

# 多维极化的测度:新的方法及其应用

洪兴建 董君

**【摘要】**鉴于现有多维极化测度方法未能体现维度之间的相关性,本文引入余弦相似度,并基于相对剥夺理论提出了新的多维极化测度方法。该方法还可以进行静态和动态分解,能够分析每个维度及其变动对多维极化的贡献率。运用中国家庭追踪调查(CFPS)数据,本文从收入、财产、交通通信支出和受教育水平四个维度实证分析了我国城乡居民家庭之间的多维极化。实证结果表明,城乡长期多维极化呈现明显下降趋势,其中教育、交通通信支出和收入三个维度的极化效应是主要原因,而财产极化效应不利于多维极化的下降。

**【关键词】**多维极化;测度方法;维度分解;城乡多维极化

**【作者简介】**洪兴建,浙江财经大学数据科学学院教授,院长,浙江财经大学和浙江工商大学统计学博士生导师,研究方向为收入分配、数字经济等宏观经济统计测度;董君(通讯作者),浙江工商大学统计与数学学院博士研究生,研究方向为国民经济统计,电子邮箱:widongjun@163.com。

**【原文出处】**《统计研究》(京),2020.9.24~33

**【基金项目】**国家社科基金重点项目“城乡共享发展成果的统计测度研究”(16ATJ002);浙江省A类一流学科(浙江财经大学统计学18A001)。

## 一、引言

收入极化研究的开拓者是 Foster 和 Wolfson (1992)以及 Esteban 和 Ray (1994),有关收入极化的文献综述可见洪兴建和李金昌(2007)、Chakravarty(2009)等。收入极化与收入不平等最大区别在于,极化重在刻画群体之间的对抗程度,它是群内差距的减函数(因为在其他条件不变的情况下,某个子群内部的差距越大,其团结一致对外的程度就越小),而收入不平等是群内差距的增函数。近年来,一些文献实证分析了我国居民收入极化、财产极化等(王朝明和李梦凡,2013;罗楚亮,2018)。鉴于多维不平等和多维贫困在实践中的广泛运用,少部分学者们开始关注多维极化的研究。目前多维极化指标的构建框架主要有两种:一种是根据不平等指标分解得到的群内不平等、群间不平等和子群

相对规模来构建多维极化指标(Gigliarano 和 Mosler, 2009);另一种则是借鉴 Esteban 和 Ray (1994)提出的框架,基于认同函数和疏远函数构建多维极化指标(Tewodaj, 2008;洪兴建, 2010)。至于多维极化的具体维度,也从收入和财富等经济指标拓展到其他非经济因素,如教育、闲暇等(Gigliarano 和 Mosler, 2009; Merz 和 Scherg, 2014)。应该说,目前多维极化测度方法的研究文献非常少,存在的问题还比较多,突出表现在没有准确反映不同维度之间的相关性。针对现有多维极化测度方法的缺陷,本文将从相对剥夺和维度相关的视角提出新的多维极化测度方法,给出静态和动态的分解方法,并运用该方法实证分析我国农村居民与城镇居民在收入、财产、交通通信支出、受教育水平四个方面的多维极化。

## 二、多维极化测度方法简述

假定总人口为  $N$ , 共有  $K$  个维度的要素, 总体被分成了  $G$  个子群, 则总的要素分布  $X$  可以表示为:

$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(K)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(K)} \end{pmatrix}_{n \times K} = \begin{pmatrix} (X_1)_{N_i \times K} \\ (X_2)_{N_i \times K} \\ \vdots \\ (X_G)_{N_i \times K} \end{pmatrix} \quad (1)$$

其中,  $x_i^{(k)}$  为个体  $i$  在维度  $k$  上的拥有量,  $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(K)})$  表示个体  $i$  在各个维度的拥有量,  $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)})^T$  为所有个体在维度  $k$  上的分布,  $X_g$  表示第  $g$  个子群且有  $\sum_{g=1}^G N_g = N$ 。

