

教育扩张与分布型教育不平等*

——复合教育基尼系数的演化性质及其检验

刘精明

提要:学校教育机会具有“得到即满足”的充要特性,教育机会供给扩大意味着机会分布不平等下降。受教育年限的复合基尼系数是各机会基尼系数的加权平均。本文以数学推导和数值模拟证明,复合教育基尼系数随教育扩张而下降是其内在的固有性质。利用 Barro-Lee 世界教育面板数据,本文以实证研究验证了这一基本性质,并建立了教育基尼系数演化形态的解释模型。结果显示,各级教育衔接时序、扩展速率是影响复合教育基尼函数演化的直接因素。

关键词:教育扩张 教育不平等 复合教育基尼系数 数值模拟

一、引言

世界范围内的近现代学校教育是在与封建贵族教育、教会教育等精英教育形式的对抗中建立起来的(Archer, 1979)。教育机会向全体民众开放是近现代学校教育内在的要求,这迎合并进一步激发了人们对教育平等权利的诉求。自联合国《世界人权宣言》将受教育权确立为基本人权以来,现代学校教育在世界各国飞速扩张(Meyer et al., 1992; Benavot et al., 2006; Barro & Lee, 2015; Hannum et al., 2019)。教育扩张对人口、健康、经济、社会、政治、文化等都产生了十分复杂的影响(Hannum & Buchmann 2003),其中最先引起关注并持续发展为相对独立的学术主题的,是教育扩张与教育不平等关系问题。

教育扩张如何影响教育不平等? 相关结论在不同研究传统之间以及各传统

* 本文受国家社科基金重大项目(项目号:19ZDA143)资助。特别感谢清华大学孙凤教授、中南大学吕鹏教授、复旦大学李煜教授在本文的多次讨论中提供宝贵意见和支持。感谢匿名评审专家的批评和建议。感谢李春玲、胡安宁、黄少伍、马元平、罗家德、洪伟等教授及其他专家所给予的批评和讨论。

内部大相径庭。如哈尔西等人在观察英国教育不平等变化时就认为,教育扩张早期教育不平等是上升的,但随后会出现下降(Halsey et al.,1980)。后来拉夫特里与豪特(Rafterly & M. Hout,1993)、卢卡斯(Lucas,2001)等一些学者先后提出了不平等最大化维持(MMI)、不平等有效维持(EMI)等理论。但更多研究发现,随着教育的逐步扩张,教育不平等要么相对稳定,要么减弱或缩小(Erikson & Jonsson,1996;Breen et al.,2009;Blossfeld et al.,2015;等等)。不平等最大化维持理论提出20年后,一项新的研究回应了相应挑战,指出上层阶级的教育即使达到饱和也不会减少教育不平等,各国追求教育扩张的努力对降低教育不平等是一种阻碍(Bar Ham & Shavit,2013)。此外,以不同指标讨论教育扩张对不平等的影响得到的结论也大不相同。这使我们意识到,区分不同的不平等研究传统对理解这一对关系至关重要,因为学者们所谈论的“教育不平等”其实不完全是同一回事。

不平等一般区分为分布型不平等和关系型不平等两种类型。所谓分布型不平等是相对系统均衡或均匀状态而言的,社会科学中特指资源或机会在人口总体中的分布状态,通常以离散统计量或某种“不平等指数”(如基尼系数等)来测量;所谓关系型不平等是指从一定的群体关系或相对结构位置来考察资源、机会的分配结果,不平等程度通常以一个群体相对另一群体的比较优势/劣势来体现。关系型不平等可进一步区分为数量关系型(如量的多寡、阶梯序次等)与具有实质联系的关系型(如剥削与被剥削关系),这是社会学由来已久的分析传统(布劳,1980;Kreckel,1982;孙治本,2001;刘欣、田丰,2018;李路路,2019;李煜,2019)。但有关“教育扩张与教育不平等关系”的讨论常常混淆或者忽略分布型不平等与关系型不平等之间的区分,如MMI理论中的“饱和效应”或“天花板”等术语就体现了这样的混杂特点。两类不平等间的相互关系是亟待拓展的研究主题,而认识分布型不平等的统计性质、特征则更为基本,因为它是我们理解群体间不平等的前提。为此,本文将集中讨论分布型教育不平等指数的建构、性质和特征,以此揭示并检验“教育扩张与教育分布不平等”之间的内在联系。

二、文献回顾

测量分布型教育不平等的指数较多,较为常见的有标准差、教育基尼系数等两类(偶尔也有采用广义熵指数等其他指标的),其中教育基尼系数已成为分析

的主流。文献梳理围绕标准差、教育基尼两类指标而展开。

(1)以标准差测量教育不平等。较早的教育不平等测量经常采用“受教育年限标准差”这一统计量,如拉姆(Rati Ram)和多利乌斯(Shawn F. Dorius)的研究可以说是利用标准差分析教育不平等的典型代表。拉姆通过研究94个国家教育发展水平与教育不平等关系发现,随着平均教育水平的提高,教育不平等(标准差)出现了先上升后下降的趋势(Ram,1990)。多利乌斯利用巴罗与李钟华(Barro & Lee)整理的76个国家和地区1870—2010年的教育扩张数据,考察了国家之间的教育不平衡状况的变化,其基本研究结论是国家间的教育标准差随着人均教育水平的提高经历了一个形似钟形曲线的变化过程(Dorius,2013)。以教育标准差讨论教育扩张与教育不平等关系的文献还有许多(Thomas et al., 2001; Wail et al., 2011; Meschi & Scervini 2014; Shukla & Mishra 2019; 刘精明, 2000),研究结论也都颇为一致,即标准差与平均受教育年限之间呈倒U形曲线关系,教育扩张使教育不平等出现先上升后下降的变化。

(2)以基尼系数测量教育不平等。人们发现,随着平均受教育水平的提高,教育不平等多呈递减趋势(Sheret, 1988; Checchi, 2000; Thomas et al., 2001; Castelló & Doménech, 2002; Wail et al., 2011; Ziesemer, 2016)。另有研究发现教育基尼与教育扩张呈反J形曲线关系,即随着平均受教育水平提高,教育基尼系数呈先下降后上升趋势(Meschi & Scervini, 2014; Shukla & Mishra, 2019)。

文献梳理表明,分布型教育不平等与教育扩张之间的关系形态,很大程度上取决于我们所采用的不平等指数形式。教育扩张对分布型不平等的影响可能并非纯粹经验问题,我们更应关注指数形式、指数的数学性质本身。

三、前提界定和概念操作化

讨论教育扩张条件下的教育不平等变化应先界定两个基本前提。首先,除非极少数例外,现代学校教育机会总是具有“充要满足”特性。也就是说若要满足个体对某种学校教育机会的需求,分配给他/她一次机会是必要的,同时也是充分的。一个人占有两次或更多完全相同的教育机会只会带来个人利益的损害,至少不会增加边际收益。这决定了任何教育机会的投入都将是机会覆盖率的提高。这是教育机会分配与收入、财富分配存在重大差异的地方。

其次,现代学校教育扩张是“循序渐进”的。近现代学校教育体系创建时所

依据的蓝本就是儿童成长过程中循序渐进的阶梯。儿童接受学校教育总是先完成较前阶段的学习,而后才有可能获得较后阶段的受教育机会。在这样的制度框架下,面向大众的现代学校教育扩张必然要求先从最低阶的基础教育开始,然后依次扩张到更高教育阶段。

本文对“教育扩张”概念的操作化定义如下。

(1)机会扩张。机会扩张是教育扩张的基本方式,获得各阶段教育机会者比重的增加都意味着整体性教育水平提高。机会扩张水平(覆盖率)以 λ_k 表示, $\lambda_k \in [0,1]$ 。 λ_k 与常见的教育程度 p_k (各教育水平人口占总体的比例)既有联系,也有很大不同。 p_k 是 k 类竞争性概率,存在 $\sum p_k = 1$,所以它是人口总体教育结构的恰当表达。 λ_k 表示的则是单一机会在总体中的覆盖率。

(2)平均受教育年限的提高。总体中平均受教育年限表示为: $\mu = \sum \lambda_k E_k$ 。 E_k 是各阶段教育的学制年限。

(3)教育全程完成率(x)。由于各国各阶段学制年限(E_k)存在一定差异,“平均受教育年限”指标并不能直接比较不同国家的教育扩张程度,故提出“教育全程完成率”指标。设 E_k 是各教育阶段学制时长, $TE = \sum E_k (k = 1, 2, \dots, K)$ 为特定社会中其教育系统设定的总教育时长。这时教育全程完成率可定义为:

$$x = \frac{\sum \lambda_k E_k}{TE}, k = 1, 2, \dots, K \quad (\text{式 } 1)$$

x 从0到1的变动实际上就是一个完整的教育扩张全程。式1包含了机会扩张 λ_k 与教育全程完成率(x)之间的关系。令 $m_k = E_k/TE$, $\sum m_k = 1$ 。 m_k 即各级教育学制年限的构成比例,因而也是教育体系的结构参数。这时式1可写成:

$$x = \sum m_k \lambda_k \quad (\text{式 } 2)$$

式2说明教育全程完成率 x 是各机会扩张水平 λ_k 的加权平均。

四、复合教育基尼系数的建构

(一)标准差的“是与非”

标准差是对绝对离散状况的测量,但它不满足指数建构的某些公理性原则,

因而较少用于衡量收入或财富不平等(万广华,2004)。然而即便如此,我们还是要问,为何教育标准差与教育扩展水平之间总是存在倒U形曲线关系?

