불평등지수(Inequality Index)와 다극화지수(Polarization Index)

김 현 호*

한 사회의 단일변수에 대한 일반적인 불평등을 측정하는 불평등 지수와 그 사회의 상대적 갈등을 측정하는 다극화 지수와는 본질적으로 서로 다르다. 예를 들면 다극화 지수들은 주어진 변수의 분포 변동에 관점을 맞추고 있지만, 대부분의 불평등지수는 분포의 전체 평균과 산포도의 차이만을 강조할 뿐 모집단이 몇 개의 집단으로 묶어지는 군집화 현상이나 분포의 변동현상 등을 설명할 수 없는 한계를 가진다. 또한 불평등 지수는 고소득층에서 저소득층으로 소득이 이전될 경우 소득의 불평등이 개선된다는 피구-달톤의 공리를 따르고 있지만, 다극화 지수는 주어진 그룹간의 동질성(identification)과 이질성(alienation) 등의 갈등의 조장이 극화를 심화시킨다는 공리를 따르고 있다는 점에서 근본적인 차이를 나타낸다. 하지만 기존의 연구들을 살펴보면 주어진 분포를 측정함에 있어서 두개의 상이한 지수를 구분 없이 사용하고 있으며, 특히 다극화 측정 시 산업별(수출과 비수출산업, IT와 비IT산업 등), 기업별(내수기업과 수출기업, 제조업과 서비스업 등), 지역별, 소득별 등 선험적, 자의적으로 구별하여 기초데이터에 의존한 주관적인 다극화의 현상을 서술하고 있는 것이 대부분이다.

따라서 본 연구에서는 주요 불평등 지수와 다극화 지수를 제시하여 주어진 단일변수에 대한 기존의 서술적인 차원의 한계에서 벗어나 보다 분석적이고 객관적인 연구방법을 제시하고자 한다. 주요 내용으로는 먼저 소득의 불평등 지수를 나타내는 지니계수를 제시하고, 다극화 지수인 Wolfson 지수, ER 지수, 그리고 Extended ER지수를 차례대로 제시하고자 한다.

1. 지니계수

지금까지 주어진 단일변수(variable)[1]의 불평등(inequality)에 대한 구체적이고 객관적이며 유일한(unique) 지수(index)를 제시하기 위하여 여러가지 방법이 개발 · 제시되었다. 하지만 하나의 지

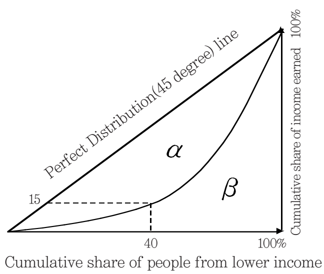
* 기술경제연구센터 부연구위원(e-mail: hyunkim@stepi.re.kr)

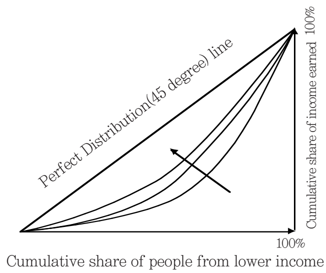
수에 불평등의 측정에 모두 필요한 속성을 포함하는 완벽한 지수를 개발한다는 것은 아마 불가능한 일일 것이다. 따라서 본 절에서는 피구-달톤의 공리(Pigou-Dalton axiom)를 만족시키는[2] 지니계수를 소개하고 그 지수의 본질과 한계점을 살펴본다. 지니계수는 한 사회의 소득의 불평등에 대하여 가장 많이 사용되는 지수로써 로렌스 곡선을 활용하여 계산한다. 먼저 로렌스 곡선을 정의하면 다음과 같다.

로렌스 곡선(Lorenz Curve): 한 사회의 구성원을 소득이 가장 낮은 사람으로 부터 높아지는 순서대로 배열한다고 할 때, 하위 몇%에 속하는 사람들이 차지하는 전체소득 중의 비율을 나타내는 점들을 모아놓은 곡선

