多维经济差异指数在省会城市不均衡性评估中的应用

献给方开泰教授 80 华诞

夏怡凡^{1*}, 张凌², 殷弘³

1 西南财经大学统计学院, 成都 611130;

2 Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA;

3 中国人民大学数学学院, 北京 100872

E-mail: xiayifan@swufe.edu.cn, zlaajj@gmail.com, yinhong@ruc.edu.cn

收稿日期: 2019-04-30; 接受日期: 2019-12-17; 网络出版日期: 2020-05-06; * 通信作者

国家社会科学基金 (批准号: 16BTJ013 和 16ZDA010)、四川省科技项目 (批准号: 2017JY0273) 和四川省社会科学基金 (批准号: SC14B091) 资助项目

摘要 本文基于多资源累积分布, 提出定义在 Lorenz 超曲面上的多维经济差异指数用以衡量多个社会资源分配的不均衡性; 进而设计用样本计算经验多维经济差异指数的算法, 证明该经验指数一致收敛于指数真值. 应用多维经济差异指数对我国各省会城市数据进行了经验分析, 结果在揭示我国省会城市社会资源不均衡特点的同时, 也说明了在复杂经济情势下, 多维经济差异指数相比 Gini 指数能够更加准确地刻画多源信息.

关键词 Gini 指数 人数分布隐函数 省会城市 资源不均衡性

MSC (2010) 主题分类 62F86, 62H86

1 引言

经济学中, 常使用 Gini 指数度量一个国家或地区收入或者财富的统计差异性. Gini 指数最初是意大利统计和社会学家 Gini^[1,2] 基于 Lorenz 曲线定义的, 如图 1 所示, 其中横坐标和纵坐标分别是人的累积比例和对应收入的累积比例. 累积收入比例可以视为累积人口比例的函数, 因此, 如果 Lorenz 曲线是 45 度对角线, 则表明收入分配是绝对均衡的, 即每个人收入都相同; 相反, 如果所有的收入都由最后一个人取得, 其他人收入为 0, 则收入分配是绝对不均衡的, 此时 Lorenz 曲线是图形中下三角的两条直角边. 当然, 实际中的 Lorenz 曲线既不会是绝对均衡的, 也不会是绝对不均衡的, 而是像图 1 显示的一样, 位于两条线之间. 因而, Gini 指数可以用 Lorenz 曲线与 45 度对角线围成的区域面积来定义, 具体可参考 Gini 的定义^[2].

英文引用格式: Xia Y F, Zhang L, Yin H. Application of the multivariate economic dispersion index in inequality evaluation of the provincial capitals (in Chinese). Sci Sin Math, 2020, 50: 679–694, doi: 10.1360/N012019-00118

图 1 (网络版彩图) Lorenz 曲线和 Gini 指数的几何定义

Gini 指数取值在 $[0, 1]$ 之间, 当 Gini 指数取值为 0 时表示收入分配是完全均衡的, 取值为 1 时表明收入分配是完全不均衡的. 之前很多学者的研究表明, Gini 指数在 0.3 至 0.4 之间对社会发展是较为适宜的, 如果指数值超过 0.5, 分配过于不均衡, 社会有不稳定的风险; 如果值低于 0.2, 分配过于均衡, 人们会缺乏创造财富的动力.

Gini 指数被提出后, 对于 Gini 指数的估算方法被不断提出. 例如, Kendall 和 Stuart^[3]、Sen^[4]、Anand^[5] 和 Jasso^[6] 分别提出了不同的估算方法, 而 Pyatt^[7] 和 Silber^[8] 应用矩阵方法改进了指数计算的效率. 现在 Gini 指数在描述社会和经济问题的均衡性方面发挥了越来越重要的作用, 而经济和社会不均衡性的指数研究也成为经济学研究领域的热点之一. 现有的文献已经讨论了这些指数的理论改进和实证应用. Atkinson^[9] 提出了 Atkinson 指数用于衡量总体分布中是左侧还是右侧对不均衡性影响更大, Imedio-Olmedo 等^[10] 对 Bonferroni 指数类在度量不均衡性方面的效率进行了比较; Sundrum^[11] 推导了分组数据下指数计算的分解方法; Glenn^[12] 则提出了用距离收入均值的平均差来度量不均衡性的指数方法, 平均差越大, 不均衡性就越大. 相比于 Gini 指数, Greselin^[13] 和 Greselin 等^[14, 15] 提出了 Zenga 指数、 L -函数和其他统计量去衡量经济的不均衡性和实际风险. 以上提到的所有指数都可以看成是 Gini 指数的改进. 另外, 也有学者提出多维贫困指数 (参见文献 [16–18]), 这类指数定义在“贫困人口”集上, 表述为各维度“贫困人口”比例的求和或乘积, 该定义虽然与 Gini 指数定义不同, 但也借鉴 Gini 指数的定义, 在每个维度上定义“贫困指数”, 然后再进行组合.

在进行理论研究的同时, 更多学者关注 Gini 指数的应用和阐释. 例如, 文献 [19] 估算 2005 年全球收入 Gini 指数值为 0.61 至 0.68; Kakwani^[20, 21] 提出 Gini 指数可用于测度贫困程度, 并与其他测度贫困的指标进行了比较. Gini 指数还被广泛用于社会学、经济学、健康学、生态学、工程科学和农学. 例如, 在社会学和经济学领域, Shorrocks^[22] 介绍了一种基于收入 Gini 指数的测度来估计收入的流动性. 这种测度后来被 Maasoumi 和 Zandvakili^[23] 推广, 现在一般被称为 Shorrocks 指数. Thomas 等^[24] 提出了教育 Gini 指数, 用于透过教育不均衡度, 了解社会发展的趋势. Roemer^[25]、Roemer 等^[26] 和 Weymark^[27] 提出了机遇 Gini 指数的概念. Sadras 和 Bongiovanni^[28] 利用 Lorenz 曲线和 Gini 指数对围场产量的不均衡性进行了评价. 进一步讲, 给定时间区间后, Gini 指数可以用于比较不同国家和地区, 或者是同一国家不同类, 例如, 不同州、教育程度、性别和种族的不均衡性. Kopczyk 等^[29] 应用

美国 1937 年以来的社会安全收入数据计算 Shorrocks 指数并推导出: 在美国, 收入的流动性有复杂的历史. 这主要是因为第二次世界大战后大量女性进入劳动力市场, 以及 1937 年至 2000 年间美国男女工人收入差异. van Mierlo 等^[30] 将 Gini 指数用于数字社交网络不均衡性的度量.

虽然 Gini 指数描述了收入或者财富的不均衡程度, 这确实是社会资源的重要环节, 但是社会资源的不均衡性并不仅仅等同于收入的不均衡性, 还需要考虑其他社会资源, 如资产、社会福利、公共服务和教育支出等. 当考虑多变量不均衡性测度时, Gini 指数变得不再适用, 原因是 Gini 指数理论上的定义是局限在一维分布上的, 这就使得该指数很难处理多维和高维数据. 如果一定要计算 Gini 指数, 那么多维数据必须通过一定的方法 (如主成分分析) 降到一维才行. Gini 指数推广的主要问题是当维数大于等于 2 时, 分布函数的逆函数不存在, 这点将在第 2 节中详细讨论. 而文献 [16–18] 中的“多维贫困指数”, 基本是基于有限集, 即样本来定义的, 缺乏统计理论上的定义和讨论. 这个问题可以通过引入一个与 Gini 指数等价的定义来解决. 如果该指数能够应用于不均衡性的测算, 那之前提及的一些测度 (如 Boferroni 指数等) 也可以推广到多维数据. 因此, 将这些指数定义域从一维扩展到多维的思想具有重要价值.

