anthony B., "On the measurement of inequality," *Journal of economic theory*, 2(3), 1970, pp. 244~263.

Benabou, Roland, "Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency?," *Econometrica*, 70(2), 2002, pp. 481~517.

Benhabib, Jess and Alberto Bisin, "Skewed wealth distributions: Theory and empirics," *Journal of Economic Literature*, 56(4), 2018, pp. 1261~1291.

Cowell, Frank A. and Emmanuel Flachaire, "Income distribution and inequality measurement: The problem of extreme values," *Journal of Econometrics*, 141(2) 2007, pp. 1044~1072.

Cowell, Frank A. and Maria-Pia Victoria-Feser, "Robustness properties of inequality measures," *Econometrica: Journal of the Econometric Society*, 64(1), 1996, pp. 77~101.

Davidson, R. and Flachaire, E., "Asymptotic and bootstrap inference for inequality and poverty measures," *Journal of Econometrics*, 141(1), 2007, pp. 141~166.

Davidson, Russell, "Reliable inference for the Gini index," *Journal of econometrics*, 150(1), 2009, pp. 30~40.

Dupor, Bill and Wen-Fang Liu, "Jealousy and equilibrium overconsumption," *American economic review*, 93(1), 2003, pp. 423~428.

Gabaix, Xavier, "Power laws in economics and finance," *The Annual Review of Economics*, 1(1), 2009, pp. 255~294.

Gibrat, R., *Les Inégalités Economiques*, Paris: Librairie du Recueil Sirey, 1931.

Gini, Corrado, *Variabilità e mutabilità*, Reprinted in Memorie di metodologica statistica(Eds. Pizetti E. et al.), 1912.

Hampel, Frank R., "The influence curve and its role in robust estimation,"

Journal of the american statistical association, 69(346), 1974, pp. 383~393.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante, "Optimal tax progressivity: An analytical framework," *The Quarterly Journal of Economics*, 132(4), 2017, pp. 1693~1754.

Hirschman, Albert O., *National power and the structure of foreign trade*, Univ of California Press, 1980.

Kolm, Serge-Christophe, "Unequal inequalities. I," *Journal of economic Theory*, 12(3), 1976, pp. 416~442.

_____, Serge-Christophe, "Unequal inequalities. II," *Journal of economic theory*, 13(1), 1976, pp. 82~111.

Lorenz, Max O., "Methods of measuring the concentration of wealth," *Publications of the American statistical association*, 9(70), 1905, pp. 209~219.

Shorrocks, Anthony F., "The class of additively decomposable inequality measures," *Econometrica: Journal of the Econometric Society*, 1980, pp. 613~625.

Wold, Herman OA, and Peter Whittle, "A model explaining the Pareto distribution of wealth," *Econometrica: Journal of the Econometric Society*, 1957, pp. 591~595.