이는 〈그림 1〉에 제시된 바와 같이 횡축에는 원점에 소득이 가장 낮은 사람부터 높은 소득 순으로 인구를 배열하여 전체 인구의 누적%를 나타내고, 종축에는 횡축에 주어진 누적인구의 %가 전체 소득에서 차지하는 비율인 누적소득의 %를 나타낸다. 그리고 그 점들을 이어서 생긴 곡선이 바로 로렌스 곡선이다. 예를 들어 〈그림 1〉 횡축의 40은 모집단의 하위소득자 40%를 가리키며, 종축의 15%는 전체 소득에서 하위 40% 소득자들이 차지하는 소득비중이 15%라는 사실을 알려주고 있다. 만약 구해진 로렌스 곡선이 그림의 대각선(45˚)과 일치한다면, 이는 모든 사람이 동일한 소득을 가지며 완전한 평등분배를 나타낸다. 반면에 전체 모집단에서 한명을 제외한 모든 사람이 소득이 0이고 한명이 모든 소득을 소유하고 있다면, 로렌스 곡선은 횡축을 계속 따라가다가 전체 인구 비율이 100%에 이르면 수직선으로 바뀌는 직각 삼각형 모양을 가지게 되는데 이 상황이 가장 불평등한 분배상태를 의미한다. 이와 같이 로렌스 곡선은 모두가 똑같은 소득을 나누어 갖는 완전 균등한 분배가 가장 평등한 분배라는 전제를 내포하고 있다.

〈그림 1〉 로렌스곡선과 지니계수





로렌스 곡선 작성 시 횡축의 원점에서 1까지 가장 적은 소득부터 순차적으로 합하기 때문에, Median을 기준으로 로렌스 곡선의 원점에 가까울수록 소득이 적은 집단이, 1에 가까울수록 고소득 집단이 위치해 있다. 따라서 로렌스곡선을 통한 소득의 불평등도가 완화되다는 것은 적은 소득 집단

의 로렌스 곡선상의 기울기와 고소득 집단의 곡선상의 기울기의 차가 적으면 적을수록 불평등도가 낮아진다. 결국 무수히 많은 소득 집단의 한계소득 누적분을 합하여 Smooth하게 작성된 로렌스 곡선은, 소득 집단 간의 기울기의 차이, 즉 소득의 차이가 적으면 적을수록 낮은 불평등도를 가져오는 것이다. 따라서, 〈그림 1〉의 오른쪽 그림에서 보듯이 로렌스 곡선이 45°선에 가깝게 그려질수록 소득의 불평등이 적고, 아래로 횡축에 가깝게 처진 곡선일수록 불평등이 심하다는 것을 알 수 있다. 이러한 로렌스곡선이 가지고 있는 단점은, 두 집단의 로렌스 곡선을 서로 비교 하는 경우, 서로에 대한 서수적인(ordinal) 평가는 내릴 수 있어도 한 집단이 다른 집단보다 어느 정도 평등하다고 말할 수는 없다는 점이다. 또한 두 로렌스 곡선이 서로 교차한다면 어떠한 객관적인 판단을 내릴 수 없다는 한계점을 가지고 있다. 따라서 이러한 로렌스 곡선의 단점을 보완하고 곡선을 이용하여 불평등 정도를 구체적이고 객관적인 수치로 나타낼 수도록 고안된 것이 바로 지니계수(Gini Index)이다.

지니계수를 구하는 방법은, 아래 식과 같이 먼저 모집단을 평균소득으로 정규화(normalized)하고 전체 인구(n)를 둘씩 조합($_nP_2$)을 통해 구해진 소득 격차의 합을 조합 수의 두배로 나누어 구한다.

$$G_1 = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| \qquad (1)$$

여기에서 y_i 는 i 번째 사람의 소득을 의미한다. 이를 그림으로 설명하면 보다 쉽게 나타낼 수 있다. 〈그림 1〉의 왼편 그림과 같이 주어진 변수의 로렌스 곡선이 구하고, 로렌스곡선과 대각선(45°)사이에 형성되는 초승달 모양의 면적(α)을 구하여 두배를 하면 바로 식 (1)의 지니계수[3)]를 얻을 수 있다. 이는 지니계수가 0에 가까울수록 더욱 평등한 분배를 1에 가까울수록 불평등한 분배를 나타냄을 알 수 있다. 좀 더 구체적으로 $2\times\alpha$의 면적을 수식으로 나타내면 다음과 같다.

$$G_2 = 2 \times [\frac{1}{2} - \frac{1}{2n} \sum_{i=0}^{n} (y_{i+1} + y_i)]. \qquad (2)$$

여기에서 $y_0 = 0$ 이며, G_1 과 G_2 를 비교하면, G_1 과 G_2 가 서로 같은 결과를 가져옴을 쉽게 증명할 수 있다.