事实上,教育年限与教育机会具有一定程度的等价性质,从而使教育年限标准差与教育机会标准差之间存在内在联系。假设一个社会中存在价值为 E 年的某种教育机会,占人口比例为 λ 的人接受了该教育机会,没有接受者的比例为 $1 - \lambda$ 。不难证明,该社会中个体受教育年限(y_i)的标准差是:

$$sd(y_i) = \sqrt{\lambda(1 - \lambda)} \times E \quad (\text{式 } 3)$$

可见,教育年限标准差是比例方差之平方根与学制年限的乘积。比例方差有这样的性质:随着 λ 由 $0 \rightarrow 1$, $\lambda(1 - \lambda)$ 是先上升后下降的,其轨迹形似“钟形”。由于 E 可以看成是一个常量,所以利用标准差讨论教育不平等变化必然可以观察到教育不平等随着教育机会供给扩大经历先上升后下降的变动过程。

(二)教育基尼系数的传统算法

一般来说,采用基尼相对平均差计算教育基尼系数是一种较基本的算法。

$$G = \frac{\Delta}{2\mu_y} = \frac{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|}{2\mu_y} \quad (\text{式 } 4)$$

这里, y_i 和 y_j 是同一分布中个体 i 和 j 的目标变量(如收入、受教育年限等等), μ_y 为这一分布的均值。它的具体计算通常需要利用个案样本数据。后来有学者利用教育年限的阶段特征以及总教育年限与各阶段教育规模的关系对式4进行了一定改造,各自独立推导出了几近相同的教育基尼系数计算方法(Checchi,2000;Thomas et al.,2001;Castelló & Doménech,2002)。其中托马斯等人(Thomas et al.,2001)提出的教育基尼系数计算方式获得了广泛认可和应用。

$$E_L = \left(\frac{1}{\mu} \right) \sum_{i=2}^n \sum_{j=1}^{i-1} (p_i | V_i - V_j | p_j) \quad (\text{式 } 5)$$

式5即托马斯等人提出的教育基尼系数 E_L 的计算公式,其中 μ 是平均受教育年限, V_i 、 V_j 是达到教育水平 i 、 j 应接受的教育年限, p_i 或 p_j 是达到相应教育水平者的百分比, $\sum p_i = 1$ 。可见,要利用式5计算教育基尼系数,只需收集或计算达到各教育水平人口比例信息即可,数据处理和计算工作大幅简化,但其局限性在于人们很难从该公式去理解教育不平等的构成及其变化方式。

(三)教育机会基尼系数与复合教育基尼系数

1. 教育机会基尼系数

在日常生活中,人们对机会不平等存在直观理解:一种机会如果只有一部分人获得而另一部分人没有获得的话,没有获得就是不平等。有趣的是,下面我们将证明,这种机会分布不平等的基尼系数正好等于没有获得该机会者的比例。

计算基尼系数有两种理论假定,一种以离散分布为基础,另一种以连续分布为基础。以连续分布看,洛伦兹曲线将三角形 OZP 分割成 A 、 B 两部分(见图1),基尼系数表示为 A 部分之面积与三角形 OZP 面积之比: $G = \frac{A}{A+B} = 2A$ 。

然而教育年限是一种特殊的离散数据,它是对教育机会进行量的转换或重新赋值,其本质仍然是“机会”,而不是连续的量。一个人若获得某种教育机会便得到固定的受教育年限(学制时长 E);若没有获得这种机会,量的转化是0。假设人口总体中获得该教育机会的比例为 λ ,未获得的比例为 $1 - \lambda$ 。针对这样一种机会的价值量(教育年限)进行洛伦兹累积,可参照离散分布的累积方式进行(徐宽,2003:762;Thomas et al.,2001:12),如图2所示。将人口按是否获得该教育机会排序,没有获得者排前面,获得者排后面;针对没有获得者,其累积人口比例为 $1 - \lambda$,累积受教育年限为0,洛伦兹累积线与横轴上的线段 OQ 重叠;获得过这种教育机会的人都得到了相同的受教育年限(E),于是,随着受教育人口的增加,受教育年限的累积百分比将匀速递增,因而这部分人的所有洛伦兹点都将落在 QZ 这一线段上。

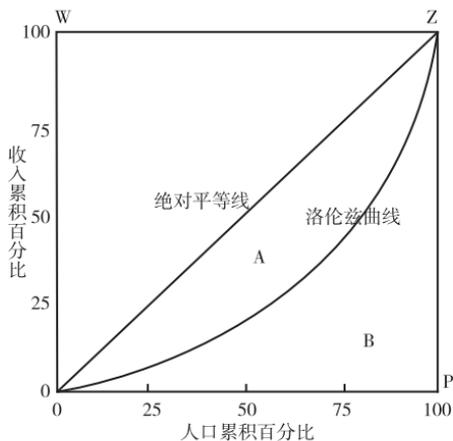


图1 连续分布洛伦兹曲线

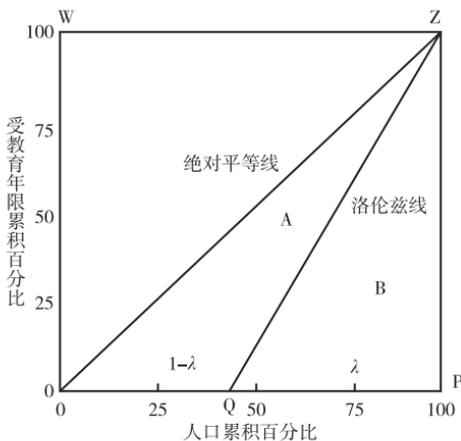


图2 教育机会不平等洛伦兹线

这样,洛伦兹曲线便钝化为由线段 OQ 、 QZ 所构成的折线,洛伦兹几何 A 的面积就是三角形 OQZ 的面积。由于 ΔOQZ 的底 $OQ = 1 - \lambda$,其高 $PZ = 1$,因此,表示该类教育机会分布不平等的基尼系数便可方便地得出:

$$G = \frac{A}{A+B} = 2A = 2 \times S_{\Delta OQZ} = 1 - \lambda \quad (\text{式 } 6)$$

式 6 表明,教育机会分布不平等之基尼系数(G)实际上就是没有获得这种机会者的比例,可称之为“教育机会基尼系数”。显然,在人口总体规模不变情况下,只要教育机会投入有所增加,没有获得该类机会者的比例 $1 - \lambda$ 就会降低。因此,从数学上讲,某种教育机会扩张同时就意味着以基尼系数衡量的机会不平等程度下降。二者实为语义上的同一反复。

式 6 还表明,如果一种教育机会在价值特征、价值量等方面都是同质的话,衡量机会不平等的基尼系数值与该机会的价值特征、价值量无关。比如,若某类高等教育机会是同质的,那么其机会不平等基尼系数就是没有获得过该机会者的比例,无关乎这种机会本身的价值高低,比如教育年限是多少,未来在收入、权力、社会地位等方面的回报有多高。总之,机会基尼系数与机会价值量无关。式 6 完美地简化了我们对教育机会不平等的理解。

2. 复合教育基尼系数:多个机会基尼系数的合成

$G = 1 - \lambda$ 还只是表现了单一机会扩张与不平等的关系,但一个人所受的教育通常由多个阶段、多种多类教育机会所组成。如何利用式 6 来测量由多机会综合而成的“量”的分布不平等,是接下来的工作重心。

(1)按构成要素分解基尼系数。尽管根据人口分组分解不平等指数似乎更为普遍,应用也更广泛,但是,从价值来源或构成来分解基尼系数也具有重要意义(Fei et al., 1978; Lerman & Yitzhaki, 1985; Pyatt et al., 1980; Shorrocks, 1982),国内学者的介绍和相关应用也较为丰富(徐宽, 2003; 万广华, 2004; 戴平生、庄赟, 2012)。不过,以此方法分解教育基尼系数者则鲜少见到。

比较而言,勒曼与伊扎基(Lerman & Yitzhaki, 1985)等提出的分解方法相对直观而简洁。他们首先利用基尼平准差的积分形式将二分之一基尼平均差转换为收入与累积分布之间的协方差,然后将总的协方差再次转换为各分量与总量累积分布之协方差的和,从而得到总量基尼系数与各分量基尼系数之间的关系:

$$G = \sum_{k=1}^K R_k G_k S_k \quad (\text{式 } 7)$$

式7即为勒曼与伊扎基提出的按要素分解的基尼系数公式。其中, G 表示总的非平等基尼系数, R_k 为各分量与总量之秩相关系数, G_k 为各要素本身的分布非平等基尼系数, S_k 则表示各分量占总量的份额。

(2)复合教育基尼系数。现代学校教育体系通常包含多阶段、多种类教育机会。个体最高受教育程度可以看成从不同教育机会中接受的教育时长(E_k)之和。设个体*i*接受的教育总时长为 Y_i ,它可分解为各阶段教育中获得的受教育年限($y_{i,k}$):

$$Y_i = \sum_0^K y_{i,k} = \sum_0^K D_{i,k} E_k \quad (\text{式 } 8)$$

这里, $y_{i,k}$ 表示个体*i*在教育阶段*k*上获得的教育时长, $D_{i,k}$ 为是否接受该教育机会的指示变量, E_k 为学制年限。这样,个体间受教育年限(Y_i)的非平等基尼系数就可以借用式7来计算,式中各符号含义和计算方法如下。

G_k 为各种各类教育形式的机会基尼系数。根据式6,有: $G_k = 1 - \lambda_k$ 。

S_k 为各类教育分量占总量的份额,其计算方式是:

$$S_k = \mu_k / \mu, \text{ 或 } S_k = \lambda_k E_k / \sum \lambda_k E_k \quad (\text{式 } 9)$$

R_k 是各教育分量($y_{i,k}$)与总量(Y_i)的秩相关系数。 $R_k \in [-1, 1]$,但 R_k 取值为1是最常见的情形。这是因为一个人接受教育总是从较低阶段开始,随后循序渐进地逐级完成。获得过较高阶段教育者必定接受过之前更低阶段的教育。因此,当按总量由低到高排序时,各分量也必定由低到高排列(各分量只有两个取值,即0和 E_k),二者之间对应关系可以完全确定,故有 $R_k = 1$ 。^①

至此,式7就简化为:

$$G_E = \sum_{k=1}^K S_k G_k \quad (\text{式 } 10)$$

式10即为本文教育基尼系数的新算法。由于 $\sum_{k=1}^K S_k = 1$,故可将 S_k 看成各类机会基尼(G_k)的权重系数,因而式10还可以表述为:教育基尼系数 G_E 等于各机会基尼(G_k)的加权平均,其权重正好是各机会的“价值量”占“价值总量”的份额(S_k)。故此, G_E 可称为“机会基尼合成系数”或“复合教育基尼系数”。

① 就特定教育机会而言,在两种极端情况下 R_k 的取值为0:(1)当一个社会没有任何人获得该教育机会时;(2)当一个社会所有个体都获得了该教育机会时。这两种情况下分量($y_{k,i}$)均为常数,它与 Y_i 的相关系数为0。不过,即使出现 $R_k = 0$,也不影响式10作为教育基尼系数公式的一般性,因为该分量为0,不参与也不影响对分布 Y_i 的基尼系数的计算。