Gigliarano 和 Mosler (2009) 按照收入极化的三个构成要素构造多维极化的测度方法, 即群间不平等  $B(X)$ 、群内不平等  $W(X)$  和子群相对规模  $S(X)$ , 给出了多维极化测度的一般公式:

$$P(X) = f(B(X), W(X), S(X)) \quad (2)$$

多维极化指标满足  $\partial f / \partial B > 0$ 、 $\partial f / \partial W < 0$  和  $\partial f / \partial S > 0$ 。虽然 Gigliarano 和 Mosler (2009) 探讨了不同的不平等指标和  $S(X)$ , 并给出了一些特定的测度方法, 但是建立在不平等分解基础上的测度指标完全受制于不平等指标及其分解方法的影响, 且  $S(X)$  的构造也有很大不确定性, 因而在实证分析中的科学性和可操作性均有待改进。

多维极化是收入极化的拓展, 一些学者尝试从收入极化测度方法中探寻多维极化的测度方法, 这其中 Duclos 等 (2004) 提出的收入极化分析思路具有较大的借鉴价值。设收入分布密度函数为  $f(x)$ , 定义任意收入  $x$  的认同 (Identification) 函数为  $I_x = f(x)^\alpha$ , 与收入  $y$  的疏远 (Alienation) 函数为  $A_{x,y} = |x - y|$ , 则 Duclos 等 (2004) 给出的收入极化测度方法为:

$$P_{DER} = \iint I_x A_{x,y} f(x) f(y) dx dy = \iint f(x)^{1+\alpha} f(y) |x - y| dx dy \quad (3)$$

其中敏感性参数  $\alpha \in [0.25, 1]$ 。Tewodaj (2008) 主

要针对二维变量探讨了多维极化测度方法, 洪兴建 (2010) 则将此方法推广到多维变量, 给出了基于矩阵  $X$  的测度方法:

$$P_H = \iint \cdots \iint f(x_i^{(1)}, \dots, x_i^{(K)})^{1+\alpha} f(x_j^{(1)}, \dots, x_j^{(K)}) \sqrt{\sum_{k=1}^K (x_i^{(k)} - x_j^{(k)})^2} dx_j^{(1)} \cdots dx_j^{(K)} dx_i^{(1)} \cdots dx_i^{(K)} \quad (4)$$

上述方法即为多维 DER 极化指标, 其中  $f(x_i^{(1)}, \dots, x_i^{(K)})$  为  $x_i$  的联合密度函数,  $f(x_i^{(1)}, \dots, x_i^{(K)})^\alpha$  表示  $x_i$  的认同函数,  $\sqrt{\sum_{k=1}^K (x_i^{(k)} - x_j^{(k)})^2}$  为疏远函数。

此外, Anderson (2010) 主要分析了贫困群体与非贫困群体的多变量两极分化问题。假设贫困与非贫困群体在所有维度的联合密度函数分别为  $f_p(x^{(1)}, \dots, x^{(K)})$ 、 $f_r(x^{(1)}, \dots, x^{(K)})$ , Anderson (2010) 提出的多维两极分化测度公式为:

$$P_A = 1 - \iint \cdots \iint \min \{ f_p(x^{(1)}, \dots, x^{(K)}), f_r(x^{(1)}, \dots, x^{(K)}) \} dx^{(1)} dx^{(2)} \cdots dx^{(K)} \quad (5)$$

式 (5) 是 Anderson (2010) 从其本人提出的一个收入极化指标拓展而来, 不过该方法并不能反映每个群体内部的聚合程度或认同程度, 且如果两个群体的多维分布不存在重叠部分, 该方法也不适用 (洪兴建, 2010)。

应该说, 现有极化指标基本以欧式距离或绝对值距离表示群体间的异质性或者疏远性, 但是在多维极化中, 该方法无法反映维度之间相关性变动的影响。假设个体  $i$  和  $j$  在两个维度上的分布分别为  $x_i = (1, 4)$ 、 $x_j = (5, 2)$ , 则以欧式距离测得两者间的疏远程度为  $\sqrt{(1-5)^2 + (4-2)^2} = 2\sqrt{5}$ ; 现若个体  $i$  和  $j$  交换第二个维度数值, 分别变为  $(1, 2)$  和  $(5, 4)$ , 此时个体  $i$  在两个维度上均低于个体  $j$ , 多维极化程度应该变大了, 但以欧式距离衡量的疏远程度为  $\sqrt{(1-5)^2 + (2-4)^2} = 2\sqrt{5}$ , 数值没有变化。以绝对值距离计算时也存在类似问题。此外, Srisuma 和 Vesperoni (2015) 分析了个体之间的多维距离, 提出将所有维度绝对值距离相乘而非相加来构造多维距离, 不过该方法的缺陷是只要有一个差距为 0, 结果就为 0, 且没能从根本上解决维度之间的相关

