

신용카드리뷰

The Credit Card Review

Vol 16-4(2022년 12월)

금융시장의 변화와 금융불평등도 및 금융극화
이론에 관한 연구

김 상 봉

한성대학교 경제학과 교수



www.kci.go.kr

금융시장의 변화와 금융불평등도 및 금융극화 이론에 관한 연구*

김 상 봉**

한성대학교 경제학과 교수

〈Abstract〉

본 연구는 금융시장의 변화하여 확대되는 것을 살펴보고, 급변하는 금융환경에서 금융소비자에 대한 금융불평등을 측정하는 방법론을 살펴본다. 주요국이나 국내에서 어떠한 신용정보가 신용평가를 위해 쓰이는 지에 대해 설명한다. 그러나 금융분야에서 가장 기본이 되는 신용등급이나 신용점수의 미시자료를 활용하는 것은 여러 가지 면에서 상당히 어렵다. 금융시장의 공급자 측면에서의 변화와 금융불평등 중에 대출시장에서 어떠한 불평등이 존재하는 지에 대한 방법론을 본 연구에 소개하였다. 또한, 금융불평등도를 연도별로 추정하는 방법론을 소개하고, 추정된 연도별 자료를 기반으로 분해방법론을 제시하였다. 또한 불평등지수들은 분포의 변동을 반영하지 못하기 때문에 극복하는 개념이 극화 또는 다극화 추정방법론과 분해방법론을 제시하고 있다.

핵심 단어: 신용정보, 신용평가, 금융불평등도, 금융극화지수, 분해방법론

* 본 연구는 한성대학교 교내학술연구비 지원과제임.

** 서울 성북구 삼선교로 16길 116 한성대학교 연구관 505호, Tel: 02)760-8038, E-mail: brainkim75@hansung.ac.kr

I. 서론

세계의 주요국의 금융환경은 매우 빠르게 변화하고 있다. 과학기술과 인터넷 등의 발전으로 인해 매일 새로운 상품이 나오고 있고 국가들 내에서도 금융의 장벽이 허물어지고 있다. 비단 금융의 장벽 뿐만 아니라 금융이 일부의 역할을 하는 비금융거래에서도 금융과 금융상품의 역할은 커지고 있다. 즉 예금, 대출, 외환 등의 일반 금융상품 뿐만 아니라 상품과 서비스의 거래에서도 금융이 도입되었고 새로운 모바일이나 금융환경에서 거래가 지속되고 있다. 이에 따라 빅테크기업이나 핀테크기업이 원래 영위하던 사업에서 벗어나 관련 금융시장에 진출하고 있다. 이전에 상품과 서비스를 플랫폼에 놓고 매매를 하였다면 상품과 서비스의 거래에 관여할 뿐만 아니라 전통적인 금융에서 취급하던 예금, 대출, 외환 등을 취급하기도 한다. 또한 보다 확대되어 은행 뿐만 아니라 증권, 카드, 보험 등의 비은행기관 또는 회사의 역할을 하기도 한다. 이러한 상황에서 기존의 금융시장에서 다루어 오던 금융기관 또는 금융회사의 역할에 대한 많은 질문이 생기게 된다. 특히 저금리 상태에서의 대출총량 등의 대출이나 대출금리 결정에 대해 많은 의문이 남게 된다. 물론 고금리 상태에서도 대출에 대한 규제나 금리 등이 발생할 수도 있다.

디지털 전환이라는 명칭으로 2019년 이후 규제 샌드박스를 도입하여 핀테크 기업 등이 새로운 서비스를 시장에 출시할 때 일정기간 기존 규제를 면제해주고 있다. 빅테크나 핀테크 등 비금융 기술기반 기업이라도 최소 자격요건을 갖추면 소액결제시스템에 참가할 수 있도록 하였다. 2020년 「개인정보보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법)」, 「신용정보의 이용 및 보호에 관한 법률(신용정보법)」의 데이터 3법 개정으로 빅데이터 분석이나 활용을 할 수 있도록 제도적 기반도 마련되었다(이천우·김상봉·김재현, 2022). 또한, 2020년 7월에 정부는 전자금융거래법 전면 개편을 추진하였고, 금융수요자 중 중요한 주체인 개인이 대출 등을 통해 자금을 조달하는 금융이 훨씬 더 활성화될 것으로 보인다.

금융환경이 급변하고 있는 상황에서 금융공급자인 금융기관이나 금융수요자 중에서 많은 부분을 차지하는 개인에 대한 불평등도 어느 정도 존재하는 것으로 보인다. 특히 대출을 중심으로 하는 금융에 신용등급 또는 신용점수라는 방법론을 통해 많은 영향을 주고 있다. 신용점수를 알고 직접적인 분석을 시행하는 것이 좋다. 그러나 개인정보 등에 대한 이유 또는 측정기관들마다 신용점수를 산출하는 것에 차이가 있다는 한계가 있고, 신용등급에서 신용점수로 기준이 바뀌면서 새로운 데이터 축적의 기간이 길지 않다. 이러한 금융시장 환경에서 이전에 전혀 연구가 되지 않았던 금융불평등도를 측정하는 연구방법론을 제시하는 것은 상당히 의미가 있다.

본 연구의 구성은 다음과 같다. 2장에서 금융시장과 신용평가 현황을 살펴본다. 3장에서 금융이 존재하는 미시자료의 존재에 대해 살펴본다. 4장에서 미시자료를 활용한 금융 불평등 측정방법론을 살펴본다. 5장에서 불평등요인 분해방법론을 살펴보고, 마지막 6장에서 결론을 다룬다.

II. 금융시장과 신용평가

2.1 금융시장의 변화

2.2.1 국내 금융시장의 변화

국내 금융시장은 상당히 빠른 발전을 하고 있다. 금융과 비금융 사이의 경계가 허물어졌으며, 비금융기업이 금융사업을 영위하는 경우도 많이 존재한다. 이른바 디지털 전환이라는 이름으로 금융시장은 변화하고 있으며, 2019년 이후 규제 샌드박스는 핀테크 기업 등이 새로운 서비스를 시장에 출시할 때 일정기간 기존 규제 면제해주고 있다. 2020년 지급결제 부문에서 혁신과 경쟁 촉진으로 핀테크 등 비금융기업의 지급결제시스템 참가를 논의하고 빅테크나 핀테크 등 비금융 ICT기업이라도 최소 갖추면 차액결제를 간접수행하는 방식으로 소액결제시스템에 참가할 수 있도록 한국은행과 금융결제원은 기준을 변경하였다(이천우·김상봉·김재현, 2022). 또한 2020년에 데이터 3법의 개정으로 빅데이터 분석이나 활용을 할 수 있도록 제도적 기반도 마련되었으며 전자금융거래법 개정도 진행되고 있다.

금융회사의 수는 2017년 이후에 업권별로 차이가 있다. 은행업의 경우 인터넷전문은행이 등장하면서 수가 증가하였다. 외은지점은 지점폐쇄 등으로 감소하고 있다. 보험업의 경우 보험회계기준(IFRS17) 등의 글로벌규제 등으로 인해 수가 감소하고 있다. 금융투자업의 수는 코로나19 이후 비대면전환과 주가상승 등으로 증가하였다. 여신전문금융업의 경우 신기술투자사 중심으로 증가하였다. 반면에 전통 금융업종인 상호금융업의 수는 감소하고 있다. 등록 대부업 및 대부금융업자는 증가하는 추세로 보이지만, 법정 최고금리 인하 등으로 인해 대부업과 대부금융업자는 감소하지만 대부 및 중개겸업이 증가하여 전체적으로 증가하는 것으로 보인다. 따라서 업권별로 기술이나 비금융업을 영위하던 회사들이 금융업으로 진출한 것을 알 수 있다.

<표 1> 국내 금융회사 현황

은행	시중은행	8	8	8	8	9
	인터넷전문은행	2	2	2	2	3
	지방은행	6	6	6	6	6
	특수은행	5	5	5	5	5
보험	외은지점	38	38	36	36	35
	생명보험	24	24	24	24	23
	손해보험	32	30	31	31	30
금융투자	증권사	55	56	56	57	58
	선물사	5	5	5	4	4
	자산운용자	215	255	297	326	350
	투자자문사	179	186	198	221	294
	부동산신탁	11	11	14	14	14
	종합금융회사	1	1	1	1	1
여신전문	신용카드사	8	8	8	8	8
	리스사	26	25	26	26	26
	할부금융사	21	22	23	23	23
	신기술금융사	42	51	58	63	74
상호금융	저축은행	79	79	79	79	79
	신용협동조합	898	888	883	879	873
	농협협동조합	1,131	1,122	1,118	1,118	1,118
	수산업협동조합	91	91	91	91	91
	산림조합	142	142	142	142	141
금융지주회사	9	9	10	10	10	
등록 대부업 및 대부중개업자	8,084	8,310	8,354	8,501	8,650	

자료: 금융감독원, 재구성

주요 금융회사에 기술기업이 진출해 있는 상황이다. 또한 2020년 7월에 정부는 전자금융거래법 전면 개편을 추진하였고, 아직 전자금융거래법 개정안은 국회에 계류 중에 있다. 주요 내용 중에 전자금융업종이 기능별로 통합되고 간소화되는데, 이전의 7개 업종에서 5개로 단순화된다. 전자자금이체업은 자금이체업, 전자화폐업과 선불전자지급수단업은 자금이체업과 대금결제업, 직불전자지급수단업은 대금결제업으로 변화된다. 또한 전자지급결제대행업, 결제대금예치업, 전자고지결제업은 결제대행업으로 통합된다(금융위원회, 2020).

전자금융업의 시작이 되는 지급지시전달업(마이페이먼트, MyPayment)은 이용자의 결제 또는 송금의 지급지시를 받아 금융회사 등이 이체를 실시하도록 전달하는 업종이다(금융위원회, 2020). 마이페이먼트는 고객의 계좌를 보유하지 않은 상태에서 고객의 동의로 결제서비스에서 필수적인 정보인 고객의 금융계좌 정보에 대해 접근할 수 있게 된다. 또한, 고객은 은행 등에 자금을 보관하면서 은행이 제공하는 금융서비스를 이용할 수 있다. 본인신용정보관리업(마이데이터, MyData) 등과 연계로 조회나 이체 또는 결제 과정에서 종합서비스가 이루어질 수 있다.

<그림 1> 전자금융업종 기능별 간소화



자료 : 금융위원회, 재정리

마이데이터는 고객의 은행 및 이용 내역 등 금융정보를 모아 분석하는 서비스를 수행한다. 마이데이터는 이전에 자유업으로 운영되었으나, 개인정보보호 및 보안 강화, 그리고 안정적인 서비스의 제공 등을 위해 2020년 「신용정보법」 개정으로 제도화되었다. 마이데이터 시행 초기 단계부터 마이데이터 산업 생태계가 조성될 수 있도록 허가제로 운영하면서도, 창의적인 기업의 시장진입을 유도하기 위하여 진입장벽을 최소자본금 5억원, 금융회사 출자요건(50% 이상) 미적용, 클라우드 전산설비 이용 허용 등으로 최소화하였다(금융위원회, 2022). 이에 따라 2022년 1월 5일 마이데이터 기업들의 마이데이터 서비스 제공이 시작되었다.

마이데이터의 허가를 받은 회사의 수는 계속하여 증가하고 있는데, 2022년 4월 13일 기준으로, 55개사에 이르며, 예비허가 회수의 수는 10개이다. 또한 허가 심의 중인 회사도 20여개사에 이르고 있다. 본허가를 받은 회사는 은행 10개사, 보험회사 2개사, 금융투자회사 7개사, 여신금융회사 9개사, 상호금융사 1개사, 상호저축은행 1개사, 핀테크회사 24개사, CB회사 등 2개사는 이미 본허가를 받은 상태이다. 이러한 은행, 지급결제, 송금 외에도 빅테크기업이나 핀테크기업은 증권, 보험업 등에 진출하여 수신 또는 여신 등을 비롯한 금융을 다루려고 하고 있다.

2.2.2 해외 금융시장의 변화

해외 금융시장도 빠르게 시장이 변화하고 있다. 기존의 전통적인 금융업은 은행, 증권, 카드, 보험 등을 포함한다. 여기에 새로운 기술기업들이 진출하고 있는 상황이다. 유럽위원회(European Commission, EU)는 디지털 금융의 중요성을 인식하고 관련 법 및 제도를 정비하고 「지급결제산업지침(Payment Service Directive, PSD)」 제정(2007) 및 시행(2009), 「전자적 신원확인 및 인증 등에 관한 법률(eDIS)」을 제정(2016)하였고, 「지급결제산업지침(Payment Service Directive 2, PSD2)」 제정(2016) 및 시행(2018), 「온라인 플랫폼의

공정성·투명성 강화를 위한 법률」을 제정(2020)하였다(이천우·김상봉·김재현, 2022).

<그림 2> PSD와 PSD2의 주요 내용

PSD의 주요 내용(6개 타이틀)	PSD2의 주요 내용(6개 타이틀, 117개 조항으로 추가수정)
EU 내 지급결제서비스를 회원국 내 동일한 수준으로 제공	EU 지급결제시장 통합을 위한 지침 범위 확대
지급결제서비스 제공업자(Payment Services Provider, PSP)에 대한 규제를 통한 서비스 경쟁력 강화	제3자 지급결제서비스 제공업자(Third Party Payment Service Provider, TPP) 도입 : 고객 계좌 미보유+지급결제서비스 수행→ 계좌정보서비스업자(AISP, 지급인이 보유하고 있는 계좌 관련 통합정보 제공), 지급지시서비스제공업자(PISP, 지급거래시 지급인의 요구로 수취인 계좌로 자금이체) 수행
소비자 권익 보호	거래투명성 제고 등 소비자 보호 강화
	강력한 고객인증 요구 등의 개별 거래의 보안 강화

자료 : 이천우·김상봉·김재현(2022)

PSD2 조문은 6개의 타이틀로 구성된다. 타이틀 I은 주제, 범위 및 용어를 정의하며 대상기관, 적용범위, 예외, 정의한다. 타이틀 II는 지급결제서비스 제공업자(PSP)에 대해 다루는데 지급결제기관(Payment Institutions)의 일반 원칙, 감독, 면제규정을 다루며 공통조항으로 지급결제시스템이나 계좌접근, 금지사항을 다룬다. 타이틀 III은 지급결제서비스 관련 거래의 투명성, 정보공개 요건을 다룬다. 타이틀 IV는 지급결제서비스 제공 및 사용에 관한 권리 및 의무를 다룬다. 공통조항으로 범위, 수수료, 적용면제 거래 다루며 지급거래 인가로 지급결제서비스 제공업자 및 사용자 의무 등, 지급거래 실행으로 지급명령, 결제시간, 책임소재, 데이터보호, 리스크 등을 다룬다. 타이틀 V은 위임규정 및 규제 기술표준(RTS)을 다루며 타이틀 VI은 최종 조항이다(이천우·김상봉·김재현, 2022).

2.2 신용평가

2.2.1 국내 신용정보와 신용평가

국내의 신용정보는 다루는 범위와 평가방식이 다양해지고 있다. 예전에 부정적인 신용정보만을 다루었다면 최근에 전통적인 신용정보를 넘어서는 긍정적인 신용정보를 다루기도 한다. 일부 핀테크기업들은 각 기업의 자료를 활용하여 대안적인 신용평가점수나 등급을 산출하기도 한다. 여기에서 전체적인 법조항을 살펴보는 것보다 「신용정보의 이용 및 보호에 관한 법률」, 소위 신용정보법에서 다루는 신용정보의 대상을 살펴보자.

2020년 12월 29일에 개정된 신용정보법은 신용정보의 개념과 범위에 대해 다루고 있는데, 개인신용정보는 개인에 관한 신용정보라고 한다. 이러한 개인신용정보는 살아있는

개인에 관한 신용정보로서, 신용거래정보, 신용도(度)정보, 신용거래능력정보, 공공기록정보, 개인신용평점정보, 그리고 이러한 정보들과 결합되는 개인식별정보 등 6가지로 구성된다.)¹⁾

신용거래정보는 신용정보주체의 거래내용을 판단할 수 있는 정보이며, 다음과 같이 크게 다섯 가지의 신용정보를 말한다. 첫째, 은행의 신용공여, 여신전문금융회사의 신용카드, 시설대여 및 할부금융 거래, 금융투자업의 신용공여 거래의 종류, 기간, 한도, 금리, 금액 등에 관한 정보를 포함한다. 둘째, 「금융실명거래 및 비밀보장에 관한 법률」에 따른 금융거래(예금거래 등)의 종류, 기간, 금액, 금리 등에 관한 정보이다. 셋째, 보험업의 보험상품의 종류, 기간, 보험료 등 보험계약에 관한 정보 및 보험금의 청구 및 지급에 관한 정보이다. 넷째, 금융투자업의 금융투자상품의 종류, 발행·매매 명세, 수수료·보수 등에 관한 정보이다. 다섯째, 「상법」에 따른 상행위에 따른 상거래의 종류, 기간, 내용, 조건 등에 관한 정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호의3).

신용도정보란 신용정보주체의 신용도를 판단할 수 있는 정보를 말한다. 첫째, 금융거래 등 상거래와 관련하여 발생한 채무의 불이행, 대위변제, 그 밖에 약정한 사항을 이행하지 아니한 사실과 관련된 정보이다. 둘째, 금융거래 등 상거래와 관련하여 신용질서를 문란하게 하는 행위와 관련된 정보로서 1) 금융거래 등 상거래에서 다른 사람의 명의를 도용한 사실에 관한 정보, 2) 보험사기, 전기통신금융사기를 비롯하여 사기 또는 부정한 방법으로 금융거래 등 상거래를 한 사실에 관한 정보, 3) 금융거래 등 상거래의 상대방에게 위조·변조하거나 허위인 자료를 제출한 사실에 관한 정보, 4) 대출금 등을 다른 목적에 유용(流用)하거나 부정한 방법으로 대출·보험계약 등을 체결한 사실에 관한 정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호의4).

신용정보법에서 신용거래능력정보란 신용정보주체의 신용거래능력을 판단할 수 있는 정보로서 개인의 직업·재산·채무·소득의 총액 및 납세실적의 신용정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호의5).

공공기록정보는 1) 신용정보주체가 받은 법원의 재판, 행정처분 등과 관련된 정보, 2) 신용정보주체의 조세, 국가채권 등과 관련된 정보, 3) 신용정보주체의 채무조정에 관한 정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호의6 가목, 나목, 다목). 법원의 재판 등과 관련된 정보로서는 법원의 개인회생이나 파산선고·면책·복권과 관련된 결정이나 경매개시결정·경락허가결정 등 경매와 관련된 결정에 관한 정보를 들 수 있다. 조세 등과 관련된 정보는 국세, 지방세, 관세, 국가채권의 납부 및 체납 관련 정보가 해당된다.

신용정보법에서 개인신용평점정보란 개인의 신용상태를 평가하기 위하여 정보를 처리

1) 보다 자세한 사항은 「신용정보의 이용 및 보호에 관한 법률」의 제2조 참조.

함으로써 새로이 만들어지는 정보로서 기호, 숫자 등을 사용하여 점수나 등급 등으로 나타낸 정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조 제1호의6 라목).

신용정보법에서 개인에 관한 식별정보란 특정 신용정보주체를 식별할 수 있는 정보를 말하는데, 살아 있는 개인에 관한 정보로서 1) 성명, 주소, 전화번호, 2) 법령에 따라 특정 개인을 고유하게 식별할 수 있도록 부여된 개인식별번호, 3) 개인의 신체 일부의 특징을 컴퓨터 등 정보처리장치에서 처리할 수 있도록 변환한 문자, 번호, 기호 또는 그 밖에 이와 유사한 정보로서 특정 개인을 식별할 수 있는 정보를 말한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호의2). 다만 이러한 식별정보가 개인신용정보에 해당하려면 위에서 언급한 신용정보와 결합되는 경우에만 해당한다(「신용정보의 이용 및 보호에 관한 법률」 제2조제1호가목).

개인신용정보는 금융거래 등의 신용도 판단에 필요하므로 이용이 되어야 할 정보이지만 반면에 개인신용정보가 유출이 되면 범죄에 이용되거나 사생활의 비밀이 보호되지 않아 해당 개인이 피해를 입는 문제가 발생하므로 개인신용정보의 보호도 상당히 중요하다. 개인신용평가회사(Credit Bureau, CB)가 개인에 대한 신용정보를 수집하고 이를 통계적 방법으로 분석하여 만든 것을 개인신용평점이라고 한다. 이를 기반으로, 향후 1년 이내에 90일 이상 장기연체과 같은 신용위험이 발생할 가능성을 수치화하여 금융회사 등에 제공하게 된다. 이렇게 산출된 개인신용평점은 금융회사들이 개인의 신용을 기초로 신용거래를 설정하거나 유지 등의 의사결정이 필요한 경우에 참고지표로 활용이 활용이 가능하다. 또한 개별 금융회사는 자체적으로 신용평점시스템(Credit Scoring System, CSS)에서 산출된 평점을 기반으로 대출승인, 대출한도, 금리결정, 신용카드발급 등 금융거래를 위한 의사결정을 하게 된다. 신용평점시스템은 두 가지로 구분된다. 신규 신용거래를 위한 신청평점시스템(Application Score System, ASS)과 현재 거래 중인 고객을 대상으로 하는 행동평점시스템(Behavior Score System, BSS)이 있다. 신청평점시스템(ASS)은 신규 거래에 대한 대출승인여부, 신용카드발급 등을 위한 의사결정에 활용된다. 행동평점시스템(BSS)은 거래 중인 고객에 대해 대출연장 시 연장여부 또는 금리변경 등의 결정이나 한도변경 등의 결정에 활용된다.

국내 CB의 개인신용평점의 주요 평가 요소 중에 조회정보는 신용평가에 활용되지 않는다. 신용평가에 활용할 신용정보가 없거나, 18세 미만의 미성년자와 100세 이상의 고령자에 대하여 신용평점을 산출하지 않는다. 국내 CB의 신용평점 산출에서 변수의 비중은 주요국과 비슷하다. 현재 연체 및 과거 채무 상환이력(30.6%), 대출 및 보증채무 등의 채무부담 정보인 부채수준(26.4%), 최초/최근 개설로부터의 신용 거래기간(13.3%), 체크/신용카드 이용정보인 신용거래 패턴과 같은 신용형태(29.7%)가 활용된다. 보다 세분화된 신용평가 요소와 변동 요인은 다음과 같다. 상환이력정보, 부채수준, 신용거래기간 순서로 비중이 높게 나타난다. 일반고객군에 비해 장기연체군은 상환이력정보, 부채수준은

더 높게 나타난다.

<표 2> 신용평가 요소 및 변동 요인

평가영역	신용평가 요소	일반고객군	장기연체군	평가개요
상환이력정보	장기연체 발생	--	--	*단기연체의 기준은 5영업일 10만원 이상 이며, 장기연체는 신용정보원 및 신용정보사 채무불이행 등재 또는 90일 이상 연체 등재를 기준으로 하고 있다. *다만, 일시적 소액연체는 신용평가에 활용되지 않는다.
	단기연체 발생	--	--	
	연체 진행일수 경과	--	--	
	연체 해제	++	+	
부채수준	연체 해제일수 경과	+	+	*보증 및 대출의 발생은 상환부담에 따른 신용위험이 있는 것으로 판단되어 부정적인 영향을 주게 되고, 반대로 상환 시에는 신용위험이 감소된 것으로 판단되어 신용평점에 긍정적인 영향을 주게 된다.
	비중	29.7%	47.8%	
	고위험 대출발생	--	--	
	고위험 외 대출 발생	-	-	
	대출잔액 증가	-	-	
	대출 부분 상환	+	+	
신용거래기간	대출 전액 상환	++	+	*신용거래 기간은 시간이 길수록 긍정적인 요인으로 분류된다.
	보증 발생	-	-	
	보증 해소	+	+	
	비중	25.5%	42.8%	
신용형태정보	신용거래기간 없음	-	-	*연체없이 사용하는 신용카드 사용은 긍정적인 요인이나, 지속/습관적인 할부 및 현금서비스의 과다 사용은 부정적인 영향을 미친다.
	신용거래기간 경과	++	+	
	비중	13.5%	9.4%	
	신용/체크카드 사용 개월	++	+	
기타	신용/체크카드 사용 금액 적정	+	+	*국세청 소득금액증명 제출 및 통신요금, 아파트관리비, 국민연금, 건강보험 성실 납부내역 제출 시(소득금액이 아닌 소득여부만 확인)
	과다 할부 사용	-	-	
	현금서비스 사용	--	--	
	비중	31.3%	0.0%	
기타	증빙소득	+	+	
	비금융거래 성실납부실적 등록	+	+	

주: 개인신용평점의 평가요소별 평균적인 변동 요인은 +++>++>+>->-->---의 순서로 긍정적 요소의 단계를 나타낸다.