复合教育基尼系数与托马斯等人(Thomas et al.,2001)提出的教育基尼系数(E_L)在计算结果上是完全等价的。这是因为教育基尼系数 E_L 是利用汇总数据计算的相对基尼系数,或者说是相对基尼系数(式4)的一个特例,而复合教育基尼系数则是按要素分解基尼系数(式7)在各分量与总量之相关系数均为1时的特例,二者本质相同,只是计算方式不同。

五、教育基尼函数的性质与演化形态

单一机会基尼系数 G_k 与机会扩张 λ_k 具有互补关系,接受某种教育机会者的比例 λ_k 从0增长到1,同时也是机会基尼 G_k 从1下降到0的过程。但复合教育基尼系数 G_E 是多个机会基尼系数的加权平均,其取值范围介乎各机会基尼系数(G_1, G_2, \dots, G_K)的最小值与最大值之间。因教育扩张符合循序渐进规律,故而一般存在 $G_1 \leq G_2 \leq \dots \leq G_K$,所以 G_E 的取值范围是: $G_1 \leq G_E \leq G_K$ 。因为 $G_k(\lambda_k)$ 是一个减函数,如果所有 λ_k 都必将扩展至1,那么作为这些减函数的加权平均, $G_E(x|x = \sum m_k \lambda_k)$ 最终也必然降至0,其长期的总体性趋势是下降的;但这并不意味着加权平均数 $G_E(x)$ 在教育扩张全程必然是减函数。各类教育机会的扩张水平、扩张速度、扩张时机共同刻画着复合教育基尼系数的演化形态。

(一)教育基尼函数的特殊形态与性质

复合教育基尼系数 $G_E(x)$ 与教育扩张(x)之间存在复杂的复合函数关系,我们称这类函数为教育基尼函数。这里以一个存在初、中、高三级教育的学校教育体系为例,从一些简单(或特殊)形式入手逐步探讨这类函数的基本性质和形态。

Type I:同步扩张模式。这种特殊模式假定三级教育的扩展进程完全相同,即在发展全程的任何时机(x)上,各级教育机会的扩张水平始终相同[$\lambda_1(x) = \lambda_2(x) = \lambda_3(x)$],因而机会基尼也都相等: $G_1(x) = G_2(x) = G_3(x)$ 。给定 x ,由于有 $S_1(x) + S_2(x) + S_3(x) = 1$,复合教育基尼系数此时可表示为:

$$\begin{aligned} G_E(x) &= S_1(x) G_1(x) + S_2(x) G_2(x) + S_3(x) G_3(x) \\ &= [S_1(x) + S_2(x) + S_3(x)] G(x) \\ &= 1 - x \end{aligned} \quad (\text{式 } 11)$$

可见,在 Type I 模式下(见图 3),若以教育全程完成率 x 为横轴,机会基尼系数 $G_k(x)$ 、复合教育基尼系数 $G_E(x)$ 完全重叠,其值正好是教育全程完成率 x 的补数。

Type II:分段扩张模式。该模式的特征是前一阶段教育全面普及完成后,后一阶段教育才开始启动和发展。如图 3 右图所示,中等教育是在所有人都完成初等教育后才开始,而高等教育又是在中等教育普及完成之后才出现的。

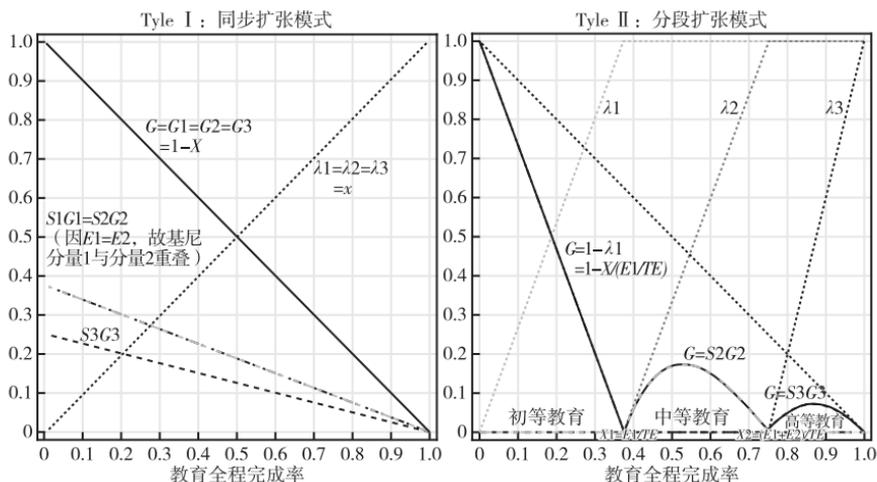


图 3 两种特殊进程中的教育不平等演化

令 $X_1 = E_1/TE, X_2 = (E_1 + E_2)/TE$ 。TE 是教育系统设定的各级教育学制年限之总和。Type II 模式下教育基尼与教育扩张的关系实际上是一个分段函数。

$$G_E(x) = \begin{cases} 1 \times G_1(x) + 0 + 0 = G_1(x), & \text{for: } 0 < x \leq X_1 \\ 0 + S_2(x)G_2(x) + 0 = G_2(x)S_2(x), & \text{for: } X_1 < x \leq X_2 \quad (\text{式 } 12) \\ 0 + 0 + S_3(x)G_3(x) = G_3(x)S_3(x), & \text{for: } X_2 < x \leq 1 \end{cases}$$

式 12 分段函数式表明,复合教育基尼系数在各教育阶段即为各自的基尼分量。如初等教育阶段,中、高等教育阶段(机会)尚未出现,总的教育基尼系数即当前阶段的机会基尼系数: $G_E = 1 - \lambda_1$ 。从教育全程来看,因有 $x = m_1 \lambda_1$ (见式 2),即有 $G_E(x) = 1 - \frac{x}{E_1/TE}$ 。中等教育开始启动并扩张时,初等教育普及化已全部完成、第三级教育尚不会出现,所以总教育基尼系数即为第二阶段教育基尼之分量 $S_2(x)G_2(x)$;同理,高等教育开始启动时,前两个阶段的教育基尼都降到了

0, 此时的复合教育基尼系数也仅是高等教育基尼分量 $S_3(x)G_3(x)$ 。

此时 $S_k(x)$ 的变动方式至关重要。由于 Type II 假定各阶段教育扩张是在前一阶段教育普及完成后才开始, 且其后更高阶段教育均未启动, 这时 λ_k 与 x 之间有着唯一确定对应关系, 因此函数 $S_k(x)$ 可以转化为 $S_k(\lambda_k)$ 来讨论。设较前阶段已完成的平均受教育年限为 E_0 (若本级教育为最低教育阶段如初等教育, 则 $E_0 = 0$), 本教育阶段的学制年限为 E , 令二者之比为 $a = E_0/E$ 。这样, 正在扩张的教育分量 k 占总体的份额可表达为与 λ_k 的函数 (为表述简洁, 舍去下标 k) :

$$S(\lambda) = \frac{\lambda E}{\lambda E + E_0} = \frac{\lambda}{\lambda + a} \tag{式 13}$$

式 13 是倒数双曲函数, 而 λ 是教育机会覆盖率, 故 $S(\lambda)$ 是该双曲线在 $0 \leq \lambda \leq 1, a \geq 0$ (λ 与 a 不同时为 0) 时的一段特殊取值。原点 $(0, 0)$ 是 $S(\lambda)$ 函数的起点, 其取值区间是 $[0, 1)$, $y = 1$ 是其渐进线, a 是其形状参数。为不失一般性, 这时我们可将式 12 各分量基尼写成教育机会扩张 (λ) 的函数:

$$f(\lambda) = S(\lambda)G(\lambda) = \frac{\lambda(1 - \lambda)}{\lambda + a} \tag{式 14}$$

可见, $f(\lambda)$ 可以看成是具有特定取值范围的倒数函数 $S(\lambda)$ 与 λ 的补数 $1 - \lambda$ 的乘积, 因而也是一种双曲函数在 $\lambda \in [0, 1]$ 时的一段特殊取值。因为 $f(\lambda)$ 在 $\lambda \in [0, 1]$ 区间二阶可导, 且 $a \geq 0$ 时 $f''(\lambda) < 0$, 所以它是一个上凸函数; 对 $f(\lambda)$ 一阶求导并寻找驻点, 可获得函数 $f(\lambda)$ 的最大值坐标。将式 14 进一步变形和整理:

$$\begin{aligned} f(\lambda) &= \frac{\lambda(1 - \lambda)}{\lambda + a} = \frac{(\lambda + a) - a}{\lambda + a} - \frac{(\lambda + a)^2 - 2a(\lambda + a) + a^2}{\lambda + a} \\ &= 1 - \frac{a}{\lambda + a} - (\lambda + a) + 2a - \frac{a^2}{\lambda + a} \\ &= (a + 1) - \lambda - a(a + 1)(\lambda + a)^{-1} \end{aligned}$$

则其导函数为: $f'(\lambda) = -1 + a(1 + a)(\lambda + a)^{-2}$

令: $f'(\lambda) = 0$, 有: $\lambda = \pm \sqrt{a(a + 1)} - a$

舍弃负数根。因此可以说, 当 $\lambda_m = \sqrt{a(a + 1)} - a$ 时, $f(\lambda)$ 有最大值。将 λ_m 值代入式 14: $\max_{0 < \lambda \leq 1} f(\lambda) = 1 - 2[\sqrt{a(a + 1)} - a] = 1 - 2\lambda_m$

由 $\max[f(\lambda)] = 1 - 2\lambda_m$ 可知, 应满足 $\lambda_m \leq 1/2$ 以使 $f(\lambda)$ 函数值不为负 (事实上, 若 $a \geq 0$, $f(a) = \sqrt{a(a + 1)} - a$ 的最大值正是 $1/2$)。这说明函数 $f(\lambda)$ 是