本文重点关注将 Gini 指数扩展使之能够处理多源数据. 本文将定义一种新指数, 命名为多维经济差异指数 (multivariate-deviation index, MI). 新指数能同时归纳不同经济变量, 例如, 收入、支出和资产等的不均衡性. 因此, 与 Gini 指数相比, MI 不仅仅是维数拓展的改进, 而且更加适用于复杂的经济问题. 本文进而设计了多维经济差异指数的计算方法. 在理论分析之后, 本文还进行了数值模拟和实证应用, 比较 Gini 指数和 MI 在中国省会城市调查数据实证结果中的差异.

2 多维经济差异指数

2.1 Gini 指数新表示

Gini 指数如图 1 所示, 令总体收入是随机变量 x , 分布函数是 $F(t)$, $L(t)$ 对应收入为 t 时累积收入的比例, 即收入低于 t 的人, 其收入总和在总收入中的占比,

$$L(t) = \frac{1}{\mu} \int_{-\infty}^t x dF(x) = \frac{1}{\mu} \int_{-\infty}^t x f(x) dx, \quad (2.1)$$

其中 $\mu = \int_0^{\infty} t dF(t)$ 表示收入均值. $F(t)$ 和 $L(t)$ 都是在 $[0, 1]$ 取值, Lorenz 曲线正是由这些 $[0, 1] \otimes [0, 1]$ 中的点 $(F(t), L(t))$ ($t \in [0, \infty]$) 构成, 可以看成隐函数 $\mathcal{L}(p)$,

$$y = \mathcal{L}(p) : \begin{cases} p = F(t), \\ y = L(t), \end{cases} \quad (2.2)$$

其中 p 和 y 分别表示累积人口和对应的累积收入比例.

现在考虑二维的情形. 记 x_1 、 x_2 和 $F(x_1, x_2)$ 为两个社会资源及其联合分布函数, 联合分布函数有两个自变量, 逆函数不存在, 因此二维情形下推导不出隐函数. 在 Gini 指数的框架下很难研究二维或更高维的经济指数问题. 然而, 可以在另外一个替代框架下考虑经济指数问题, 为了阐释这个新的替代框架, 首先在此框架下重新定义 Gini 指数. 如图 2 所示, 交换图 1 中横坐标和纵坐标, 用 $\mathcal{F}(y)$ 记由

图 2 (网络版彩图) Lorenz 曲线和 Gini 指数的另一种表示

点 $(L(t), F(t))$ ($t \in [0, \infty]$) 构成的人数分布隐函数 (implicit population proportion function, IPPF),

$$\pi = \mathcal{F}(y) : \begin{cases} \pi = F(t), \\ y = L(t), \end{cases} \quad (2.3)$$

$\mathcal{F}(y)$ 描述了按收入由低到高排序, 累积收入占总收入的比例为 y 时, 对应的累积人口的比例为 π , 也就是说, 总共 π 的人口拥有的累积收入占总收入的比例为 y . $\mathcal{F}(y) = F(t)$, 因为它们表达的是同一个累积概率, 只是自变量定义域分别为 $[0, 1]$ 和 $\mathbb{R}$. 事实上, 因为 $F(t)$ 和 $L(t)$ 都是严格递增函数, t 和 y 是一一对应的, 如 (2.1) 所示.

因为图 2 中 B 区域位于左上部分, 所以, Gini 指数可以如下推导:

$$\text{Gini} = 2 \int_0^1 \mathcal{F}(y) dy - 1. \quad (2.4)$$

从图 1 和 2 中可看出, (2.4) 提供了一个新的框架来阐释 Gini 指数.

2.2 二维经济差异指数

本小节可以在上述框架下将 Gini 指数拓展成多维经济差异指数. 为简单说明思想, 仍然考虑二维的情形, 记 $F(t_1, t_2)$ 为两个社会资源的联合分布, 联合人数分布隐函数记为 $\mathcal{F}(y_1, y_2)$, 其自变量为第一种社会资源的累积比例 $y_1 = L_1(t_1)$ 和第二种社会资源的累积比例 $y_2 = L_2(t_2)$. $\mathcal{F}(y_1, y_2)$ 是指同时满足两个条件的人数累积比例. 联合人数分布隐函数可以表示为

$$\pi = \mathcal{F}(y_1, y_2) : \begin{cases} \pi = F(t_1, t_2), \\ y_1 = L_1(t_1) = \frac{1}{\mu_1} \int_{-\infty}^{t_1} \int_{-\infty}^{\infty} x_1 dF(x_1, x_2), \\ y_2 = L_2(t_2) = \frac{1}{\mu_2} \int_{-\infty}^{t_2} \int_{-\infty}^{\infty} x_2 dF(x_1, x_2), \end{cases} \quad (2.5)$$

其中 μ_1 和 μ_2 分别是两个社会资源的均值. 在此框架下, 比例 L_1 和 L_2 分别是 t_1 和 t_2 的一元函数, $\mathcal{F}(y_1, y_2)$ 描述了定义在单位矩形上的 Lorenz 曲面, 其取值与 $F(t_1, t_2)$ 相同, 其中 t_1 和 t_2 分别对应 y_1 和 y_2 . 多维经济差异指数可以按下式定义:

$$\mathbf{MI}^2 = \int_0^1 \int_0^1 \mathcal{F}(y_1, y_2) dy_1 dy_2, \quad (2.6)$$

其中上标“2”表示 2 维指数. 自变量为比例 y_1 和 y_2 , 接着利用二维 Lorenz 曲面上的积分完成二维经济差异指数的定义. 从几何上直观看, 二维经济差异指数就是在单位立方体内曲面 $\mathcal{F}(y_1, y_2)$ 下的体积. 显然, 多维经济差异指数是 Gini 指数在多维的延伸和拓展. 当只考虑一维时, 经济差异指数与 Gini 指数是等价的, 根据 (2.4), 它们之间存在如下转化关系:

$$\mathbf{Gini} = 2 * \mathbf{MI} - 1. \quad (2.7)$$

当资源随机变量为离散型分布且其均值有限时, 可类似得到二维经济差异指数的定义. 令 $f(t_{1,i}, t_{2,j})$ ($i = 1, 2, \dots, r_1, j = 1, 2, \dots, r_2$) 是随机向量 $(x_1, x_2)^T$ 的联合分布律, x_1 和 x_2 的边际分布律由下式确定:

$$f_1(t_{1,i}) = \sum_{j=1}^{r_2} f(t_{1,i}, t_{2,j}), \quad i = 1, 2, \dots, r_1, \quad (2.8)$$

$$f_2(t_{2,j}) = \sum_{i=1}^{r_1} f(t_{1,i}, t_{2,j}), \quad j = 1, 2, \dots, r_2. \quad (2.9)$$

对应 $t_{m,i}$ 的人数比例 $y_{m,i}$ ($m = 1, 2$) 由下式定义:

$$y_{m,i} = \frac{1}{\mu_m} \sum_{k=1}^i t_{m,k} f_m(t_{m,k}), \quad m = 1, 2, \quad j = 1, 2, \dots, r_m, \quad (2.10)$$

其中 μ_1 和 μ_2 是随机变量的均值. 人数分布隐函数的定义如 (2.5), 则离散分布下二维经济差异指数可由下式计算:

$$\mathbf{MI}^2 = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \mathcal{F}(y_{1,i}, y_{2,j}) \Delta y_{1,i} \Delta y_{2,j}, \quad (2.11)$$

其中 $\Delta y_{m,i} = y_{m,i} - y_{m,i-1}$, $y_{m,0} = 0$, $m = 1, 2$.