자료: NICE평가정보, 김상봉·강경우(2022)

이러한 개인신용평가 관련 인원수의 통계자료는 신용평점별 인원수, 장기연체가능성, 대출보유고객, 신용카드보유자수로 나타난다.

<표 3> 신용평점별 인원수(2021년말 기준)

900~	20,358,081	0.06%	8,530,246	19,696,583
800~	11,864,489	0.20%	6,638,556	7,809,189
700~	12,595,487	0.45%	2,687,916	2,638,510
600~	729,594	0.62%	640,997	565,255
500~	110,631	0.72%	103,659	87,262
400~	46,037	0.81%	44,607	37,545
300~	1,872,619	0.83%	866,123	633,619
200~	112,709	0.83%	108,259	101,772
~199	3,164	0.83%	3,144	3,164
계	47,692,811	0.83%	19,623,507	31,572,899

주: 장기연체가능성 : 2021.01~2021.12까지 신규로 한국신용정보원에 3개월 이상 연체 등록으로 신용평점 구간은 시작점수~1000점임

자료: NICE평가정보

최근에 위와 같은 전통적인 신용정보를 활용한 부분을 넘어서 대안적인 평가정보가 활용되기도 한다. 예를 들어, 카드는 금융정보 이력이 부족한 금융이력부족자(thin filer)의 공과금, 세금, 통신비 납부 등 다양한 카드 사용 내역을 추가적인 신용평가 데이터로 활용해 대출 상품의 문턱을 낮추고 있다. 통신업과 금융업의 협업도 활발히 이루어지고 있다. 예를 들어, SK텔레콤은 하나금융그룹과 손잡고 핀테크 업체 핀크(Finnq)를 선보였다. 핀크 앱을 이용해 오픈뱅킹 기반의 편리한 금융서비스를 제공하고, 고객의 휴대폰 이용 정보를 바탕으로 한 비금융정보 신용평가서비스인 T스코어를 제공한다. 이 같은 T스코어는 기존에 신용평가에서 불이익을 받은 썬파일러의 신용평가 등급을 개선시킴으로써 금융 소외계층을 포용하는데 기여했다. 또한 한화생명명은 2016년 중금리 연구를 위하여 4백만 고객의 고객등급 21개 항목 정보를 SKT의 고객 연체정보 등의 21개 항목과 결합을 시도하였다. 금융정보가 없는 고객도 통신사 데이터와 보험데이터만으로 신용평가가 가능해졌다. 비금융정보로 신용평가하는 대안 모델이 될 수 있으며, 청년층 신용평가도 가능해졌다. 추정소득, 보험계약 해지, 보험료 연체, 보험약관 대출 등이 통신요금 연체를 설명할 수 있기 때문이다.

<그림 3> SKT와 한화생명의 정보 결합

	직업	신용대출건수	최초대출날짜	최초연체날짜	총신용대출금액	총상환금액	신용대출연체율
한화생명	최근1년 신용대출연체율	30일이나 신용대출연체율	최초신용등급	최근신용등급	보험료연체율	최근1년 보험료연체율	실요해지건수
	기납입보험료	월납입보험료	직업기반 추정소득금액	가구단위 추정소득금액	평균약관대출율	약관대출금액	자동차 실패율수
	나이	성별	사용개월수	멤버십등급	월평균통화시간	월평균통화빈도	APRU
SKT	결합상품가입여부	단말기출고가	이용정지간	당월 통신료연체금액	최근1년 최대통신료연체금액	납부방법	회선상태
	남은단말기 할부원금	가입회선수	테블릿PC 보유여부	스마트워치 보유여부	멤버십 당월사용금액	멤버십 당년사용금액	통신료미납회수

주 : 주황색(직업, 나이, 성별) 바탕은 준식별자, 회색바탕은 민감정보, 빨간색 글씨는 신용정보임
 자료 : 김상봉(2021a), 재구성

비슷한 시기에 금융-통신부문 간 이중 데이터 결합을 통해 신용도가 낮은 서민에게 중금리 대출 제공할 수 있거나 대출금리 인하 기회를 제공하여 금융사, 통신사, 고객 3자에 대해 모두 이익이 될 수 있는 협업 모델이 개발되었다. 금융데이터 결합을 통해 고객 금융서비스 기회를 확대하고 통신 고객에게는 로열티 강화가 가능하며, 고객 입장에서는 금리인하 효과가 있을 수 있다.

2.2.2 해외 신용정보와 신용평가

국내에도 공적 신용집중기관(PCR)과 민간 신용정보회사(CB)가 존재한다. PCR은 법규에 근거하여 금융기관이 일정 금액의 개인대출에 관한 정보를 중앙은행이나 금융감독당국에 집중한 후, 이러한 정보를 금융 감독당국과 개별 금융기관에 제공하는 역할을 한다. 반면, CB는 이윤창출을 위한 목적으로 운영되고 있다. 신용정보를 가지고 있는 금융기관등으로부터 개인신용정보를 수집한 후, 정보를 금융기관 및 신용공여기관에 제공한다. 주요국들도 개인정보를 집중하거나 정보를 공유하는 범위가 다양하다.

<표 4> 주요국의 개인신용정보 집중기관 및 정보공유 범위

CB	불량 및 우량 신용정보	대규모 CB 존재
CB	불량 및 우량 신용정보	대규모 CB 존재
CB	불량 및 우량 신용정보	금융업권별 대규모 CB
CB	불량 신용정보	우량 신용정보 공유 금지
PCR	불량 신용정보	CB 설립 불가
PCR, CB	불량 및 우량 신용정보	대형 CB 존재
PCR, CB	불량 신용정보	
PCR, CB	불량 및 우량 신용정보	
PCR, CB	불량 및 우량 신용정보	PCR에 다양한 우량정보 등록 의무화
PCR, CB	불량 및 우량 신용정보	PCR에 다양한 우량정보 등록 의무화

자료: 김상봉외(2012), 재구성

미국의 경우 이퀴팩스(Equifax), 익스페리언(Experian), 트랜스 유니언(TransUnion) 등 3대 CB가 존재한다. 이러한 CB는 개인과 기업에 대한 신용조회, 평가 등의 서비스를 제공하고, 신용조사와 조회 업무를 하고 있으며 대출을 신청 또는 사용하거나 공공기록을 보유한 개인에 대한 신용보고서를 축적한다. 이러한 신용보고서에 소비자의 신용정보(credit report), 채권추심정보, 직장소득정보, 부동산가치평가 정보가 포함된다. 신용평점(Credit Score) 서비스는 표준평점서비스, 업종 및 상품별 평점서비스를 제공하며 조기경보(Early Watch) 서비스는 소비자의 신용정보 변동 시 조기에 대처할 수 있도록 변동사항 통보서비스가 있다. 사기방지(Fraud Prevention)는 신청사기방지 서비스, 신청사기 평점모형을 서비스하며, 추심위치서비스(Collection Locator Tools)는 채권추심을 위한 채무자 주소 연락처, 소재 미과약자에 대한 주소, 연락처 등을 제공한다. 선별(Prescreening)은 신용거래를 위한 정책, 거절리스트 정보, 직원채용을 위한 대상자를 점검한다. 의사결정 솔루션(Decision Solution)은 신용거래 의사결정체 필요한 정보처리, 데이터. 의사결정지원 시스템이며 데이터분석 서비스(Data Analytical Service)는 데이터 분석을 통한 비교분석 보고서를 제공한다. 미국의 CB는 소비자의 이름이나 주소부터 계좌개설일자, 대출관련 자료, 상환관련 자료 등을 수집하고 가공한다. 이러한 3대 CB들은 대부분 최근에 전화요

금 등과 같은 비전통적 정보를 포함하는 신용평점 모형을 개발을 발표하였다.

페어아이작(Fair Isaac and Company)의 FICO Expansion Scoring은 미국에서 많이 사용된다. 데빗, 회원가입, 전기료나 전화료, 저당권 설정 등의 자료, 파산 선고 등의 정보를 이용하여 왔다. 2007년 말 이후 페어아이작은 PRBC(Payment Reporting Build Credit)와 제휴를 맺어 만든 PRBC Credit Report with FICO Expansion Score를 통해, 렌트 및 요금청구서(bill) 납부 자료 등의 비금융신용이력정보, 기존 3개 CB로의 전통적 신용정보 등을 포함하는 포괄적인 신용리포트를 결합하였다. 현재에도 FICO의 신용평점은 상환이력(payment history) 35%, 채무액(credit and debt) 30%, 신용거래기간(length of credit history) 15%, 대출한도(new line of credit) 10%, 대출유형(type of credit) 10% 등의 비중으로 구성 및 산출된다. 이러한 FICO 신용평점은 기본적인 상환정보에 초점이 맞추어져 있으며 은행이용정보(banking activity data)로 보완하는 시도(ultra FICO score)도 이루어지고 있다. 미국 은행의 소비자대출의 약 90%는 FICO 신용평점에 근거하여 대출이 이루어지며, 신용평점이 높을수록 높은 대출한도와 낮은 대출금리가 부과된다. 이외에도 LexisNexis는 통신기록 등의 비전통적정보인 공공정보를 활용하기도 한다. L2C는 전화료 납부 자료, 데빗이나 당좌계좌(checking account) 사용 내역, 임차료 납입 정보 등과 같은 대출자료를 활용하여, 금융거래이력이 없는 집단(no-filer)의 80~90%를 평점화하였다. 이제는 Bank of America와 같은 금융기관들도 자체적으로 일부 고객에 대해 기존 신용평가에 비전통적 정보를 포함하기도 하였다. 이러한 미국 은행들의 신용정보 이용과 모형에 개선 움직임이 보인다. 2021년말 상반기 기준으로 약 5,300만명의 미국 성인 인구가 신용정보 부족이나 부채로 FICO 평점을 부여받지 못하고 있다. 미국 은행들은 자체적으로 모형을 개발하고 있거나 금융당국은 신용평점모형이 금융소외계층의 신용접근을 원천적으로 차단함으로써 초단기 소액대출(payday loan) 등 약탈적 금융(predatory lending)으로 내몰리도록 만드는 요인으로 작용하고 있는 것으로 인식하고 있다(이광상, 2021).

독일의 대표적인 CB인 Schufa는 8,300만명의 인구 80% 이상의 인구에 대해 신용보고서를 제공한다. 그러나 이러한 데이터베이스(DB)는 신용의 이용목적, 소득, 국적, 종교에 대한 정보를 담고 있지 않다. 또한, DB에 있는 인구의 90%이상에 대해 긍정적 정보를 보유하고 있다.

Ⅲ. 미시자료 접근방법론

기존의 자료를 사용하여 불평등도를 측정하는 방법을 사용하여 적용할 수도 있다. 이러한 방법의 문제점은 미시적인 개인정보가 필요하며, 이러한 미시적인 데이터를 바탕으로 지수를 산출할 수 있다. 불평등도를 측정하기 위해 신용정보가 포함된 미시자료 데이

터를 얻어 소득, 자산, 직업 등으로 금융불평등에 미치는 영향을 파악할 수 있다. 그러나 문제는 자료 획득이 용이하지 않다. 다만 통계청 등의 자료를 활용할 수 있으나 신용불평등을 살펴보기는 어려우며 금융불평등은 파악이 가능하다.

먼저 이용가능한 자료로 통계청이 제공하는 가계금융복지조사를 활용할 수 있다. 가계금융복지조사는 가구구성, 자산, 부채 소득, 가계지출, 노후생활 등에 대해 설문한다. 2010년부터 하나의 시계열로 제공하던 가계금융복지조사(2010~)를 가계금융조사(2010~2011, 표본1만가구)와 가계금융복지조사(2012~, 표본2만가구)로 각각 분리하였으며, 2011년과 2012년의 시계열 분석 및 일부 세부계층자료 이용에 유의하여야 한다. 2013년 이후 조사표의 조사부문별 항목 내역이 변경되고 있다.

<표 5> 조사항목별 조사목적

구분	주요 조사목적
I. 가구구성	가구주 및 가구원 인구학적 특징, 가구원들의 산업, 직업, 가구주 관계 등을 파악하여 가구특성별 세부사항 분석
1. 실물자산	가구의 부(wealth)의 정도 파악 부동산 이경제상황에 미치는 정도 파악 - 거주주택 종류, 주거용 면적, 입주형태, 현재시가, 임대보증금, 월세, 거주주택의 보유 부동산, 부동산 구입계약 중도금, 자동차, 권리금, 기타 실물자산
	가구의 부(wealth)의 정도 파악 - 현금 및 입출금이 자유로운 저축, 적립식 저축, 펀드, 저축성 보험 및 보장성보험, 예치식저축, 펀드, 주식 및 채권, 기타 금융자산(별려준 돈, 임차보증금, 갯돈), 공적연금, 사적연금, 퇴직금 및 퇴직연금 형태의 자산
II. 자산	
2. 금융자산	금융자산 운용
3. 금융자산 운용	금융자산 투자 목적 금융자산 운용방법 및 고려사항
4. 부동산 운용	1년후 주택가격 변화 등 투자여부 파악 여유자금의 부동산 투자 여부 파악 부동산 투자 목적 부동산 운용 방법 부동산 투자하지 않는 이유
	가구의 재무건전성 측정(DTI, LTV, DSR 등) 하우스푸어 등 기초자료 제공 - 담보대출, 신용대출(마이너스통장 포함), 신용카드 관련 대출, 외상 및 할부구입 - 연금형부채(주택연금, 농지연금)
III. 부채	
1. 금융부채	부채규모 변화, 부채증가 사유 및 상환방법 납부기일 경과 여부 및 사유 원리금 상환의 생계부담 정도 가계지출 절감 여부 부채상환 능력 가구재무 스트레스 정도 파악
2. 부채상환 능력	
IV. 소득	
1. 소득	가구특성별 소득분위별 소득규모 및 자산분포 파악 경제적 어려움을 겪고 있는 인구정보 파악 정부보조금이나 세금정책 등 새로운 정부효과 분석 가계생활 수준파악 및 웰빙측정 빈곤율 및 소득분배 측정
V. 가계지출	
1. 가계지출	가구의 소비지출 및 비소비지출 파악 적자가구, 흑자가구 파악 소득대비 소비지출 규모 파악
VI. 노후생활	
노후 생활	은퇴여부 향후 은퇴연령 은퇴후 최소 및 적정생활비 노후준비 상황 정도 은퇴연령 은퇴후 생활비 마련 정도 및 방법

자료 : 통계청(2020)

통계에 대한 문제는 다음과 같다. 먼저, 가계금융복지조사는 표본조사이기 때문에, 표본오차와 조사과정에서 나타날 수 있는 비표본오차가 존재할 수 있다. 둘째, 조사단위가

경제적 가족이기 때문에 가계동향조사 등의 가구단위 조사와 다를 수 있다. 셋째, 2010~2011년에 1만 가구 조사인 가계금융조사와 2012년 이후 2만 가구 조사인 가계금융복지조사의 결과에 차이가 있을 수 있다. 넷째, 가구 특성은 가구주를 기준으로 작성된다. 예를 들어 자영업자 가구는 가구주가 자영업자인 경우이다. 다섯째, 자산 및 부채를 보유하지 않은 가구까지 포함한 전체가구느냐와 자산 및 부채를 보유한 가구만을 대상으로 통계는 다를 수 있다. 여섯째, 통계표의 수는 반올림되었기 때문에, 세목과 전체 합계가 같지 않을 수 있다. 일곱째, 부동산의 평가방법은 조사기준시점 현재 시가(시장가격)로 조사된다. 여덟째, 2018년부터 소득 및 처분가능소득, 비소비지출은 국세청, 보건복지부 등의 행정자료와 면접조사 자료를 활용하여 새롭게 작성된 것이다(통계청, 2020).

조사표 변경이 발생하면 현저한 답변의 변화가 생길 수 있다. 예를 들어, 기타 소비지출 금액이 조사표 변경으로 인해 이전 대비 57.6% 증가하였다(신인석·한민, 2021). 또한, 다른 통계와 비교하면 가계금융복지조사의 통계가 일치하지 않을 수 있다. 2016년 가계금융복지조사 결과와 행정자료로 보정된 결과를 비교하면, 피조사자가 소득과 자산을 응답할 때 편향이 발생하고 있다. 소득 부분은 과소 보고하게 되는 문제가 존재한다. 예를 들어, 근로소득이 높을수록 과소보고 하고, 금융소득에는 응답을 하지 않거나 응답을 하는 경우에도 금융소득을 크게 과소보고하였다. 반대로 소득이 낮은 경우에 응답한 소득이 행정자료의 소득보다 커지는 경향이 있고, 응답한 소득과 행정자료 소득 사이의 괴리가 큰 경우도 나타난다. 둘째, 응답자료와 과세자료와 비교하면 금융소득이 최상층에서 누락이 많은 것으로 나타난다. 이는 일종의 샘플링 오류라고 볼 수 있다. 셋째, 자산의 경우 응답한 결과를 직접 과세 자료와 비교하면, 주택의 경우 두 자료의 자산 분포가 근접한 것으로 나타났지만, 자산 편중이 보다 심한 토지의 경우에는 최상층의 일부가 누락되었으며, 금융자산의 경우 이러한 문제가 더욱 심각한 것으로 나타난다(김낙년, 2020).

IV. 미시자료를 활용한 금융불평등 측정방법론

4.1 미시자료를 활용한 측정

가계금융조사는 설문조사의 방식을 취하고 있고, 다양한 오류들이 존재하지만 설문항목에 따라 소득, 자산 등의 변화에 대한 부분을 살펴볼 수 있다. 지금까지의 연구는 소득 및 자산에 대한 연구였다면 대출 또는 금융에 대한 부분을 알기 어렵다. 금융불평등이 발생하는 원인은 대출자의 정보가 금융기관에 의해 제공되면 알 수 있으나 개인정보

등의 이유로 미시자료를 구하기 어렵다.

이러한 상황에서 가계금융복지조사는 소득분과 관련된 자료로 다음과 같은 통계를 산출한다. 소득분배지표 관련 주요 용어의 정의나 개념 등은 OECD 국제기준을 따르고 있다. 소득분배를 나타내는 지표를 금융불평등에 응용하는 지표로 사용될 수도 있다. 예를 들어, 금융불평등도를 나타내기 위하여 상대적 불평등도를 대표하는 지니계수나 소득5분위배율, 소득점유율, 팔마비율이 사용될 수 있다.

<표 6> 소득분배지표 관련 주요 용어 및 정의

균등화소득	가구원수가 다른 가구간의 후생(복지)수준을 비교할 수 있도록 가구소득을 가구원수로 나눈 소득
균등화시장소득	시장소득(=근로소득+사업소득+재산소득+사적이전소득-사적이전지출)을 균등화한 소득 *가구간 및 비영리단체 이전지출
균등화 처분가능소득	처분가능소득(=시장소득+공적이전소득*+공적이전지출**)을 균등화한 소득 *공적이전소득 : 공적연금(국민연금 등), 기초연금, 양육수당, 장애수당 등 **공적이전지출 : 세금, 공적연금 기여금 사회보험료 등
지니계수	소득불평등도를 나타내는 지표로써 '0'이면 완전평등, '1'이면 완전불평등을 의미함
상대적빈곤율	전체 인구 중에서 소득수준이 빈곤선(균등화 처분가능소득의 중위소득 50% 또는 60%) 이하에 속한 인구가 차지하는 비율
근로연령층빈곤율	전체 근로연령층(18세~65세) 인구중 소득수준이 빈곤선이하인 근로연령층 인구가 차지하는 비율
은퇴연령층빈곤율	전체 은퇴연령층(66세이상) 인구중 소득수준이 빈곤선이하인 은퇴연령층 인구가 차지하는 비율
평균 빈곤갭(%)	빈곤인구*의 평균소득과 빈곤선의 차이를 빈곤선으로 나눈 값 *소득수준이 빈곤선(균등화처분가능소득의 중위소득 50%)보다 낮은 인구
소득5분위	전체인구(가구)의 소득을 오름차순으로 정렬하여 한 그룹에 20%의 인구(가구)수가 포함되도록 5개의그룹으로 나눈 것
소득5분위배율	소득상위 20% 계층의 평균소득을 소득하위 20% 계층의 평균소득으로 나눈 값
소득10분위	전체인구(가구)의 소득을 오름차순으로 정렬하여 한 그룹에 10%의 인구(가구)수가 포함되도록 10개의 그룹으로 나눈 것
소득10분위배율	소득상위 10% 계층의 평균소득을 소득하위 10% 계층의 평균소득으로 나눈 값
소득점유율	전체 소득총액에서 해당 소득분위별 소득총액이 차지하는 비율 예)소득5분위점유율=소득5분위 가구들의소득총액/전체가구들의 소득총액
팔마(Palma)비율	소득상위 10%인구의 소득점유율을 하위 40% 인구의 소득점유율로 나눈 값
소득경계값	소득분위별로 구간을 나눌 때 각 구간의 상한값 예) P20은 소득하위 20% 구간의 상한값을 의미함

자료: 통계청(2020)

기존의 측정 방법과 더불어 새로운 측정방법론은 크게 2단계로 진행될 수 있다. 대출 등의 금융시장의 기본이 되는 신용평가시스템의 모형을 역추적하여 금융불평등을 만들어 내는 변수를 생각할 수 있다. 예를 들어, 주요국들은 직업, 산업, 나이 등을 신용평가시에 참고를 할 뿐이며, 변수로 취급하지 않는 경우가 많다. 직업이 금융부채 중의 각종 대출 등에 어떠한 영향을 미치는 지에 대해 알 수도 있다. 따라서 미시자료를 활용하여 기본적인 소득, 자산 외에도 가계금융조사에서 이루어지는 인구학적 특징, 가구원이 속해 있는 산업, 직업, 가구주와의 관계 등을 파악하여 가구특성별 세부사항으로 분석이 가능하다. 이러한 분석을 위하여 각종 불평등도지수와 극화지수 등과 계량경제학 모형, 그리고 최근 금융시장에서 이용되는 인공지능 등이 활용될 수 있다.

4.2 불평등도지수와 극화지수의 공리체계

4.2.1 불평등도지수의 개념

불평등도(inequality), 빈곤(poverty), 그리고 후생(welfare)은 동일한 개념이 아니다. 소득을 기준으로 생각해 보자. 빈곤율이란 빈곤선 이하의 소득을 가진 개인 또는 가구의 비율을 의미한다. 그러나 소득불평등도는 소득 전체의 분포를 정의하므로 빈곤율보다 광범위하다고 볼 수 있다. 소득분포의 하위 부분뿐만 아니라, 중간이나 상위 부분도 중요하게 볼 수 있다. 또한 소득불평등도를 후생과 비교하면, 소득불평등도는 좁은 개념으로 볼 수 있다. 후생과 소득불평등도 모두 전체 분포에 관심을 둔다는 점에서 비슷하다. 그러나 소득불평등도는 분포의 퍼짐(dispersion) 또는 분산만을 고려하지만, 후생은 분포의 퍼짐과 동시에 평균도 고려하기 때문이다. 그러나 이러한 약간의 차이에도 불구하고 불평등도, 빈곤, 후생은 서로 연관되어 있다고 할 수 있다.