一个峰值出现在前 1/2 扩张进程的左偏态倒 U 形曲线, 具有较强典型意义, 可称之为“教育基尼函数”(参见图 4)。

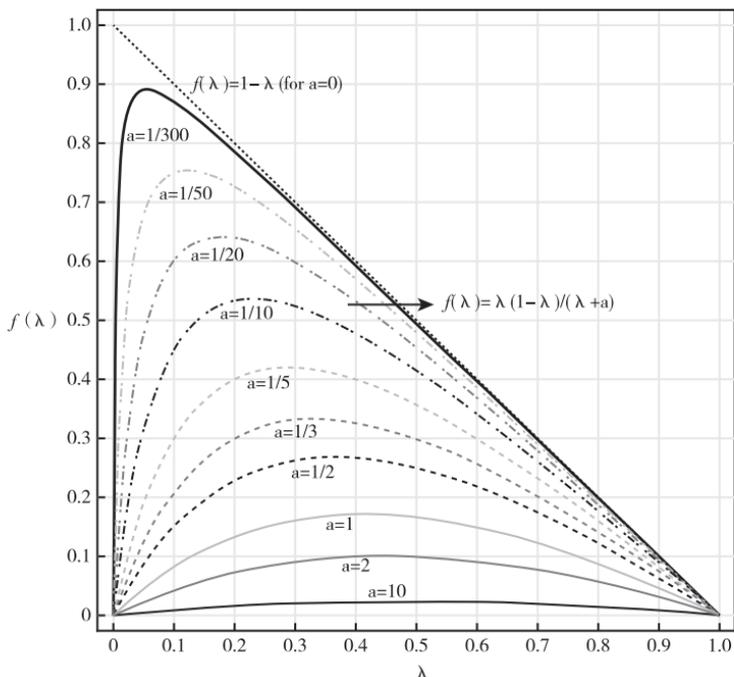


图 4 参数 a 不同取值下的教育基尼函数

结合上面的讨论, 可对式 14 教育基尼函数的性质作进一步归纳。

性质 1: 函数 $f(\lambda)$ 是先上升后下降的左偏态倒 U 形曲线, 其极大值坐标为 $(\sqrt{a(a+1)} - a, 1 - 2[\sqrt{a(a+1)} - a])$, 极大值出现在该阶段教育扩张的前半程; 函数在区间 $(0, \sqrt{a(a+1)} - a]$ 单调递增, 在区间 $[\sqrt{a(a+1)} - a, 1]$ 单调递减。

性质 2: 函数 $f(\lambda)$ 取值范围是 $0 \leq f(\lambda) \leq 1 - \lambda$, 由于它与横轴有 2 个交点 $(0, 0)$ 和 $(1, 0)$, 因此所有函数值均落在直线 $f(\lambda) = 1 - \lambda$ 与两坐标轴所框定的三角区域。这一性质决定了函数 $f(\lambda)$ 总体趋势是下降的, 且最终将归于 0。

性质 3: $a = E_0/E$ 是函数 $f(\lambda)$ 的形状参数, a 的取值越大, 不平等的反弹幅度越小, 反弹峰值出现的时间越晚; 反之, a 的取值越小, 达到的峰值越大, 且峰值越容易出现在本阶段教育扩张的早期。

(二) 复合教育基尼函数的一般形式与性质

Type II 假定在教育全程某个给定时点只存在一种教育机会扩张, 而其他类别教育机会维持原有水平。这既是复合教育基尼函数 $G_E(x)$ 的一种特殊形态, 同时也是 $G_E(x)$ 之分量 $S_k(x) G_k(x)$ 的内在演化轨迹。接下来需要讨论的是, 如果放开对 Type II 的这种约束, 即一个分量演化的同时, 其他分量也按一定方式协同变化, 这时复合教育基尼函数 $G_E(x)$ 的演化会是什么情况?

回顾前文式 2, 有: $m_k = E_k/TE$, $\sum m_k = 1$, 且 $x = \sum m_k \lambda_k(x)$ 。在某个较为确定的时期内, 可假定社会体系中各级教育的学制时长 E_k 相对稳定, 故可将作为结构参数的 m_k 视为一常数。同时应注意到, 针对教育全程完成率 x 的第 k 个基尼分量函数 $f_k(x)$, 与给定 x 条件下相对 λ_k 的基尼分量函数 $f(\lambda_k | x = \sum m_k \lambda_k)$ 是完全等价的。为方便书写, 我们将 $f(\lambda_k | x = \sum m_k \lambda_k)$ 记为 $f(\lambda_k)$ 。这时有:

$$\begin{aligned} f_k(x) &\equiv f(\lambda_k) = S_k(x) G_k(x) \\ &= \frac{\lambda_k(x) E_k}{\mu(x)} \times (1 - \lambda_k(x)) = \frac{\lambda_k(x) E_k}{x \times TE} \times (1 - \lambda_k(x)) \\ &= \frac{m_k [\lambda_k(x) (1 - \lambda_k(x))]}{x} \end{aligned} \tag{式 15}$$

按照教育基尼函数中参数 a 的构造方式, 当放开“其他分量不变”这一约束条件时, 各分量基尼中的参数 a_k 可进一步作如下构造, 令:

$$a_k = (u - u_k) / E_k = (\sum \lambda_k E_k - \lambda_k E_k) / E_k \tag{式 16}$$

a_k 是分量 k 扩张时系统中其他各教育阶段已经完成的教育存量 $(u - u_k)$ 与本阶段预期的教育投入 E_k 的比值, 其解释与参数 a 相同。这时式 15 可写成:

$$\begin{aligned} f(\lambda_k) &= \frac{\lambda_k(x) (1 - \lambda_k(x)) m_k}{\sum m_k \lambda_k(x)} = \frac{\lambda_k(x) (1 - \lambda_k(x)) m_k}{m_k \lambda_k(x) + \frac{u - \lambda_k E_k}{TE}} \\ &= \frac{\lambda_k(x) (1 - \lambda_k(x))}{\lambda_k(x) + \frac{u - \lambda_k E_k}{m_k \times TE}} = \frac{\lambda_k(x) (1 - \lambda_k(x))}{\lambda_k(x) + \frac{u - u_k}{E_k}} \\ f(\lambda_k) &= \frac{\lambda_k(x) (1 - \lambda_k(x))}{\lambda_k(x) + a_k} \end{aligned} \tag{式 17}$$

比较教育基尼函数式 14 可知, 放开约束条件后各分量基尼仍然具有与教育

基尼函数 $f(\lambda)$ 相同的函数形式。又, 由于式 15 和式 17 是等价的:

$$m_k \times \frac{\lambda_k(x)(1-\lambda_k(x))}{x} = \frac{\lambda_k(x)(1-\lambda_k(x))}{\lambda_k(x) + a_k} \quad (\text{式 18})$$

因为结构参数 m_k 是常数, 所以式 18 左边的 $\frac{\lambda_k(x)(1-\lambda_k(x))}{x}$ 与右边的教育基尼函数 $f(\lambda_k)$ 具有相同的性质、形状和形式, 只是取值发生了 m_k 倍的伸缩变化。据此可构造一个新教育基尼函数 $\tilde{f}(\lambda_k)$, 使 $\tilde{f}(\lambda_k) = \frac{1}{m_k} f(\lambda_k)$, 并将式 17 改写成:

$$f(\lambda_k) = \frac{\lambda_k(x)(1-\lambda_k(x))}{\lambda_k(x) + a_k} = m_k \tilde{f}(\lambda_k) \quad (\text{式 19})$$

式 18 两边求和, 复合教育基尼函数 $G_E(x)$ 就存在如下两种等价表达:

$$G_E(x) = \begin{cases} \sum f_k(x) = \frac{\sum [\lambda_k(x)(1-\lambda_k(x))] m_k}{x} & (\text{式 20a}) \\ \sum f(\lambda_k) = \sum \tilde{f}(\lambda_k) m_k & (\text{式 20b}) \end{cases}$$

由式 18~20 可知, $G_E(x)$ 既是给定 x 条件下 k 个比例方差 $\lambda_k(x)(1-\lambda_k(x))$ 的加权平均数与 $\frac{1}{x}$ 的乘积 (式 20a), 也是 k 个教育基尼函数 $\tilde{f}(\lambda_k)$ 的加权平均 (式 20b)。这说明不仅分段函数 $f_k(x)$ 受族函数 $f(\lambda)$ 之性质的约束, $G_E(x)$ 本身也是各水平教育基尼函数 $\tilde{f}(\lambda_k)$ 的加权平均, 因此它的取值、性质和演化形态也都服从 $f(\lambda)$ 函数的约束。也就是说, $G_E(x)$ 就是 $f(\lambda)$ 函数族的特殊成员。

有意思的是, 式 20a 及 $f(\lambda)$ 函数本身 (式 14) 还表明, 复合教育基尼系数、基尼分量与机会分布的比例方差之间具有内在关联。以式 20a 来说, 给定 x 取值, 复合教育基尼系数 $G_E(x)$ 是对各机会比例方差加权平均后的“标准化”。

作为 $f(\lambda)$ 函数族的一员, $G_E(x)$ 至少同样具有如下几点重要性质。

性质 4: 作为多个 $f(\lambda)$ 的加权平均, $G_E(x)$ 与 $f(\lambda)$ 取值范围相同: $0 \leq G_E(x) \leq 1-x$, 即 $G_E(x)$ 所有取值均落在直线 $G_E(x) = 1-x$ 与两坐标轴交叉所构成的三角形区域内。这一性质决定了 $G_E(x)$ 的总体趋势是下降的。

性质 5: 由于 $G_E(x)$ 是各分量函数 $\tilde{f}(\lambda_k)$ 的加权平均, $G_E(x)$ 函数的单调性区间需要根据分量函数之极值点的组合来判断和讨论。

1. $G_E(x)$ 在特定区间的单调性和特定取值特征

根据 $f(\lambda)$ 函数的单调性特征, 给定 $x = \sum m_k \lambda_k(x)$, $k = 1, 2, \dots, K$, 复合基尼函数 $G_E(x)$ 存在如下特殊的单调性变化区间和特定取值。

- (1) 当 $\forall \lim_{x \rightarrow 0} (\lambda_k(x)) = 0$ 时, $\lim_{x \rightarrow 0} (G_E(x)) = 1$;
- (2) 当 $\forall \lim_{x \rightarrow 1} (\lambda_k(x)) = 1$ 时, $\lim_{x \rightarrow 1} (G_E(x)) = 0$;
- (3) 当 $0 \leq \forall (\lambda_k(x)) \leq [\sqrt{a_k(a_k + 1)} - a_k]$ 时, $G_E(x)$ 单调递增;
- (4) 当 $[\sqrt{a_k(a_k + 1)} - a_k] \leq \forall (\lambda_k(x)) \leq 1$ 时, $G_E(x)$ 单调递减。