性问题。

其实,在多维不平等的测度中,部分学者早就注意到了维度相关性问题,比如 Tsui (1999) 提出了相关增进优化 (Correlation Increasing Majorization) 的概念,但是该方法并不能直接运用到多维极化的测度中。针对上述缺陷,本文将利用余弦相似度并结合相对剥夺理论改进多维框架下疏远程度的测度方法,从而给出新的多维极化指标,并将多维极化分解为各个维度极化之和的形式。

### 三、一个新的多维极化测度方法及其分解方法

#### (一) 一个新的多维极化测度方法

对于人群之间的疏远或者对抗而言,一般是资源拥有量较少的一方首先感受到不公,进而产生疏远或者对抗行为,因而引入相对剥夺度量极化中的疏远程度是比较合适的。参照 Runciman (1966) 和 Yitzhaki (1979) 对相对剥夺的定义,假设个体  $i$  在维度  $k$  上的拥有量  $x_i^{(k)}$  低于个体  $j$  在维度  $k$  上的拥有量  $x_j^{(k)}$ ,则在维度  $k$  上个体  $i$  相对于个体  $j$  的相对剥夺为  $x_j^{(k)} - x_i^{(k)}$ ;反之,若  $x_i^{(k)} \geq x_j^{(k)}$ ,那么在维度  $k$  上个体  $i$  相对于个体  $j$  的相对剥夺为 0。因此,定义个体  $i$  相对于个体  $j$  在维度  $k$  上的相对剥夺为  $D_{i,j}^{(k)} = \max \{x_j^{(k)} - x_i^{(k)}, 0\}$ ,进而个体  $i$  相对于个体  $j$  的平均相对剥夺  $D_{i,j}$  可以表示为:

$$D_{i,j} = \sum_{k=1}^K \varphi_k D_{i,j}^{(k)} = \sum_{k=1}^K \varphi_k \max \{x_j^{(k)} - x_i^{(k)}, 0\} \quad (6)$$

其中  $\varphi_k$  为第  $k$  个维度的权重。余弦相似度是用向量空间中两个向量夹角的余弦值反映维度之间的相关性。记  $\theta_{i,j}$  为向量  $x_i = (x_i^{(1)}, \dots, x_i^{(K)})$  和向量  $x_j = (x_j^{(1)}, \dots, x_j^{(K)})$  的夹角,当  $\theta_{i,j}$  越接近 0 度,即  $\cos \theta_{i,j}$  越接近 1 时,说明两个向量越相似,同等条件下多维极化就越大。因此,  $x_i$  与  $x_j$  之间的疏远函数可表示为:

$$A_{i,j} = D_{i,j} \cos \theta_{i,j} = \sum_{k=1}^K \varphi_k \max \{x_j^{(k)} - x_i^{(k)}, 0\} \cos \theta_{i,j} \quad (7)$$

对于前述(1,4)和(5,2)、(1,2)和(5,4),如果不考虑权重,虽然两者的平均相对剥夺  $D_{i,j}$  均为 6,

但是疏远函数  $A_{i,j}$  分别为  $78/\sqrt{493}$ 、 $78/\sqrt{205}$ ,前者小于后者,因而式(7)较为准确地刻画了维度相关性的影响。设  $x_i$  的认同函数为  $I(x_i) = f(x_i^{(1)}, \dots, x_i^{(K)})^\alpha$ ,从而可得多维极化的测度方法:

$$P = \int \dots \int f(x_i^{(1)}, \dots, x_i^{(K)})^{1+\alpha} f(x_j^{(1)}, \dots, x_j^{(K)}) D(x_i, x_j) \cos \theta_{i,j} dx_j^{(1)} \dots dx_j^{(K)} dx_i^{(1)} \dots dx_i^{(K)} \quad (8)$$