불평등도(inequality)와 극화(polarization)의 개념도 약간 상이하다. 상대적 소득불평등도지수로 불평등도를 측정하면, 어떠한 사회의 구성원의 소득이 모두 두 배로 증가하였다면 불평등도의 변화가 없게 된다. 2인으로 구성된 사회의 예를 들어 보자. 구성원 1의 소득이 1,000만원에서 2,000만원으로, 구성원2의 소득이 2,000만원에서 4,000만원으로 증가하였다면, 지니계수로 계산되는 소득불평등도는 같다. 그러나 소득불평등도의 개념을 절대적 소득불평등도인 소득격차의 개념으로 생각하면, 소득격차는 1,000만원에서 2,000만원으로 두 배 증가하였으므로 소득불평등도는 두 배가 되었다고 할 수도 있다. 극화의 개념은 소득분포 상에서 서로 다른 집단이 다수 등장하는 경우에 불평등도는 증가하게 된다. 이러한 극화지수는 경우에 따라 상대적 소득불평등도지수와 비슷한 경우도 있고 다른 경우도 존재한다.

앞에서 가계금융복지조사와 같이 기존의 지니계수나 5분위배율과 같은 소득에 대한 불평등지수를 금융불평등도에 활용할 수 있다. 그러나 많은 불평등도지수가 개발되었지만 어떤 지수가 바람직한 속성을 모두 갖추고 있는 것에 대해 아직 발견되지 못했다. 따라서 각 불평등도지수가 어떤 장단점을 갖고 있는지를 파악하고 적절한 것을 선택하여 사용할 수 밖에 없다.

소득과 같은 하나의 특성변수의 분포와 같이 개별 분포를 하나의 수치로 나타내는 불평등도지수는 다양하게 정의될 수 있다. 그러나 두 개 이상의 불평등도를 비교하는 경우에 공리(axiom)에 맞는 불평등도지수를 찾고, 다양한 불평등도에 대한 비교를 할 수 있다. 예를 들어, 소득에 대한 지니계수와 같은 상대적 소득불평등도지수는 공리체계 하에서 소득불평등도의 크기나 비교 불가능성을 제시할 수 있다. 상대적 불평등도지수가 주

어진 공리하에서 측정이 이루어졌다면, 절대적 불평등도 또는 양극화나 다극화와 같은 극화지수는 공리의 일부를 위배하지만 불평등도를 측정하는 지수이다. 따라서 불평등도의 개념은 정해진 것이 아니고 주어진 공리하에서 불평등의 정도를 비교하게 되므로, 가정을 바꾸면 비교되는 불평등도가 다른 결과가 될 수도 있다.

4.2.2 불평등도지수 공리체계

많은 연구들에서 불평등이나 극화에 대한 공리체계를 설명하고 있지만, 여기서 보다 넓게 정리하도록 한다. 상대적 불평등도지수와 절대적 불평등도지수 또는 극화지수의 차이점을 이해하기 위하여 이러한 기준이 되는 공리체계가 존재한다. 상대적 불평등도지수는 네 가지 공리(axiom)인 익명성(anonymity), 동차성(homogeneity), 인구동차성(population homogeneity), 이전원칙(transfer principle)인 피구-달튼 원칙(Pigou-Dalton principle)을 만족해야 한다. 여기에 추가로 순위로 불평등도의 비교를 가능할 수 있는 지수는 분해성(decomposability)을 만족해야 한다.

익명성(anonymity)의 공리 또는 대칭성(symmetry)의 원리는 다음과 같다. 만약 $X \in \Omega$ 가 $Y \in \Omega$ 의 치환(permutation)으로부터 얻어진 것이라면 $X \sim Y$ 이다. 즉, 사회구성원의 변화가 없을 때, 구성원들 사이에 소득과 같은 변수가 서로 바뀌는 경우(permutation)에 불평등도의 변화는 없다. 다시 말하면, 사회구성원의 소득과 같은 변수의 분포 외에 다른 특성이 변한다고 하더라도 불평등도는 변화하지 않는다. 따라서 사회구성원의 변화는 없고 사회구성원의 변수(소득)만 서로 바뀌는 경우(permutation)에 불평등도는 변화가 없다. 사회구성원의 소득상의 위치가 바뀌어도 불평등지수는 변하지 않는다. 예를 들어, 한 사회의 변수인 소득분포가 $X = (100, 200, 300)$ 에서 $Y = X' = (300, 200, 100)$ 로 변화하였다면, 상대적 소득불평등도는 같다. 익명성 또는 대칭성의 공리가 성립하지 않으면, 불평등도를 비교하는 경우에 기본 전제가 성립하지 않으므로 모든 불평등도지수는 이 원칙을 만족시켜야 한다. 범위를 극단적으로 좁혀, 두 가구만 존재하고 가구원수가 다르지만, 가구원의 소득분포가 같다고 하더라도 두 집단의 소득불평등도는 같지 않다.

둘째, 동차성(homogeneity) 또는 규모비의존성(scale independence)의 공리는 다음과 같다. 만약 $X \in \Omega$ 가 $Y \in \Omega$ 의 소득에 양(+)의 수 λ 를 곱하여 얻은 것이라면 $X \sim Y$ 이다. 즉, 사회구성원의 소득이 비례적으로 변한다면 불평등도는 변화가 없다. 다시 말하면, 사회구성원의 소득이 같은 비율로 증가하거나 감소하면 불평등도에는 변화가 없다. 예를 들어, 한 사회의 구성원의 소득분포가 $X = (100, 200, 300)$ 에서 $Y = 2X = (200, 400, 600)$ 로 두 배 증가하였다면 소득불평등도에는 변함이 없다. 다만, 비율적인 곱으로 변하는 경우의 동차성을 규모비의존성(scale independence)이라고 하며, 더하거나 빼는 것으로 변하는 경우를 변환비의존성(translation independence)이라고 한다 (Amiel and Cowell, 1999).

셋째, 인구동차성(population homogeneity)의 공리 또는 인구비의존성(population independence)의 공리는 다음과 같다. 만약 $X \in \Omega$ 가 $Y \in \Omega$ 를 복제하여 증가시킨 방법으로 얻어진 것이라면 $X \sim Y$ 이다. 즉, 기존 사회구성원들의 소득분포와 동일한 소득분포를 가진 사회구성원이 등장하여 소득분포가 합해지는 경우에도 소득불평등도의 변화는 없다. 예를 들어, 소득분포가 $X = (100, 200, 300)$ 인 사회에서 동일한 소득을 가진 사회구성원들이 등장하여 $Y = (100, 100, 200, 200, 300, 300)$ 로 바뀐다고 하더라도 소득불평등도의 변화는 없다.

넷째, 이전원칙(transfer principle)의 공리 또는 피구-달튼 원칙(Pigou-Dalton principle)은 다음과 같다. 모든 다른 소득은 일정하다고 가정하자. 만약 $X \in \Omega$ 가 $Y \in \Omega$ 에서 상대적으로 부자인 a 로부터 상대적으로 가난한 b 에게 양의 소득을 이전함을 통해, a 와 b 의 소득순위는 변함이 없이 얻어진 것이라면 $X \succ Y$ 이다. 즉, 사회구성원의 평균소득은 변화하지 않은 상태에서, 부자가 빈자에게 소득의 일부를 이전(mean preserving income transfer)하면 소득불평등도는 감소한다. 다시 말하면, 사회구성원의 전체의 평균소득은 변하지 않은 상태에서, 부자가 빈자에게 소득을 이전하면 소득불평등도는 감소한다. 따라서 소득이 높은 사람으로부터 소득이 낮은 사람에게로 소득이 이전되면, 소득분배가 개선되고 소득불평등지수는 감소한다. 예를 들어, 소득분포가 $X = (100, 200, 300)$ 인 사회에서 $Y = (150, 200, 250)$ 인 사회로 소득의 일부의 이전이 발생하면 소득불평등도는 감소하게 된다. 다만, 분포 중에 분산(variance)과 로그분산(log variance)은 이러한 원칙을 위배하게 되며, 극화지수(polarization index)들도 이 원칙을 위배하며 등장한 지표들이다.

위와 같은 네 가지 공리를 만족하는 경우에 여러 가지 소득불평등도지수가 산출된다. 이렇게 산출된 소득불평등도지수들에서 나온 소득분포를 비교할 때, 어느 하나의 소득분포가 다른 소득분포와 비교하여 평등하다, 같다, 불평등하다고 나눌 수 있게 된다. 위 네 가지 공리 외에 하위 집단(group)으로 나누는 분해성(decomposability)이 추가로 포함되기도 한다.²⁾

다섯째, 분해성(decomposability)의 공리는 다른 조건이 일정한 경우에 전체 소득불평등도는 전체를 구성하는 일부분의 소득불평등도와 일치한다는 것이다. 즉, 전체에 포함되는 일부분의 집단의 불평등도가 변하면, 전체의 불평등도도 같은 방향으로 변해야 한다. 예를 들어, 두 집단의 소득분포가 $X = (300, 1000, 1100)$, $Y = (500, 600, 1300)$ 이고, 새로운 제3의 집단의 소득분포가 $Z = (700, 900)$ 이라고 하자. 이제 제3의 집단 Z 가 X 와 Y 에 각각 통합되어 $X' = (300, 700, 900, 1000, 1100)$, $Y' = (500, 600, 700, 900, 1300)$ 으로 바뀐다면, 바뀐 X' 과 Y' 의 소득불평등도 순위는 통합되기 이전의 소득불평등도

2) 그러나 서로 다른 소득불평등도의 크기를 비교하는 경우에 분해성의 공리에 대한 만족 여부는 상관 없다. Cowell(1995)은 분해성을 포함하여 다섯 가지 공리를 모두 만족하는 모든 불평등도지수는 이후에 다룰 일반화된 엔트로피(Generalized Entropy, GE)지수에 포함되는 것을 보였다.

크기인 X 와 Y 와 일치하여야 한다. 일반화된 엔트로피지수(Generalized Entropy, GE)의 종류와 같은 지수는 위 다섯 가지 원칙을 모두 지키는 것으로 증명된다(Cowell 1995).³⁾ 일반화된 엔트로피지수는 집단내(within-group) 불평등도와 집단간(between-group) 불평등도로 분해된다. 또한, 분해된 각 불평등도의 합계는 전체 불평등도와 같다. 그러나 앳킨슨(Atkinson)지수와 같은 종류의 불평등도지수는 집단간과 집단내 불평등도로 분해가 가능하지만, 합계가 전체 불평등도와 일치하지 않는다. 지니계수는 하위집단의 인구분포가 겹치지 않는 경우에 한해 분해가 가능하지만, 하위집단의 인구분포가 겹치면 분해가 불가능하기 때문에 분해성의 원칙을 부분적으로 만족시키지 못한다.

분해성을 제외한 앞의 익명성, 동차성, 인구동차성, 피구-달튼 원칙인 이전원칙이 로렌츠곡선을 통해 산출된 상대적 소득불평등도지수 등의 불평등지수의 크기 비교가 가능하기 위해 반드시 만족해야 한다. 그러나 마지막의 분해성 공리에 위배되더라도, 네 가지 공리를 만족하는 불평등도지수가 산출된다면, 로렌츠일치(Lorenz-consistence)와 동일한 소득불평등도의 크기 비교를 할 수 있다.

4.3 불평등도지수의 종류

불평등도를 측정하는 지수는 많지만, 상대적 불평등도의 기본 공리를 만족하며 로렌츠곡선상에서의 불평등도를 비교하는 것과 같이, 순위로 불평등도의 비교할 수 있는 지수는 한정적이다. 이렇게 순위로 불평등을 비교할 수 있는 불평등도지수를 로렌츠일치(Lorenz-consistent)지수라고 한다. 지니계수와 타일(Theil)의 엔트로피지수는 로렌츠일치지수 중에 대표적인 지수라고 할 수 있다.

4.3.1 분위배율(quantile ratio) 및 배율(share ratio) 불평등지수 추정

가장 간단한 분위배율은 다음과 같이 추정된다.

$$\widehat{QR}(p_1, p_2) = \frac{\widehat{Q}(p_1)}{\widehat{Q}(p_2)} \quad (1)$$

여기서 $Q(p)$ 는 p 분위수를 나타내고, p_1 과 p_2 는 백분위수이다.

배율은 다음과 같이 추정된다.

3) 인공지능이론 중의 의사결정나무나 마코프모형, 정보이론 등에서 엔트로피나 지니계수도 여전히 산출가능하다. 이에 관한 부분은 김상봉(2021b)를 참고하라.

$$\widehat{SR}(p_1, p_2, p_3, p_4) = \frac{\widehat{GL}(p_2) - \widehat{GL}(p_1)}{\widehat{GL}(p_4) - \widehat{GL}(p_3)} \quad (2)$$

여기서 $GL(p)$ 는 일반화된 로렌즈곡선이며 p_1, p_2, p_3, p_4 는 백분위수이다.

4.3.2 상대적 및 절대적 지니계수

지니계수(Gini coefficient)는 다양한 형태로 표시되고 산출된다. 절대적 불평등도를 나타내는 절대적 지니계수나 극화를 나타내는 극화지수를 상대적 지니계수와 비교할 필요가 있기 때문에 평균격차(mean difference)를 알아 본다. 평균격차(Δ)는 확률분포에서 나온 두 독립된 소득(Y) 등에 대한 값에 대한 차이의 절대값에 대한 평균으로, 이산적인 형태의 경우 다음과 같이 정의된다.

$$\Delta = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| \quad (3)$$

평균격차는 절대적 평균격차 또는 지니 평균격차라고도 한다. 절대적 평균격차는 절대적 불평등도를 측정하는 절대적 지니계수의 두 배가 된다.⁴⁾ 절대적 지니계수를 AG 로 나타내면 $\Delta = 2AG$ 가 된다.

$$AG = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| \quad (4)$$

절대적 지니계수는 로렌즈곡선과 소득 등의 균등분배 균등선인 45도 직선의 면적을 분모로 하고 균등선 이하를 분자로 하여 산출된다. 만약 완전히 균등하게 분배되는 경우에 값은 지니계수는 0이며, 완전 불평등할 경우에 1이 된다.

$$G = \frac{AG}{\mu} = \frac{1}{2n(n-1)\mu} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| \quad (5)$$

상대적 평균격차(relative mean difference)는 평균격차(Δ)를 그 집단의 평균(μ)으로 나눈 값이다. 상대적 평균격차도 상대적 지니계수의 두 배의 값이 된다. Gini(1912)는 상대

4) 이렇게 두 배가 되는 이유는 다음과 같다. 지니계수에서 두 가지 값의 차이값은 한 번만 계산하여 산출하지만, 평균격차는 정의에 의해 두 번 모두 계산하기 때문이다. 이는 상대적 지니계수에서 정의할 때도 그대로 적용된다.

적 소득불평등도를 측정하는 상대적 지니계수는 절대적 지니계수를 평균값으로 나누어 정규화한 값으로 정의하였다.

AG 에 관한 식과 G 에 관한 식에서, 절대적 지니계수는 상대적 지니계수에 평균을 곱한 값으로 정의된다. 따라서 경제가 성장하고 국민의 전체평균소득(μ)이 증가한다면, 절대적 지니계수는 증가하게 된다.

절대적 지니계수는 최빈층의 소득변화에 민감하게 변하지 않는다는 단점이 있다. 정부가 다양한 방식으로 소득 재분배 정책을 도입하여 소득불평등 정도가 개선되었다고 주장하기도 한다. 그런데 다른 불평등도에 대한 설명없이, 절대적 지니계수의 변화만으로 소득불평등 정도를 설명하는 경우에 최빈층의 소득점유율이 낮아지는 경우가 발생하기도 한다. 상대적 지니계수는 지니계수의 이러한 단점을 보완하고, 최빈층의 소득점유율 변화에 높은 가중치를 주게 된다.

불평등도를 나타내는 변수를 소득으로 가정하자. 어떤 사회구성원들의 소득이 y_1, \dots, y_n 이라면 i 번째 구성원의 소득점유율은 다음과 같으며 확률밀도함수로도 나타낼 수 있으며, 연속함수로도 정의할 수 있다.

$$z_i = \frac{y_i}{\sum_{i=1}^n y_i} \quad (6)$$

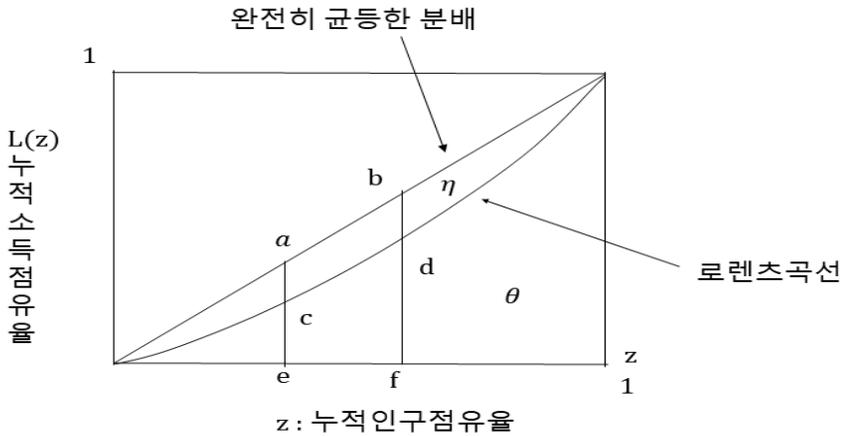
위에서 개인의 능력에 따라 $z \in [0, 1]$ 을 축으로 정하고, 구성원 i 에게 소득이 돌아갈 확률 또는 소득점유율이 z_i 이라고 하면 $f(z)$ 는 확률밀도함수로 해석할 수 있다.

$$AG = 1 - 2 \int_0^1 L(z) dz \quad (7)$$

$$L(z) \equiv \int_0^1 f(z') dz' \quad (8)$$

여기서 $L(z)$ 로 정의된 로렌즈곡선은 한 사회의 구성원에 대해 소득이 가장 낮은 사람부터 시작하여 높아지는 순서대로 배열하는 경우에, 하위 몇 퍼센트(%)에 속하는 사람들이 전체 소득 중에서 몇 퍼센트(%)를 차지하고 있는지에 대해 나타내는 점들을 모아 놓은 곡선이다. 로렌즈곡선을 나타내는 방법은 다음과 같다. 불평등도 분석의 대상인 소득에 대한 로렌즈곡선은 소득이 낮은 사람부터 높은 사람까지 순서대로 나타낸 후, 누적인구백분위율에 대한 누적소득점유율을 그래프로 나타낸 것이다.

<그림 4> 로렌츠곡선과 지니계수



따라서 상대적인 소득불평등도를 측정하는 상대적 지니계수는 절대적 지니계수를 평균값으로 나누므로 정규화한 값이 된다.

$$G \equiv 1 - \int_0^1 \frac{L(z)}{z} dz = \frac{\eta}{\eta + \theta} \tag{9}$$

앞에서 언급한 바와 같이, 상대적 지니계수는 z 가 작은 경우인 최빈층의 영향을 강조한 것이다. 만약 모든 사람들의 소득이 같아 $L(z) = z$ 이면, 상대적 지니계수는 0이 된다. 만약 극단적으로 모든 소득을 최상층이 모두 차지하면, $z = 1$ 인 점을 제외하고, $L(z) = 0$ 이므로 상대적 지니계수는 1이 된다. <그림 8>에서 절대적 지니계수를 산출할 때, 로렌츠곡선 위의 $abcd$ 의 넓이를 이용한다. 그러나 상대적 지니계수는 (면적 $cdef$)/(면적 $abef$)의 넓이의 비율을 이용한다. 만약 $z = 0.1$ 인 점에서 $L(z) = 0.025$ 라고 가정하면, 최빈층의 소득이 절대적 지니계수에 큰 영향을 미치지 못하지만, 상대적 지니계수는 $L(z)/z = 0.25$ 을 사용하게 되므로, 불평등지수에 큰 영향을 주게 된다.

4.3.3 타일의 엔트로피지수

타일(Theil, 1967)지수는 타일의 엔트로피지수(Theil's entropy index)로 불리며, 각 개인의 금융이나 소득 등을 모집단의 평균과 비교하여 불평등도를 산출한다. 엔트로피는 무질서한 정도 또는 어떤 사건이 발생했다는 정보의 가치를 의미한다(김상봉, 2021b).⁵⁾

5) 이러한 의미에 따라 엔트로피 극대화 방식이란 주어진 정보하에서 가급적 확률밀도함수를 평평하게 만든다는 것이다. 만약 우리에게 아무런 정보가 없다면 확률밀도함수는 수평이 된다. 만약 1계와 2계 조건이 주어졌다면 확률밀도함수는 정규분포함수가 된다. 엔트로피를 극대화시킨다는 것

어떤 사건 i 가 발생할 확률을 p_i 라 하면, p_i 가 낮을수록 해당 사건이 발생했다는 정보의 가치인 $h(p_i)$ 는 커진다. 따라서 확실히 발생하는 사건인 $p_i = 1$ 에서 정보의 가치는 무의미하므로 $h(p_i) = 0$ 이 된다. 또한, 독립적인 두 사건의 결합한 정보가치 $h(p_1 p_2)$ 는 개별 사건에서 정보가치의 합 $h(p_1) + h(p_2)$ 과 같다는 특징을 가진다. 이러한 정보가치의 속성은 $h(p_i) = -\ln p_i$ 로 표현된다. 엔트로피 H 는 어떠한 상황에서 정보가치의 기대값을 의미하므로, 각 사건의 정보가치를 발생확률 p_i 로 가중하여 계산된다.

$$H = \sum p_i h(p_i) = - \sum p_i \ln(p_i) \quad (8)$$

이제 소득분배에 응용하자. p_i 를 대체하기 위해, 전체 n 명의 소득에서 개인 i 가 차지하는 소득의 점유율을 y_i 라고 하면 소득불평등지수인 엔트로피지수 T 는 다음과 같다.

$$T = \sum \frac{1}{n} h\left(\frac{1}{n}\right) - \sum y_i h(y_i) \quad (10)$$

만약 모든 사람에게 동일한 소득이 $y_i = \frac{1}{n}$ 로 배분되었다면 엔트로피 값은 최대가 된다. 위 식 우변의 첫째 항이 이러한 것을 나타낸다. 그리고 둘째 항은 실제 엔트로피이다. 결과적으로 타일지수는 소득이 균등하게 분배되는 경우에 실제 엔트로피 값과의 차이를 통하여 불평등한 정도를 의미하게 된다. 타일의 엔트로피지수는 모든 소득이 균등하게 배분되면 0의 값을 가지고, 한 명에게 소득이 집중되면 $\ln n$ 의 값을 갖는다. 즉, 모든 개인이 동일한 평균소득을 가지면 엔트로피지수는 0이 되고, 한 명이 전체 소득을 가진다면 엔트로피지수는 $\ln n$ 이 된다.

$$T = \sum_{i=1}^n \left(\frac{y_i}{\sum_{j=1}^n y_j} \times \ln \frac{y_i}{y} \right) \quad (11)$$

타일의 엔트로피지수는 전체가 여러 하위집단(group)으로 구성된다면, 개별 하위집단에서 불평등도의 합으로 분해할 수 있다. 다음 식은 소득 전체를 m 개의 집단으로 구분

은 최소한의 정보인 우리에게 주어진 정보만을 이용한다는 것이고, 그 외의 정보는 이용하지 않는다는 것이다. 엔트로피가 낮으면, 우리에게 주어지지 않은 정보를 이용하여 확률밀도함수를 덜 평평하게 만들게 된다. 가장 보수적으로 확률밀도함수를 구하는 방식은 주어진 조건하에서 정의된 엔트로피를 극대화하는 방식으로 확률밀도함수를 구하는 것이다.