以上说明, 现代学校教育兴起之初机会分布不平等最大(逼近于 1), 而各阶段教育均趋于普及完成时, 复合基尼最低(逼近于 0); 教育全程内尽管 $G_E(x)$ 可能经历起伏波动, 但总体趋势是下降的, 并会最终完成 $1 \rightarrow 0$ 的演化。此外, 当各分量机会覆盖率均处于函数 $f(\lambda_k)$ 最大值左侧时, 整个教育系统必出现短暂的不平等上升; 各分量基尼均越过函数 $f(\lambda_k)$ 最大值时, $G_E(x)$ 必处于持续下降过程。

2. $G_E(x)$ 随扩张而下降趋势中的反弹和波动

复合函数 $G_E(x)$ 是各分量函数 $f_k(x)$ 之加权平均, 而 $f(\lambda)$ 函数有 1 个极大值和 2 个极小值, 在权重 S_k 与 G_k 的组合方式不断变化的情况下, $G_E(x)$ 理论上就存在出现多个峰值或拐点的可能。 $G_E(x)$ 由 $1 \rightarrow 0$ 的下行过程充满波动起伏。不过现代学校教育有自己的发展规律、惯性和特点, 因此不平等下降过程中的反弹又是可预见的。对此我们将运用蒙特卡洛仿真与数值模拟技术予以呈现。

(三) 复合教育基尼系数的演化形态: 数值模拟

1. 数值模拟模型

由机会基尼系数(式 6)、复合教育基尼系数(式 10)的定义, 以及后来对函数式 14、式 17 ~ 式 20 等的讨论, 均可看出决定复合教育基尼函数演变的核心是各教育机会扩张本身, 而这又涉及 λ_k 的扩展速率、扩展时机及其与其他各阶段教育的组合方式等多种复杂情形。因此, 若要描述复合教育基尼函数的演化形态, 首先应对各级教育机会扩张过程加以恰当设定。机会扩张 λ_k 属于典型的累积概率, 利用 Logistic 分布函数, 我们建构数值模拟的算法模型如下:

$$\lambda_{j,k}(x'_{j,k}) = \frac{1}{1 + \exp(-r_{j,k} \times [K \times x'_{j,k} - (\mu_0 + C_{j,k})])} \quad (\text{式 21})$$

这里下标 j 表示仿真模拟顺序号, $\lambda_{j,k}(x'_{j,k})$ 表示第 j 次模拟时, 教育阶段 k

之累积概率的 Logistic 分布($k=1,2,3$, 分别表示初等、中等、高等教育之层级)。 $x'_{j,k} \in [0,1]$ 是模拟的教育全程完成率, 由使用均匀分布的伪随机函数生成。 K 为教育全程, 按照我国主体学制将其取值为 16 年。 μ_0 是 $\lambda_{j,k}(x'_{j,k})$ 函数中心位置之初始度量值, 根据 Logistic 函数的特征可设定 $\mu_0 = 5$ 。 $C_{j,k}$ 是函数的位置变动参数, 即第 j 次模拟时各分布函数中心位置的移动幅度。这里, $C_{j,k}$ 也是教育扩张的时序, 它取值越大, 表明该阶段教育起步越迟。此外, 若要使分布函数在 $x'_{j,k} \in [0,1]$ 内得到较完整的观察, C 的取值范围应设定为 $0 \leq C_{j,k} < (K - \mu_0)$ 。 $r_{j,k}$ 是形状参数, 也是表示增长速率的近似估计, 其值越大意味着扩张的速率越快。

任意第 j 次模拟时, 各阶段分布函数 $\lambda_{j,k}(x'_{j,k})$ 最初都各自独立, 若要使之成为同一体系内部同时存在的教育扩张, 需要借助前文式 2 来建立它们的联合分布:

$$x_j = \sum m_k \lambda_{j,k}(x'_{j,k}) \quad (\text{式 } 22)$$

最后, 利用式 20a 或式 20b 计算复合教育基尼函数 $GE_j(x)$ 的相应取值:

$$\left. \begin{aligned} GE_j(x) &= \sum \frac{\lambda_{j,k}(x_j)(1 - \lambda_{j,k}(x_j))}{\lambda_{j,k}(x_j) + a_{j,k}}, \text{或} \\ GE_j(x) &= \frac{\sum [\lambda_{j,k}(x_j)(1 - \lambda_{j,k}(x_j))] m_k}{x_j} \end{aligned} \right\} \quad (\text{式 } 23)$$

2. 几种典型形态的参数设置

复合教育基尼函数 $GE_j(x)$ 的演化形态取决于各 Logistic 分布函数 $\lambda_{j,k}(x'_{j,k})$ 的形状参数 $r_{j,k}$ 和位置参数 $C_{j,k}$ 的取值。不同参数设置及其组合存在复杂的系列性流变, 为简明起见, 我们根据理论假设预先设置固定参数, 从而在函数 $GE_j(x)$ 变化流中截取若干帧理论片段, 选取不平等演化的典型形态来加以讨论。

由于 $GE_j(x)$ 是各基尼分量之和, 且 $GE_j(x) \leq 1 - x$, 所以只要各阶段教育扩张并非完全“步调一致”, $GE_j(x)$ 就会呈现“波动”下降。这意味着各级教育之间的发展衔接是影响 $GE_j(x)$ 形态的关键, 而位置参数 $C_{j,k}$ 的设置能较好地反映这种衔接性特征。所以, 在片段截取时, 我们先将速率参数 $r_{j,k}$ 设置为 1, 以观察各级教育在常规发展速度下, 只是由于教育起步时间 ($C_{j,k}$) 不同而产生的形态差异; 特殊形态模式的讨论则同时调整参数 $r_{j,k}$ 。在表 1 中我们尝试列举了 8 种形态的参数设置 (8 次模拟), 每个单元格中有 2 个数字, 分别对应 $r_{j,k}$ 和 $C_{j,k}$ 的取值。

表 1 特定的复合教育基尼函数演化形态的参数设置 ($r_{j,k}, C_{j,k}$)

典型形态 序号	不平等的演化形态 [$GE_j(x)$ 理论片段的模拟]	教育阶段 ($k=1,2,3$)		
		初等教育	中等教育	高等教育
1	Type I:同步扩张模式	1,0	1,0	1,0
2	Type II:分段扩张模式	3,0	3,5	3,8
3	Type IIIa:三级教育分离模式	1,0	1,5	2,8
4	Type IIIb:初等教育分离模式	1,1	1,5	1,6
5	Type IIIc:三级教育有序衔接模式	1,0	1,2	1,4
6	Type IVa:高等教育优先发展(早期)	1,2	1,3	3,0
7	Type IVb:高等教育优先发展(中期)	1,0	1,1	3,1
8	Type IVc:高等教育优先发展(后期)	1,0	1,2	3,5

3. 典型演化形态的模拟结果

根据循序渐进规律,作为由多个偏态抛物线函数 $f(\lambda)$ 加权平均所形成的复合函数, $GE_j(x)$ 首先就是指在有初等教育之不平等下降前提下的变化趋势。根据 $f(\lambda)$ 函数特点,参数 a 越小,极大值(拐点)出现越早,极大值也越高。由于初等教育最初的发展是在没有任何教育存量前提下开展的 ($E_0 = 0$),因而当一个社会按照循序渐进原则发展教育时,初等教育将率先带动 $G_E(x)$ 跨入持续下降的轨迹中。现代学校教育机会扩张都不是一蹴而就的,而是依序释放、逐渐扩大,因此在教育全程完成率 (x) 最初扩张的一段时期内,因 E_k 一定而 E_0 较小,相对所有的 $f_k(x)$ 来说,都会存在较小的 a_k 、较早出现的拐点、较长距离的协同下降。因此,在大多数社会中,当 x 较小时, $G_E(x)$ 都会呈现加速下降模式。如图 5 之各图中,当 x 约小于 0.35 时,其左侧 $G_E(x)$ 的走势大体都呈线性下降形式。这也是当前平均受教育年限较低的大多数低度发展中国家正经历的不平等演化形态。

从教育发展全程看,前文图 3 左图显示的是 Type I 同步扩张模式,三级教育起步时序相同,发展过程中机会累积速率也都相同。因而符合这一条件的参数取值都能使 $GE_j(x) = 1 - x$ (如表 1 中第 1 行参数设置)。图 3 右图的 Type II 分段扩张模式也是一种十分极端的形态,各阶段教育的扩展时时间隔较大,较之初等教育而言,中等、高等教育的起步时间要分别晚 5 年和 8 年,而这种间隔效应同时又被形状参数 $r_{j,k} = 3$ 加以放大(见表 1 中第 2 行参数设置)。

相对 Type II 而言,三级教育分离模式 (Type IIIa) 是一种相对缓和的分段发展形态,其位置参数与 Type II 相同,但形状参数为 1,这是一般扩展速率下先将初等教育发展到接近普及完成水平,然后再依次推动中等教育、高等教育由起

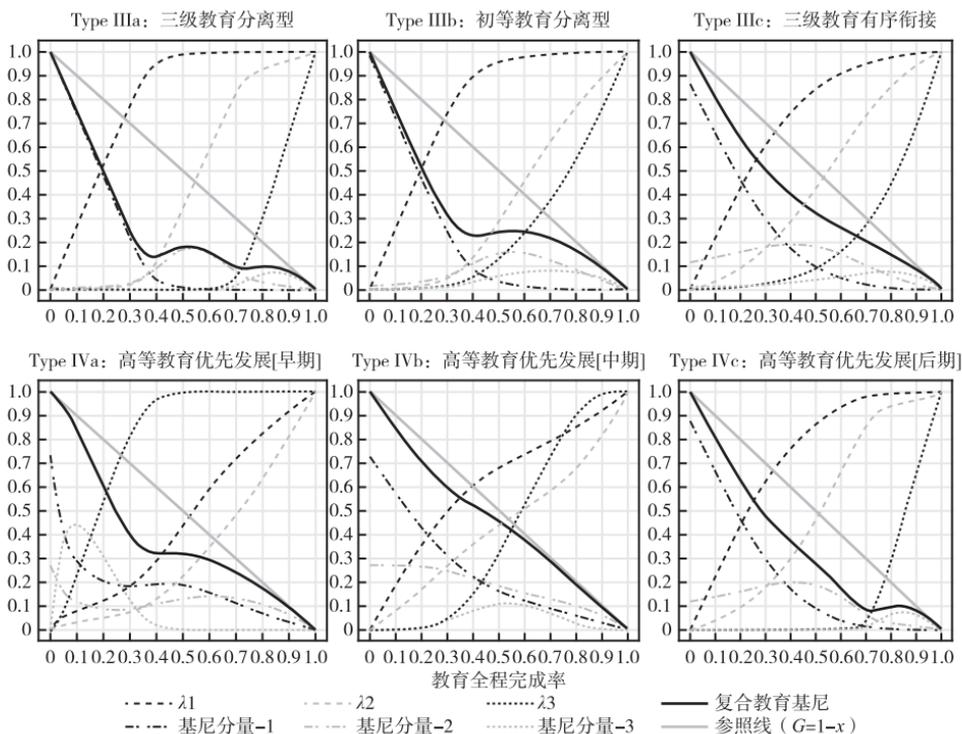


图5 复合教育基尼系数演化典型形态(数值模拟)