如果总体外生地被分成了若干个子群,本文可以从群体对抗的角度构造多维极化指标 (Esteban 和 Ray, 1994; Lasso 和 Urrutia, 2006)。就城乡两个子群之间多维极化而言,假设  $i$  为农村中的任意一个家庭,  $j$  为城镇中的任意一个家庭,分布函数分别为  $f(x_i) = f(x_i^{(1)}, \dots, x_i^{(K)})$ 、 $g(x_j) = g(x_j^{(1)}, \dots, x_j^{(K)})$ ,则与式(8)相对应的城乡多维极化测度公式为:

$$P = \sum_i \sum_j f(x_i)^{1+\alpha} g(x_j) \sum_{k=1}^K \varphi_k \max \{x_j^{(k)} - x_i^{(k)}, 0\} \cos \theta_{i,j} + \sum_j \sum_i g(x_j)^{1+\alpha} f(x_i) \sum_{k=1}^K \varphi_k \max \{x_i^{(k)} - x_j^{(k)}, 0\} \cos \theta_{j,i} \quad (9)$$

#### (二) 分解方法

为了进一步说明不同维度在多维极化中的贡献,本部分将探讨多维极化的分解方法,主要思路是将多维极化分解为不同维度极化之和的形式。

##### 1. 静态分解。

由于疏远函数可进行如下分解:

$$A(x_i) = \int \dots \int f(x_j^{(1)}, \dots, x_j^{(K)}) \sum_{k=1}^K \varphi_k \max \{x_j^{(k)} - x_i^{(k)}, 0\} \cos \theta_{ij} dx_j^{(1)} \dots dx_j^{(K)} \\ = \sum_{k=1}^K \varphi_k \int \dots \int f(x_j^{(1)}, \dots, x_j^{(K)}) \max \{x_j^{(k)} - x_i^{(k)}, 0\} \cos \theta_{ij} dx_j^{(1)} \dots dx_j^{(K)} \triangleq \sum_{k=1}^K \varphi_k A^{(k)}(x_i) \quad (10)$$

其中  $A^{(k)}(x_i) = \int \dots \int f(x_j^{(1)}, \dots, x_j^{(K)}) \max \{x_j^{(k)} - x_i^{(k)}, 0\} \cos \theta_{ij} dx_j^{(1)} \dots dx_j^{(K)}$ ,为  $x_i$  在第  $k$  个维度的疏远函数。可以看出,多维极化指标中的疏远函数为各个维度疏远函数之和,因此总的多维极化可表示为:

$$P = \sum_{k=1}^K \varphi_k \int \dots \int A^{(k)}(x_i) f(x_i^{(1)}, \dots, x_i^{(K)})^{1+\alpha} dx_i^{(1)} \dots dx_i^{(K)} \triangleq \sum_{k=1}^K \varphi_k P^{(k)} \quad (11)$$

其中  $P^{(k)} = \int \dots \int A^{(k)}(x_i) f(x_i^{(1)}, \dots, x_i^{(K)})^{1+\alpha} dx_i^{(1)} \dots dx_i^{(K)}$ ,为第  $k$  个维度的拟极化指数 (Pseudo - Polarization Index)<sup>①</sup>。进一步,由于第  $k$  个维度的极化指数为  $P^{(k)} = \int f(x_i^{(k)})^{1+\alpha} f(x_j^{(k)}) \max \{x_j^{(k)} - x_i^{(k)}, 0\}$

$$dx_j^{(k)} dx_i^{(k)}, \text{因而式(12)可以变为: } P = \sum_{k=1}^K \varphi_k R_k P^{(k)} \quad (12)$$

$$\text{其中 } R_k = \frac{\iint \dots \iint f(x_i^{(1)}, \dots, x_i^{(K)})^{1+\alpha} f(x_j^{(1)}, \dots, x_j^{(K)}) A_{i,j}^{(k)} \cos\theta_{i,j} dx_j^{(1)} \dots dx_i^{(1)} \dots dx_i^{(K)}}{\iint f(x_i^{(k)})^{1+\alpha} f(x_j^{(k)}) A_{i,j}^{(k)} dx_j^{(k)} dx_i^{(k)}} = \frac{P^{(k)}}{P^{(k)}}$$