했을 때 타일의 엔트로피지수의 분해이다.

$$T = \sum_{k=1}^m s_k T_k + \sum_{k=1}^m s_k \ln \frac{\bar{y}_k}{\bar{y}} \quad (12)$$

여기서 s_k 는 집단 k 의 소득비율, T_k 는 집단 k 의 타일의 엔트로피지수, \bar{y} 는 전체 평균소득, \bar{y}_k 는 집단 k 의 평균소득을 의미한다. 즉, 타일의 엔트로피지수는 하위집단별 불평등도(within-group inequality)와 집단간 불평등도(between-group inequality)의 합으로 나타낼 수 있다. 각 집단이 동일한 소득규모를 가진다는 가정 하에서 집단간 불평등도가 나타난다.

타일의 엔트로피지수를 보다 보편적인 공식으로 바꾼 일반화된 엔트로피(Generalized Entropy, GE) 지수의 함수형태는 다음과 같이 나타낼 수 있다(Jenkins, 1995).

$$GE(a) = \frac{1}{a^2 - a} \left[\frac{1}{n} \sum_i \left(\frac{y_i}{\mu} \right)^a - 1 \right] \quad (13)$$

여기서 y_i 는 개인(또는 가구) i 의 소득을 의미하며, μ 는 평균소득을 의미하고 a 는 각 소득계층에 대한 가중치를 나타낸다. 일반적으로 엔트로피지수 함수에서 가장 많이 사용되는 a 값은 0, 1, 2이다. 만약 $a=0$ 이면 소득 등의 변수가 낮은 계층의 소득 변화에 큰 비중이 주어지고 $a=1$ 이면 모든 소득분포에 고르게 비중이 주어진다. $a=2$ 이면 소득이 높은 계층의 소득 변화에 보다 더 큰 비중이 주어진다.

만약 $a=0$ 이면, 로피탈(L'Hopital) 정리를 사용하여 다음과 같이 정리되며 이러한 경우에 로그평균편차(MLD)와 동일하게 된다.

$$GE(0) = \frac{1}{n} \sum_{i=1}^n \log \frac{\mu}{y_i} \quad (14)$$

만약 $a=2$ 이면, $GE(2)$ 는 다음과 같이 변동계수제곱(Squared Coefficient Variation, SCV)의 절반이 된다.

$$GE(2) = \frac{1}{2} \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{\mu} \right)^2 - 1 \right] = \frac{1}{2} SCV \quad (15)$$

y_i 는 집단 i 의 소득, $f(y_i)$ 는 집단 i 의 확률, k 가 집단의 수를 나타낼 때 일반화된 엔트로피지수는 다음과 같이 나타낼 수 있다(Shorrocks, 1980, 1984).

$$GE(y, a) = \begin{cases} \sum_{i=1}^k f(y_i) \left[\left(\frac{y_i}{\mu} \right)^a - a \right], & \text{if } c \neq 0, 1 \\ \sum_{i=1}^k f(y_i) \left(\frac{y_i}{\mu} \right) - \ln \left(\frac{y_i}{\mu} \right), & \text{if } c = 1 \\ \sum_{i=1}^k f(y_i) - \ln \left(\frac{y_i}{\mu} \right), & \text{if } c = 0 \end{cases} \quad (16)$$

4.3.4 앳킨슨지수

앳킨슨(Atkinson, 1970)지수는 기존의 불평등도를 측정하는 지수가 특정한 가치판단 또는 사회후생함수를 내포하고 있다. 그러나 그러한 측정의 배경이 되는 사회후생함수를 정의하기 힘들다고 보았다. 따라서 사회후생함수가 명확하게 잘 드러날 수 있는 성격을 가진 새로운 불평등도지수로 앳킨슨지수를 제안하였다.

앳킨슨지수를 이해하기 위하여 균등분배대등소득(equally distributed equivalent income)에 대해 알아보자. 1인당 평균소득이 \bar{y} 로 주어진 사회가 있다고 하자. 소득이 어떤 특정한 상태로 사람들에게 분배된 결과로 인해 \hat{W} 의 사회후생수준이 달성되었다고 가정하자. 모든 사람들에게 y_e 만큼의 소득을 균등하게 분배하여 주는 경우에 \hat{W} 의 사회후생수준이 달성될 수 있다고 하자. 그리고 이러한 소득수준 y_e 을 균등분배대등소득이라고 정의한다.

$$A = 1 - \frac{y_e}{\bar{y}} \quad (17)$$

앳킨슨지수는 0에서 1까지의 값을 나타낸다. 앳킨슨지수가 1에 가까울수록 보다 불평등한 상태를 뜻하며 소득분배가 완전히 균등하다면 $y_e = \bar{y}$ 이므로 $A = 0$ 이 된다. 똑같은 분배상태라고 하더라도, 보는 사람에 따라 균등분배대등소득이 다를 수 있으므로, 앳킨슨지수의 값도 여러 개가 나올 수 있다. $A = 1 - \left[\sum_{i=1}^n \left(\frac{y_i}{\bar{y}} \right)^{1-\epsilon} f_i \right]^{1/1-\epsilon}$ 이다. 여기서 y_i 는 i 번째 소득집단에 속하는 사람들의 평균소득이고, f_i 는 이러한 i 번째 소득집단에 속하는 사람들이 전체 인구에서 차지하는 점유율(비율)을 나타내며, ϵ 은 분배매개변수

(distributive parameter)이다. ϵ 에 따라 앳킨슨지수의 값이 달리 나오게 되며 값이 커질수록 저소득층의 소득에 부여하는 가중치가 커진다. 앳킨슨지수는 불평등성에 대한 명확한 가치판단을 전제로 산출된다는 특성을 가진다.

4.3.5 척도에 영향을 받지 않는 불평등도지수와 변동

척도에 영향을 받지 않는(scale invariant measures) 불평등지수로 변동계수(coefficient of variation) V , 상대평균편차(relative mean deviation) RD , 지니계수(Gini index) G , 엔트로피지수(entropy index) T , 로그분산(variance of logarithms) L 등이 있다. 변동계수 V 는 평균에 대한 표준편차의 상대적 크기로 $V = \frac{\sigma}{\mu}$ 로 정의된다. Schutz(1951)의 상대평균편차 또는 Schutz계수 RD 는 다음과 같이 정의된다.

$$RD = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu)}{2\mu} \quad (18)$$

여기서 Y_i 는 소득 등의 변수이다. 자연로그를 취한 변수값의 분산으로 계산되는 로그분산 L 은 다음과 같이 정의된다.

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (19)$$

여기서 y 는 변수의 자연로그를 취한 값이다. Allison(1978)은 이들 다섯 가지 불평등도지수 중에서 상대평균편차 RD 는 소득이 낮은 사람으로부터 높은 사람으로 이전될 때 불평등도지수는 감소하여야 한다는 Dalton(1920)의 이전법칙(principle of transfers)을 만족하지 않는 문제가 있다고 보았다. 자연로그분산 L 은 소득이 낮은 수준에서 이전법칙이 지켜지지만 소득이 아주 높은 수준(소득에 대한 기하평균의 2.718배보다 높은 수준)에서는 (상대적)으로 낮은 소득수준의 사람으로부터 더 높은 소득수준의 사람으로 이전하는 것은 오히려 불평등도지수를 감소시키는 문제가 있음을 증명하였다. Atkinson(1970)에 의하면 변동계수 V , 지니계수 G , 타일지수 T 는 이전법칙을 만족하는데 이들은 이전(transfer)에 대한 민감성(sensitivity to transfer)에서 서로 다른 특징을 가지고 있다고 보았다.

낮은 소득 Y_i 로부터 높은 소득 Y_j 로 h 만큼의 소득이 이전된다고 할 때 이전(transfer)의 변동계수를 V_1 , 이후의 변동계수를 V_2 라고 하면 변동계수의 변화는 다음과 같다

(Dalton, 1920).

$$V_2^2 - V_1^2 = c \times h \times (Y_j - Y_i) + c \times h^2 \quad (20)$$

여기서 c 는 양(+)의 상수이다. 이 식에 의하면 변동계수는 V 는 모든 소득수준에서의 이전에 동일한 민감성을 갖는다. 한편 후술할 지니계수는 각 변수의 소득수준보다 각 변수의 순위(rank)에 의하여 영향을 받기 때문에 위의 조건이 발생하는 경우의 지니계수 변화는 다음과 같다.

$$G_2 - G_1 = c' \times h \times (j - i) \quad (21)$$

여기서 c' 는 양(+)의 상수이다. 이 식에 의하면 지니계수는 관측치들이 집중되어 있어 동일한 규모의 소득이전에 의하여 순위변동이 크게 일어나는 중위수 부근의 변동에 민감하게 반응하며 양극단에서의 변동에는 민감하게 반응하지 않는 특징을 가진다. 극한이론(limit theory)을 적용하여 표현한 타일(Theil)의 엔트로피지수의 변화는 h 가 0의 값에 근사할 때 다음과 같다.

$$\Delta T = c'' \times h \times \ln\left(\frac{Y_j}{Y_i}\right) \quad (22)$$

여기서 c'' 는 양(+)의 상수이다. 이에 따르면 변동계수 V 의 변화가 소득이 이전되는 Y_i 와 Y_j 간의 소득격차에 의하여 표현되는 반면에 타일의 엔트로피지수 T 의 변화는 이들의 비율(ratio)에 의하여 표현된다. 따라서 T 는 소득수준이 낮을수록 지수가 소득이전에 민감하게 반응하게 된다.

4.4 극화지수의 종류

4.4.1 불평등지수와 극화지수의 차이

초기의 소득불평등의 연구에서 소득불평등에 대한 논의는 지니계수, 일반화된 엔트로피지수 등을 기준으로 한 소득불평등의 추세나 그 원인분석을 넘어 분포의 변동으로 그 초점이 다양화되었다. 이러한 원인은 중산층의 몰락(disappearing middle class)이라는 시대적 문제가 자리 잡고 있다. 흔히 중산층의 몰락 또는 쇠퇴 현상은 소득불평등의 심화 현상과 동일시되고 있으나, Wolfson(1994)는 두 가지 개념상의 상당한 차이를 가지고 있

다고 보았다. 또한, Esteban and Ray(1994)은 불평등지수 분석을 위해 자주 사용되는 지니계수 등은 분포의 변동을 반영하지 못하는 것을 지적하였다. 따라서 분포의 변동을 반영하기 위해 등장한 개념이 극화(polarization)이다. 극화 중에 n 개의 집단(group)으로 나누어지는 것을 다극화(polarization)라고 하고, 두 개의 집단으로 나누어지는 극화를 양극화(bi-polarization)라고 한다.

소득불평등과 극화 중 양극화의 개념의 차이를 이해하기 위해 극단적인 예로 살펴보자. 예를 들어, 한 명의 개인이 전체 소득을 차지하고, 한 명을 제외한 나머지 모든 사람들은 소득이 없어 소득의 합이 영(0)이 되는 경우에 가장 완벽한 소득불평등이 된다. 그러나 양극화의 개념에서 이러한 사회는 양극화되었다고 보기 힘들다. 왜냐하면 한 명의 개인을 제외하고, 나머지 모든 사람들이 소득이 없이 동질적이기 때문이다.

소득불평등과 양극화의 개념에서, 소득불평등에 대한 일반적인 척도는 주어진 모집단에서 개인 간 소득 차이를 하나의 수치로 나타낸다(Cowell, 1995). 즉 대부분의 소득불평등지수는 근본적으로 분포의 퍼짐(dispersion)의 정도를 나타낸다. 이러한 소득 불평등지수들의 바탕에는 소득의 순위가 불변이라는 조건도 포함하여 다른 조건이 같을 경우, 부자로부터 빈자에게 소득을 이전하는 것은 소득불평등도를 감소시키는 피구-달톤의 공리(Pigou-Dalton principle)인 이전원칙이 성립한다.

Esteban and Ray(1994)는 불평등지수들의 공리들은 전체 평균(global mean)으로 수렴(convergence)하고 그러한 평균으로부터 떨어진 만큼의 차이만을 강조하고 있으나, 국지적인 평균(local mean)에서 군집화(clustering)가 발생하는 현상을 식별하지 못하고 무시하는 단점을 가지고 있다. 만약 전체 평균에 수렴하게 되면 불평등도지수와 극화지수 둘 다 감소하지만, 군집화하면 불평등도지수는 감소하나 극화지수는 증가한다.

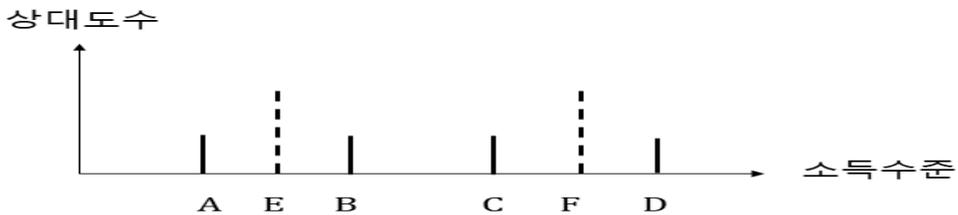
한 가지 소득분포에서 산출된 불평등지수가 매우 낮지만, 그 분포에서 분리된 몇 개의 극점들을 중심으로 군집화되어 있다고 하자. 그러한 극점들 사이의 거리가 가깝다고 하더라도, 소득분포는 다극화되었다고 할 수 있으며 다극화된 집단 사이에 갈등이 발생할 가능성은 높아진다. 지니계수와 같은 소득불평등지수들이 이러한 다극화를 포착할 수 없다.⁶⁾

이와는 달리 극화지수는 분포의 군집화(clustering)에 큰 비중을 두고, 분포의 변동이 발생하면 지수의 변동에 반영할 수 있다. 따라서 중산층(middle class)의 몰락과 같은 현상들이 바로 극화와 관련된 개념이다. 이러한 극화의 개념은 전통적인 소득불평등지수

6) Esteban and Ray(1994)와 Wolfson(1994)은 극화에 대해 연구를 시작하였다. Esteban, Gradin, and Ray(1999)는 연속적인 분포에 대해 여러 개의 극점들로 표시함에 따라 발생하는 근사(approximation)에 의한 오차를 교정하는 방법을 개발하고, 이러한 개발된 지수를 이용하여 소득의 다극화 추세를 국가간 비교에 응용하였다. Esteban et al.(1999)과 Duclos, Esteban, and Ray(2004)는 Wolfson의 W지수는 자신들이 개발한 지수의 한 가지 특별한 경우라는 것을 밝히고, 다양한 지수들 간 관계를 나타내고 있다.

공리 중 네 번째인 피구-달톤(Pigou-Dalton)의 이전원칙 공리를 위배할 수 있다. 예를 들어, 아래의 그림에서 A, B, C, 그리고 D의 네 가지 소득수준으로 이루어진 소득의 균일분포(uniform distribution)를 생각하자. 여기서 A와 B 사이와 C와 D 사이에 소득재분배를 시행한다고 하자. 소득재분배의 결과로 가운데 점선으로 표시된 E와 F의 두 소득수준으로 이루어진 소득분포로 변화되었다고 하자. 확실히 처음 네 가지 소득수준보다 전반적인 소득불평등은 감소하였다. 그러나 소득재분배 이전과 재분배 이후에 소득분포가 보다 균집화되었으며, 중산층은 사라졌다. 이러한 사회는 양극화되었다고 할 수 있다.

<그림 5> 소득불평등과 극화 예시



극화지수의 최종적인 목표는 그 사회의 내적 갈등의 수준을 나타내는 것이다. 극화지수 값이 높을 때, 그 사회의 집단간 충돌 가능성이 커진다. 반대로 그 값이 낮을 때, 그 사회의 집단간 충돌 가능성은 낮아진다. 그러한 충돌 가능성은 집단내의 동질성이 강할수록 커지며, 집단간의 이질성이 강할수록 커지고, 큰 규모를 가지는 집단의 수가 적을수록 커진다. 극화지수는 전통적인 불평등지수와 비교와 더불어, 사회 갈등 또는 사회불안정성을 직접적으로 나타낼 수 있게 된다.

사회의 다극화나 양극화 문제는 소득에만 국한되는 것은 아니다. 금융, 교육, 직업군, 종교 집단 등 소득 이외의 여러 특성변수들을 이용하여 다극화 문제를 다룰 수도 있다. 이러한 다극화나 양극화의 문제는 정치, 경제, 사회학 분야에서 양극화 또는 다극화 문제를 다룬 연구들은 매우 많다.

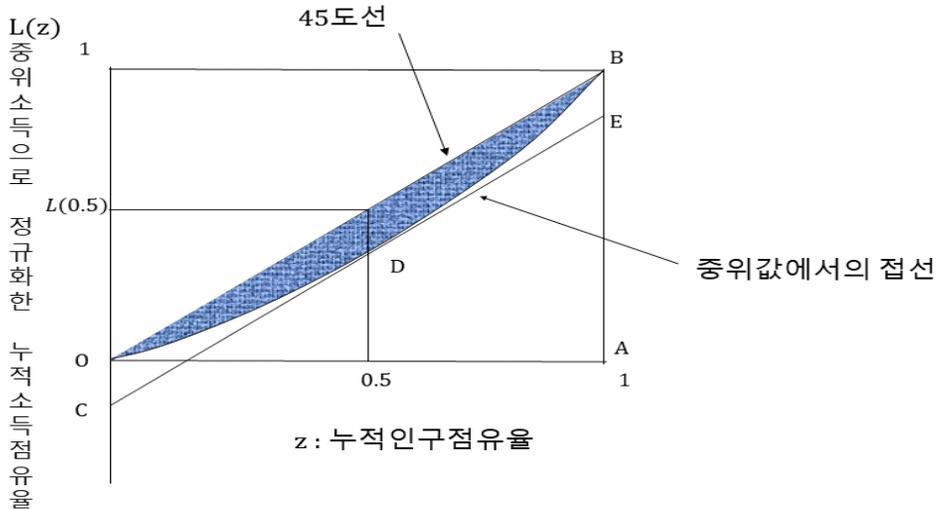
4.4.2 울프슨지수(W지수)

울프슨(Wolfson, 1994)지수 또는 W지수는 중산층의 소멸정도를 나타낸다. Wolfson(1994)은 체계적인 공리체계로부터 출발하는 것이 아니라, 상대적인 불평등도를 측정하는 것에 사용되는 로렌츠곡선에서 극화지수를 유도하였다. 중산층을 중위소득 부근의 인구계층으로 정의하였다. 이러한 중위소득으로부터 소득의 퍼짐 정도가 클수록 소득의 양극화가 커진다는 정의를 통하여 중산층의 몰락을 양극화(bi-polarization)로 보았다. 즉, 중위소득으로부터 소득의 퍼짐 정도인 분산이 커질수록 중산층 규모도 줄어든다는 가설을 전제로 하여, 중위소득과 상위계층과 하위계층간 소득 차이의 절대값을 이용

하여 산출한다.

로렌즈곡선을 이용한 양극화지수 유도방법을 살펴보자. 아래 그림에 로렌즈곡선의 형태의 양극화곡선(polarization curve)이 나타나 있다. 가로축의 누적인구점유율에 대응되는 세로축의 누적소득점유율의 평면에서, 세로축의 값은 로렌즈곡선의 경우 평균(μ)으로 나누어 정규화한다. 그러나 양극화지수는 중산층에 초점을 두기 때문에, 일반적인 로렌즈곡선이 나타나는 세로축의 값에 평균을 곱하고, 다시 중위값(median, m)으로 나누어 정규화한다는 점에서 차이가 있다. 세로축은 로렌즈곡선의 경우와 같이 상대적인 누적인구점유율로 일반적인 로렌즈곡선과 같다. 그리고 중위값 또는 중위인구에서 그은 수직선과 로렌즈양극화 곡선이 교차하는 점에서 접선을 그으면, 로렌즈양극화곡선과 접선 사이의 면적(W^* , 면적 $OCDEB$)이 Wolfson의 양극화지수가 된다.

<그림 6> W지수의 척도



T 를 45도선과 중위값에서의 접선 사이의 사다리꼴 면적($OCEB$)으로 정의하자. 접선과 로렌즈양극화곡선 사이의 면적(W^*)은 T 에서 반원모양의 면적을 빼는 것으로 계산할 수 있다. 그런데 이 반원모양의 면적은 다음과 같이 구할 수 있다. 그림에서 지니계수를 G 라고 하면, G 는 반원면적을 삼각형 OAB 면적으로 나누면 된다. 삼각형 OAB 의 면적은 $1/2$ 이므로 반원의 면적은 $G/2$ 가 된다. 따라서 접선과 로렌즈곡선 사이의 면적 $W^* = T - G/2$ 가 된다. 여기서 W^* 는 중앙값으로 정규화된 것이므로, 원래의 값으로 환원하기 위해 중앙값을 곱하고 평균을 나누면 된다. Wolfson(1994)에 따르면 $mtangent$ ($median tangent = m/\mu$)로 W^* 를 나누는 것과 같은 값이 된다.

사다리꼴 T 의 면적은 밑변과 높이로 구할 수 있다. 밑변은 중위값에서 그은 수직선과 45도선이 만나는 값에서 점점의 값을 뺀 값이 되고, 높이는 원점 O 에서 A 까지의 거리가 된다. 예를 들어, $L(0.5)$ 은 하위 50% 소득계층의 소득점유율이라 하면, 면적 T 는 $0.5 - L(0.5)$ 로 표시할 수 있다. 여기서 완전히 평등한 소득분포인 경우 0이 된다. 만약 인구의 절반이 0의 소득을 가지고, 나머지 인구의 절반이 2μ 의 소득을 가지는 완전히 양극화된 소득분포를 가지면, 면적 T 는 0.25가 된다. 이에 따라 Wolfson(1994)은 W^* 에 4를 곱한 $W = 4W^* = 2[2T - G]/mtangent$ 을 양극화지수로 제시하였다.

$$W = 4W^* = 2[2(0.5 - L(0.5) - G)] \frac{\mu}{m} \quad (23)$$

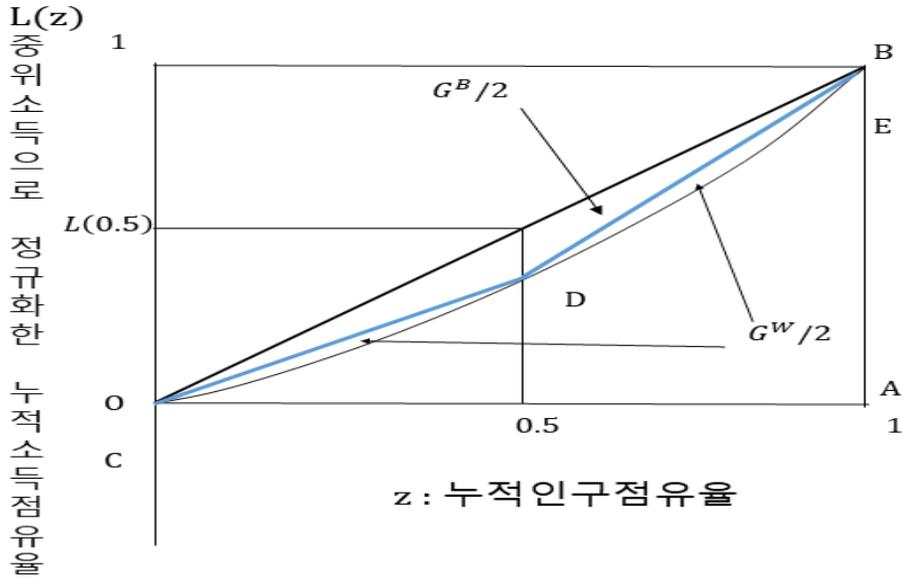
집단간 불평등도와 집단내 불평등도를 나누어보자. 사다리꼴 $OCEB$ 의 면적 T 는 $0.5 - L(0.5)$ 이고, 삼각형 ODB 면적의 두 배가 된다. 그림에서 삼각형 OEB 의 면적은 집단간(between group)의 지니계수를 의미한다. 또한, 두 집단 구성원의 소득범위가 겹치지 않기 때문에, 지니계수는 집단간(between group) 불평등도와 집단내(within group)의 불평등도로 정확히 나눌 수 있다. 이러한 W 지수는 다음과 같이 쓸 수 있다.

$$W = 2 \frac{\mu}{m} (G^B - G^W) \quad (24)$$

여기서 G^B 는 중앙값에 의해 분리되어 계산된 집단간(between group) 지니계수이고, G^W 는 집단내(within group) 지니계수이다.⁷⁾ 집단간 지니계수는 집단 사이의 이질성을 나타낸다. 즉, G^B 는 이질적인 집단 사이의 소외감을 나타내며 소외감은 집단간의 불평등도와 양(+)의 관계가 있다. 반면에 G^W 는 유사한 개인 사이의 동질성을 나타내므로 동질성은 집단내 불평등도와 음(-)의 관계가 있다.