2.3 多维经济差异指数

多维经济差异指数可以类似定义, 考虑多维经济资源向量 $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$, 如果向量的联合分布函数 $F(t_1, t_2, \dots, t_p)$ 和相应的第 j 个分量的边际分布函数 $F_j(t_j)$ 已知, 那么 (2.12) 可推导出联合人数分布隐函数, 描述超立方体内的 Lorenz 超曲面

$$\pi = \mathcal{F}(y_1, y_2, \dots, y_p) : \begin{cases} \pi = F(t_1, t_2, \dots, t_p), \\ y_j = L_j(t_j) = \frac{1}{\mu_j} \int_{-\infty}^{t_j} x_j dF_j(x_j), \quad j = 1, 2, \dots, p, \end{cases} \quad (2.12)$$

其中 μ_j 表示第 j 个分量的均值, y_j 表示对应资源 t_j 的累积值在总资源中的占比. 多维经济差异指数的定义可由下式得出:

$$\mathbf{MI}^p = \int_{[0,1]^{\otimes p}} \mathcal{F}(y_1, y_2, \dots, y_p) dy_1 dy_2 \cdots dy_p, \quad (2.13)$$

其中上标 “ p ” 表示 p 维指数, $[0,1]^{\otimes p}$ 表示在实数空间 $\mathbb{R}^p$ 中的 p 维超立方体. 类似地, 在离散分布情形下, 可以由下式计算多维经济差异指数:

$$\mathbf{MI}^p = \sum_{i_1, \dots, i_p} \mathcal{F}(y_{1,i_1}, y_{2,i_2}, \dots, y_{p,i_p}) \Delta y_{1,i_1} \cdots \Delta y_{p,i_p}. \quad (2.14)$$

多维经济差异指数具有良好的统计性质, 下面进行讨论.

性质 1 当资源向量 $\mathbf{x}$ 能被分解为 m 个相互独立的子向量, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)^T$, 子向量 $\mathbf{x}_l$ 含 k_l ($l = 1, 2, \dots, m, \sum_{l=1}^m k_l = p$) 个随机分量时, 整个向量的多维经济差异指数等于各子向量的多维经济差异指数的乘积:

$$\mathbf{MI}^p = \prod_{l=1}^m \mathbf{MI}_l^{k_l}. \quad (2.15)$$

特别地, 如果资源向量中各分量是独立同分布的, 则向量的多维经济差异指数等于单个分量经济差异指数的 p 次方:

$$\mathbf{MI}^p = (\mathbf{MI})^p. \quad (2.16)$$

性质 1 的证明可以从定义显然得到. 这是因为在独立性条件下, 联合分布函数可以写成各边际分布函数的乘积, 从而联合人数分布隐函数也可以写成边际人数分布隐函数的乘积. 根据定积分的性质, 可以得出性质 1.

性质 2 在独立同分布条件下, 若资源向量各分量都处于完全均衡情形, 多维经济差异指数取值为 $(\frac{1}{2})^p$; 若各资源分量都处于完全不均衡情形, 多维经济差异指数取值为 1, 即多维经济差异指数在各分量独立同分布条件下, 取值区间为 $[(\frac{1}{2})^p, 1]$.

如图 2 所示, 在完全均衡情形下, 单个分量的经济差异指数等于 $\frac{1}{2}$; 在完全不均衡情形下, 单个分量的指数等于 1. 因此, 性质 2 可以由性质 1 得到.

性质 3 在同分布完全相依条件下, 若资源向量各分量都处于完全均衡情形, 多维经济差异指数取值为 $\frac{1}{p+1}$; 若各资源分量都处于完全不均衡情形, 多维经济差异指数取值为 1, 即多维经济差异指数在各分量同分布完全相依条件下, 取值区间为 $[\frac{1}{p+1}, 1]$.

性质 3 也容易证明, 因为在同分布完全相依条件下, 资源向量联合分布等于各边际分布的最小值, 即

$$F(t_1, t_2, \dots, t_p) = \min_{j=1,2,\dots,p} F_j(t_j).$$

而在完全均衡情形下, 各分量的边际人数分布隐函数都是 $\mathcal{F}_j(y_j) = y_j$, 从而,

$$\mathbf{MI}^p = \int_{[0,1]^{\otimes p}} \min\{y_1, y_2, \dots, y_p\} dy_1 \cdots dy_p = \frac{p!}{(p+1)!} = \frac{1}{p+1}. \quad (2.17)$$

文献 [16, 17] 提出了用多维贫困指数测度多个资源的贫困不均衡性. 比较而言, 经济差异指数具有以下优势. 第一, 经济差异指数是基于分布定义而非有限样本集, 因而, 无论资源是以连续分布、离

散分布或者样本形式给出都可以计算, 而多维贫困指数只能是基于样本计算. 第二, 经济差异指数考虑联合分布函数, 包含了多个资源分布的全部信息, 包括相关性、尾部相依性和偏态等, 而这些在多维贫困指数中都没有涉及. 更重要的是, 经济中资源变量 (如收入、资产、消费等) 的分布多显现“尖峰厚尾”特征, 采用多维经济差异指数能捕捉不均衡性的更多信息, 可以整合经济社会资源更多信息, 充分考虑资源之间的相关性及其他高阶矩性质. 因此, 该指数更加适合在复杂经济问题中探索不均衡性, 特别是多变量问题的不均衡性.

3 样本经验多维经济差异指数及性质

3.1 样本多维经济差异指数

本小节主要讨论基于样本, 定义经验多维经济差异指数, 并给出估计算法. 假定随机抽取社会资源向量 $\mathbf{x}$ 的 n 个独立样本, 得到样本资料矩阵 $X = (x_{ij})_{n \times p}$, 向量 $\mathbf{x}$ 的经验联合分布函数由下式给出:

$$\hat{\pi} = \hat{F}_n(t_1, t_2, \dots, t_p) = \begin{cases} \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^p \mathbf{1}_{(-\infty, t_j)}(x_{ij}), \\ 0, \quad \text{其他}, \end{cases} \quad (3.1)$$

其中 $\mathbf{1}_A$ 表示事件 A 的示性函数. 对于给定的 t_j , 如果样本 i 的第 j 个分量值 $x_{ij} \leq t_j$, 那么示性函数 $\mathbf{1}_{(-\infty, t_j)}(x_{ij})$ 等于 1, 而当条件 $x_{i1} \leq t_1, x_{i2} \leq t_2, \dots, x_{ip} \leq t_p$ 同时满足时, 示性函数的乘积 $\prod_{j=1}^p \mathbf{1}_{(-\infty, t_j)}(x_{ij})$ 等于 1. 经验联合人数分布隐函数 $\hat{\mathcal{F}}_n(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_p)$ 由 (3.2) 定义, $\hat{y}_j$ 表示社会资源向量各分量的样本比例,

$$\hat{\pi} = \hat{\mathcal{F}}_n(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_p) : \begin{cases} \hat{\pi} = \hat{F}_n(t_1, t_2, \dots, t_n), \\ \hat{y}_j = \hat{L}_{nj}(t_j) = \frac{1}{n\hat{\mu}_j} \sum_{i=1}^n x_{ij} \mathbf{1}_{(-\infty, t_j)}(x_{ij}), \quad j = 1, 2, \dots, p, \end{cases} \quad (3.2)$$

其中 $\hat{\mu}_j$ 是第 j 个分量的样本均值,

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = 1, 2, \dots, p. \quad (3.3)$$

基于样本, 为计算经验多维经济差异指数的值, 首先需要确定由样本构成的网格点上的取值. 令 $x_{(i)j}$ ($i = 1, 2, \dots, n$) 记样本资料阵 X 第 j 个分量 n 个样本值升序排列,

$$x_{(1)j} \leq x_{(2)j} \leq \dots \leq x_{(n)j}, \quad j = 1, 2, \dots, p.$$

这些升序排列后的样本值构成 $\mathbb{R}^p$ 中网格上的点, 可以用之前定义的 (3.2) 计算每个网格点上的经验比例 $\hat{y}_{i,j}$,

$$\hat{y}_{i,j} = \hat{L}_{nj}(x_{(i)j}), \quad \Delta y_{i,j} = \hat{y}_{i,j} - \hat{y}_{i-1,j}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p, \quad (3.4)$$

其中 $y_{0,j} = 0, j = 1, 2, \dots, p$. 这样, 可以得到 p 维超立方体 $[0, 1]^{\otimes p}$ 中由 np 个节点构成的网格 $(\hat{y}_{i_1,1}, \hat{y}_{i_2,2}, \dots, \hat{y}_{i_p,p})$.