反映了第  $k$  个维度与所有维度联合分布的某种相关性,可以理解为第  $k$  个维度与所有维度的极化相关系数<sup>②</sup>。当所有个体要素分布的夹角为 0,即所有列向量  $x^{(k)}$  两两之间的相关性均为 1 时,系数  $R_k = 1$ 。式(12)表明,多维极化可以分解为所有维度极化的加权之和,每个维度极化的

权重等于每个维度的权重与其极化相关系数的乘积。

类似地,关于式(9)的分解形式为:

$$P = \sum_{k=1}^K \varphi_k \sum_i \sum_j f(x_i)^{1+\alpha} g(x_j) \max\{x_j^{(k)} - x_i^{(k)}, 0\} \cos\theta_{i,j} + \sum_i \varphi_k \sum_j \sum_i g(x_j)^{1+\alpha} f(x_i) \max\{x_i^{(k)} - x_u^{(k)}, 0\} \cos\theta_{j,t} \triangleq \sum_{k=1}^K \varphi_k R_k P^{(k)} \quad (13)$$

$$\text{其中 } R_k = \frac{(\sum_i \sum_j f(x_i)^{1+\alpha} g(x_j) \max\{x_j^{(k)} - x_i^{(k)}, 0\}) + \sum_j \sum_i g(x_j)^{1+\alpha} f(x_i) \max\{x_i^{(k)} - x_u^{(k)}, 0\}) \cos\theta_{i,j}}{\sum_i \sum_j f(x_i^{(k)})^{1+\alpha} g(x_j^{(k)}) \max\{x_j^{(k)} - x_i^{(k)}, 0\} + \sum_j \sum_i g(x_j^{(k)})^{1+\alpha} f(x_i^{(k)}) \max\{x_i^{(k)} - x_j^{(k)}, 0\}} = \frac{P^{(k)}}{P^{(k)}}$$

为极化相关系数;第  $k$  个维度的极化指数为  $P^{(k)} = \sum_i \sum_j f(x_i^{(k)})^{1+\alpha} g(x_j^{(k)}) \max\{x_j^{(k)} - x_i^{(k)}, 0\} + \sum_j \sum_i g(x_j^{(k)})^{1+\alpha} f(x_i^{(k)}) \max\{x_i^{(k)} - x_j^{(k)}, 0\}$ 。

和极化效应。在动态分解结果中,当某项的影响效应为正时,表明该项加剧了多维极化;反之则对多维极化起抑制作用。将各项分别除以  $\Delta P$  得到的百分数,表示的是该项对多维极化变动的贡献率。

## 2. 动态分解。

## 四、中国城乡居民之间的多维极化分析

利用静态分解结果还可以对多维极化的变动进行分解。设  $P_t, P_{t+1}$  分别为第  $t$  期和  $t+1$  期的多维极化,  $\Delta P = P_{t+1} - P_t$  表示两期多维极化的变化,  $\Delta\varphi_k = \varphi_{k,t+1} - \varphi_{k,t}$  表示各个维度的权重变化,  $\Delta R_k = R_{k,t+1} - R_{k,t}$  表示第  $k$  个维度与所有维度极化相关系数的变化,  $\Delta P^{(k)} = P_{t+1}^{(k)} - P_t^{(k)}$  表示第  $k$  个维度极化的变动。一般来说,先有维度权重的结构变动,再有极化相关系数的变动,最后是各维度极化的变动,因此基于式(12)的多维极化变动有如下分解形式:

我国是典型的城乡二元社会,虽然近年来一直致力于打破城乡壁垒,但仍然没有完全摆脱体制束缚,农村居民与城镇居民在共享发展成果的程度仍有差异。因此,测度城乡在收入、财产等多个维度的极化问题,不仅可以刻画城乡两个群体之间多维极化的变动趋势及其成因,也能从一个角度反映当前城乡共享发展的变化。

### (一) 数据来源与处理

$$\Delta P = \sum_{k=1}^K (\varphi_{k,t+1} - \varphi_{k,t}) R_{k,t} P_t^{(k)} + \sum_{k=1}^K \varphi_{k,t+1} (R_{k,t+1} - R_{k,t}) P_t^{(k)} + \sum_{k=1}^K \varphi_{k,t+1} R_{k,t+1} (P_{t+1}^{(k)} - P_t^{(k)}) = \sum_{k=1}^K \Delta\varphi_k R_{i,t} P_t^{(k)} + \sum_{k=1}^K \varphi_{k,t+1} \Delta R_k P_t^{(k)} + \sum_{k=1}^K \varphi_{k,t+1} R_{k,t+1} \Delta P_t^{(k)} \quad (14)$$