<그림 7> 집단간과 집단내(between and within group) 요인의 구분

7) 이러한 로렌즈곡선 개념을 이용하면, Wolfson지수의 유도나 개선된 ER(Esteban and Ray)지수의 개념을 이해하는데 도움이 된다.



4.4.3 ER지수

ER(Esteban and Ray, 1994)지수는 서수적인 다극화 측정법이다. Esteban and Ray(1994) 류의 연구는 동질성-이질성 접근법(identification-alienation approach)에 기초하고 있다. 극화(polarization)는 서로 다른 집단(group)에 속하는 개인들 사이에 적대감(antagonisms)의 합계로 표시된다. 적대감은 같은 집단내에서 동질감(identification)과 더불어 다른 집단간에 소외감 또는 이질성(alienation)이 합해진 결과이고, 이것이 극화지수로 표현된다. 따라서 극화는 집단내에서 동질감이 증가하면 커지고, 적대감이 증가할 때도 커진다. 따라서 극화의 공리체계는 개별 집단의 중앙값 사이에 역진적인 소득이전이 발생하면 극화가 증가한다. 반대로, 중앙값으로 구분된 개별 집단에서 누진적인 소득이전이 발생하면 극화는 감소한다. Esteban and Ray(1994)는 상대적 불평등도지수가 가지고 있는 공리 중에 네 번 째인 피구-달톤의 이전원칙이 성립되지 않는 공리체계를 네 가지 원칙을 만족하고 다음과 같은 네 가지 특징을 가지는 ER지수를 제시하였다. 첫째, 개별 집단내에서 높은 동질성이 있다. 둘째, 집단과 집단 사이에 높은 이질성이 있다. 셋째, 적당한 규모의 집단은 소수로 존재한다. 넷째, 이러한 점은 집단들의 문제이며, 분리된 개인은 중요하지 않다.

$$ER(\alpha, \rho) = \sum_{i=1}^n \sum_{j=1}^n \pi_i^{1+\alpha} \pi_j |y_i - y_j|, \quad 0 \leq \alpha \leq 1.6 \quad (25)$$

여기서 n 은 특정 사회의 구성원수, y_i 는 집단 i 의 소득의 자연로그, π_i 는 상대규모를 나타낸다. ρ 는 n 개의 극점들로 구성된 집합으로, $\rho = \{y_0, y_1, \dots, y_n; \pi_0, \pi_1, \dots, \pi_n\}$, $y_0 < \dots < y_n$ 이다.

단순 ER지수는 두 행태함수(behavioral function)인 동질성함수(identification function)와 이질성함수(alienation function)에 의존한다. $|y_i - y_j|$ 는 개인들 사이의 소외감 또는 거리를 나타내는 이질성함수(alienation function)이다.

π_i^α 는 집단 i 의 개별 구성원이 자신의 집단에 대해 가지는 동질성을 나타내는 동질성함수(identification function)이며, α 는 동질성의 중요성을 나타내는데, 극화지수의 민감도(polarization sensitivity)를 나타내며, 극화지수가 불평등지수와 구별되게 하는 모수값이다.

어떠한 개인이 자신과 같은 집단에 속한 개인들의 수가 많을수록 그 개인이 느끼는 동질감은 커진다. 따라서 동질감과 소외감으로 곱해져 있는 적대감 $\pi_i^\alpha |y_i - y_j|$ 는 집단 i 의 구성원이 집단 j 의 구성원에게 느끼는 유효적대감(effective alienation)이 된다. 즉, 극화지수는 동질성과 이질성의 증가함수이며, ER지수는 두 함수를 곱하여 모든 그룹에 대해 가중평균한 값이 된다.⁸⁾ 소득집단간 괴리가 커지면, 집단간 갈등은 커지며, 집단내 구성원들의 동질성이 강화될수록 그응집력에 의해 러한 정도는 더욱 커진다.

이러한 ER지수는 특수한 경우의 지니계수와 매우 비슷하다. 만약 $\alpha = 0$ 이면, 지니계수에서 등장하는 평균격차 Δ (절대적 지니계수의 2배)와 유사하다. α 의 값이 클수록 극화지수는 상대적 불평등도지수와 달라지게 되며, 극화의 주어진 공리체계를 만족시키기 위해서 0보다 크지만 1.6 이하의 범위에 있어야 한다(Esteban and Ray, 1994).

보다 일반적인 이산형인 함수로 나타내자. 특정 사회의 구성원 수 n , 소득의 자연로그 y , 개인의 분포함수 $(\pi, y) = (\pi_1 \dots \pi_n; y_1 \dots y_n)$ 의 집합이라고 하자. 단, π_i 는 개인 y_i 의 소득집단에 속한 사람들의 척도이고 모집단에서 차지하는 비율이다. (π, y) 와 관련된 총인구를 $\sum_{i=1}^n \pi_i$ 을 D 라고 할 때 양극화지수 ER 은 D 에서 양(+의 실수의 집합 R_+ 로 정의한다.

$$ER(\alpha, \rho) = \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j T(I(\pi_i), A(\delta(y_i, y_j))) \quad (26)$$

8) 다극화지수가 집단내 동질성의 증가함수라고 하였다. 따라서 집단내의 소득이전은 바로 다극화지수의 값을 증가시킨다. 다극화지수와 관련하여 피구-달톤의 공리는 집단간에는 성립하지 않지만, 집단내에서는 여전히 성립한다.

여기서 $T(I, A)$ 를 유효반감함수이다. y_i 가 y_j 에 느끼는 유효적대감, $I(\pi_i)$ 는 y_i 가 느끼는 동질성(indentification)을 느끼는 동질성함수, $A(\delta(y_i, y_j))$ 는 소외감의 크기를 나타내는 이질성함수라고 하자. 즉, $\frac{\partial T}{\partial I} > 0$ 이고 $\frac{\partial T}{\partial A} > 0$ 이다. 즉, 소득수준이 y_i 인 개인은 소득수준이 y_j 인 개인에게 이질감을 크게 느낄수록, 적대감의 크기는 커진다. 동시에 자신이 속한 집단에 대해 보다 강하게 동질감으로 느낄수록 역시 적대감의 정도는 커진다. 이러한 상황에서 한 사회 내의 소득분포가 얼마나 다극화되어 있는가의 정도는 그 사회 내의 모든 구성원들이 느끼는 적대감의 합계로 표시된다. Esteban and Ray(1994) 모형의 두 가지 단점은 모집단이 이미 n 개의 소득집단으로 구분되어 있어야 한다. 또한, 집단 i 에 속한 소득수준을 가지는 개인이 느끼는 동질성은 그 개인이 속하는 집단이 모집단에서 차지하는 비율 π_i 로 표시될 수 있어야 한다는 점이다.

앞의 ER지수를 연속함수로 표현하면 다음과 같다.

$$ER(F) = \iint T(I(x|F), A(x, y))dF(x)dF(y) \quad (27)$$

이산적인 부분과 마찬가지로, 구체적인 함수의 형태로 나타내기 위해 어떻게 $T(\cdot)$, $I(\cdot)$ 과 $A(\cdot)$ 함수를 설정해야 한다. Esteban and Ray(1994)은 사전적으로 정해진 소득 집단의 수와 개인이 느끼는 동질성은 그 개인이 속하는 집단이 모집단에서 차지하는 비율로 표시되어야 한다는 제약을 어느 정도 완화시켰다. 우선 서로 다른 세 소득 수준, x , y , z 가 있다고 하자. 임의의 양수 E 에 대해 $(y-x) < E$ 일 때 서로 동질성을 느낀다고 하자. 여기서 구간 $[0, E]$ 에서 정의되는 가중치 함수 $\omega(E)$ 를 생각하고, $\omega'(\cdot) < 0$ 이며 $\omega(E) = 0$ 이라고 하자. 여기서 주어진 소득분포 F 하에서 동질성함수를 다음과 같이 정의할 수 있다.

$$I(x, F) = \int_{\{y: (y-x) < E\}} \omega(|y-x|)dF(y) \quad (28)$$

즉, 이러한 점은 집단내에서 동질감을 느끼는 사람들 사이에 동질감의 정도(가중치)를 합하는 것이다. 이러한 상황에서 두 소득 수준 x 와 y 사이에 이질성의 정도는 다음과 같다.

$$A(x, y) = \max\{|x-y| - E, 0\} \quad (29)$$

만약 특정 개인이 다른 소득집단에 대해 느끼는 적대감의 정도가 동질성함수와 이질성함수의 곱으로 표시된다면, 다극화지수는 다음과 같이 정리된다.

$$ER(\alpha, F) = \iint \left[\int_{\{y: (y-x) < E\}} \omega(|y-x|) dFy \right]^\alpha \max\{|x-z| - D, 0\} dF(x) dF(z) \quad (30)$$

이산형과 마찬가지로, α 는 동질성함수의 중요성을 나타내는 양(+)의 상수가 된다. 만약 동질성함수가 극화지수에서 전혀 중요하지 않다면 $\alpha = 0$ 이 되며, 소득불평등도를 나타내는 절대적 지니계수의 2배(평균격차)가 된다. Esteban and Ray(1994)의 극화지수가 지니계수와 구별되는 것이 동질성함수가 존재하기 때문이다.

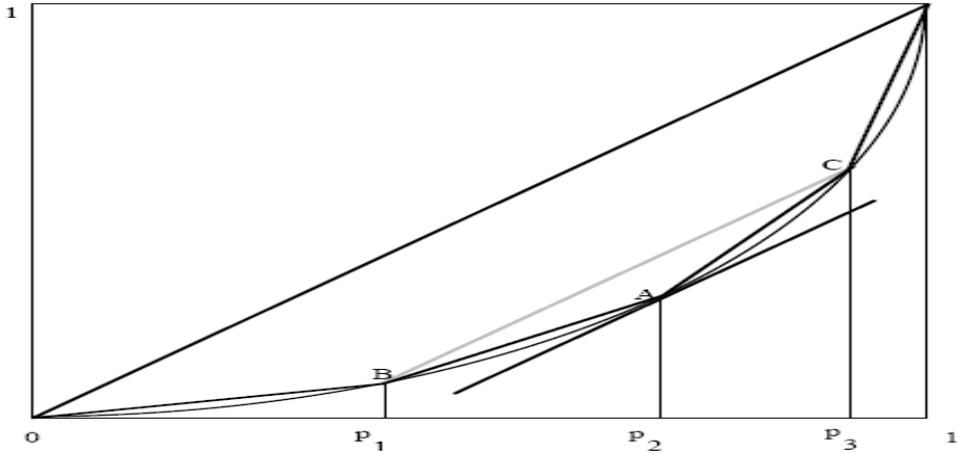
또한, 두 집단간 갈등이 발생할 가능성은 두 집단의 소득격차로 나타나는 이질성의 정도에 따라 다르지만, 동질성을 느끼는 집단의 규모에 따라서도 다르다. 만약 모든 소득 변수의 값들이 구간 $[0, E]$ 에 존재한다면, 극화지수의 값은 0이 된다. 왜냐하면 만약 E 값이 충분히 작다면, 모든 개인들의 소득수준이 완전히 같지 않더라도 전체 집단의 모든 구성원들은 서로 동질감을 느끼기 때문이다. 또한, 문제는 가중치 함수의 형태와 가중치를 어떻게 줄 것인가 지에 대한 문제가 남는다.

4.4.4 EGR지수 : 일반화된 ER지수의 기하학적 의미

Esteban, Gradin, and Ray(1999, 2007)는 위에서 정의된 ER지수를 개선하였다. ER지수는 전체 모집단이 이미 여러 개의 집단으로 나누어져 있다는 가정하고 있다. 따라서 ER지수는 성, 인종, 종교 등과 같이 이미 외생적으로 집단이 분리되어 있다면, 극화정도를 추정할 수 있다. 그러나 외생적으로 집단이 분리되지 않는 경우에 분리하는 기준이 없다. 예를 들어, 많은 인구의 소득분포가 연속적으로 존재할 때 얼마나 많은 집단으로 분리해야 할지 또는 그 집단간의 분리는 어떤 기준으로 할 지에 대한 문제가 존재한다. Esteban, Gradin, and Ray(1999, 2007)는 집단의 수가 결정되면, 집단간의 규모를 결정하는 방법에 대한 원칙을 제시한다.

Esteban, Gradin, and Ray(1999, 2007)의 방법을 그래프를 이용하여 쉽게 접근해 보자. 예를 들어, 소득다극화를 추정하려면 집단의 수가 정하고, 집단내 불평등도를 추정해야 한다. 지니계수는 로렌즈곡선을 통해 추정되며, 집단이 겹치지 않도록 구성된 불평등도의 추정치이다. 따라서 지니계수는 겹치지 않는 집단간의 불평등도와 집단내 불평등도로 구성된다. 집단들의 위치를 설정하는 원칙은 집단내 불평등도를 최소화하도록 결정하는 것이다. 예를 들어, 다음과 같은 4극 분포의 경우에 세 개의 극점, 네 개의 극화에 따른 분포에 대응하는 로렌즈곡선이 존재한다. 아래의 그림에서 로렌즈곡선과 B , A , C 3개의 극점을 연결하는 두 번째 로렌즈곡선이 나타난다.

<그림 8> 세 개의 극점, 네 개의 극화와 두 개의 로렌츠곡선



자료: Esteban, Gradin, and Ray(1999, 2007)

세 번째 극점(spike)을 찾는 방법을 설명하자. 먼저 p_1 과 p_3 가 임의로 정해진 두 개의 극점이라 하면, 오차를 최소화하는 p_2 의 위치를 찾아야 한다. 원호 BA 가 집단 2의 내부 불평등도를 나타내고 원호 AC 가 집단 3의 내부 불평등도를 의미하므로, 두 면적의 합을 최소화하는 방법을 찾아야 한다. 따라서 삼각형 BAC 면적을 최대화하여야 하며, BC 선과 기울기가 같아지는 로렌츠곡선상의 접선을 그리면 된다. 따라서 오차를 최소화하는 최적의 극점은 두 집단 전체소득의 평균이 된다. 이러한 점을 반영한 Esteban, Gradin, and Ray(1999, 2007)의 일반화된 ER 지수는 다음과 같다.

$$EGR(F; \alpha, \beta, \rho) = ER(\alpha, \rho) - \beta \epsilon(F; \rho) \quad (31)$$

여기서 β 는 ER 지수를 산출할 때 발생하는 오차에 두게 되는 가중치를 나타내는 모수 값이다. ρ 는 n 개의 극점으로 구성된 집합이다. 또한 $\epsilon(f, \rho)$ 는 연속된 소득분포이며, n 개 집단의 분포에 대해 근사하면서 발생하는 오차를 의미한다. 위 식에 최적의 ρ^* 값을 대입하여 구한 최적오차는 다음과 같다.

$$\epsilon(F; \rho^*) = G(F) - G(\rho^*) \quad (32)$$

최적오차를 ER 지수에 대입하면, 일반화된 ER 지수인 EGR 식은 다음과 같다.

$$EGR(F; \alpha, \beta, \rho^*) = ER(\alpha, \rho) - \beta [G(F) - G(\rho^*)] \quad (33)$$

Esteban, Gradin, and Ray(1999, 2007)는 소득이 두 집단으로 나누어지는 양극화를 가정하고, 일반화된 ER지수의 유도과정을 보여준다. 특히, $\alpha = \beta = 1$ 인 특별한 경우에 일반화된 ER지수는 원래의 Wolfson지수와 같다고 증명하였다. 즉, Wolfson지수 $W = 2 \frac{\mu}{m} (G^B - G^W)$ 에서 G^B 는 $G(F)$ 를, G^W 는 $G(\rho^*)$ 를 의미하며 $G(\cdot)$ 은 괄호안의 분포에 해당하는 지니계수이다.

극화에 대한 많은 연구들은 소득계층(소득집단)을 양분화하여 사용한다. 왜냐하면 EGR지수를 적용할 때 소득계층에 대한 구분에 대한 명확한 근거가 없으며, 소득계층이 3개 이상으로 구분되면 계산하기 어렵기 때문이다. 따라서 다극화를 산출하는 ER지수나 EGR지수는 평균소득을 기준으로 양분화된 지수를 주로 산출하므로, 양극화(bi-polarization)를 계산한 수치가 된다.

4.4.5 EGR지수의 통계적 접근

여기서 Esteban, Gradin, and Ray(1999, 2007)에 따라, 확률밀도함수 f 는 다음과 같은 n 개의 극점들로 구성된 집합(ρ)으로 나타낼 수 있다. 확률변수인 소득 y 는 유한 폐구간 $[a, b]$ 에서 확률밀도함수 f 로서 표시하자. 소득 y 는 기대소득인 평균(μ) 1로 정규화(normalize)되었다고 하자.

$$\rho = \{y_0, y_1, \dots, y_n; \pi_0, \pi_1, \dots, \pi_n; \mu_0, \mu_1, \dots, \mu_n\}, \quad a = y_0 < \dots < y_n = b \quad (34)$$

$\pi_i = \Pr(y_{i-1} < y < y_i) = \int_{y_{i-1}}^{y_i} f(y)dy, \quad i = 1, \dots, n$ 으로 확률변수값이 y_{i-1} 과 y_i 구간에 놓일 확률이다.

$\mu_i = E(y | y_{i-1} < y < y_i) = \frac{1}{\pi_i} \int_{y_{i-1}}^{y_i} yf(y)dy$ 로, 각 구간 내에서의 조건부 평균값을 의한다. 따라서 Esteban and Ray(1994)의 ER지수는 다음과 같다.

$$ER(\alpha, \rho) = \sum_i \sum_j \pi_i^{1+\alpha} \pi_j |\mu_i - \mu_j| \quad (35)$$

이러한 단순 ER지수는 원래의 소득분포에 대해 여러 개의 극점을 표시하는 과정에서 문제가 발생한다. 즉, 연속확률밀도함수를 n 개의 극점으로 표시하는 과정에서 오차가 발생할 수 있다. 오차 $\epsilon(F; \rho)$ 는 원래의 확률밀도함수의 형태와 극점들을 표시하는 방식에 따라 달라진다. Esteban, Gradin, and Ray(1999, 2007)는 단순 ER지수가 근사에 의해 극점

으로 표시하는 과정에서 발생하는 오차를 조정하기 위해 개선된 다극화지수를 다음과 같은 만들었다.

$$EGR(F; \alpha, \beta) = ER(\alpha, \rho) - \beta(F, \rho) \tag{36}$$

여기서 β 는 단순 ER 지수를 산출하는 과정에서 발생하는 근사오차에 두는 가중치를 나타내는 모수값이다. 이제 소득분포를 어떻게 n 개의 극점(spike)들로 나타내야 하는 지에 두 가지 문제가 존재한다. 문제는 극점들의 수와 그 극점들의 위치(location)이다. 먼저, 극점들의 수는 모형에서 외생적으로 결정된다. 극점들은 어떠한 기준에 따라 내생적으로 결정되는 것이 아니라, 연구자의 자의적인 판단으로 구분되는데, 연구자들 모두 동의하는 기준은 아직 없다. 다만, 집단(group)내 동질성이 어느 정도 유지되어야 한다. 즉, 그룹 내의 퍼짐(산포도)는 전체 퍼짐과 비교하면 상대적으로 작아야 한다. 따라서 근사에 따른 오차를 정의하고 오차를 최소화하는 ρ 를 선택해야 한다.

$$\epsilon(F, \rho) = 0.5 \sum_i \int_{y_{i-1}}^{y_i} \int_{y_{i-1}}^{y_i} |x - z| F(x) F(z) dx dz \tag{37}$$

이는 집단내에서 두 소득 값의 차이를 평균한 것과 같다. 앞에서 기하학적으로, n 개의 극점의 집합으로 ρ 는 원래의 로렌즈곡선을 n 개의 선분으로 이루어진 선형 로렌즈곡선으로 변형시킨 것이다. 위 식에서 오차를 최소화하는 것은 원래의 로렌즈곡선과 선분으로 이루어진 로렌즈곡선 사이의 면적을 최소화하는 것과 같다. 따라서 최적점 ρ^* 에서 다음과 같다.

$$\epsilon(F; \rho) = G(F) - G(\rho^*) \tag{38}$$

$G(\cdot)$ 는 팔호 안의 분포에 해당하는 지니계수이다. 위의 식을 $EGR(F; \alpha, \beta) = ER(\alpha, \rho) - \beta(F, \rho)$ 에 대입하면 다극화지수의 값은 다음과 같다.

$$EGR^*(F; \alpha, \beta) = ER(\alpha, \rho^*) - \beta(F, \rho^*) \tag{39}$$

Esteban, Gradin, and Ray(1999, 2007)는 오차를 최소화하는 최적 ρ^* 는 다음의 조건을 만족함을 밝혔다.

$$y_i^* \int_{y_{i-1}^*}^{y_{i+1}^*} F(x) dx = y \int_{y_{i-1}^*}^{y_{i+1}^*} x F(x) dx, \quad i = 1, \dots, n-1 \quad (40)$$

이를 재정리하면 다음과 같다.

$$y_i^* = \lambda \mu_i^* + (1-\lambda) \mu_{i+1}^*, \quad \lambda = \frac{\pi_i^*}{\pi_i^* + \pi_{i+1}^*} \quad (41)$$

따라서 극점들의 위치를 선택할 때, 서로 인접한 두 구간을 분할하는 소득수준은 두 구간의 조건부 평균값들의 가중평균이어야 하며, 이때 오차가 최소화된다.

4.4.6 EGR지수 : 양극화지수의 예

불평등이나 극화의 연구대상에서 많이 사용되는 예는 한 사회의 양극화이다. 분석의 대상이 되는 특성변수가 지지정당이나, 노조 및 비노조 등 쉽게 양분화되는 경우이다. 소득이라는 특성변수에서도 중산층이 몰락하는 경우에 빈민층, 중산층, 부유층으로 삼극화하는 것보다 빈자와 부자라는 양극화하는 것이 상대적으로 더 설득력있다. 한 사회의 양극화의 정도를 분석하기 위해서 극점을 두 개로 정하면 된다. 이 경우에 소득 분할점은 하나이고, 이를 y 라고 하고, 이 분할점에서의 분포함수값을 π 라고 하면 다음과 같다.

$$\pi = \int_a^y f(x) dx \quad (42)$$

π 에 대응하는 로렌츠곡선의 세로축 길이를 $L(\pi)$ 라고 하면 $\mu_1 = L(\pi)/\pi$ 이고 $\mu_2 = \{1 - L(\pi)\}/(1 - \pi)$ 이다. 이를 이용하며 단순 $ER(\alpha, \rho) = \sum_i \sum_j \pi_i^{1+\alpha} \pi_j |\mu_i - \mu_j|$ 은 다음과 같다.

$$\begin{aligned} ER(\alpha, \rho) &= (\mu_2 - \mu_1) \{ \pi^{1+\alpha} (1-\pi) + (1-\pi)^{1+\alpha} \pi \} \\ &= \{ \pi^\alpha + (1-\pi)^\alpha \} \{ \pi - L(\pi) \} \end{aligned} \quad (43)$$

연속변수의 분포를 극값으로 표시하여 발생하는 오차는 다음과 같다.

$$\epsilon(F, \rho) = G - \{ \pi - L(\pi) \} \quad (44)$$

따라서 $EGR(F; \alpha, \beta, \rho) = ER(\alpha, \pi) - \beta\epsilon(F; \rho)$ 은 다음과 같다.

$$EGR(F; \alpha, \beta, \rho) = \{\pi^\alpha + (1 - \pi)^\alpha\} \{\pi - L(\pi)\} - \beta[G - \{\pi - L(\pi)\}] \quad (45)$$

위 식에서 다극화지수는 π 값 또는 y 값에 의존한다. 여기서 근사에 의한 오차를 최소화하기 위하여, 로렌즈곡선과 45도선 사이의 거리를 최대화하는 y 값을 찾아야 한다. 즉, $\{\pi - L(\pi)\}$ 을 최대화하는 π 값을 찾아야 한다. 만약 로렌즈곡선이 강볼록(strictly convex) 함수라면, 이때 유일한 해가 존재하며 이는 $y = \mu$ 가 된다. 즉, 근사에 의한 오차를 가장 최소화시키는 최적의 소득분할점은 전체 소득의 평균값이 된다. 한편 평균편차(mean deviation, D)가 다음과 같이 표시된다.

$$D = \frac{1}{2\mu} \int |\mu - y| f(y) dy = \pi_\mu - L(\pi_\mu), \quad \pi_\mu = F(\mu) \quad (46)$$

$\mu = 1$ 로 정규화하였으므로, 오차를 최소화한 상태에서 EGR식은 다음과 같다.

$$EGR(F; \alpha, \beta) = \{\pi^\alpha + (1 - \pi)^\alpha\} D - \beta[G - D] \quad (47)$$

예를 들어, 분석을 위해 근사오차에 두는 가중치인 $\beta = 1$ 로 고정시키고, α 값을 변화시키면서 양극화지수 값을 추정하고, 각종 불평등지수 등과 비교가능하다.