步到完成的过程。因此,从图5的Type IIIa可以观察到初等与中等、中等与高等之间均存在一定的不平等反弹。Type IIIb初等教育分离模式与Type IIIa的区别在于初等教育与中等、高等教育间的启动时间分别相差5年、6年,但中等与高等教育间启动间隔相对较小(仅1年),因而在教育扩张全程的1/3至2/3处能观察到 $GE_j(x)$ 的一段上升形态,但随后又单调递降。一些受新教育改革影响较大的西欧国家、北欧五国可能就属于这样的情形。这些国家的初等教育早在19世纪初就已基本完成普及,但中等、高等教育发展直到第二次世界大战结束后才启动,所以这些国家在平均受教育年限不高于中等教育程度时都会经历较大的不平等上升。Type IIIc则为三级教育有序衔接模式,各级教育启动间隔相对较小而教育发展速度大体相当, $GE_j(x)$ 的演化通常不会出现较大的起伏波动。

Type IV呈现了人们更关心的高等教育扩张。扩张时机不同,所引发的复合教育基尼演化形态也有所差异。Type IVa、IVb分别是整个教育发展的早期、中期优先发展高等教育的模型,它的起步时机早于初等或中等教育,并以超常规速

度发展(设定 $r_{j,k}=3$)。在早期发展模型(IVa)中,教育全程完成一半($x=0.5$)左右时,其高等教育覆盖率已接近100%,其基尼分量(S_3G_3)经历了先上升、后下降的一个较大起伏,并拉动了早期总的教育基尼系数的快速下降。Type IVb 则显示,高等教育基尼分量所经历的反弹之幅度相对较小,总的教育基尼系数更贴近参照线 $G=1-x$ 。不过人们应该了解,这两类高等教育发展通常不能持续太久,特别是早期优先发展型。如违背教育的“循序渐进”规律,初等、中等教育不能为高等教育提供充足的合格毕业生,教育扩张必然出现倒退和回缩。印度教育发展历史中反复出现的教育退缩现象就是其鲜明例证。Type IVc 是高等教育扩张相对较晚的情形,这使复合教育基尼系数在教育全程后期出现一定的反弹。

六、复合教育基尼演化特征的实证检验

本文实证研究的主要目标是考察世界各国现代学校教育发展过程中的教育不平等随教育扩张而演化的形态差异,以验证前述推导和仿真模拟所得到的主要研究结论。为此,本文采用一份包含111个国家和地区、跨度140年的教育发展面板数据展开相关实证分析。利用复合教育基尼公式,本文首先计算了各时点教育基尼系数,后续相关分析策略和步骤可简述如下:首先是关于“演化形态”的操作化测量。从数学形式看,演化形态可由基尼函数的曲线特征如曲率、面积、偏斜率、峰度值等来共同概括。但经过前期比较研究,本文最终选取教育基尼函数曲线所分割的面积作为演化形态的关键测量。要得到对面积的有效估计,先须获得教育基尼曲线的精确函数式,为此,利用分数幂多项式回归技术拟合了每个国家和地区教育基尼 GE_j 随教育全程完成率(x)而变动的精确表达式。其次,利用仿真算法模型的变化形式拟合各水平教育的扩展过程(λ_k),从而建立不同国家的教育演进特征变量,如教育起步时点(时序)、教育扩张速率等。最后,以面积为因变量,以教育扩展特征、发展水平等为自变量,建立教育基尼演化形态的解释模型,进而对隐含在仿真模型中的理论判断(影响教育基尼演化形态的主要因素在于教育扩展时序与速率)作出相应的实证检验。

(一) 利用 Barro-Lee 数据计算复合教育基尼系数

本文采用的是巴洛与李钟华在世界银行支持下整理和形成的一份国家与地区教育发展的面板数据(参见 Barro & Lee, 2015)。该数据出现过多个版本,本

文采用的版本为李钟华 2016 版 (Lee & Lee, 2016)。^① 用于本文分析的数据包含 111 个国家和地区 29 个数据时点 (时间跨度为 1870—2010 年, 以 5 年为间隔), 主要变量包括各时点 15~64 岁人口总体的受教育状况。该数据将教育程度区分出四个层级、七个水平。这四个层级分别是未上过学和小学、中学、大学三个教育阶段, 并将三个教育阶段均划分出未完成、完成两个水平 (“未完成”者之学制时长设定为原学制的二分之一)。原数据不仅提供了各教育水平人口的比例, 同时还按初等、中等、高等三级教育分别计算了各学段的平均受教育年限。

利用分组数据计算复合教育基尼系数, 关键在于获得过各水平教育机会覆盖率 (λ_k) 以及各教育阶段的学制时长 (E_k)。李钟华 2016 版数据中七个教育水平人口的比例 p_k 是已知的, 覆盖率可通过公式 $\lambda_k = 1 - \sum_{k=0}^{K-1} p_k$ 加以换算 (p_0 表示未上过学人口比例); 但数据没有直接提供不同国家在各时点的学制信息, 而是给出了人口总体在各阶段的平均受教育年限。本研究根据巴洛与李钟华论文 (Barro & Lee, 2013) 提供的计算公式将学制时长 (E_k) 反算出来。由此, 数据中 111 个国家/地区在 29 个数据时点上的复合教育基尼系数就能方便地计算出来。

图 6 呈现了 111 个国家和地区 29 个数据时点全部的复合教育基尼系数值与教育完成率 (x) 的散点图。图 6 显示, 所有复合教育基尼系数值都位于参照线与两坐标轴所围成的三角形中, 这是对教育基尼函数性质最直观反映。

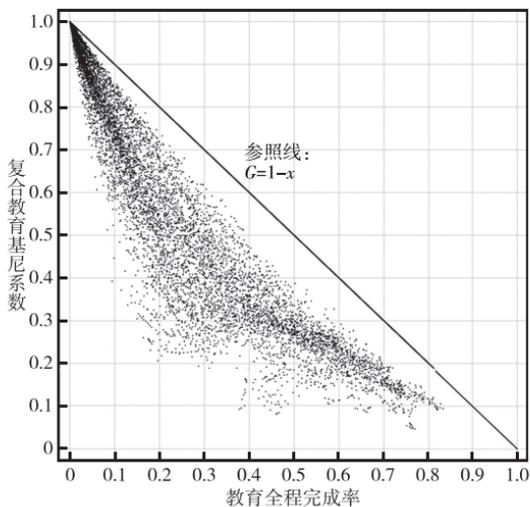


图 6 复合教育基尼系数的演化

^① 数据的详细情况可访问相关介绍网页获取 (<http://www.barrolee.com>)。

(二) 教育基尼演化轨迹的精确拟合

1. 分数幂多项式回归 (fractional polynomial regression)

为计算基尼函数的曲线特征(如曲率、面积等),需要获得基于教育基尼观测数据与教育完成率 x 之间较精确的函数关系。为此,本文运用罗伊斯顿与奥尔特曼提出的分数幂多项式回归技术(Royston & Altman, 1994)来解决此类问题。分数幂多项式回归是一种数据驱动的模式拟合方法,它允许负数次幂、对数、非整数次幂和重复性多项式等参与回归,对精确地拟合非线性关系数据具有很大优势。以本研究来说,各国家或地区均以 GE 为因变量、教育全程完成率 x 为自变量设立分数多项式 OLS 模型(各模型均包含 29 个数据时点),其统一形式为:

$$g(x) = \beta_0 + \sum_{m=1}^M \beta_m H_m(x) \quad (\text{式 24})$$

为与前述基尼函数相区别,分数幂模型对 GE 的估计以 $g(x)$ 表示; $H_m(x) = x^{(p_m)}$ 表示 x 的第 m 个幂次项, $m = 1, 2, \dots, M$ 是幂次项(也称维度)的顺序标记。当多项式 $H_m(x)$ 存在幂次重复项($p_m = p_{m-1}$)时,其构建方式为:

$$H_m(x) = \begin{cases} x^{(p_m)} & \text{if } p_m \neq p_{m-1} \\ H_{m-1}(x) \ln(x) & \text{if } p_m = p_{m-1} \end{cases} \quad (\text{式 25})$$

拟合分数多项式回归时通常需要对 x 进行去中心化处理,这时的分数多项式表述为: $(x^{(p_1, p_2, \dots, p_m)} - c^{(p_1, p_2, \dots, p_m)})' \beta$ 。就本文数据而言,由于各样本地区教育发展差异较大,对教育完成率 x 去中心化处理更有利于相互比较。本研究在拟合所有 111 个国家和地区的模型时,均设定 x 的中心值为 $c = 0.5$ 。利用上述模型技术对 111 个国家(地区)进行幂次项筛选,最终均获得较为理想的模型拟合,108 个模型的拟合优度 R^2 大于 0.98,另外 3 个模型的 R^2 也都大于 0.93。

2. 面积的含义及其计算

分数幂多项式回归使我们获得了较精确的教育基尼函数表达式,从而可产生诸多曲线特征变量,其中之一便是函数曲线对参照线与两坐标轴所围成的三角形面积的分割。前面已指出,参照线($G = 1 - x$)是教育基尼演化的最大边界,而参照线与坐标轴所围成的三角形区域内的点均可以视为因各分量非同步扩张而产生的不平等降低效应。因此,我们将基尼曲线 $g(x)$ 与 x 轴之间的面积(sh)称为基尼函数面积,它表示演化过程中仍然维持的不平等水平,参照线与基尼曲线 $g(x)$ 之间的面积(ss)则称为基尼偏移幅度。可见,教育基尼演化路径偏离参