经验联合人数分布隐函数在网格点上的取值 $\hat{\mathcal{F}}_n(\hat{y}_{1,j_1}, \hat{y}_{2,j_2}, \dots, \hat{y}_{p,j_p})$ 可以用于计算经验多维经济差异指数 $\widehat{\mathbf{MI}}$. 将 (3.1)–(3.4) 代入 (2.13) 中, 得到计算 p 维该指数 $\widehat{\mathbf{MI}}^p$ 的算法,

$$\widehat{\mathbf{MI}}^p = \sum_{i_1=1}^n \cdots \sum_{i_p=1}^n \hat{\mathcal{F}}_n(\hat{y}_{i_1,1}, \hat{y}_{i_2,2}, \dots, \hat{y}_{i_p,p}) \Delta y_{i_1,1} \cdots \Delta y_{i_p,p}. \quad (3.5)$$

3.2 经验多维经济差异指数的变异性 and 一致性

令 $\Theta(\mathcal{F}) = \int_{[0,1]^{\otimes p}} \mathcal{F}(\mathbf{y}) dy_1 dy_2 \cdots dy_p$, 其中 $\mathbf{y} = (y_1, y_2, \dots, y_p)^T$, 并应用积分中值定理, 可得

$$\widehat{\mathbf{MI}}^p - \mathbf{MI}^p = \Theta(\hat{\mathcal{F}}_n) - \Theta(\mathcal{F}) = \hat{\mathcal{F}}_n(\mathbf{t}) - \mathcal{F}(\mathbf{t}), \quad (3.6)$$

其中 $\mathbf{t} = (t_1, t_2, \dots, t_p)^T$ 是超立方体 $(0, 1)^{\otimes p}$ 中的一个固定值. 根据中心极限定理, 可得

$$\frac{\hat{\mathcal{F}}_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})}{\sqrt{\hat{\mathcal{F}}_n(\mathbf{t})(1 - \hat{\mathcal{F}}_n(\mathbf{t}))/n}} \rightarrow N(0, 1), \quad n \rightarrow \infty. \quad (3.7)$$

给定显著水平 α , 可以推导出

$$P(|\widehat{\mathbf{MI}}^p - \mathbf{MI}^p| \leq Z_{\frac{\alpha}{2}} \delta_n) = 1 - \alpha, \quad (3.8)$$

其中 $Z_{\frac{\alpha}{2}}$ 是标准正态分布的右侧 $\frac{\alpha}{2}$ 分位数, δ_n 是 $\widehat{\mathbf{MI}}^p$ 的标准误,

$$\delta_n = \sqrt{\frac{\hat{\mathcal{F}}_n(\mathbf{t})(1 - \hat{\mathcal{F}}_n(\mathbf{t}))}{n}} = \sqrt{\frac{\widehat{\mathbf{MI}}^p(1 - \widehat{\mathbf{MI}}^p)}{n}}. \quad (3.9)$$

根据上述定义, 应用经验经济差异指数 $\widehat{\mathbf{MI}}^p$, 可以给出多维经济差异指数置信水平为 $1 - \alpha$ 的置信区间 $(\widehat{\mathbf{MI}}^p - Z_{\frac{\alpha}{2}} \delta_n, \widehat{\mathbf{MI}}^p + Z_{\frac{\alpha}{2}} \delta_n)$.

经验多维经济差异指数具有良好的收敛性, 可以证明在一些简单的条件下, 样本经验经济差异指数 $\widehat{\mathbf{MI}}$ 收敛到总体多维经济差异指数, 具有强相合性.

定理 3.1 假定联合分布函数 $F(x_1, x_2, \dots, x_p)$ 在每个分量上都是连续函数, 并且每个分量的均值 μ_j 是有限的, 则由 (3.5) 定义的经验多维经济差异指数 $\widehat{\mathbf{MI}}$ 将依概率 1 收敛到总体多维经济差异指数.

定理的证明是显然的 (见附录 A), 可以用 Glivenko-Cantelli 定理来完成.

4 数据模拟分析

本节将运用模拟仿真来寻找多维经济差异指数的影响因素. 为简化计算, 考虑 $p = 2$, 即两个社会资源. 仿真中变换分布类型、分布偏度、分布标准差 σ 和两个社会资源的相关系数 ρ 来检验这些因素对多维经济差异指数的影响.

令 x 和 y 分别记两个社会资源对应的随机变量. 仿真中, 令样本容量 $n = 1,000$ 并且考虑不同的分布类型, 进行 1,000 次 Monte Carlo 模拟, 得到平均的样本多维经验差异指数及相应的标准差.

数值模拟一: 该数值模拟主要验证经济差异指数的取值范围, 考虑两个社会资源的不同设定, 计算二维指数.

(1) 主要是验证多维经济差异指数在社会资源处于极端情况下的取值, 设定 $x \sim U(x_0 - \delta, x_0 + \delta)$, $y \sim U(y_0 - \delta, y_0 + \delta)$, 其中 $x_0 = 6.82$, $y_0 = 0.75$. δ 取值越小, 社会资源越接近完全均衡状况, 设定 $\rho = 0$ 和 $\rho = 1$;

(2) 主要考察变量间相关性对多维指数的影响, 设定 $\log(x) \sim N(\mu_0, \sigma_0^2)$, $\log(y) \sim N(\mu_0, \sigma_0^2)$, 其中 $\mu_0 = 1.92$, $\sigma_0 = 0.81$, 设定相关系数 ρ 的取值为从 -1 到 1 的 10 个不同取值.

表 1 中结果表明, 在社会资源绝对均衡的条件下, 指数值趋于理论下界, 在 $\rho = 0$ 下趋于 $\frac{1}{2^2} = 0.25$, 而在 $\rho = 1$ 条件下趋于 $\frac{1}{3}$, 当 δ 增大时, 社会不均衡性逐渐增大, 则指数值也增大. 从仿真 (2) 中可以看出, 资源间的相关性对社会资源的不均衡性影响较小, 正相关性会加大社会不均衡性, 而负相关性会减小社会不均衡性.

数值模拟二: 该数值模拟主要验证分布的变异性 and 偏态对多维指数值的影响, 进行如下设定, 分别计算二维和三维指数.