EGR식은 Wolfson(1994)의 지수와 밀접하다. 이러한 관계를 증명하기 위해, y 값으로 중앙값(median)을 선택하면 $y = m$ 이 된다. 따라서 $\pi = 0.5$ 이므로 EGR식은 다음과 같다.

$$EGR(F; \alpha, \beta) = (2^{1-\alpha} + \beta) \{0.5 - L(0.5)\} - \beta G \quad (48)$$

$\alpha = \beta = 1$ 로 제약을 부과하면 다음과 같다.

$$EGR(F; 1, 1, m) = 2\{0.5 - L(0.5)\} - G = 0.5mER^W(F) \quad (49)$$

여기서 ER^W 는 양극화에 대한 Wolfson의 W지수를 나타낸다.⁹⁾ Wolfson의 W지수는

9) Wolfson(1994)의 양극화지수는 다음과 같은 과정으로 도출된다. ① 소득변수의 누적분포함수(CDF)의 축을 전치시켜 가로축이 모집단의 누적비율(population percentile)이 되게 하고 세로축이 소득이 되게 한다. ② 소득변수는 평균이 아닌 중앙값 정규화한다. ③ 가로축의 0.5가 되는 점을 표시한다. ④ 가로축을 곡선과 0.5를 표시한 점이 만날 때까지 상향으로 이동시킨다. ⑤ 0.5로 분할된 곡선의 왼쪽부분을 가로축을 중심으로 대칭이동시킴으로써 함수값이 모두 양이 되도록 한

개선된 ER 지수의 하나의 특별한 경우가 된다.¹⁰⁾

4.4.7 Duclos, Esteban and Ray(2004) 극화지수(DER)

Duclos, Esteban and Ray(2004)의 극화지수(DER)는 다음과 같이 표현할 수 있다.

$$DER(\alpha) = \int \int f(x)^{1+\alpha} f(y) |y-x| dy dx \quad (50)$$

여기서 $f(x)$ 는 x 의 밀도함수이다. 이러한 지수를 추정하기 위해 이산적인 형태가 다음과 같이 사용된다.

$$DER(\alpha) = \frac{\sum_{i=1}^n w_i f(y_i)^\alpha a(y_i)}{\sum_{i=1}^n w_i} \quad (51)$$

정규화된 DER은 다음과 같이 정의된다.

$$\overline{DER}(\alpha) = \frac{DER(\alpha)}{2\mu^{(1-\alpha)}} \quad (52)$$

다. 이 0.5에서 꺾인 모양의 함수는 바로 중앙값으로부터 소득분포가 얼마나 퍼져있는가를 나타낸다. ⑥ 0.5를 출발점으로 좌우 방향으로 곡선을 따라 적분하면 Foster and Wolfson(1992)이 말하는 양극화곡선(polarization curve)가 생긴다. 이 양극화곡선과 가로축 사이의 거리를 0부터 1까지 적분하면 바로 Wolfson의 양극화지수 값이 된다. ⑦ Wolfson(1994)은 소득불평등지수와 양극화지수의 쌍대성(duality)을 보이는 과정에서 이 양극화지수의 값이 전통적인 로렌즈곡선을 이용하여 $\{(0.5 - L(0.5)) - 0.5G\} \times (\mu/m)$ 로 표시됨을 증명하였고, 이는 0부터 0.25 사이의 값만을 가질 수 있음을 보였다. 따라서 지니계수와 같은 스케일의 지수를 만들기 위해서 임의로 4를 곱하여 최종적인 양극화지수 $ER^W(F)$ 로 사용하였다. 한편 평균이 1이 되도록 정규화하면, 위의 식 우변은 $0.5 \times m \times 4 \times \{(0.5 - L(0.5)) - 0.5G\} / m$ 이 되고 이를 정리하면 바로 좌변의 ER의 개선된 양극화지수가 된다.

10) 위의 식에서 m 은 평균이 1로 표준화된 분포에서의 중앙값을 말하며 표준화되지 않은 분포에서는 m/μ 로 조정해야 한다.

$$\text{여기서 } a(y_i) = \mu + y_i \left[\left(\frac{2 \sum_{j=1}^i w_j - w_i}{\sum_{i=1}^n w_i} \right) - 1 \right] - \left(\frac{2 \sum_{j=1}^i w_j y_j + w_i y_i}{\sum_{i=1}^n w_i} \right) \text{이다.}$$

가우스 커널 추정치(Gaussian kernel estimator)는 밀도함수를 추정할 때 사용된다. Araar(2008)은 Duclos, Esteban and Ray(2004) 극화지수를 인구집단에 따른 분해를 다음과 같이 제안하였다.

$$P_g^{DER} = \sum_g \psi_g^{1+\alpha} \phi_g^{1-\alpha} R_g P_g + \tilde{P} \tag{53}$$

여기서 우변의 첫항은 집단내(within group) 불평등도이고, 둘째항은 집단간(between group) 불평등도를 나타낸다. $R_g = \frac{\int a_g(x) \pi_g(x) f(x)^{1+\alpha} dx}{\psi_g \int a_g(x) f_g(x) f_g(x)^{1+\alpha} dx}$ 이다. ψ_g 는 그룹 g 의 인구점유율이고 ϕ_g 는 소득점유율이다. $\pi_g(x)$ 는 집단 g 에 속하고 소득 x 를 가지는 개인의 국지적 비율이다. \tilde{P} 는 집단내 극화나 불평등도가 무시될 때 DER 극화지수이다.

4.4.8 Inaki(2008) 극화지수

인구가 $n_i > 0$ 의 크기로 N 개의 집단으로 나누어진다고 하자. 밀도함수를 $f_i(x)$, 집단 i 의 평균은 μ_i , 인구점유율을 π_i 로 표시하자. μ 는 전체 평균이다. 따라서 $\int f_i(x) = 1$, $\sum_{i=1}^N \pi_i \mu_i = \mu$ 이고 $\sum_{i=1}^N \pi_i = 1$ 이다. Inaki(2008)의 사회적 극화지수는 다음과 같이 정의된다.

$$P(F) = \sum_{i=1}^N [P_W(i, F) + P_B(i, F)] \tag{54}$$

여기서 $P_W(i, F) = \mu^{\alpha-1} \pi_i^{2+\alpha} \iint f_i^{1+\alpha}(x) f_i(y) |x-y| dy dx$ 이고

$$P_B(i, F) = \mu^{\alpha-1} \pi_i^{1+\alpha} \left[(\mu - \pi_i \mu_i) \int f_i^{1+\alpha}(x) dx + (1 - \pi_i) \int f_i^{1+\alpha}(x) x dx \right]$$

사회적 극화지수 $P(F)$ 는 집단 요인으로 분해할 수 있으며 다음과 같은 결과가 나타

난다.

추정된 하위집단 i 의 인구점유율 π_i , 추정된 하위집단 i 의 소득비율 $\pi_i \mu_i / \mu$, 추정된 하위집단 i 의 $P_W(i, F)$ 지수, 추정된 하위집단 i 의 $P_B(i, F)$ 지수, 추정된 $P_W = \sum_i P_W(i, F)$ 지수, 추정된 $P_B = \sum_i P_B(i, F)$ 지수, 추정된 전체 지수 P_F 이다.

4.4.9 TW지수

Wang and Tsui(1998)가 제안한 지수는 다음과 같다.

$$TW = \left(\frac{\theta}{n} \right) \sum_{j=1}^k f(y_j) \left| \frac{\bar{y}_j - m}{m} \right|^r \quad (55)$$

여기서 \bar{y}_j 는 집단 j 의 평균, θ 는 양(+)의 상수, n 은 전체 표본의 수, k 는 집단의 수, $f(y_j)$ 는 전체에서 차지하는 집단 j 의 비중, m 은 중위소득, r 은 (0, 1)의 수를 나타낸다. Tsui and Wang(1998)은 Wolfson의 W지수를 기초로, 극화의 증가와 분포확산의 증가라는 두 개의 공리체계를 따로 제시하며 새로운 지수를 개발하였다. 그러나 TW지수는 W지수나 ER지수와 큰 차이가 없다.

4.4.10 ZK지수

Zhang and Kanber(2000)는 타일의 엔트로피지수를 이용하여, 집단간과 집단내 불평등도의 비율을 이용하여 양극화를 측정하였다. 즉, 타일의 엔트로피지수

$T = \sum_{k=1}^m s_k T_k + \sum_{k=1}^m s_k \ln \frac{\bar{y}_k}{y}$ 는 집단내 불평등도 T^W 와 집단간 불평등도 T^B 의 합으로 표시되므로 ZK지수는 다음과 같이 정의된다.

$$T = \sum_{k=1}^m s_k T_k + \sum_{k=1}^m s_k \ln \frac{\bar{y}_k}{y} = T^W + T^B \quad (56)$$

$$ZK = T^B / T^W \quad (57)$$

Zhang and Kanber(2000)의 ZK지수는 집단간 불평등도와 집단내 불평등도의 비율을 새로운 양극화지수로 제시한 이유는 중국의 지역간(도시와 농촌, 내륙과 해안) 소득불평등도 및 양극화의 정도를 지니계수, 타일의 엔트로피지수, W지수, ER지수 등을 이용하면, 지수들의 변화에 차이가 크지 않기 때문이다. 즉, 중국은 도시와 농촌, 내륙과 해안으로

소득계층이 극명하게 분리되지만, 양극화지수가 불평등도지수에 비해 이러한 변화를 제대로 반영하지 못한다.

4.4.11 다차원 불평등지수(multidimensional inequality index)

한 사회의 불평등 구조를 이해하기 위해 다차원적 고려가 필요하다. 예를 들어, 소득 불평등 논의에서 다양한 소득원천을 다차원으로 고려하는 확장과정이 존재할 수 있다. 대체로 소득불평등 논의에서 소득은 개인의 후생수준 수준을 실현할 수 있는 구매능력의 대리지표(proxy)가 될 수 있는데, 처분가능소득 개념이 가장 근접한 것으로 볼 수 있다. 그러나 처분가능소득을 단순히 노동에 대한 보상이라는 차원에서 임금이나 자영업 소득에 한정하기 어렵다. 왜냐하면 임금 이외에 다양한 차원의 소득원천이 존재하기 때문이다. 예를 들어, 금융자산거래로부터 나오는 배당금, 이자, 거래소득이 존재할 수 있고, 비금융자산자산으로부터 나오는 유입소득이 있고, 정부가 개입함으로써 발생하는 소득인 각종 조세 및 보조금과 사회간접자본으로부터의 소득 등이 포함될 수 있다.

물론 직접적으로 관찰할 수 없는 소득흐름의 보정(imputation) 방법에 대한 여러 논란이 있다. 그러나 어떤 소득개념을 활용하느냐에 따라 소득불평등 수준의 측정과 평가에 있어 상당한 차이가 나타날 수 있다. 점차 넓은 소득개념을 이용한 불평등 수준의 측정이나 이에 대한 다양한 소득원천의 기여도를 파악할 수도 있다. 그러나 이러한 방식은 원천자료를 정보손실이나 사후보정없이 활용한다는 장점은 있으나, 차원별 집계정보만을 이용함으로써 차원간 관계, 하위집단별 특성 등 개인이나 사회후생수준에 대한 충분한 정보를 제공하지 못할 수도 있다.

또 다른 접근방법으로, 보다 직접적으로 다차원 불평등 수준을 단일지수(single index)로 제시할 수 있다. 즉, 다양한 경제적 정보나 비경제적 차원의 정보를 개인 또는 하위 집단 수준에서의 효용함수로 결합하고, 이를 기반으로 불평등지수를 측정한다. 이를 수행하기 위해 크게 두 가지 접근이 이루어지고 있다. 먼저, 개인 및 가구소득의 산정에서 가구원수, 지역, 성별, 연령 등에 따른 규모의 경제정도를 산입하여 추정하는 가치균등화(equivalence scale) 접근과 비슷하게 비경제적 차원의 각 차원별 가치정도를 일종의 결손 정보로 간주하고 이를 보정(imputation)하여 소득규모를 다시 추정하는 방식이다. 그러나 이러한 경우에 차원별 가중치 산정이 쉽지 않고, 개인 및 가구의 특성이 고려되지 못한다.

또 다른 접근방법은 다차원불평등지수를 구성하는 것이다. 다차원 불평등지수의 산출은 대체로 2단계로 이루어진다. 1단계에서 각 차원들의 분포를 가장 잘 반영할 수 있는 집합적 수준의 후생함수를 산출한다. 2단계에서 1단계 결과를 이용하여 통상적인 소득불평등지수를 활용하여 다차원 불평등지수를 산출한다. Maasoumi(1986, 1999)는 1단계에서 소득, 교육, 건강 등 각 차원들이 서로 상이한 분포형태(distributional shape)를 갖고 있는

상황에서 이들의 분포를 가장 잘 대표하는 하나의 거리함수(distance function)를 구성하고, 이 거리를 최소화할 수 있는 최적의 사회후생함수(social welfare function, S_i)를 구성하였다. 즉, 한 사회의 구성원들의 후생을 결정하는 K 개의 차원이 존재한다고 가정할 때, 사회구성원 개인의 후생이 아래의 S_i 의 식과 같이 결정된다고 하자. 따라서 각 차원의 분포를 가장 잘 대변하는 최적의 거리함수 D 를 다음 식과 같이 구성할 수 있다.

$$D_\beta(S, X; w) = \sum_{k=1}^K d_k \left\{ \sum_{i=1}^n S_i \left[\left(\frac{S_i}{x_{ik}} \right)^{-\beta} - 1 \right] / \beta(\beta-1) \right\} \quad (58)$$

$$S_i \propto \begin{cases} \left(\sum_{k=1}^K w_k x_{ik}^{-\beta} \right)^{-\frac{1}{\beta}}, & \beta = 0 \\ \prod_{k=1}^K x_{ik}^{w_k}, & \beta \neq 0 \end{cases} \quad (59)$$

여기서 w_k 는 k 번째 차원의 가중치이며, 합은 1이 되고, β 는 차원 간의 대체 가능성을 나타낸다.

2단계에서 위의 비례식과 같이 산출된 사회구성원들의 후생정보를 정보이론에 기초한 소득불평등지수 중 하나인 일반화된 엔트로피지수(GE)를 이용하고, 아래의 식과 같이 계산하여 다차원 불평등지수를 산출하게 된다. 무엇보다 사회 구성원의 후생을 결정하는 여러 차원(예를 들어, 소득, 건강, 교육 등)의 분포형태를 대표할 수 있는 사회후생함수를 구성했다는 점을 특징으로 들 수 있지만 피구-달톤의 이전원칙을 만족시키지 못하는 상황이 발생할 수도 있다.

$$I_M = \frac{1}{\alpha(1-\alpha)} \frac{1}{n} \sum_{i=1}^n \left[1 - \left\{ \frac{\left(\sum_{k=1}^K w_k x_{ik}^{-\beta} \right)^{-\frac{1}{\beta}}}{\bar{S}} \right\}^\alpha \right] \quad (60)$$

여기서 α 는 그 사회가 갖는 불평등에 대한 회피(aversion)정도를 나타내는데, 값이 작을수록 후생분포의 낮은 부분의 변화에 보다 민감하다.

Bourguignon(1999)도 다음과 같은 다차원 불평등지수를 제시하였다. Maasoumi(1986, 1999)의 지수가 각 개인으로부터 도출된 사회후생의 평균에서 지수를 정규화한다. 그러나 Bourguignon(1999)는 평균적인 개인(각 차원의 평균)으로 정규화함으로써, 불평등지수가 갖추어야 하는 다른 특성들과 함께 피구-달톤의 이전의 원칙을 만족시키고 있다.

$$I_B = 1 - \frac{1}{n} \frac{\sum_{i=1}^n \left[\sum_{k=1}^K (w_k x_{ik}^{-\beta}) \right]^{-\frac{\alpha}{\beta}}}{\left[\sum_{k=1}^K (w_k \mu_k^{-\beta}) \right]^{-\frac{\alpha}{\beta}}} \quad (61)$$

Araar(2009)는 Maasoumi(1986, 1999)나 Bourguignon(1999)의 지표들이 갖는 문제점을 지적하면서 이를 보완한 새로운 다차원 불평등지수를 제시하였다. 이전의 두 지수는 개인들의 효용함수 또는 집합적으로 사회후생함수의 구성이 경험적 사실보다 공리에만 의존하고 있으며, 기존의 다차원 불평등지수의 각 차원들이 양(+)의 값으로만 구성되어 있다고 가정함으로써, 상대적인 불평등 정도만을 측정할 수 있다는 한계와 제약이 있다는 것이다. 또한, 이러한 지표들이 다양한 차원과 결합하여 개인의 후생함수를 측정하고, 이를 이용하여 엔트로피지수 등의 단일차원의 불평등지수 등을 사용하고 있기 때문에 후생함수의 특성에 따라 개별 차원의 특성들이 일부분 상쇄되고 있다고 하였다. 따라서 남녀나 세대, 그리고 직업군 등의 인구구성상 하위집단에 따른 다차원 불평등지수의 분해(decomposition)를 할 수 있으나, 사회후생함수를 구성하는 소득이나 건강 등과 같은 개별 차원의 기여도(contribution)로 분해하기 어렵다.

다차원적 불평등 정도를 측정하고 다양한 차원들 간의 구조적 관계를 파악해야 한다는 점에서 다차원 불평등지수가 차원별 기여도를 파악하는데 제약이 있다는 점은 그 유용성 측면에서 중요한 한계일 수 밖에 없다. 따라서 Araar(2009)는 이러한 기존 지표들의 문제점들을 보완하고, 각 차원별 기여도를 제시하기 위한 다차원 불평등지수를 다음의 식을 제시하였다.¹¹⁾

$$I_A = \sum_{k=1}^K \phi_k [\lambda_k I_k + (1 - \lambda_k) C_k] \quad (62)$$

여기서 ψ_k 는 가중치이고, C_k 는 요인 K 에 대한 집중도계수이며, λ_k 는 불평등의 사회적 선호를 나타낸다.

11) Araar(2009)가 제시한 다차원 불평등 지수의 구성요인 중 λ_k 는 사회 내 차원 K 의 불평등 감소와 사회 구성원 개인의 불평등 감소 중에 어떤 요인을 더 선호하는 지에 대한 교환비율로 이해할

수 있다. 예를 들어 각 차원별 분포가 다음과 같이 변한다고 하자. $\begin{vmatrix} 111 \\ 222 \\ 333 \end{vmatrix} \Rightarrow \begin{vmatrix} 211 \\ 222 \\ 233 \end{vmatrix}$. 첫 번째 차

원의 불평등 정도가 감소함으로써 다차원 불평등 정도는 감소하게 되지만, 첫 번째 개인의 경우 각 차원별 분포가 상이해짐으로서 개인수준의 차원간 불평등 정도는 증가하게 된다. 이러한 상황에서 λ_k 는 각 차원별로 어떤 상황을 선호하는 지에 대한 사회적 합의정도라고 할 수 있다.

V. 불평등요인 분해방법론

5.1 분석적 접근방법

각종 불평등의 원인이나 요인별 기여도를 밝히기 위한 방법론은 다양하다. 첫째 소득원천의 유형에 따른 소득불평등의 분석으로써, 소득원천별 분해법이 사용된다. 이러한 불평등이 가구의 사회경제적 특성에 의해 발생한 것으로 가정하고, 이를 이용하여 소득 불평등을 분석하기 위해 일반화된 엔트로피지수 요인분해법을 적용한다. 둘째, horrocks(1984) 등과 같은 집단별 분해를 통한 분석이 있다. 셋째, 회귀분석, 분위별 회귀 분석, 패널회귀분석 등과 같은 전통적인 계량경제학적 분석기법을 이용하여 분석하기도 한다.

또한, 분석적 접근방법은 전체 불평등을 소득원천의 불평등 기여도의 합으로 나타내는 수학적 방식에 기반하고 있다. 또 다른 접근방법은 전체 불평등에 대한 소득원천의 예상 한계적 기여도(marginal contribution)에 기반하고 있다.

5.1.1 로렌츠곡선

지니계수를 구하기 전에 로렌츠곡선부터 구해야 한다. 일반적인 방법으로 로렌츠곡선을 구하기 위해 X 축에 누적인구비율과 Y 축에 누적소득비율의 점을 연결하면 된다. 그러나 다양하게 시뮬레이션이 가능하다.

1단계에서 소득과 분위수의 초기 분포를 만든다. 정규분포와 로그정규분포인 경우를 살펴보자. x 가 평균 μ 이고 분산 σ^2 을 가지는 로그정규분포라고 하자. 로렌츠곡선은 다음과 같이 정의된다.

$$L(p) = \Phi\left(\frac{\ln(x) - (\mu - \sigma^2)}{\sigma}\right), \quad p = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \quad (63)$$

$\mu = 1$ 이라고 가정하고 Shorrocks and Wan(2008)이 제시한 과정을 따라 분산을 추정한다. 로그소득의 표준편차 σ 는 $\sigma^k = \Phi^{-1}(p_k) - \Phi^{-1}L((p_k))$, $k = 1, \dots, m-1$ 의 $m-1$ 개의 추정치를 평균하여 얻을 수 있다. 여기서 m 은 계급(class)의 수이고 Φ 는 표준정규분포함수이다.