照线的幅度或基尼函数面积都可当作衡量演化形态差异的辨识变量。

图 7 呈现了捷克和法国教育基尼的演化形态(面积图)。^① 需要注意的是,捷克作为新教运动最早发生的地区之一,其初等教育普及化在 18 世纪中期就达到了较高水平,但 Barro-Lee 数据观测时点是从 1870 年开始的,之前的教育扩张过程并未观察到。基于同样原因,还有其他许多国家如北欧五国、大多数西欧国家等都出现了类似情况。为了比较这些国家教育扩展全程内的不平等演化,作者根据前面的理论逻辑对未观察到的教育基尼系数值以线性回归方法予以估计,如图 7 中虚线所示。这样,面积计算就可求定积分办法来获得:

$$sh = \begin{cases} \int_0^b g(x) dx, & \text{for } a = 0 \\ \int_0^a f_2(x) dx + \int_a^b g(x) dx, & \text{for } a > 0 \end{cases} \quad (\text{式 } 26)$$

$$ss = \int_0^b f_1(x) dx - sh \quad (\text{式 } 27)$$

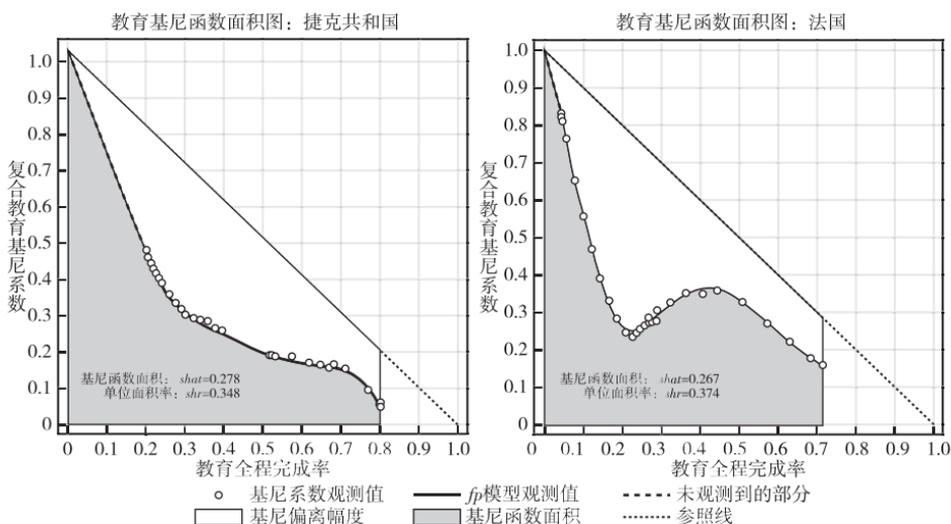


图 7 面积的不平等含义及其比较

这里, a 表示 x (教育扩展全程) 的最小观测值, b 表示 x 的最大观测值, $f_1(x) = 1 - x$ 是参照线, $f_2(x)$ 是对未观察到 GE 的线性估计函数; $g(x)$ 则由分数幂模型(式 24)估计而来。 $g(x)$ 函数中分数幂多项式的原函数可方便获得, 从

① 对其他国家和地区的教育基尼演化图感兴趣的读者可向作者索取: sociu@hotmail.com

而利用定积分方法计算面积也相对容易。面积的计算也可采用蒙特卡洛仿真模拟的方法来估计,事实上本研究同时采用了这两种方法以相互验证其计算结果。

在观察结束时点,不同国家的教育完成率各不相同,而面积大小又与之直接相关。为消除这类影响,使演化形态变量具有可比性,本文进一步构建单位面积率指标作为替代: $shr = sh/b$; $ssr = ssb$ 。 shr 和 ssr 是基尼函数分割参照线下方三角形时同时产生的,都用来定义基尼函数的演化形态,二者面积上互补,但单位面积率的含义相反。一个国家或地区基尼函数的单位面积率(shr)越大,表明其所经历过的教育不平等程度越深;反之,基尼偏移幅度之单位面积率(ssr)越大,历史过程中消减的教育不平等也就越多。

(三) 基于 Logistic 回归估计教育扩展特征

基于 Logistic 模型的数据拟合可使我们获得各国教育扩张的一些重要特征变量,如各教育阶段的起步时序($C_{j,k}$)、扩张速率($r_{j,k}$)等。为利用 Barro-Lee 数据,我们对式 21 进行了一定的形式转换,经过移项、对数转化等处理后有:

$$\ln\left(\frac{\lambda_{j,kt}(x'_{j,t})}{1 - \lambda_{j,kt}(x'_{j,t})}\right) = -r_{j,k}(\mu_0 + C_{j,k}) + r_{j,k}(K \times x'_{j,t}) \quad (\text{式 } 28)$$

这里, $\lambda_{j,kt}$ 表示国家/地区 j 第 k 教育层次在时点 t 的教育覆盖率,它是一个取值为 $0 \leq \lambda_{j,kt} \leq 1$ 的累积概率值。这时式 28 形成了对式 21 的线性转化; $x'_{j,t}$ 为对应的教育完成率观测值;模型中各符号之含义如前节所述。

$$\text{令: } L_{j,kt} = \ln\left(\frac{\lambda_{j,kt}(x'_{j,t})}{1 - \lambda_{j,kt}(x'_{j,t})}\right), X_{j,t} = K \times x'_{j,t}, \alpha_{j,k} = -r_{j,k}(\mu_0 + C_{j,k}), \beta_{j,k} = r_{j,k}, \text{有}$$

$$L_{j,kt} = \alpha_{j,k} + \beta_{j,k} X_{j,t} + \varepsilon_j \quad (\text{式 } 29)$$

式 29 本质上是 $\lambda_{j,kt}$ 的 Logistic 方程,转化后获得了线性模型形式,可用 OLS 估计。利用式 29 对 111 个国家/地区、6 个层次教育的覆盖率进行独立的 OLS 回归(共拟合 666 个模型),即得到各水平教育的起步时序、扩张速率的估计值:

$$\begin{cases} \widehat{C_{j,k}} = -\left(\widehat{\mu_0} + \frac{\widehat{\alpha_{j,k}}}{\widehat{\beta_{j,k}}}\right) \\ \widehat{r_{j,k}} = \widehat{\beta_{j,k}} \end{cases} \quad (\text{式 } 30)$$

OLS 回归后还可通过事后估计公式 $\lambda_{j,kt}^{\wedge} = \exp(L_{j,kt}^{\wedge}) / 1 + \exp(L_{j,kt}^{\wedge})$ 来还原机会覆盖率的拟合值,进而利用式 22、式 23 间接估计出复合教育基尼函数的演化轨迹,不过这已不是此处关注的重点。

(四) 教育基尼演化形态的解释模型

至此,我们利用分数幂回归、Logistic 回归等模型技术,获得了对教育基尼演化形态及解释变量的取值,包括 6 个扩张时序变量 $C1 \sim C6$,六个扩展速率变量 $r1 \sim r6$,它们分别对应于六个层次的教育水平。教育基尼演化形态则以基尼函数曲线与横轴之间的面积 sh 及单位面积率 shr 表示。这些变量的相关系数矩阵显示,演化形态 shr 与扩展时序变量(除 $C1$ 以外)均呈负向关系,与扩展速率呈正相关关系,且相关系数值较大;扩展时序、扩展速率二类参数内部的相关性也较强。6 个扩展速率变量两两间相关系数介乎 $0.88 \sim 0.98$ 。这些特征表明,同一教育体系内部各级教育发展的政策模式具有高度相似性,特别是在各级教育发展速率的选择上高度关联且具有内在的稳定性。

正由于扩展时序、扩展速率这两组变量内部存在高度相关,以此建立演化形态 shr 之解释模型时就容易产生多重共线性问题。为此作者对上述 12 个变量作了进一步归纳,构建了 3 个概念化程度更高的变量,其建构方式如下:

(1) 中等教育与初等教育的衔接时序: $d21 = [(C3 + C4) - (C1 + C2)] / 2$

(2) 高等教育与中等教育的衔接时序: $d32 = [(C5 + C6) - (C3 + C4)] / 2$

(3) 教育扩展速率因子($facR$):这一变量是从 $r1 \sim r6$ 六个变量中析取出 1 个因子而得到的(单一因子析取获得 96.3% 的方差解释)。

这样,我们最终建立教育基尼演化形态的解释模型如下:

$$shr_j = \alpha_j + \beta_1 xmax_j + \beta_2 d21_j + \beta_3 d32_j + \beta_4 facR_j \quad (\text{式 } 31)$$

这里下标 j 表示不同的国家和地区,变量 $xmax$ 是已经达到的教育全程完成率之最大值(也是对教育发展水平的测量)。其他变量的含义如前所述。

表 2 记录了解释模型(式 31)的分步回归结果。拟合优度显示,教育发展水平、教育衔接时序、教育扩展速率对教育基尼演化形态都具有较强的独立解释能力。模型 M4 显示,三个因素对教育基尼演化形态的综合解释能力可达 92.8%,标准化系数反映了各解释变量的重要性顺序。从模型整体结果看,各国/地区复合教育基尼演化形态取决于各级教育发展过程中的衔接时序与扩展速率。

表 2 教育基尼演化形态的解释模型 N = 111

	M1	M2	M3	M4		
				回归系数	标准化系数	VIF
教育发展水平	-0.755 ***			-0.410 ***	-0.513 ***	5.93
中等教育衔接时序		-0.058 ***		-0.023 ***	-0.310 ***	3.40
高等教育衔接时序		-0.022 ***		-0.009 **	-0.093 **	2.72
教育扩展速率			0.122 ***	0.026 ***	0.166 ***	1.34
截距	0.969 ***	0.824 ***	0.566 ***	0.888 ***		
调整 R ²	0.888	0.792	0.605	0.928		

注: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ 。

七、结论与讨论

教育扩张是教育机会分布的一种运动形式。本文主要以基尼系数为教育不平等的测度指标,推导和论证了教育扩张与教育分布不平等之间的函数关系。总体来说,教育扩张即意味着分布型教育不平等下降,这是一种基本规律。复合教育基尼演化形态上的差异形式上源于各级教育发展时序性的不同,但内在的原因还应从一个国家或地区的宗教、历史、文化和经济上的差异中去寻找。

复合教育基尼函数对扩大教育投入、缩小教育不平等方面具有非常明确的政策信号。随着教育的扩张,分布型教育不平等尽管可能出现一定幅度的波动,但从长期趋势看总是处于下降的运动过程中。所以,只要有教育投入的增加,就必然会收到教育不平等下降的功效;教育发展过程中分布不平等的反弹是教育扩张的自然反应,不应归于所谓社会竞争方面的原因。这对世界各国,特别是广大发展中国家制定教育发展政策、实施教育扶贫来说是非常重要的。