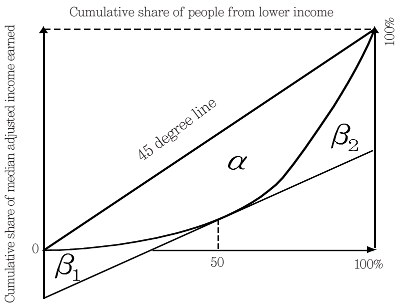
2. Wolfson 지수

Wolfson(1994)은 중산층 소멸(middle class collapse)과 양극화(bipolarization)를 동의어로 간주하고 양극화에 대한 서수적 척도를 제안하였다. 즉 중산층을 중위소득(median income)을 가진 인구계층으로 규정하고 중위소득으로부터 소득의 분산정도가 클수록, 즉 중위소득을 기준으로 중위소득과 여타 계층의 소득의 차이가 커질 수록 소득의 양극화가 커진다고 정의하였다. 따라서 본 지수는 어떠한 기준 공리에서 출발하기 보다는 소득불평등도를 측정하는 로렌스 곡선을 이용하여 중

산층 소멸을 통한 양극화를 유도한 것이다.

먼저 횡축으로는 가장 낮은 소득부터 고소득 순으로 누적 소득 인원의 비율을 정렬하고, 종축에는 소득을 중위수(median)/평균(mean)으로 곱한 소득의 누적비율을 횡축의 누적인원에 맞추어 배열하여 변형된 로렌스 곡선을 그린다. 그리고 횡축의 중위수(50%)에서 잇은 수직선과 로렌스 곡선이 맞는 곳에서 접선(중위 Tangent)을 긋는다. 이제 종축의 0, 1과 대각선인 45%선, 그리고 중위수의 접선으로 이루어진 사다리꼴의 면적을 구한다. 그리고 사다리꼴 면적을 대각선과 로렌스 곡선사이의 면적으로 차감하면 중산층 소멸에 대한 Wolfson 지수를 구할 수 있다.

〈그림 2〉 Wolfson 지수



〈그림 2〉를 통하여 구체적으로 설명하면, 중위수(50) 접선을 통한 사다리꼴의 면적, $\alpha+\beta_1+\beta_2$에서 α을 빼고 다시 중위수와 평균의 비율(median/mean)로 재조정을 하고 여기에 4를 곱하면[4] 0과 1 사이의 Wolfson 지수를 얻을 수 있다는 것이다. 따라서 Wolfson의 양극화 지수는 다음과 같이 나타낼 수 있다.

$$W = 4 \times (\alpha + \beta_1 + \beta_2 - \alpha)/(median/mean) \tag{3}$$

이러한 Wolfson 지수는 중위소득자의 기울기(중위 Tangent)로 중위소득자의 한계소득누적분을 나타내고 이 중위자의 한계누적분을 하위소득자와 상위계층자의 한계소득분과 비교하여 중산층과는 상이한 기울기를 가진 계층과의 거리로 중산층의 몰락을 측정한 것이다. 따라서 중위소득 범위의 인구비중을 기준으로 양극화 및 중산층 소멸을 분석하는 지수라고 할 수 있다.

3. ER 지수

본 절에서는 Esteban & Ray(1994)(이하, ER)가 제시한 동질성-이질성 접근법(identification-alienation approach)을 통한 다극화 지수를 소개하고자 한다. 모집단이 어떤 한 변수에 대하여 N

개의 소집단으로 구분된다고 하자. 이때 어느 한 집단에 속한 특정 개인은 그 집단 내의 구성원들과는 동질성(identification)을 가질 것이고, 타 집단 내의 구성원들에 대해서는 이질성(alienation) 을 느끼게 될 것이다. 따라서 주어진 소득분포(F)에 대하여, 소득수준이 x인 개인이 느끼는 동일집단 내의 동질성 정도는 $I(x, F)$ 로, 개인소득 수준이 y인 타 집단 소속의 개인들에게 느끼는 이질성의 정도는 $r(x, y)$ 로 나타내면, 그 개인이 느끼는 자신의 소득 수준에 따른 유효반감(effective antagonism)함수인 T(I(x, F), r(x, y))는 동질성과 이질성의 증가함수로 다음과 같이 표기할 수 있다.