(3) 主要验证分布标准差 σ 对指数的影响, 设定 $\log(x) \sim N(\mu_0, \sigma^2)$ 并且 $\log(y) \sim N(\mu_0, \sigma^2)$, 标准差 σ 从 $\sigma_0/4$ 到 $4\sigma_0$ 共 5 个取值, 分别计算 $\rho = 0$ 和 $\rho = 0.6$ 时的指数值;

(4) 主要考察偏度对多维经济差异指数值的影响, 设定 $\log(x^{(\lambda)}) \sim N(\mu_0, \sigma_0^2)$, $\log(y^{(\lambda)}) \sim N(\mu_0, \sigma_0^2)$, 引入 Box-Cox 变换 λ , $\mathbf{a}^{(\lambda)} = \frac{\mathbf{a}^\lambda}{\lambda}$, λ 越大分布偏度越大.

在仿真中应用 9 种不同的 λ 取值. 在 (3) 和 (4) 中, $\mu_0 = 1.92$, $\sigma_0 = 0.81$.

表 2 结果表明, (3) 中当两个分布的方差都增加时, 指数从 0.3106 显著增加到 0.9657, 已经接近指数的上界 1, 这说明当社会资源整体变得极端不均衡时, 指数值将接近上界, 变异性对均衡指数的影响大于相关性.

(4) 中当 $\lambda = 0.1$ 时, 指数值趋于理论下界 (指数下界二维时为 0.25, 三维时为 0.125), 当 $\lambda = 10$ 时, 指数值几乎达到理论上界 1, 这表明社会资源分布的偏度会极大影响均衡性, 导致严重的社会不均衡性. 相依性对指数值的影响不大, $\rho = 0.6$ 时指数值相对 $\rho = 0$ 时有所增加, 但增加幅度不高, 说明正相关性会增大不均衡性.

总的来说, 多维经济差异指数能够较客观地评价多个社会资源的不均衡性, 指数具有相对稳健性, 与 Gini 指数相比, 真实地反映了社会实际情况. 指数的影响因素中, 变异性和偏度对指数值影响更大, 资源间正相关性会加大指数值, 负相关性会减小指数值.

表 1 数值模拟一的经济差异指数值

(1)	δ	0.001	0.01	0.05	0.2	0.75
均匀分布	指数	0.2506	0.2517	0.2567	0.2751	0.3459
$\rho = 0$	标准误	(0.0027)	(0.0026)	(0.0026)	(0.0026)	(0.0034)
均匀分布	指数	0.3340	0.3351	0.3400	0.3587	0.4299
$\rho = 1$	标准误	(0.000002)	(0.00002)	(0.00009)	(0.00037)	(0.0018)
(2)	ρ	0.25	0.5	0.75	0.9	1
对数正态分布	指数	0.5285	0.5435	0.5605	0.5706	0.5778
	标准误	(0.0061)	(0.0065)	(0.0066)	(0.0062)	(0.0061)
	ρ	-0.25	-0.5	-0.75	-0.9	-1
对数正态分布	指数	0.4998	0.4862	0.4732	0.4659	0.4608
	标准误	(0.0057)	(0.0059)	(0.0067)	(0.0072)	(0.0075)

表 2 数值模拟二的经济差异指数值

(3)	σ	$\sigma_0/4$	$\sigma_0/2$	σ_0	$2\sigma_0$	$4\sigma_0$
二维	指数	0.3106	0.3760	0.5137	0.7626	0.9657
	标准误	(0.0028)	(0.0033)	(0.0055)	(0.0129)	(0.0105)
三维 $\rho = 0$	指数	0.1851	0.2595	0.4381	0.7744	0.9743
	标准误	(0.0028)	(0.0036)	(0.0071)	(0.0182)	(0.0077)
三维 $\rho = 0.6$	指数	0.2633	0.3383	0.5044	0.7979	0.9764
	标准误	(0.0033)	(0.0046)	(0.0083)	(0.0177)	(0.0077)
(4)	λ	0.1	0.2	0.3	0.5	1.5
二维	指数	0.2741	0.2983	0.3234	0.3758	0.6473
	标准误	(0.0026)	(0.0027)	(0.0028)	(0.0033)	(0.0095)
三维 $\rho = 0$	指数	0.1474	0.1719	0.1989	0.2594	0.6230
	标准误	(0.0026)	(0.0028)	(0.0029)	(0.0036)	(0.0132)
三维 $\rho = 0.6$	指数	0.2227	0.2496	0.2776	0.3384	0.6661
	标准误	(0.0026)	(0.0031)	(0.0034)	(0.0045)	(0.0137)
	λ	2	3	5	10	
二维	指数	0.7622	0.9080	0.9860	0.9984	
	标准误	(0.0136)	(0.0152)	(0.0061)	(0.0012)	
三维 $\rho = 0$	指数	0.7739	0.9296	0.9874	0.9956	
	标准误	(0.0181)	(0.0131)	(0.0044)	(0.0011)	
三维 $\rho = 0.6$	指数	0.7981	0.9351	0.9883	0.9959	
	标准误	(0.0182)	(0.0142)	(0.0042)	(0.0011)	

5 我国省会城市社会资源不均衡性的实证分析

本节将应用多维经济差异指数, 基于民生调查数据集进行经验分析. 该数据集来源于西南财经大学一项全国省会城市大型调查. 该调查关于中国民生满意度指数的调查报告已经于 2017 年出版^[31]. 本文考虑从社会资源均衡性的角度重新研究该数据集, 考虑 3 个重要的社会资源: 收入、教育性支出和资产. 作为比较, 本文计算 4 个指数: 基于收入的 Gini 指数 (缩写为 GI)、基于收入与教育性支出的 2 维经济差异指数 (缩写为 MI1)、基于收入与资产计算的 2 维经济差异指数 (缩写为 MI2) 和基于 3 个资源计算的 3 维经济差异指数 (缩写为 MI3). 删除缺失数据后, 本文得到关于 MI1、MI2 和 MI3 有效样本, $n_1 = 4646$, $n_2 = 5863$, $n_3 = 3314$.

经过计算, 全样本的 MI1 等于 0.4955, 95% 置信区间为 (0.4811, 0.5099); 全样本的 MI2 等于 0.5280, 95% 置信区间为 (0.5152, 0.5408). 后者较前者大, 意味着在我国省会城市中, 收入与资产的社会资源不均衡程度相比收入与教育性支出不均衡程度更大. 全样本的 MI3 等于 0.3658, 95% 置信区间为 (0.3438, 0.3878).

本文主要比较我国各区域和各省会城市社会资源的不均衡程度. 将全国省会城市按区域分为东北 (哈尔滨、长春、沈阳)、华北 (北京、天津、石家庄、太原、呼和浩特)、华东 (上海、济南、南京、杭州、南昌、福州、合肥)、华中 (郑州、武汉、长沙)、华南 (广州、南宁、海口)、西南 (重庆、成都、昆明、贵阳、拉萨) 和西北 (西安、兰州、西宁、银川、乌鲁木齐) 7 个区域, 每个区域计算了 Gini 指数

和多维经济差异指数 MI1、MI2 和 MI3, 对指数取值进行比较. 指数结果见表 3 和 4. 需要指出的是, Gini 指数和多维经济差异指数因为计算方法不同, 数值直接比较不能说明问题, 因此, 我们主要比较两个指数各区域和城市间的差异, 以此来揭示多维经济差异指数的优势.

表 3 列出了以上 7 个区域的 Gini 指数和 3 种多维经济差异指数值. Gini 指数最高的是西南, 其次是华南, 然后是华东, 西北最低. 这与我国实际基本符合, 西北地区地广人稀, 收入普遍不高, 因此不均衡性相对较低; 西南地区由于民族众多, 城市之间发展差距较大, 既有相对发达的重庆、成都等新一线城市, 又有经济收入相对落后的拉萨, 因此贫富差距大, 不均衡性最大; 而华南和华东地区经济发达, 大量外来人口流入, 因此收入不均衡性也较大.