分布型不平等是关系型不平等变化的前提和情境条件。以往社会学关于教育扩张与教育不平等关系的讨论并未对两类不平等进行明确区分,教育扩张对教育不平等的影响也较少得到有效测量,有的学者甚至认为教育扩张不必然是降低教育机会不平等的有效工具(Bar Haim & Shavit, 2013)。更多学者发展新的测量工具来分析教育供给扩大与教育机会不平等之间、分布不平等与阶级不平等之间的关系,为此展开的研究在方法和实证上都有先导性作用(Sørensen, 1971; Hu, 2015; Jackson, 2021; Katrňák & Hubatková, 2022)。本文认为,分布型不平等是关系型不平等的前提:一旦机会供给的增加确定下来,复合教育基尼系数就确定了。由于现代学校教育机会具有“得到即满足”的充要满足特性,因此若

不改变教育体系结构,教育机会总量一定,那么无论机会如何在两个不同人群之间分配,都不会改变分布型不平等程度的大小。两种不平等的关系是单向的,只存在“分布型教育不平等的增强或减弱对关系型教育不平等会产生何种影响”的问题,而不会是相反。这种特殊性是将来可以继续探讨的研究议题。

与此相关的另一个研究传统在经济学领域,即有关教育扩张、教育不平等与经济增长、收入不平等之间关系的讨论。显然,如果说教育扩张与教育不平等之间存在一种规律性的变化模式的话,有些逻辑性关联就需要重新审视。经济学文献一般认为,由于存在教育的收入回报差异,教育不平等程度越高的地区收入不平等程度也相对更大(Psacharopoulos, 1985; Knight & Sabot, 1983; Stiglitz, 1973)。但是,对发展中国家来说,教育扩张是以经济增长为背书的,理论上存在“经济增长降低教育不平等”,进而可推导出“经济增长降低收入不平等”这与库兹尼茨理论不同的研究假设。这些理论命题背后的逻辑基础与本文讨论的基本统计规律不无直接关联。

本文讨论的教育基尼系数公式($G_E = \sum S_k G_k$)实际上是将复合教育基尼系数视为各级教育机会基尼的加权平均,其权重即各级各类机会之价值(如教育年限)所占份额。这种理解拓展了我们分析教育不平等问题的思路。比如,以教育年限来计算当前中国社会中的教育不平等,得到的结果是非常低的(1990年出生组中的复合教育基尼系数约为0.12)。但结合当前的社会感受,我们也许不禁会问,如此低的教育不平等为何会产生如此激烈的教育竞争?显然以教育年限为权重标准的计算不足以衡量老百姓心中真实感受到的教育不平等程度。那么,如果将权重的系数体系 S_k 切换到教育在收入回报、地位获得、社会流动等方面的功效的时候,结果又会如何呢?期待未来有更进一步的研究和讨论。

参考文献:

- 布劳,彼得,1991,《不平等的异质性》,王春光、谢圣赞译,北京:中国社会科学出版社。
- 戴平生、庄赞,2012,《基尼系数按收入来源分解的一种新方法》,《统计与决策》第24期。
- 李路路主编,2019,《社会分层与社会流动》,北京:中国人民大学出版社。
- 李煜,2019,《代际社会流动:分析框架与现实》,《浙江学刊》第1期。
- 刘精明,2000,《教育不平等与教育扩张、现代化关系之初探》,《浙江学刊》第4期。
- 刘欣、田丰,2018,《社会结构研究40年:中国社会学研究者的探索》,《江苏社会科学》第4期。
- 孙治本,2001,《社会阶级的终结——水平式不平等、社会不平等的个人主义化》,《政治大学社会学报》第10期。
- 万广华,2004,《收入分配的度量与分解:一个对于研究方法的评介》,《世界经济文汇》第1期。
- 徐宽,2003,《基尼系数的研究文献在过去八十年是如何拓展的》,《经济学(季刊)》第3期。
- Archer, Margaret 1979, *Social Origin of Educational System*. Beverly Hills, Calif.: Sage.
- Bar Haim, Eyal & Yossi Shavit 2013, “Expansion and Inequality of Educational Opportunity: A Comparative

- Study." *Research in Social Stratification and Mobility* 31.
- Barro, Robert J. & Jong - Wha Lee 2013, "A New Data Set of Educational Attainment in the World, 1950 - 2010." *Journal of Development Economics* 104.
- 2015, *Education Matters: Global Schooling Gains from the 19th to the 21st Century*. London: Oxford University Press.
- Benavot, A., J. Resnick & J. Corrales 2006, *Global Educational Expansion: Historical Legacies and Political Obstacles*. Cambridge, MA: American Academy of Arts and Sciences.
- Blossfeld, Pia N., Gwendolin J. Blossfeld & Hans-Peter Blossfeld 2015, "Educational Expansion and Inequalities in Educational Opportunity: Long-Term Changes for East and West Germany." *European Sociological Review* 31(2).
- Breen, Richard, R. Luijkx, W. Müller & Reinhard Pollak 2009, "Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries." *American Journal of Sociology* 114(5).
- Castelló, Amparo & Rafael Doménech 2002, "Human Capital Inequality and Economic Growth: Some New Evidence." *The Economic Journal* 112(478).
- Checchi, Daniele 2000, "Does Educational Achievement Help to Explain Income Inequality?" Working Paper No. 208. UNU-WIDER (<https://www.wider.unu.edu/sites/default/files/wp208.pdf>).
- Dorius, Shawn F. 2013, "The Rise and Fall of Worldwide Education Inequality from 1870 to 2010: Measurement and Trends." *Sociology of Education* 86(2).
- Erikson, Robert & Jan O. Jonsson (eds.) 1996, *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, USA: Westview Press.
- Fei, John C. H., Gustav Ranis & Shirley W. Y. Kuo 1978, "Growth and the Family Distribution of Income by Factor Components." *Quarterly Journal of Economics* 92(1).
- Halsey, H., A. Heath & J. M. Ridge 1980, *Origins and Destinations: Family, Class, and Education in Modern Britain*. Oxford: Oxford Clarendon Press.
- Hannum, E. & C. Buchmann 2003, *The Consequences of Global Educational Expansion: Social Science Perspectives*. American Academy of Arts and Sciences (<https://www.amacad.org/sites/default/files/publication/downloads/Ubase.pdf>).
- Hannum, Emily, Hiroshi Ishida, Hyunjoon Park & Tony Tam 2019, "Education in East Asian Societies: Postwar Expansion and the Evolution of Inequality." *Annual Review of Sociology* 45(1).
- Hu, Anning 2015, "Evaluating Educational Inequality within Educational Expansion: A Formal Comparison between Odds Ratio and the Educational Gini Coefficient." *The Journal of Mathematical Sociology* 39(4).
- Jackson, Michelle 2021, "Expansion, Enrollment, and Inequality of Educational Opportunity." *Sociological Methods and Research* 50.
- Katriák, Tomáš & Barbora Hubatková 2022, "Transitional Inequality Offset: Educational Expansion and Inequality of Educational Opportunity in European Countries between 2000 and 2018." *European Sociological Review* 38(1).
- Knight, J. B. & R. H. Sabot 1983, "Educational Expansion and the Kuznets Effect." *The American Economic Review* 73(5).
- Kreckel, Reinhard 1982, "Class, Status and Power? Begriffliche Grundlagen Für Eine Politische Soziologie Der Sozialen Ungleichheit." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 34.

- Lee, Jong-Wha & Hanol Lee 2016, "Human Capital in the Long Run." *Journal of Development Economics* 122.
- Lerman, Robert I. & Shlomo Yitzhaki 1985, "Income Inequality Effects by Income Source: A New Approach and Applications to the United States." *The Review of Economics and Statistics* 67(1).
- Lucas, Samuel R. 2001, "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." *American Journal of Sociology* 106.
- Meschi, Elena & Francesco Scervini 2014, "Expansion of Schooling and Educational Inequality in Europe: The Educational Kuznets Curve Revisited." *Oxford Economic Papers* 66(3).
- Meyer, John W., Francisco O. Ramirez & Yasemin Nuhoglu Soysal 1992, "World Expansion of Mass Education, 1870 - 1980." *Sociology of Education* 65(2).
- Psacharopoulos, G. 1985, "Returns to Education: A Further International Update and Implications." *The Journal of Human Resources* 20(4).
- Pyatt, Graham, Chau-nan Chen & John C. H. Fei 1980, "The Distribution of Income by Factor Components." *The Quarterly Journal of Economics* 95(3).
- Raftery, Adrain A. & Michael Hout 1993, "Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921 - 75." *Sociology of Education* 66(1).
- Ram, Rati 1990, "Educational Expansion and Schooling Inequality: International Evidence and Some Implications." *The Review of Economics and Statistics* 72(2).
- Royston, Patrick & Douglas G. Altman 1994, "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling." *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 43(3).
- Sheret, Michael 1988, "Using Census Data to Investigate Trends in the Distribution of Education among Individual Persons." *Social Indicators Research* 20(2).
- Shorrocks, A. F. 1982, "Inequality Decomposition by Factor Components." *Econometrica* 50(1).
- Shukla, Vachaspati & Udaya Mishra 2019, "Educational Expansion and Schooling Inequality: Testing Educational Kuznets Curve for India." *Social Indicators Research* 141.
- Sørensen, Aage Bøttger 1971, "Equality of Educational Opportunity in an Expanding Educational System." *Acta Sociologica* 14(3).
- Stiglitz, Joseph E. 1973, "Education and Inequality." *The ANNALS of the American Academy of Political and Social Science* 409(1).
- Thomas, Vinod, Yan Wang & Xibo Fan 2001, "Measuring Education Inequality: Gini Coefficients of Education. SSRN." World Bank Group Workingpapers(<https://doi.org/10.1596/1813-9450-2525>).
- Wail, Benaabdelaali, Said Hanchane & Abdelhak Kamal 2011, "A New Data Set of Educational Inequality in the World, 1950 - 2010: Gini Index of Education by Age Group." *SSRN Electronic Journal* (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1895496).
- Ziesemer, Thomas 2016, "Gini Coefficients of Education for 146 Countries, 1950 - 2010." *Bulletin of Applied Economics* 3(2).

作者单位:清华大学社会科学学院、中国经济社会数据研究中心
责任编辑:杨可