$$T(I(x,F),r(x,y)),\quad st.\ \frac{\alpha T}{\alpha I}>0, \frac{\alpha T}{\alpha r}>0 \tag{4}$$

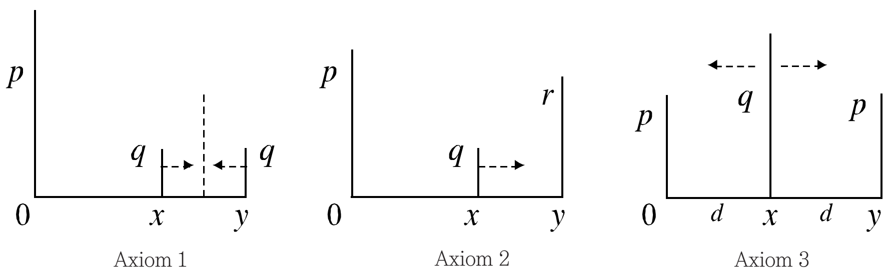
식 (4)의 주된 내용은 주어진 사회에 갈등의 본질은 첫째, 소속 집단 내부의 동질성이 강하면 강할수록, 둘째 집단 간의 이질성이 크면 클수록 사회내 갈등이 커짐을 나타내고 있다. 따라서 사회 전체의 유효반감은 개인의 동질성과 이질성으로 측정된 유효반감의 전체 합으로 나타낼 수 있다.

$$P(F)=\int\int T(I(x|F),r(x,y))dF(x)dF(y) \tag{5}$$

이제 주어진 문제는 함수 T, I, r 을 어떻게 정의하느냐이다.

이에 대하여 Esteban & Ray(1994)는 식(5)의 내용을 만족시키며 상이한 소득 분포에 대한 다극화 지수값을 서로 비교할 수 있는 지수를 제시하고자 다음과 같이 몇가지 공리(axiom)을 설정하였다(〈그림 3〉 참조).

〈그림 3〉 ER 지수 공리(Axiom)



Axiom1: $p, q \gg 0$, $p > q$, $0 < x < y$ 일때, p보다 작은 두개의 집단이 p와의 평균거리를 일정하게 유지하면서 $(x+y)/2$로 가까이 가면 갈수록 다극화(polarization)는 증가한다.

Axiom2: (p, q, r)>>0, p>r, x>|$y-x$|일때, 중간집단인 q가 p보다 작은 r집단에 가까이 가면 갈수록 다극화(polarization)는 증가한다.

Axiom3: (p, q)>>0, x=$y-x$≡d 일때, 중앙의 집단 q가 같은 거리만큼 떨어진 양극의 집단으로 동일한 크기로 이동하면 할수록 다극화(polarization)는 증가한다.

ER은 이러한 공리를 만족시키는 다극화 지수를 다음과 같이 소개하였다. 먼저 단일변수 y가 유한 폐구간 ⟨a, b⟩ 에서 밀도함수 f 로 표시된다고 하고, y 는 전체평균이 1이 되도록 정규화 한다. 이때 밀도함수 f 가 N 개의 소집단으로 외생적으로 구분되었다면, 단일변수 y 에 대한 구분되는 집단에 대하여 다음과 같이 형태로 정리하여 나타낼 수 있다.

$$\rho = (y_0, y_1, ...y_n; \pi_1, ...\pi_n; \mu_1, ...\mu_n), st.(y_0 = a, y_n = b)$$

여기에서 y는 각 소집단간의 구간경계(critical points)를 나타내는 값이며,

$$\pi = \int_{y_{i-1}}^{y_i} f(y)dy, \quad \mu = \frac{1}{\pi_i}\int_{y_{i-1}}^{y_i} yf(y)dy, \quad \text{for all} \quad i = 1, ...n$$

π는 각 소집단별 확률을, μ는 각 구간 내에서의 조건부 평균값을 의미한다. ER은 이렇게 외생적으로 주어진 소집단 수와 각 집단별 확률과 조건부 평균값을 이용하여 앞에서 제기한 다극화 공리에 부합하는 간단한 다극화 지수를 다음과 같이 제시하였다.

$$ER\ Index \equiv ER(\alpha, \rho) = \sum_i \sum_j \pi_i \pi_j \pi_i^{\alpha} |\mu_i - \mu_j| \qquad (6)$$