일반화된 2차형 로렌츠곡선(Generalised Quadratic Lorenz Curve)은 다음을 가정한다.

$$L(1-L) = a(p^2 - L)bL(p-1) + c(p-L) \tag{64}$$

$L(1-L)$ 을 $(p^2 - L)$, $L(p-1)$, 그리고 $(p-L)$ 에 회귀를 하되, 상수항은 없고 좌표 (1,1)을 통과하는 곡선을 만들어내는 함수를 선택하기 위해 마지막 관측치를 제거한다. 따라서 다음과 같다.

$$Q(p) = -\frac{b}{2} - \frac{(2mp+n)(mp^2 + np + e^2)^{-0.5}}{4} \tag{65}$$

여기서 $e = a + b + c + 1$ 이고, $m = b^2 - 4a$ 이며 $n = -2be - 4c$ 이다. 베타 로렌즈곡선(Beta Lorenz Curve)는 다음을 가정한다.

$$\log(p-L) = \log(\theta) + \gamma \log(p) + \delta \log(1-p) \tag{66}$$

모수들을 추정된 후, 다음과 같은 분위수를 생성할 수 있다.

$$Q(p) = \theta + p^\gamma (1-p)^\delta \left[\frac{\gamma}{p} - \frac{\delta}{(1-p)} \right] \tag{67}$$

Singh-Maddala(1976) 분포는 다음과 같은 형태를 취한다.

$$F(x) = 1 - \left[\frac{1}{1 + (x/b)^a} \right]^q \tag{68}$$

단 $a \geq 0$, $b \geq 0$, $q \geq 1/a$ 는 추정된 모수이다. 소득 x 는 0보다 크거나 같다고 가정한다. 밀도함수는 다음과 같이 정의된다.

$$f(x) = (aq/b)(1 + (x/b)^a)^{-(q+1)} (x/b)^{(a-1)} \tag{69}$$

분위수는 다음과 같이 정의된다.

$$Q(p) = b((1-p)^{-1/q} - 1)^{1/a} \tag{70}$$

모수에 대한 추정 접근방법으로, 우도함수를 극대화하고 개별 계급(class)의 평균소득에서 평가되는 밀도함수의 곱이 된다.

2단계에서 누적데이터와 일치시키기 위해 초기분포를 조정한다. 두 번의 연속적인 조정 과정을 거친다. 첫째 조정에서 각 소득집단이 원래 평균소득을 갖도록 초기소득 벡터를 수정한다. 다음 조정으로 계급 간 분포를 평활화(smoothing) 한다.

5.1.2 분석적 접근방법 : Rao(1969)의 분해방법론

Rao(1969)는 지니계수에 불평등요인으로 분해하였는데, 지니계수가 다음 식과 같이 분해 가능하다고 보았다.

$$I(\tilde{y}) = \frac{\mu_k}{\mu_{\tilde{y}}} \tilde{C}_k \quad (71)$$

여기서 \tilde{y} 가 소득 등의 순위변수(ranking variable)일 경우 $\mu_{\tilde{y}}$ 는 \tilde{y} 의 평균, \tilde{C}_k 는 s_k 의 집중계수(coefficient of concentration)이다

5.1.3 분석적 접근방법 : Lerman and Yitzhaki(1985), Podder and Tran-Nam(1991)의 지니계수 분해방법론

Lerman and Yitzhaki(1994)는 지니계수를 분해하였는데, 이러한 방법은 불평등도 분석에 지니계수 분해방법이 사용될 수 있다. 지니계수는 다음과 같은 연속함수의 형태로 도출될 수 있다.

$$AG = \int_a^b F(x)[1 - F(x)]dx \quad (72)$$

여기서 x 가 소득, a 가 최저소득, b 가 최고소득이고, F 가 소득의 분포함수이다. $u = F(x)[1 - F(x)]$, $v = x$ 로 놓고 부분적분하면 다음과 같다.

$$AG = \int_a^b x[F(x) - 1/2]F(x)dx \quad (73)$$

$x(F)$ 를 $F(x)$ 의 역함수하면, 변수변환에 의해 다음과 같이 쓸 수 있다.

$$AG = 2 \int_0^1 x(F)[F(x) - 1/2]dF \quad (74)$$

함수 F 가 $[0, 1]$ 사이의 균일분포(uniform distribution)을 가지면 평균이 $1/2$ 이 되므로 다음과 같이 쓸 수 있다.

$$AG = 2Cov[x, F(x)] \tag{75}$$

이를 평균소득 μ 로 나누면 상대적 지니계수가 된다. x_1, \dots, x_K 를 소득원천별 소득이라고 하자. 공분산과 $x = \sum_{k=1}^K$ 의 관계를 적용하면 $Cov(x_k, F)$ 가 소득원천 k 의 집중도지수(concentration index)를 나타낼 때, 전체 지니계수는 다음과 같이 나타낼 수 있다.

$$G = \frac{2 \sum_{k=1}^K Cov(x_k, F)}{\mu} \tag{76}$$

R_k 가 소득원천 k 내에서의 순위와 총소득의 순위 간의 지니상관계수, G_k 가 소득원천 k 의 상대지니계수 또는 요인 k 에 대한 집중도지수, S_k 가 소득원천 k 의 총소득에 대한 비중이라고 하자. 이때 위의 식에서 각 소득원천 k 에 $Cov(x_k, F)$ 를 곱하고 μ_k 로 나누면 분해된 소득원천별 요인들의 합계를 다음과 같이 산출할 수 있다(Lerman and Yizhaki, 1994)

$$\begin{aligned} G &= \sum_{k=1}^K \left[\frac{Cov(x_k, F)}{Cov(x_k, F_k)} \cdot \frac{2Cov(x_k, F_k)}{\mu_k} \cdot \frac{\mu_k}{\mu} \right] \\ &= \sum_{k=1}^K R_k G_k S_k \end{aligned} \tag{77}$$

예를 들어, R_k 가 총소득과 근로소득과의 순위상관계수(지니상관계수), G_k 가 근로소득의 지니계수, 그리고 S_k 가 총소득에서 근로소득이 차지하는 비중을 나타낼 때 근로소득이 총소득의 불평등(지니계수)에 기여하는 정도를 C_k 라고 하면 C_k 는 다음과 같이 계산된다.

$$C_k = R_k \times G_k \times S_k \tag{78}$$

즉 근로소득의 총소득불평등에 대한 기여도 C_k 는 근로소득의 지니계수 G_k 에 근로소득과 총소득 간의 지니상관계수 R_k , 그리고 근로소득이 총소득에서 차지하는 비중 S_k 를 곱한 값이 된다. 이와 같이 각 소득원천을 k, l, m, n 이라고 하면 이러한 소득원천이 총소득불평등에 대한 기여도를 계산할 수 있으며, 이러한 소득원천별 소득의 기여도를 모두 합하면 다음과 같이 총소득의 지니계수와 동일한 값이 된다.

$$\text{총소득 지니계수} = C_k + C_l + C_m + C_n \quad (79)$$

한편, μ_k 가 유형 k 소득의 평균이고, μ 가 모든 유형의 소득합계의 평균, C_k 가 유형 k 소득의 집중도지수, G 가 지니계수일 때 다음과 같이 지니계수가 소득유형별 집중도지수의 합으로 나타낼 수 있다(Kakwani, 1980, 1986).

$$G = \sum_{k=1}^K \frac{\mu_k}{\mu} C_k \quad (80)$$

Podder(1993)도 Kakwani(1980, 1986)와 유사하게 Rao(1969)가 도출한 지니계수 분해방법을 이용하여 소득유형별로 지니계수를 분해하여 각 유형별 소득이 전체 소득불평등도에 미치는 효과를 분석하였다. 그러나 Podder(1993)는 Kakwani(1980, 1986)의 분해방법이 집중도지수를 지니계수와 동일하게 각 유형별 소득의 서열을 이용하여 산출한 오류가 있었다고 지적하면서 총소득(각 유형별 소득의 합)을 낮은 것에서부터 높은 순서로 나열하고 그 서열로써 각 소득 유형별 집중도지수를 구해야 한다고 주장하였다.

또한 Podder(1993)는 어떤 유형의 소득이 총소득에 대하여 일정하다면, 그 유형의 소득은 총소득불평등도에 어떤 영향도 주지 않는다는 소득불평 등도의 특성에 주목하였다. 실제로 총소득에 대하여 일정한 비율로 분포된 소득유형이 거의 없기 때문에 총소득에 어떤 유형의 소득을 가감하면 총소득불평등도가 변할 수밖에 없는데, Podder and Tran-Nam(1991)은 그것을 반영하여 위의 식을 경제학적 의미해석이 용이한 형태로 다음과 같이 변환하였다.

$$\sum_{k=1}^K \frac{\mu_k}{\mu} (C_k - G) = 0 \quad (81)$$

이 식은 어떤 유형의 소득이 총소득에 대하여 비례적으로 분포되면, 그 소득은 총소득 불평등도에 전혀 영향을 주지 않음을 보여준다. 이 식의 각 소득유형의 부호가 양(+)이면 그 유형의 소득이 총소득에 대해 비례적인 것 이상으로 증가하여 그 소득의 집중도

지수가 총소득의 지니계수보다 커서 전체 소득불평등도지수를 증가시킴을 의미한다. 부호가 음(-)이면 어떤 유형의 소득이 총소득에 대해 비례적인 것보다 작게 증가하여 전체 소득불평등도지수를 감소시키는 역할을 하고 있음을 의미한다. 이는 Kakwani(1980)이 제시한 조세집중도지수와 지니계수와의 차이로써 누진·역진·비례적인 조세임을 판정하는 것과 동일한 논리이다. 지니계수와 달리 집중도지수는 음(-)의 부호를 가질 수도 있다 (Podder, 1993).

또 다른 해석은 $C_k - G$ 는 어떤 유형의 소득의 총소득에 대한 탄력성이 1(단위 탄력적)에서 얼마나 괴리되어 있는가를 가중 평균한 것이다(Lerman and Yitzhaki, 1985). 이러한 탄력성은 전체 소득불평등도에 대한 각 유형별 소득의 상대적 중요성을 나타내는데, 그 합은 다음과 같이 항상 0이 된다.

$$\eta_k = \frac{1}{G} \left[\frac{\mu_k}{m} (C_k - G) \right] \tag{82}$$

$$\sum_{\eta} \eta_k = \sum_k \frac{1}{G} \left[\frac{\mu_k}{\mu} (C_k - G) \right] = 0 \tag{83}$$

여기서 η_k 는 유형 k 소득의 탄력성을 나타낸다. 이러한 식 역시 $(C_k - G)$ 항 때문에 모든 유형의 소득변화가 총소득에 대하여 비례적이라면 지니계수가 변하지 않는다는 것을 보여준다. 한편 $\frac{\mu_k}{\mu} (C_k - G) = s_k (C_k - G)$ 는 어떤 유형의 소득 10% 변화가 이러한 수치에 10%를 곱한 값만큼 지니계수를 변화시킨다. 따라서 이러한 수치를 가지고 어떤 유형의 소득이 총소득불평등도에 대해 얼마나 효과를 미치는 지에 대해 알 수 있다. 이때 탄력성 부호가 음(-)이면 지니계수를 탄력성×10% 만큼 감소시킨다. 반대로 양(+)이면 지니계수를 탄력성×10% 만큼 증가시킨다. 따라서 어떤 소득유형이 전체 소득불평등도에 얼마나 영향을 미치는지를 보여주는 지를 나타낼 수 있다.

5.1.4 타일의 엔트로피지수 분해방법론

지니계수는 비선형성의 특징을 가지고 있기 때문에, 소득불평등을 소득원천별 또는 인구집단별로 요인분해를 실시할 때 불편하다. 따라서 소득불평등의 요인분해에 주로 타일의 엔트로피지수를 보다 보편적인 공식으로 바꾼 일반화된 엔트로피(Generalized Entropy, GE) 계열의 지표들이 사용된다.

집단간 불평등도분해에 주로 GE(0)이 사용되며 이때 전체 불평등도는 아래와 같이 집단간 불평등도와 집단내 불평등도로 분해될 수 있다.

여기서 $GE(0)_k$ 는 k 집단의 소득불평등도, v_k 는 k 집단이 모집단에서 차지하는 비율 ($\equiv N_k/N$), λ_k 는 k 집단 평균소득의 모집단 평균소득에 대한 비율($\equiv \mu_k/\mu$)을 나타낸다.

$$GE(0) = \sum_k v_k GE(0)_k + \sum_k v_k \log\left(\frac{1}{\lambda_k}\right) \quad (84)$$

또한, 시간의 경과에 따른 엔트로피지수의 전체 불평등도 변화를 분석하고 정태적 요인분해를 위해 위의 식을 시간에 대해 차분하고, 테일러 전개를 이용한 근사치를 이용하여 정리하면 다음과 같다.

$$\begin{aligned} \Delta GE(0) &\equiv GE(0)_{t+1} - GE(0)_t \\ &= \sum_k \overline{v_k} \Delta GE(0)_k + \sum_k \overline{GE(0)_k} \Delta v_k \\ &\quad - \sum_k [\overline{\log(\lambda_k)}] \Delta v_k - \sum_k \overline{v_k} \Delta \log(\lambda_k) \\ &\approx \sum_k \overline{v_k} \Delta GE(0)_k + \sum_k \overline{GE(0)_k} \Delta v_k \\ &\quad + \sum_k [(\overline{\lambda_k}) - \overline{\log(\lambda_k)}] \Delta v_k + \sum_k [\overline{\theta_k} - \overline{v_k}] \Delta \log(\mu_k) \quad (85) \end{aligned}$$

이 식에서 Δ 는 차분연산자이며 $\overline{v_k}$ 는 비교대상이 되는 두 기간간의 평균을 의미한다. 마지막 식에서 좌변은 모집단 전체의 불평등 변화를 의미하며 우변 첫째 항은 불평등 변화의 순효과(pure inequality effect), 둘째 항 및 셋째 항은 집단구성 변화의 효과(allocation effect), 넷째 항은 집단간 상대소득 변화의 효과(income effect)를 의미한다.

소득불평등도를 요인분해 하는 방법은 크게 두 가지로 나눌 수 있다. 첫째, 소득원천별 요인분해 방법이다. 예를 들어, 농가소득이 농업소득과 농외소득(겸업소득, 사업외소득), 이전수입 등으로 구성되어 있다고 하자. 이와 같은 소득원천별 소득의 불평등도가 전체 농가소득 불평등도에 얼마나 영향을 미쳤는지를 요인분해할 수 있다. 둘째, 농가의 특성별 요인분해 방법으로, 경영주의 연령별 분해 또는 영농규모별로 요인을 분해를 할 수 있다.¹²⁾

변이계수 제곱의 1/2인 GE(2)를 이용하여 전체 소득 불평등도를 소득원천별로 요인을

12) 예를 들어, 박준기·문한필·김용택(2004)은 농가 특성별 요인분해를 위해 GE(0) 지표를 사용하였으며, 소득원천별 요인분해를 위해 GE(2) 지표를 사용하였다. 소득원천별 분해에서 소득원천에 따라 소득이 전혀 없는 경우가 많이 발생하는데 GE(2) 지표는 이러한 경우의 분해에 이용되기에 적합하다.

분해하면, 다른 지표들보다 요인의 분해와 분석결과에 대한 직관적 해석이 용이하다 (Jenkins, 1995). 소득원천별 요인분해 방법을 정리하면, s_f 를 소득원천 f 가 전체 불평등도에 미치는 절대적인 기여분이라고 할 때 전체 불평등도는 다음과 같이 소득원천별 기여분의 합계로 나타낼 수 있다.

$$GE(2) = \sum_f S_f \tag{86}$$

여기서 $\sum_f s_f = 1$ 일 때 S_f 를 전체 불평등도 $GE(2)$ 로 나누면 상대적 기여도인 s_f 를 다음과 같이 도출할 수 있다.

$$s_f = \frac{S_f}{GE(2)} \tag{87}$$

s_f 는 다양한 방식으로 정의할 수 있으나, Shorrocks(1982)의 방법을 적용하면 s_f 는 개별 소득원천 x_f 를 전체 소득 x 에 회귀분석한 회귀계수로 다음과 같이 표현할 수 있다.¹³⁾

$$s_f = \frac{Cov(x_f, x)}{Var(x)} \tag{83}$$

여기서 $s_f = \frac{\Delta x_f}{\Delta x} + \mu_i$ 이다. 따라서 S_f 는 다음과 같이 정의할 수 있다.

$$\begin{aligned} S_f &= s_f \cdot GE(2) \\ &= \rho_f \cdot x_f \sqrt{GE(2) \cdot GE(2)_f} \end{aligned} \tag{88}$$

여기서 ρ_f 는 소득원천 x_f 와 전체소득 x 의 상관계수, x_f 는 소득원천 x_f 의 전체소득 x 에 대한 비율($\equiv \mu_i/\mu$)이다.

다음으로 불평등 요인분해는 모집단의 연령, 규모 등 농가 특성을 기준으로 세분화하여 전체 불평등도를 집단간 불평등도와 집단내 불평등도로 분석하는 농가특성별에도 적용될 수 있다(박준기·문한필·김용택, 2004). 전체 불평등도 $GE(0)$ 는 다음 식과 같이 집단간 불평등도로 분해된다. 다음 식에서 $GE(0)_k$ 가 k 집단의 소득불평등도, v_k 가 k 집단

13) Shorrocks의 방법은 지수(index) 형태와 분석배열에 중립적(neutral) 모형으로 많이 응용된다.

이 모집단에서 차지하는 비율($\equiv N_k/N$), λ_k 가 k 집단 평균소득의 평균소득의 모집단 평균소득에 대한 비율($\equiv \mu_k/\mu$)일 때 좌변은 모집단 전체의 불평등도이고, 우변 첫째 항은 개별집단내의 불평등도의 가중평균, 둘째 항은 집단간의 불평등도를 나타낸다.

$$GE(0) = \sum_k v_k GE(0)_k + \sum_k v_k \log\left(\frac{1}{\lambda_k}\right) \tag{89}$$

5.1.5 분석적 접근방법 : 로그편차평균지수 분해방법론

로그편차평균(Mean Log Deviation, MLD)지수는 가구소득이 일반적으로 로그 정규분포를 따르는 특징을 반영하여 다음과 같이 자연대수로 전환된 소득에 대하여 그 편차를 평균한 수치로 정의할 수 있다. 식에서 측정된 로그편차평균은 소득분배가 완전하게 평등한 경우 최소값은 0이 된다. 이 지수는 선형이므로 소득불평등도를 요인분해하는 데 적합하며, 다른 지수들에 비해 저소득계층의 분포에 크게 영향을 받는다.

$$\begin{aligned} MLD &= \frac{\sum_i \ln\left(\frac{\mu}{Y_i}\right)}{N} \\ &= \ln\mu - \frac{1}{N} \sum_i \ln Y_i \end{aligned} \tag{90}$$

전체 소득불평등을 가구 특성(가구주 연령/성별/학력 및 가구의 취업자 수)에 따라 집단을 분류하여 각 집단내에서 발생하는 소득불평등과 집단간에 발생하는 소득불평등의 합으로 표시할 수 있는 불평등지수는 일반화된 엔트로피지수인데, 엔트로피지수 가운데 동태적 분석을 하기 위해서 로그편차평균(Mean Log Deviation, MLD)이 사용된다. 대수편차평균을 집단내 소득불평등(MLD_W)과 집단간 소득불평등(MLD_B)으로 구분하면 다음과 같이 정리할 수 있다.

$$\begin{aligned} MLD &= \frac{1}{N} \sum_{i=1}^N \ln \frac{\mu}{Y_i} \\ &= \sum_{k=1}^K (v_k) MLD_k + \sum_{k=1}^K \left(\frac{N_k}{N}\right) \ln\left(\frac{1}{\lambda_k}\right) \\ &= MLD_W + MLD_B \end{aligned} \tag{91}$$

여기서 MLD 는 로그편차평균, MLD_k 는 k 집단내의 로그편차평균, $v_k = \frac{N_k}{N}$ 는 k 집단의 가구가 전체 모집단 가구에서 차지하는 비율, $\lambda_k = \frac{\mu_k}{\mu}$ 는 k 집단 평균소득(μ_k)의 모집단 평균소득(μ)에 대한 비율이다.

5.2 하위집단의 회귀방법을 활용한 소득불평등 분해방법론

5.2.1 Shorrocks(1982), Fields(2003), Wan(2004)의 분해방법론

전통적인 계량경제학적인 회귀분석에 기반한 불평등분해가 가지는 장점은 불평등의 다양한 결정요인을 식별하고 분석할 수 있다는 점이다. 다음과 같은 전통적인 일반선형 회귀모형으로 회귀계수를 추정했다고 하자.

$$y = \hat{\beta}_0 + \hat{\beta}_1 y_1 + \dots + \hat{\beta}_k y_k + \hat{\epsilon} \quad (92)$$

여기서 y 는 소득이나 임금, 그리고 대출 등의 금융이 될 수 있으나 논의를 간단하게 하기 위하여 소득이라고 하자. $y = [y_1, \dots, y_k]$ 는 소득 등을 결정하는 일련의 설명변수들이다. $\hat{\beta}_k x_k$ 는 소득 등에 대한 변수 x_k 의 몫(share)이다. $\hat{\epsilon}$ 은 잔차항이다. Fields(2003)은 변수에 대한 몫인 $\hat{\beta}_k x_k$ 과 잔차가 Shorrocks(1982)의 소득원천(f)의 동일항으로 간주할 수 있다고 하였다. 따라서 변수 x_k 의 소득불평등에 대한 상대적 기여도는 다음식과 같다. 다만, 상수항은 상대적 기여도에 포함되지 않는다.

$$s_k = \frac{Cov(\hat{\beta}_k x_k, y)}{\sigma_y^2} \quad (93)$$

Shorrocks(1982)의 분해방법은 두 시점 ($t, t+1$)에서 소득불평등 변화에 대한 적용이 가능하다. 따라서 두 시점간 불평등도 변화에 대한 x_k 의 기여도는 다음과 같다.

$$I_y^{t+1} - I_y^t = \sum_k (s_k^{t+1} I_k^{t+1} - s_k^t I_k^t) \quad (94)$$

Fields(2003)에서는 상수항이 상대적 기여도에 포함되지 않는다. Wan(2004)은 이를 보완 방법을 제시하였다. 이러한 추정된 회귀식은 $y = s_0 + s_1 + \dots + s_k + s_\epsilon$ 로 쓸 수 있다.

여기서 s_0 는 추정된 상수이고, $s_k = \hat{\beta}_k x_k$ 이며 s_e 은 잔차이다. 관심변수의 평균이 0일 경우(즉, 잔차의 경우) 상대적인 불평등도 지수는 정의되지 않는다. 또한 관심 변수가 상수일 경우(즉, 추정된 상수의 경우) 불평등도 지수는 0이 된다. Wan(2004)은 잔차와 상수항에 대한 문제를 해결하기 위해, 다음의 원칙을 제시하였다. $\hat{y} = s_0 + s_1 + \dots + s_k$ 이고 $\tilde{y} = s_1 + s_2 + \dots + s_k$ 라고 가정하면, $I(y) = cs_0 + I(\tilde{y}) + cs_r$ 이 된다. 따라서 상수의 기여도는 $cs_0 = I(y) - I(\hat{y})$ 이고 잔차의 기여도는 $cs_r = I(\hat{y}) - I(\tilde{y})$ 가 된다.

Shorrocks(1982)은 다음 식과 같이 변동계수제공(Squared Coefficient of Variation, SCV)을 적용하였다.

$$SCV(\tilde{y}) = \sum_{k=1}^K \frac{Cov(y, s_k)}{\mu_{\tilde{y}}^2} \quad (95)$$

그러나 변동계수제공과 달리 지니계수는 위의 식이 적용될 수 없다. 지니계수에 적용하기 위하여, Rao(1969)의 지니계수 분해방식을 살펴볼 필요가 있다. Rao(1969)에 따르면, 지니계수는 다음과 같이 분해가능하다. 여기서 \tilde{y} 가 순위변수(ranking variable)일 경우 $\mu_{\tilde{y}}$ 는 \tilde{y} 의 평균, \tilde{C}_k 는 s_k 의 집중계수(coefficient of concentration)이다.

$$I(\tilde{y}) = \frac{\mu_k}{\mu_{\tilde{y}}} \tilde{C}_k \quad (96)$$

5.2.2 샤플리(Shapley) 분해방법론

샤플리값은 협조적 게임이론에서 게임의 참여자간 협조로 얻은 총보수(total payoff)에 대해 각 참여자의 한계적 기여(marginal contribution)만큼 나누는 균형배분규칙이다.¹⁴⁾ 샤플리값은 협조적 게임의 균형개념으로 개발되었으나, 불평등의 원천별 한계적 기여를 계산하는 경우에도 사용된다. 예를 들어, 전체 소득의 원천으로 근로소득, 자산소득, 기타 소득 세 가지로 고려하자. 이러한 경우에 소득원천 각각을 게임의 참여자로 보고, 소득원천의 다양한 조합으로 이루어진 부분집합(S)과 그에 따른 불평등도를 연결시키는 함수를 보수함수(v)로 간주하자. 이때 계산된 샤플리값은 전체 소득의 불평등에 대한 각 소득원천의 한계적 기여를 나타낸다. 그러나 불평등도 기여도 분석에서 샤플리값을 이용하는 경우에 각각의 소득원천별로 전체 소득에 대한 점유율이 다르다는 점을 고려해야 한다. 샤플리값은 일반적으로 각 참여자를 한 사람으로 보기 때문에 동일 비중으로 계산

14) 보다 자세한 샤플리값의 의미에 대해 이성재·이우진(2017)을 참조하라.