结合收入和教育性支出来看社会资源的不均衡性 (MI1), 情况有所不同, 华东地区不均衡性居第一, 西南排第二, 两个区域的不均衡性都显著高于其他地区 (置信区间在其他几个地区指数值之上), 其他几个区域差异较小, 东北地区显著低于其他区域. 这一现象值得研究, 华东地区的教育性支出的差异主要原因是, 相比其他地区, 华东地区拥有全国最好的教育资源, 但是城市分布不均衡, 主要集中在上海和南京两座城市, 这造成了社会资源地域间不均衡性; 另外, 由于收入的差距, 造成外来人和本地人对教育支出的差距, 少数民族和汉族之间教育支出也加大了差距, 这是社会资源的人群不均衡性. 加之人们由于教育理念不同, 对教育的投入也不同, 这造成了社会资源主观投入的不均衡性. 以上三个不均衡性是华东地区多维经济差异指数值偏高的主要原因. 对于西南地区, 本来收入的不均衡性就很大, 加之上面三个原因都在起作用, 因此指数值也偏高. 华南地区由于地域间不均衡性相对较小, 因此不均衡性适中, 指数值不高. 东北地区指数值偏低, 是因为这三个因素都不太起作用. 相比其他地区, 东北地区教育资源分布比较均衡, 人口不断流出, 外来人口较少, 而且因为是老工业基地, 人们教育理念基本类似, 造成了指数明显偏低.

表 3 我国各区域 Gini 指数和多维经济差异指数比较

区域	东北	华北	华东	华中
GI	0.4478	0.4446	0.4607	0.4387
95% 置信区间	(0.4351, 0.4605)	(0.4319, 0.4574)	(0.4479, 0.4735)	(0.426, 0.4514)
MI1	0.4529	0.4795	0.5336	0.4794
95% 置信区间	(0.4386, 0.4672)	(0.4651, 0.4939)	(0.5193, 0.5479)	(0.4650, 0.4938)
MI2	0.5214	0.5181	0.5162	0.5150
95% 置信区间	(0.5086, 0.5342)	(0.5053, 0.5309)	(0.5034, 0.5290)	(0.5022, 0.5278)
MI3	0.3419	0.3491	0.3671	0.3577
95% 置信区间	(0.2883, 0.3955)	(0.3132, 0.3849)	(0.3296, 0.4046)	(0.3145, 0.4009)
区域	华南	西南	西北	全国
GI	0.4708	0.5164	0.4020	0.4454
95% 置信区间	(0.458, 0.4836)	(0.5036, 0.5292)	(0.3894, 0.4145)	(0.4327, 0.4581)
MI1	0.4722	0.5029	0.4801	0.4955
95% 置信区间	(0.4578, 0.4866)	(0.4885, 0.5173)	(0.4657, 0.4945)	(0.4811, 0.5099)
MI2	0.5465	0.5508	0.5092	0.5280
95% 置信区间	(0.5338, 0.5592)	(0.5381, 0.5635)	(0.4964, 0.5220)	(0.5152, 0.5408)
MI3	0.3704	0.3792	0.3657	0.3658
95% 置信区间	(0.3119, 0.4289)	(0.3365, 0.4219)	(0.3224, 0.4090)	(0.3438, 0.3878)

表 4 我国各省会城市 Gini 指数和多维经济差异指数比较

城市	哈尔滨	长春	沈阳	北京	天津
GI	0.3004 (7)	0.5159 (24)	0.3229 (9)	0.5028 (23)	0.3018 (8)
MI1	0.4307 (4)	0.4503 (9)	0.4552 (10)	0.4696 (17)	0.4936 (21)
MI2	0.5095 (12)	0.5362 (23)	0.5127 (13)	0.5348 (22)	0.5005 (11)
MI3	0.2830 (1)	0.3413 (12)	0.3520 (18)	0.3385 (10)	0.3289 (7)
城市	石家庄	郑州	太原	济南	南京
GI	0.4129 (16)	0.3994 (14)	0.3679 (12)	0.5195 (25)	0.2973 (6)
MI1	0.4623 (12)	0.4960 (22)	0.4433 (6)	0.5592 (30)	0.4649 (13)
MI2	0.4651 (1)	0.5248 (19)	0.5371 (25)	0.5205 (15)	0.4971 (8)
MI3	0.3064 (2)	0.3505 (16)	0.3418 (13)	0.3126 (3)	0.3674 (24)
城市	上海	杭州	南昌	合肥	武汉
GI	0.4857 (22)	0.5351 (28)	0.5338 (27)	0.2826 (4)	0.6043 (31)
MI1	0.5691 (31)	0.5112 (25)	0.4470 (7)	0.4295 (3)	0.4667 (15)
MI2	0.5365 (24)	0.4967 (7)	0.5307 (20)	0.4816 (4)	0.5208 (17)
MI3	0.3812 (27)	0.3503 (16)	0.3547 (20)	0.3554 (22)	0.3442 (15)
城市	长沙	广州	福州	海口	重庆
GI	0.3735 (13)	0.3451 (10)	0.5427 (29)	0.2904 (5)	0.4810 (21)
MI1	0.5306 (28)	0.4366 (5)	0.4798 (18)	0.4602 (11)	0.5002 (24)
MI2	0.5001 (9)	0.5327 (21)	0.5177 (14)	0.5948 (31)	0.5543 (30)
MI3	0.3712 (25)	0.3185 (5)	0.3150 (4)	0.3437 (14)	0.3778 (26)
城市	成都	昆明	贵阳	南宁	拉萨
GI	0.5274 (26)	0.5892 (30)	0.4076 (15)	0.3461 (11)	0.4294 (18)
MI1	0.4833 (19)	0.4665 (14)	0.4864 (20)	0.5290 (27)	0.5391 (29)
MI2	0.5492 (28)	0.5413 (27)	0.5396 (26)	0.4914 (6)	0.5188 (15)
MI3	0.3533 (19)	0.3911 (31)	0.3548 (21)	0.3406 (11)	0.3660 (23)
城市	西安	兰州	银川	西宁	呼和浩特
GI	0.4418 (20)	0.4133 (17)	0.2447 (2)	0.4436 (19)	0.2411 (1)
MI1	0.4489 (8)	0.4271 (2)	0.4694 (16)	0.5124 (26)	0.4162 (1)
MI2	0.4903 (5)	0.5219 (18)	0.4810 (3)	0.5001 (10)	0.4743 (2)
MI3	0.3845 (29)	0.3332 (8)	0.3239 (6)	0.3885 (30)	0.3812 (28)
城市	乌鲁木齐				全国
GI	0.2692 (3)				0.4454
MI1	0.4980 (23)				0.4955
MI2	0.5502 (29)				0.5280
MI3	0.3342 (9)				0.3658

结合收入和资产来看社会资源的不均衡性 (MI2), 有如下结果: 第一, 我国收入和资产的不均衡性更大, MI2 的值整体上显著大于 MI1, 表现为 7 个区域的 MI2 的置信区间都显著高于 MI1 的置信区间. 第二, 从区域来看, 收入与收入和资产的不均衡性有较强的关联性, 7 个区域上两者的排序基本是

一致的, 西南地区不均衡性最高, 西南和华南地区显著高于其他地区, 而其他几个地区没有显著差异, 西北地区不均衡性最低.