이와 같이 ER지수는 연속함수인 y 를 외생적으로 주어진 소집단의 수만큼 쪼개어서 각 집단별 불연속적인(discrete) 주요 값 – 평균(μ_i)과 확률(π_i)– 을 이용하여 다극화 지수를 측정한 것으로, 같은 집단간의 동질성과 타 집단과의 이질감의 증가함수로 측정하였다. 따라서 식 (6)의 ER지수는 동일집단의 동질성 함수(μ_j^{α})는 그 구간확률에 동질성의 강도를 나타내는 α[5])를 제곱하여 나타내고, 이질성 함수(|μ_i– μ_j|)는 각 집단을 대표하는 구간평균간의 절대적 차이로 나타낸다. 그리고 두 함수를 독립적인 실행으로 서로 곱한 후 모든 각 구간별 확률(π_i 와 π_j)을 곱하여 합한 가중평균 값을 다극화의 지수로 제시한 것이다.

4. Extended-ER 지수

ER 지수는 원래의 소득분포를 외생적으로 결정된 몇 개의 극점(N)으로 나누고 각 그룹간 소득을

선형화(μ_i)하여 상대 비교($|\mu_i - \mu_j|$)함으로써, 각 그룹간의 다극화 측정을 다소 과대포장하는 문제점을 발생한다. 이러한 문제점은 소그룹의 구간경계가 외생적으로 주어졌다는 것과 주어진 소그룹의 집단별 선형화된 대표값(평균 및 확률)을 사용하여 동질성과 이질성을 평가한다는 점이 근본적인 한계인 것이다. 따라서 Esteban, Gradin, Ray(1999)는 단순한 ER지수로 인한 다극화(polarization) 지수가 과장되는 문제점을 오차($\varepsilon(f, \rho)$)의 조정을 통하여 개선하고자 하였다.

$$P(f; \alpha, \beta) = ER(\alpha, \rho) - \beta\epsilon(f, \rho) \tag{7}$$

여기에서 β는 오차에 두는 가중치를 나타내는 값이다.

이러한 과장된 다극화 조정을 위한 오차($\varepsilon(f, \rho)$)측정에 대한 문제는 전체를 N 개의 소그룹으로 구분하는 문제와 서로 연관되어 결정된다. 결국 오차는 소그룹내의 두 소득간 차이를 평균한 값으로 정의하고 이 값을 최소화 하는 기준으로 전체를 소그룹하기 때문이다. 따라서 N의 소그룹이 주어졌다는 가정 하에 오차값을 수식으로 나타내면 다음과 같다.

$$\epsilon(f, \rho) = \sum_i \int_{y_{i-1}}^{y_i} \int_{y_{i-1}}^{y_i} |x - z| f(x) f(z) dx dz. \tag{8}$$

이 오차식은 다음과 같은 전개를 통하여 지니계수의 식으로 전환될 수 있다.

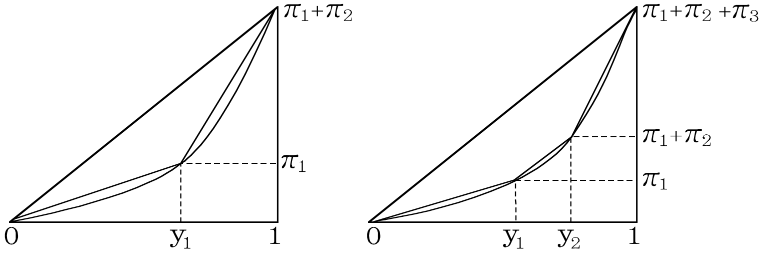
$$\begin{aligned} G(f) &= \frac{1}{2}\int_a^b \int_a^b |x - z| f(x) f(z) dx dz = \frac{1}{2}\sum_i \int_{y_{i-1}}^{y_i} \int_a^b |x - z| f(x) f(z) dx dz \\ &= \frac{1}{2}\sum_i \int_{y_{i-1}}^{y_i} \int_{y_{i-1}}^{y_i} |x - z| f(x) f(z) dx dz + \frac{1}{2}\sum_i \sum_j |\mu_i - \mu_j| \pi_i \pi_j \\ &= \epsilon(f, \rho) + G(\rho). \end{aligned}$$

여기에서 G(.)는 지니계수를 의미한다. 이렇게 구해진 오차값은 다음과 같다.