하지만, 소득원천의 경우에 각각의 소득원천별로 점유율이 다를 수 있다.

S 에 속하지 않는 소득원천의 가구별 소득값의 처리방법에 따라 점유율에 대한 고려가 달라진다. 영점 샤플리값 분해(zero Shapley value decomposition method)는 S 에 속하지 않는 소득원천에 대하여 모든 가구에 0의 소득을 부여하여 S 에 속한 소득원천의 불평등도를 계산한다. 평균 샤플리값 분해(mean equalized Shapley value decomposition method)는 S 에 속하지 않는 소득원천에 대해서는 모든 가구가 각 원천의 평균값을 동일하게 가지고 있다고 가정한 후 이를 S 에 속한 소득원천의 값들에 더하여 불평등도를 계산한다(Chantreuil and Trannoy, 1999; Sastre and Trannoy, 2002).

하위집단인 원천소득의 수의 범위를 확대하자. 가구들의 전체 집합을 $I = \{1, \dots, i, \dots, N\}$ 하고, 소득원천 종류의 전체집합을 $J = \{1, \dots, j, \dots, k\}$ 라 하고 x_i^j 를 가구 i 의 j 번째 원천소득이라 하자. 따라서 j 번째 원천소득의 소득분포는 $x^j = [x_1^j, \dots, x_N^j]$ 으로 N 차원 벡터가 된다. 전체 소득원천 종류의 부분집합을 $S \subset J$ 라 하면, K 개의 원천소득으로 만들어지는 부분집합 S 는 공집합을 포함하여 총 2^N 개가 된다.

집합 S 에 속하는 소득원천들의 합으로 만들어지는 소득벡터를 $y(S)$ 라 하면, 이러한 소득분포는 다음과 같은 N 차원 벡터가 된다.

$$y(S) = \left[\sum_{j \in S} x_1^j, \dots, \sum_{j \in S} x_N^j \right] \tag{97}$$

집합 S 에 속하지 않은 소득원천들은 모든 가구가 가지고 있지 않다고(0을 가지고 있다고) 가정한 후, 이를 더하여 처리하는 것과 동일한 방식이다. 반면에 집합 S 에 속한 소득원천들의 값에 대해 집합 S 에 속하지 않는 소득원천들의 평균값을 더하면 다음과 같다.

$$y^e(S) = \left[\sum_{j \in S} x_1^j + \sum_{j \notin S} \mu(x^j), \dots, \sum_{j \in S} x_N^j + \sum_{j \notin S} \mu(x^j) \right] \tag{98}$$

$\mu(x^j)$ 는 j 번째 소득원천의 평균값이다. 이제 소득분포 $y = [y_1, \dots, y_N]$ 에 대한 불평등도를 $G(y)$ 라고 하면, j 번째 원천소득의 영점 샤플리값은 다음과 같다.

$$Sh_j(I, J, G) = \sum_{S \subset J, j \in S} \frac{(s-1)!(k-s)!}{k!} [G(y(S)) - G(y(S - \{j\}))] \tag{99}$$

j 번째 원천소득의 평균 샤플리값은 다음과 같다.

$$Sh_j^e(I, J, G) = \sum_{S \subset J, j \in S} \frac{(s-1)!(k-s)!}{k!} [G(y^e(S)) - G(y^e(S - \{j\}))] \quad (100)$$

앞의 Wan(2004)의 회귀모형을 생각하자. $\hat{y} = s_0 + s_1 + \dots + s_k$ 이고 $\tilde{y} = s_1 + s_2 + \dots + s_k$ 라고 가정하면, $I(y) = cs_0 + I(\tilde{y}) + cs_r$ 이 된다. 따라서 상수의 기여도는 $cs_0 = I(y) - I(\hat{y})$ 이고 잔차의 기여도는 $cs_r = I(\hat{y}) - I(\tilde{y})$ 가 된다. 선형모형에 대해 샤플리 분해방법론은 다음과 같은 형태를 취하게 된다.

$$I(y) = I(\hat{y}) + cs_r \quad (101)$$

여기서 잔차의 기여도는 $cs_r = I(y) - I(\hat{y})$ 이다. 반로그선형모형(Semi-log linear model)에 대해 샤플리 분해방법은 상수와 잔차를 포함한 모든 요인에 대해 적용된다.

5.2.3 DER 분해방법론

소득원천별 DER지수의 분해는 다음과 같다.

$$P_{source} = \sum_k \phi_k CP_k \quad (102)$$

여기서 $CP_k = \frac{\int f(x)^{1+\alpha} a_k(x) dx}{\phi_k^\alpha \mu_k^{\alpha-1}}$ 로 가상집중도지수(pseudo-concentration index)이고, ϕ_k 는 소득원천 k 의 소득점유율이다.

VI. 결론

본 연구는 금융시장의 변화하여 확대되는 것을 살펴보고, 급변하는 금융환경에서 금융소비자에 대한 금융불평도를 측정하는 방법론을 살펴보았다. 소득, 자산, 교육 등 다양한 분야에서 오래동안 불평등도에 대한 측정은 시행되어 왔다. 주요국이나 국내에서 어떠한 신용정보가 신용평가를 위해 쓰이는 지에 대해 설명한다. 그러나 금융분야에서 가장 기본이 되는 신용등급이나 신용점수의 미시자료를 활용하기는 상당히 어렵다. 신용등

급이나 신용점수는 개인정보이기도 하고, 신용평가회사나 금융기관들이 자체적으로 개인에 대해 신용등급이나 신용점수를 산출하기 때문에 자료의 일치성도 확인하기 어렵다. 또한, 신용등급이나 신용점수 기준의 변화로 연도별 자료를 구하는 것에도 한계가 존재한다.

금융시장의 공급자 측면에서의 변화와 금융불평등, 특히 대출시장에서 어떠한 불평등이 존재하는 지에 대한 방법론을 본 연구에 소개하였다. 대출한도나 대출금리 등도 신용점수에 의해 영향을 받는다. 그러나 대출에 영향을 주는 변수는 훨씬 많다. 나이, 소득이나 자산 등의 정량적 변수도 존재하고 직업 등의 정성적 변수도 존재한다. 그러나 금융불평등도를 연도별로 추정하는 방법론을 소개하고, 추정된 연도별 자료를 기반으로 분해방법론을 제시하였다.

이러한 금융불평등도에 추정 및 분해방법론은 국내의 금융불평등 현황을 연도별로 제시할 수도 있고, 금융불평등에 영향을 미치는 주요한 개별 변수들에 대해서도 알 수가 있다는 장점이 있다. 다양한 금융불평등도 추정방법론으로 여러 가지 결과가 나올 수 있다. 그러나 비슷한 방향으로 나오는 금융불평등도 외에 다른 결과가 나오는 경우에 분해방법론을 통해 비교할 수 있다는 장점이 있고, 해석할 수 있다는 장점이 있다.

또한 불평등에 대한 지니계수나 일반화된 엔트로피지수 등의 불평등 추세나 원인분석을 넘어 분포의 변동으로 다양화되었다. 중산층의 몰락 또는 쇠퇴 현상은 소득불평등의 심화 현상과 동일시되고 있는데, 전통적인 불평등지수는 분포의 변동을 반영하지 못한다. 이를 극복하는 개념이 극화 또는 다극화이며, 본 연구는 현재까지 나온 추정방법과 분해방법론을 제시하고 있다.

금융불평등, 특히 개인에 대한 대출총량이 많을수록 좋다는 것이 아니다. 당연히 금리가 높아지는 경우에 대출총량은 감소하는 경향이 있고, 금리가 낮아지는 경우에 대출총량은 증가하는 경향이 있다. 또한 과도한 대출은 금리인상 시기에 큰 리스크로 발생하게 되고, 확대되면 금융시스템에도 영향을 미칠 수 있다. 그러나 개인에 대한 금융불평등이 소득이나 자산 등과 같이 전통적인 변수에 영향을 받는 것이 아니라, 다른 정성적인 변수가 영향을 받는다면 사회적 리스크는 보다 커질 수 밖에 없다. 이러한 점에서 본 연구는 금융불평등도 및 극화지수를 소개하고, 불평등도와 극화지수 분해방법론을 다루고 있다.

참 고 문 헌

- 금융위원회(2018), 「금융분야 마이데이터 산업 도입 방안」, 보도자료.
- 금융위원회(2020), 「4차 산업혁명 시대의 디지털금융 종합혁신방안」, 보도자료.
- 금융위원회(2022), 「'22년 본인신용정보관리업(마이데이터) 허가심사방향」, 보도자료.
- 김낙년(2020), “가계조사의 행정자료에 의한 보정: 2016년 가계금융복지조사를 중심으로”, 「한국사회정책」, 27(1), 39-61.
- 김상봉(2021a), 「빅데이터세상」, 지필미디어.
- 김상봉(2021b), 「인공지능이론」, 지필미디어.
- 김상봉·강경우(2022), “금융불평등도 측정을 위한 방법론 연구”, 「기술금융연구」, 11(2), 43-66.
- 김상봉·김시연(2018), “딥러닝을 이용한 경제성장 모형에 대한 실증연구”, 「신용카드리뷰」, 12(4), 67-88.
- 김상봉·김정렬·박덕배·손영범·심현섭·이보우·전인구·조경준(2012), 「신용정보의 이해와 활용」, 지필미디어.
- 김은미·김상봉·조은서(2020), “기계학습을 활용한 주택매도 결정요인 분석 및 예측모델 구축”, 「지적과 국토정보」, 50(1), 181-200.
- 김은미·김상봉(2020), “주택시장의 손실회피성향에 관한 연구 : 머신러닝 방법을 중심으로”, 「신용카드리뷰」, 14(3), 52-74.
- 박준기·문한필·김용택(2004), “농가소득 불평등도의 요인분해”, 「농촌경제」, 27(4), 15-27.
- 백승주·김현섭(2013), “불평등의 다차원 접근 : 측정 및 활용”, 「한국정책학회보」, 22(2), 283-312.
- 신동균·전병유(2005), “소득분포의 양극화 추이”, 「노동경제논집」, 28(3), 77-109.
- 신인석·한민(2021), “2017년 「가계금융복지조사」 조사표 변경의 가계담변에 대한 영향”, 「금융정보연구」, 10(3), 19-39.
- 유경준(2007a), “소득불평등 개념과 실태”, 「노동경제논집」, 30(3), 103-138.
- 유경준(2007b), 「소득불평등도와 양극화」, 정책연구시리즈, 2007-01, 한국개발연구원.
- 이계임·김성용·한혜성(2008), “식품소비지출의 불균등도와 양극화 분석”, 「농촌경제」, 31(3), 1-14.
- 이광상(2021), 「미국 은행들의 FICO 신용평점모델 의존도 축소 및 보강 노력 강화」, 금융브리프, 30(16), 18-20.
- 이성재·이우진(2017), 샤플리값을 이용한 한국의 소득 및 자산 불평등의 원천별 기여도 분석, 「한국경제의 분석패널」, 23(1), 57-109.
- 이천우·김상봉(2019), “최근 고용현황과 고용전망 모형을 이용한 예측”, 「신용카드리뷰」, 13(3), 32-53.
- 이천우·김상봉·김재현(2022), “전자금융거래법 개정과 금융시장의 변화”, 「신용카드리뷰」, 16(2), 38-60.
- 임규민·송명중·김상봉(2021), “코로나19에 의한 국내 경제 파급효과”, 「신용카드리뷰」,

- 15(2), 16-39.
- 임형준(2021), 「신용평가산업 규제 역사와 경쟁정책 개선방안」, KIF 금융분석보고서, 2021-04.
- 통계청(2020), 「가계금융복지조사」 통계정보보고서.
- 홍희주·김상봉(2018), “최고금리 변동에 따른 비은행금융기관 이용자의 배제 규모에 대한 연구”, 「신용카드리뷰」, 12(1), 29-50.
- Aghevli, B. B. and Mehran, F.(1981), Optimal grouping of income distribution data, *Journal of the American Statistical Association*, 76(373), 22-26.
- Altman, E. I.(1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23(4), 189-209.
- Altman, E. I. and Haldeman, R. G.(1995), *Corporate credit-scoring models: approaches and tests for successful implementation*, NYU Working Paper, No. FIN-95-001.
- Altman, E. I., Haldeman, R. G., and Narayanan, P.(1977), ZETA analysis: a new model to identify bankruptcy risk of corporations, *Journal of Banking and Finance*, 1(1), 29-54.
- Amiel, Y. and Cowell, F. A.(1999), *Thinking about inequality*, Cambridge University Press, Cambridge.
- Asako, K. and Uchino, Y.(1987), *Bank loan market of Japan-a new view on the disequilibrium analysis*, BOJ Monetary and Economic Studies, 169-215.
- Allison, P. D.(1978), Measures of inequality, *American Sociological Review*, 43, 865-880.
- Araar, A.(2008), *On the decomposition of polarization indices: illustrations with Chinese and Nigerian household surveys*, Cahiers de recherche 08-06, CIRPEE.
- Araar, A.(2009), *The hybrid multidimensional index of inequality*, Cahier de recherche Working Paper 0945.
- Atkinson, A. B.(1970), On the measurement of inequality, *Journal of Economic Theory*, 2(3), 244-226.
- Barro, R. J.(2008), *Inequality and growth revisited*, ADB Working Paper Series on Regional Economic Integration, No.11.
- Bessis, J.(2002), *Risk management in banking*, 2nd Ed., England, John Wiley & Sons Ltd.
- Bourguignon, F. (1999). *Comment to multidimensional approaches to welfare analysis by Maasoumi, E., in Silber, J. (Ed.)*, Handbook of Income Inequality Measurement, Kluwer Academic, Boston, 477-484
- Capon, N.(1982), Credit scoring system : a critical analysis,” *Journal of Marketing*, 46(2), 82-91.
- Chantreuil, F. and Trannoy, A.(1999), *Inequality decomposition values: the trade-off between marginality and consistency*, THEMA Discussion Paper 99-24.
- Chantreuil, F, and Trannoy, A.(2011), Inequality decomposition values, *Annals of Economics and Statistics*, 13-36.
- Cheney, J. S.(2008), *Alternative data and its use in credit scoring thin- and no-file consumers*, Discussion Paper Payment Cards Center, Federal Reserve Bank of Philadelphia.

- Cowell, F.(1995), *Measuring inequality*, 2nd edition, Prentice Hall/Harveste Wheatsheaf, London.
- Dalton, H.(1920), The measurement of the inequality of incomes, *The Economic Journal*, 30(119), 348-361.
- D'Ambrosio, C.(2001), Social distance between workers, *Economics of Transition*, 9(2), 463-486.
- Davies, J. and Shorrocks, A. F.(1989), Optimal grouping of income and wealth data, *Journal of Econometrics*, 42(1), 97-108.
- Duclos, J. Y., Abdelkrim, A., and Giles, J.(2006), *Chronic and transient poverty: measurement and estimation, with evidence from China*, Working Paper 0611, CIRPEE.
- Duclos, J. Y., Abdelkrim, A., and Fortin, C.(2010), DAD : A software for distributive analysis/analyse distributive, MIMAP programme, Universite Laval, Quebec, Canada.
- Duclos, J. Y., Esteban, J. M., and Ray, D.(2004), Polarization: concepts, measurement, estimation”, *Econometrica*, 72(6), 1737-1772.
- Durand, D.(1941), *Risk elements in consumer installment financing, financial research program, studies in consumer installment financing*, NBER New York.
- Esteban, J. M., Gradin, C., and Ray, D.(1999), *Extensions of a measure of polarization with an application to the income distribution of five OECD countries*, Mimeo, Instituto de Analisis Economico.
- Esteban, J. M., Gradin, C., and Ray, D.(2007), *An Extension of a Measure of Polarization, with an application to the income distribution of five OECD countries*, *The Journal of Economic Inequality*, 5(1), 1-19.
- Esteban, J. M. and Ray, D.(1994), On the measurement of polarization, *Econometrica*, 62(4), 819-851.
- Esteban, J. M. and Ray, D.(1999) Conflict and distribution, *Journal of Economic Theory*, 87(2), 379-415.
- Fields, G. S.(2003), *Accounting for income inequality and its change: am new method, with application to the distribution of earnings in the United States*, in Polachek, S. W. (Ed.), *Research in Labor Economics*, Vol.22, Emerald Group Publishing Limited, Bingley, 1-38.
- Fisher, R. A.(1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, 7(2), 179 - 188.
- Foster, J. and Wolfson, M.(1992), *Polarisation and the decline of the middle Class: Canada and the U.S.*, mimeo, Vanderbilt University.
- Gini, C.(1912), *Variabilita e mutabilita*, Bologna.
- Gini, C.(1936), *On the measure of concentration with special reference to income and statistics*, Colorado College Publication, General Series, 208, 73-79.
- Gradin, C.(2000), Polarization by sub-populations in Spain, 1973-91, *Review of Income and Wealth*, 46(4), 457-474.
- Jenkins, S. P.(1995), Accounting for inequality trends: decomposition analyses for the UK,

- 1971-86, *Economica*, 62(245), 29-63.
- Kakwani, N. C.(1980), *Income inequality and poverty: methods of estimation and policy applications*, Oxford University Press.
- Kakwani, N. C.(1986), *Analyzing redistributive policies: a study using Australian data*, Cambridge University Press, Cambridge.
- Lerman, R. I., and Yitzhaki, S.(1985), Income inequality effects by income source: a new approach and applications to the United States, *Review of Economics and Statistics*, 67(1), 151-156.
- Leonard, K. L.(1995), The development of credit scoring quality measures for consumer credit applications, *International Journal of Quality and Reliability Management*, 12(4), 79-85.
- Lopez, J. H.(2004), *Pro-growth, pro-poor: is there a tradeoff?*, Policy Research Working Paper Series 3378, The World Bank.
- Maasoumi, E.(1986), The measurement and decomposition of multi-dimensional inequality, *Econometrica*, 54(4), 991-997.
- Maasoumi, E.(1999), *Multidimensioned approaches to welfare analysis*, in Silber, J. (Ed.), Handbook of Income Inequality Measurement, Recent Economic Thought Series, Vol.71. Dordrecht, Springer, 437-484.
- Myers, J. H. and Cordner, W.(1957), *Increase credit operations profit*, Credit World, February, 12-13.
- Myers, J. H. and Forgy, E. W.(1962), The development of numerical credit evaluation systems, *Journal of American Statistical Association*, 58(303), 799-806.
- Permanyer, Iñaki(2008), *The measurement of social polarisation in a multi-group context*, UFAE and IAE Working Papers 736.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).
- Podder, N.(1993), The disaggregation of the Cini Coefficient by factor components and its application to Australia, *Review of Income and Wealth*, 39(1), 51-61.
- Podder, N. and Tran-Nam, B.(1991), *Uses and abuses of the decomposition of Cini index by factor components*, The University of New South Wales, mimeo.
- Rao, V. M.(1969), Two decompositions of concentration ratio, *Journal of the Royal Statistical Society*, 82: 418-425.
- Sastre, M. and Trannoy, A.(2022), Shapley inequality decomposition by factor components: some methodological Issues, *Journal of Economics*, 9(1), 51-89.
- Schutz, R. R.(1951), On the measurement of income inequality, *American Economic Review*, 41(1), 107-122.
- Shapley, L. S.(1953), *A Value for n-Person Games*. In: Kuhn, H. and Tucker, A.(Ed.), Contributions to the Theory of Games II, Princeton University Press, Princeton, 307-317.
- Shannon, C.(1948), *The mathematical theory of communication*, Bell System Technical Journal July-Oct.; reprinted in: C. Shannon and W. Weaver, The mathematical theory of communication, 3-91, University of Illinois Press, Urbana, IL.

- Shorrocks, A. F.(1980), The class of additively decomposable inequality measures, *Econometrica*, 48(3), 613-625.
- Shorrocks, A. F.(1982), Inequality decomposition by factor components, *Econometrica*, 50(1), 193-211.
- Shorrocks, A. F.(1984), Inequality decomposition by population subgroup, *Econometrica*, 52(6), 1369-1385.
- Shorrocks, A. F.(1989), Optimal grouping of income and wealth data, *Journal of Econometrics*, 42(1), 97-108.
- Shorrocks, A. F. and Wan, G.(2008), *Ungrouping income distributions: synthesising samples for inequality and poverty analysis*, WIDER Working Paper Series RP2008-16, World Institute for Development Economic Research.
- Singh, S. K. and Maddala, G. S.(1976), A function for size distribution of incomes, *Econometrica*, 44(5), 963-970.
- Theil, H.(1967), *Economics and information theory*, Chicago, Rand McNally and Company.
- Theil, H.(1996), *Studies in global econometrics*, Dordrecht, Kluwer Academic Publishers.
- Tsui, K. Y. and Wang, Y. Q.(1998), *Polarization orderings and new classes of polarization indices*, The Chinese University of Hong Kong University, Working Paper No. 102.
- Turner, M. A. and Agarwal, A.(2008), Using non-traditional data for underwriting loans to thin-file borrowers: evidence, tips and precautions, *Journal of Risk Management in Financial Institutions*, 1(2), 165-180.
- Wan, G.(2004), Accounting for income Inequality in rural China: a regression-based approach, *Journal of Comparative Economics*, 32(2), 348-363.
- Wang, Y. Q. and Tsui, K. Y.(2000) Polarization orderings and new classes of polarization indices, *Journal of Public Economic Theory*, 2(3), 349-363.
- Wolbers, H. L.(1949), *The use of the biographical data blank in predicting good and potentially poor credit risks*, Unpublished MA thesis, University of Southern California.
- Wolfson, M. C.(1994), When inequalities diverge?, *American Economic Review*, 84(2), 353-358.
- Wolfson, M. C.(1997), Divergent inequalities: theory and empirical results, *Review of Income and Wealth*, 43(4), 401-421.
- Zhang, X. and Kanber, R.(2000), What difference do polarization measures make? an application to China, *Journal of Development Studies*, 37(3), 85-98.

A Study on Changes in Financial Markets and Theories of Financial Inequality and Financial Polarization

Sang Bong Kim*

Professor, Department of Economics, Hansung University, Seoul, Korea

—〈Abstract〉—

This study examines the change and expansion of the financial market, and examines the methodology for measuring the financial inequality of financial consumers in the rapidly changing financial environment. It explains what kind of credit information is used for credit evaluation in major countries or in Korea. However, it is quite difficult in many ways to use micro data of credit ratings or credit scores, which are the most basic in the financial field. Changes in the supplier side of the financial market are introduced, and the methodology of what kind of inequality exists in the loan market among financial inequalities is introduced in this study. In addition, methodologies for estimating financial inequality by year are introduced, and decomposition methodologies are presented based on the estimated year-by-year data. Also, since inequality indices do not reflect distribution fluctuations, polarization estimation methodologies and decomposition methodologies are presented as concepts to overcome.

Keywords: Credit information, Credit Scoring System, Financial inequality index, Financial polarisation index, Decomposition

<최초 투고일: 2022년 10월 10일>, <수정일: 2022년 12월 7일>

<게재 확정일: 2022년 12월 7일>

* Address: 505 Research Building, 116 Samseongyo-ro 16gil, Seongbuk-gu, Seoul 02876, Korea, E-mail: brainkim75@hansung.ac.kr, Tel: +82-2-760-8038