结合 3 个社会资源看不均衡性 (MI3), 有如下结果: 第一, 3 维指数值并没有显著的区域差异, 每个地区 3 维指数值都被其他地区指数的置信区间覆盖, 说明从整体看, 我国的三资源不均衡性相对较小, 这与两资源的不均衡性不同. 第二, 西南地区、华南地区和华东地区 3 维指数值较大, 而东北地区和华北地区 3 维指数值较小, 说明整体看我国南方的资源不均衡性可能要略高于北方. 第三, 西北地区在收入 Gini 指数和收入和资产 2 维指数值都最低, 而 3 维指数值反而居中. 这既说明西北地区三资源有较强的正相依性, 影响了指数的取值, 也说明多维指数在描述多资源不均衡性上的优势.

表 4 列出了我国大陆 31 个省会城市 (台北、香港和澳门没有进行调查, 因此没有列出) 的 Gini 指数和多维经济差异指数, 表中分别按照 MI1、MI2 和 MI3 的取值从低到高排序. 从表中可以看出, 我国省会城市收入均衡性和收入和教育性支出整体均衡性大体是一致的, 但是有些省会城市差异较大, 三资源的均衡性也有所不同, 下面分别阐述.

西北地区的兰州和西安是西北地区经济发达的大城市, 收入差距相对较大, 因此, GI 值较大, 分别居全国第 17 和 20 位; 但是这两个城市 MI1 取值较小, 分别列全国第 2 和 8 位, 说明这两个城市教育性支出相对均衡, 这也与当地教育资源较为丰富和均衡, 以及教育实力较强的情况相吻合; 两个城市的 MI3 有很大差异, 兰州的指数值小于西安, 两者分列全国第 8 和 29 位, 这说明西安和兰州三资源在分布的偏度、相依性等有较大差异. 西北地区另外两个城市银川和乌鲁木齐, 经济相对落后, 收入较低, 贫富差距不大, 因此 GI 值较小, 列全国第 2 和 3 位, 但是由于这两个地区都是民族地区, 人群不均衡性和主观投入不均衡性对教育性支出的影响很大, 因此这两个地区的 MI1 值偏大, 分别居全国第 16 和 23 位, MI3 值较小, 分列第 6 和 9 位, 西宁的 MI3 较大, 列全国第 30 位, 可能是受民族、收入和宗教等因素综合影响.

东北地区哈尔滨和沈阳, GI 和 MI1 排名基本接近, 说明两者的不均衡性相当, 哈尔滨 MI3 值为全国最小, 可能与气候、人口流出和教育实力强都有关系, 沈阳和长春 MI3 值基本相当. 长春的 GI 值列全国第 24 位, 但是 MI1 列全国第 9 位, 说明长春收入差距大, 但是教育性支出不均衡性小, 这或许与长春人口流出有关, 流出的都是中高收入人口, 造成贫富差距较大, 而长春的教育资源相对较好, 也比较均衡, 因此, MI1 值不大.

西南地区昆明的 GI 值较大, 列全国 30 位, 但是 MI1 值较小, 列全国第 14 位, MI3 值全国最高, 这一点与昆明的民族和地理位置有关, 云南是我国民族最多的省份, 而且与东南亚接壤, 因此各民族贸易往来频繁, 造成收入不均衡性加大, 但近年来国家和省政府对教育加大投入, 昆明和贵阳的教育资源相对较平均, 因此 MI1 值不大, 但是三资源的不均衡性仍然较大. 拉萨的 GI 值不大, 但是 MI1 值较大, 列全国第 29 位, MI3 值也较大, 列全国 23 位. 这说明拉萨的教育性支出不均衡性高, 主要原因是西藏教育资源较差, 且分布不均衡, 汉族人和藏族人对教育的重视程度不同, 在此影响下, 人群不均衡性和主观投入不均衡性都较大, 导致 MI1 的值大.

华中地区武汉和长沙对比, 武汉的 GI 值偏大, 长沙的 MI1 值较大, MI3 值也是长沙大于武汉; 武汉是因为交通发达, 流动人口较多, 造成收入差距较大; 长沙的 GI 和 MI2 都适中, MI1 和 MI3 值较大, 说明长沙三资源的相依性可能较强, 因此综合不均衡性较强.

华东地区资源不均衡性整体较高, 例如, 上海、济南、杭州、福州和南昌的 GI 值都较大, 说明这几个城市收入差距大, 但是 MI1 值南昌较小, MI2 值杭州较小, MI3 值济南和福州较小, 说明济南和福州可能资源间相依性相反. 另外一个城市南京, GI、MI1 和 MI2 值都较小, MI3 值却较大, 说明资源之间可能存在正相依性.

上述结果显示出多维经济差异指数描述社会不均衡性比 Gini 指数具有明显的优势, 因为该指数的来源变量更多, 能更全面和综合地反映社会整体的不均衡性, 而不仅仅是收入的不均衡性。

6 结论及讨论

本文运用一种新的框架, 将 Gini 指数的定义从一维拓展到多维, 得到多维经济差异指数。在本文的讨论中, 新定义的不均衡性指数更加适合在复杂经济数据中分析和阐述经济不均衡性。本文设计了算法计算样本经验多维经济差异指数, 并证明其在一定条件下收敛到总体指数值。该指数与 Gini 指数一样, 也能应用到自然和社会科学各个领域, 仿真和实证分析表明该指数能够综合多个变量, 形成指数, 因而, 新指数对于解释多维经济数据的不均衡性, 是一种更一般且先进的方法。

当然, 仍然有不少有价值 and 需要深入研究的问题, 本文只是将指数拓展到多维, 需要考虑在高维或稀疏情形下该指数是否依然有效。本文的算法是基于联合分布函数的一般算法, 对分布没有特殊要求, 也许针对某些特殊的分布, 算法可以进一步简化。

致谢 作者感谢编委会和两名审稿人对论文提出的宝贵意见, 对于论文的改进起到了非常重要的作用。

参考文献

- 1 Gini C. On the measurement of concentration and variability of characters (English translation from Italian by Fulvio de Santis). *Metron*, 2005, 63: 1–38
- 2 Gini C. Measurement of inequality of incomes. *Econom J*, 1921, 31: 124–126
- 3 Kendall M G, Stuart A. *The Advanced Theory of Statistics vol. 1: Distribution Theory*, 4th ed. Philadelphia: Lippincott Company, 1977
- 4 Sen A. *On Economic Inequality*. Oxford: Clarendon Press, 1973
- 5 Anand S. *Inequality and Poverty in Malaysia: Measurement and Decomposition*. Oxford: Oxford University Press, 1983
- 6 Jasso G. On Gini's mean difference and Gini's index of concentration. *Amer Sociol Rev*, 1979, 44: 867–870
- 7 Pyatt G. On the interpretation and disaggregation of Gini coefficients. *Econom J*, 1976, 86: 243–255
- 8 Silber J. Factor components, population subgroups and the computation of the Gini index of inequality. *Rev Econom Stat*, 1989, 71: 107–115
- 9 Atkinson A B. On the measurement of inequality. *J Econom Theory*, 1970, 2: 244–263
- 10 Imedio-Olmedo L J, Bárcena-Martín E, Parrado-Gallardo E M. A class of Bonferroni inequality indices. *J Public Econom Theory*, 2011, 13: 97–124
- 11 Sundrum R M. *Income Distribution in Less Developed Countries*. London: Routledge, 1991
- 12 Glenn F. Empirics of world income inequality. *Amer J Sociology*, 1999, 104: 1597–1630
- 13 Greselin F. More equal and poorer, or richer but more unequal? *Econom Qual Control*, 2014, 29: 99–117
- 14 Greselin F, Pasquazzi L, Zitikis R. Contrasting the Gini and Zenga indices of economic inequality. *J Appl Stat*, 2013, 40: 282–297
- 15 Greselin F, Puri M L, Zitikis R. *L-functions, processes, and statistics in measuring economic inequality and actuarial risks*. *Stat Interface*, 2009, 2: 227–245
- 16 Tsui K. Multidimensional poverty indices. *Soc Choice Welfare*, 2002, 19: 69–93
- 17 Bourguignon F, Chakravarty S R. The measurement of multidimensional poverty. *J Econom Inequal*, 2003, 1: 25–49
- 18 Alkire S, Santos M E. Acute multidimensional poverty: A new index for developing countries. *SSRN J*, 2010, doi: 10.2139/ssrn.1815243
- 19 United Nations. *The Real Wealth of Nations: Pathways to Human Development*. New York: United Nations Development Program, 2011
- 20 Kakwani N C. *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. New York: Oxford University Press, 1980
- 21 Kakwani N. On a class of poverty measures. *Econometrica*, 1980, 48: 437–446
- 22 Shorrocks A. Income inequality and income mobility. *J Econom Theory*, 1978, 19: 376–393