$$\epsilon(f, \rho) = G(f) - G(\rho). \tag{9}$$

결국, 이러한 오차식이 나타내는 것은 소득분포에 대한 로렌스 곡선과 각 그룹별 선형화된 로렌스 선과의 차이를 최소화 하는 기준으로 그룹화(grouping)하는 것이 최적임을 보이고 있다. 즉 〈그림 4〉의 왼편에서는 전체를 두개의 그룹으로 나누는 y_1의 위치는 곡선의 로렌스 곡선과 그룹내부의 선형 로렌스 곡선 사이의 면적을 최소화 하는 횡축의 위치를 정하여 그룹을 양분하며, 한편 오른편 그림은 전체 그룹을 세 개의 그룹으로 나눌 때 y_1과 y_2의 위치결정은 그룹간 선형 로렌스와 전체 로렌스 곡선 사이의 면적을 최소화 하는 y_1과 y_2의 위치결정을 통하여 내생적으로 결정된다는 것이다.

〈그림 4〉 Extended-ER지수



이러한 논리에 대하여 Esteban al, (1999)과, Duclos at(2004)는 근사오차를 최소화 하는 ρ^* (각 구간경계)의 조건이 다음을 만족시킴을 증명하였다.

$$y_i^* \int_{y_i^*}^{y_i^*+1} f(x)dx = \int_{y_i^*}^{y_i^*+1} xf(x)dx, \quad for \quad i=1,2,..,n-1$$
$$y_i^*(\pi_i+\pi_{i+1}) = \mu_i^* + \mu_{i+1}^*\pi_{i+1}, \quad for \quad i=1,2,..,n-1$$

이것은 극점들의 위치를 선정함에 있어서 인접한 두 구간을 분할하는 소득 수준은 두 구간의 조건부 평균값들의 가중평균이어야 한다는 것이다. 다시 말하면 주어진 구간의 구간경계(critical point)는 두 구간에 대한 평균과 확률의 가중평균 값과의 차이가 최소되는 경계지점을 구해야 한다.

$$y_i^* = \frac{\pi_i^*}{\pi_i^*+\pi_{i+1}^*}\mu_i^* + \frac{\pi_{i+1}^*}{\pi_i^*+\pi_{i+1}^*}\mu_{i+1}^*$$

따라서 식(7)에 대한 개선된 다극화지수(extended ER)는 다음과 같이 정리할 수 있다.

$$P(f;\alpha,\beta) = ER(\alpha,\rho^*) - \beta(G(f)-G(\rho^*)) \qquad (10)$$

[주]

1) 본문에서는 편의상 '소득변수'만을 논의의 대상으로 간주한다.
2) 불평등 지수를 측정 시 만족해야 하는 대표적인 조건 중의 하나로 주어진 소득분포에서 고소득층으로 부터 저소득층으로 소득이 이동한다면, 소득의 순위를 포함한 다른 조건이 동일할 때, 불평등은 감소한다는 원리이다.
3) 여기에 초승달의 면적을 두 배를 하는 이유는 만약 소득이 완전하게 균등배분 되어 로렌스곡선이 대각선과 일치하면 지니계수는 0이 될 것이고, 반면에 소득이 가장 불균등하면 초승달의

면적이 최대 1/2이 값을 가지기 때문에 편의상 2를 곱하여 지니계수가 0과 1사이의 값을 가지게 하기 위해서이다.

4) 이것은 양극단의 경우, 즉 완전 동등 소득은 0을 완전 양극단 소득은0.25의 지수값을 가지기 때문에, 4배를 하여 편의상 0과 1 사이의 지수를 제시한다.

5) 동질집단의 강도를 나타내는 민감도이며, 다극화 지수를 불평등 지수와 구별되게 하는 값이다.

[참고문헌]

신동균, 전병유(2005), "소득 분포의 양극화 추이," 「노동경제논집 28(3)」, pp. 77-109.

Duclos, J. Y., J. Esteban and D. Ray(2004), "Polarization: Concepts, Measurement, Estimation," *Econometrica* 72, pp. 1737-1772.

Esteban, J. and D. Ray(1994), On the Measurement of Polarization," *Econometrica* 62, pp. 819-851.

Esteban, J., C. Gradin and D. Ray(2007), "An Extension of a Measure of Polarization, with an Application to the Income Distribution of Five OECD Countries", *Journal of Economic Inequality* 5, pp. 1-19.

Wolfson, M.C.(1994), "When Inequalities Diverge," *American Economic Review Papers and Proceedings* 84, pp. 353-358.