- 23 Maasoumi E, Zandvakili S. A class of generalized measures of mobility with applications. *Econom Lett*, 1986, 22: 97–102
- 24 Douhéret G, Davis M I. Measuring education inequality: Gini coefficients of education. *Chem Soc Rev*, 1993, 22: 43–50
- 25 Roemer J E. Economic development as opportunity equalization. *World Bank Econom Rev*, 2013, 28: 189–209
- 26 Roemer J E, Aaberge R, Colombino U, et al. To what extent do fiscal regimes equalize opportunities for income acquisition among citizens? *J Public Economics*, 2003, 87: 539–565
- 27 Weymark J A. Generalized Gini indices of equality of opportunity. *J Econom Inequal*, 2003, 1: 5–24
- 28 Sadras V, Bongiovanni R. Use of Lorenz curves and Gini coefficients to assess yield inequality within paddocks. *Field Crops Res*, 2004, 90: 303–310
- 29 Kopczuk W, Saez E, Song J. Earnings inequality and mobility in the united states: Evidence from social security data since 1937. *Quart J Econom*, 2010, 125: 91–128
- 30 van Mierlo T, Hyatt D, Ching A T. Employing the Gini coefficient to measure participation inequality in treatment-focused Digital Health Social Networks. *Netw Model Anal Health Inform Bioinforma*, 2016, 5: article number 32
- 31 西南财经大学中国民生指数研发中心. 中国省会城市民生满意度指数报告. 成都: 西南财经大学出版社, 2017

附录 A 定理 3.1 的证明

因为联合经验人数分布函数 $\hat{\mathcal{F}}_n(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_p)$ 和联合经验分布函数 $\hat{F}_n(t_1, t_2, \dots, t_p)$ 除了定义域不同外, 具有相同的统计性质, 因此在证明中主要考虑联合经验分布函数 $\hat{F}_n(t_1, t_2, \dots, t_p)$ 的相合性. 定理的证明可以利用 Glivenko-Cantelli 定理. 为简便起见, 考虑向量每个分量是连续变量, 对第 i 个分量, 固定 $-\infty = T_{i,0} < T_{i,1} < \dots < T_{i,m_i} = \infty$, 满足

$$F(t_1, \dots, t_{i-1}, T_{i,j}, \dots, t_p) - F(t_1, \dots, t_{i-1}, T_{i,j-1}, \dots, t_p) = \frac{1}{m_i}, \quad (\text{A.1})$$

$i = 1, 2, \dots, p, j = 1, 2, \dots, m_i$, 则对于任一 $t_i \in \mathbb{R}$, 必然存在 $j \leq m_i$ 满足 $t_i \in [T_{i,j-1}, T_{i,j}]$. 注意到

$$\begin{cases} \hat{F}_n(t_1, \dots, t_i, \dots, t_p) - F(t_1, \dots, t_i, \dots, t_p) \\ \leq \hat{F}_n(t_1, \dots, T_{i,j}, \dots, t_p) - F(t_1, \dots, T_{i,j}, \dots, t_p) + \frac{1}{m_i}, \\ \hat{F}_n(t_1, \dots, t_i, \dots, t_p) - F(t_1, \dots, t_i, \dots, t_p) \\ \geq \hat{F}_n(t_1, \dots, T_{i,j-1}, \dots, t_p) - F(t_1, \dots, T_{i,j-1}, \dots, t_p) - \frac{1}{m_i}, \end{cases} \quad (\text{A.2})$$

因此, (A.3) 几乎必定成立 (almost surely, 简记为 a.s.),

$$\begin{aligned} \|\hat{F}_n - F\|_\infty &= \sup_{\mathbf{t} \in \mathbb{R}^p} |\hat{F}_n(\mathbf{t}) - F(\mathbf{t})| \\ &\leq \sum_{i=1}^p \frac{1}{m_i} + \sum_{i=1}^p \max_{j_i \in \{1, \dots, m_i\}} |\hat{F}_n(T_{1,j_1}, \dots, T_{i,j_i}, \dots, T_{p,j_p}) \\ &\quad - F(T_{1,j_1}, \dots, T_{i,j_i}, \dots, T_{p,j_p})|. \end{aligned} \quad (\text{A.3})$$

既然 $\max_{j \in \{1, 2, \dots, m_i\}} |\hat{F}_n(t_1, \dots, T_{i,j}, \dots, t_p) - F(t_1, \dots, T_{i,j}, \dots, t_p)| \rightarrow 0$, a.s., 由强大数定律知, 对任一 $\mathbf{m} = (m_1, m_2, \dots, m_p)^T$, 必然存在 N , 使得对所有 $n > N$, 有

$$\|\hat{F}_n - F\|_\infty \leq \sum_{i=1}^p \frac{1}{m_i} \quad \text{a.s.}, \quad (\text{A.4})$$

这就保证了联合经验分布函数 $\hat{F}_n$ 的强收敛性.

一方面, 既然 $\hat{F}_n$ 收敛于 F , a.s., 联合经验人数分布隐函数 $\hat{\mathcal{F}}_n(\mathbf{y})$ 是关于联合经验分布函数 $\hat{F}_n$ 的严格单调变换, 因此收敛到总体人数分布隐函数 $\mathcal{F}(\mathbf{y})$, a.s. 另一方面, 经验多维经济差异指数是经验联合人数分布隐函数的 Riemann 和, 将在几乎一定意义下, 当样本容量趋于无穷时, 收敛到总体人数分布隐函数的积分, 即总体多维经济差异指数. 因此, 经验多维经济差异指数 $(\widehat{\mathbf{MI}})^p$ 依概率 1 收敛于总体多维经济差异指数 $(\mathbf{MI})^p$:

$$\widehat{\mathbf{MI}}^p \rightarrow \mathbf{MI}^p, \quad \text{a.s.} \quad (\text{A.5})$$

定理 3.1 证明完毕.

Application of the multivariate economic dispersion index in inequality evaluation of the provincial capitals

Yifan Xia, Ling Zhang & Hong Yin

Abstract In this paper, we consider a natural extension of Gini index, a widely-used measure of inequality of wealth, from one dimensional economic source to multi-dimensional economic sources, and propose a new index termed as the multi-dimensional economic dispersion index (MI). We propose an estimator of the MI, and study its asymptotic property. We illustrate the usefulness of the new index by an application of evaluating the social inequality level of Chinese major geographical regions and provincial capital cities.

Keywords Gini index, implicit population proportion function, provincial capitals, inequality of resources

MSC(2010) 62F86, 62H86

doi: 10.1360/N012019-00118