本部分数据主要来自中国家庭追踪调查(CFPS)<sup>③</sup>。CFPS 重点关注中国居民的经济与非经济福利,于 2010 年正式开展调查,追踪收集了个体、家庭、社区三个层次的数据。基于当前人们的现实需求和 CFPS 的调查内容,本文选择家庭人均收入、人均财产、人均交通通信支出<sup>④</sup>和成人平均受教育年限四个指标。为消除价格影响,分别利用城镇和农村居民消费价格指数将历年收入、财产和交通通信支出数据折算为 2010 年不变价,并将各个指标分别除以算术平均数进行标准化处理。考虑到数据的可获得性和完整性,本文选取 2010 年、2012 年、2014 年和 2016 年四个年份的匿名数据,并匹配得到 4674 个家庭的追踪数据。

若各期的维度权重固定不变,则动态分解可简化为如下形式:

$$\Delta P = \sum_{k=1}^K \varphi_{k,t} \Delta R_k P_t^{(k)} + \sum_{k=1}^K \varphi_{k,t} R_{k,t+1} \Delta P_t^{(k)} \quad (15)$$

式(14)右边第一项中反映每个维度权重变化对多维极化变动的的影响,此谓结构效应;第二项反映每个维度极化相关系数变化对多维极化变动的的影响,称之为相关效应;第三项反映每个维度极化变化对多维极化变动的的影响,称之为维度极化效应。简化后的式(15)仅包括各个维度的相关效应

## (二) 城乡极化的测度

基于 CFPS 数据,可以从三个角度测度城乡极化。一是依据各年 CFPS 数据分别测度各年的城乡极化,并不考虑数据对应的家庭代码,此谓匿名性分析方法;二是依据追踪家庭数据测度各年城乡极化,此谓非匿名性分析方法;三是根据非匿名性计算所有家庭在一段时期内各个维度的长期数值,然后依据长期数值测算极化,此谓长期极化。大多数经验分析都是匿名性分析方法,这一方面是由于很多数据库不是追踪调查数据,只能根据匿名数据进行分析;另一方面是一些分析者没有意识到匿名性分析方法的缺陷,忽视了家庭在不同年份可能存在的流动性。依据长期数据测算的长期极化,能够反映一段较长时期的极化表现,可以弥补单一年份简单对比的不足。因此,本文将分别测算三种形式的城乡极化,重点是通过长期多维极化说明城乡之间极化的变动趋势及其原因。

### 1. 城乡单维极化。

假设  $i$  为农村中的任意一个家庭,  $j$  为城镇中的任意一个家庭,第  $k$  个维度的密度函数分别为  $f(x_i^{(k)})$ 、 $g(x_j^{(k)})$ ,则城乡居民之间第  $k$  个维度的极化为:

$$P^{(k)} = \sum_i \sum_j f(x_i^{(k)})^{1+\alpha} g(x_j^{(k)}) \max \{ x_j^{(k)} - x_i^{(k)}, 0 \} + \sum_j \sum_i g(x_j^{(k)})^{1+\alpha} f(x_i^{(k)}) \max \{ x_i^{(k)} - x_j^{(k)}, 0 \} \quad (16)$$

其中,参数  $\alpha \in [0.25, 1]$ ,两个密度函数均以 CFPS 中每个家庭人口规模与总人口的比重表示。此外,本文对参数  $\alpha$  的不同取值进行了测算,发现极化随着  $\alpha$  增加而下降,但同一时期四个维度极化的位次不发生改变,且除了匿名性极化变动趋势略有变动外,非匿名和长期极化的变动趋势不随参数的改变而发生变化,因此下面以  $\alpha = 0.25$  的测算结果进行分析,具体结果如表 1 所示。为便于分析极化的变动程度,将基期 2010 年极化数值看成 100,其他年份与之对比计算相对数。

可以看出,对于四个维度的极化而言,匿名性和非匿名性均显示极化程度最高的为财产,最低的为受教育程度,收入极化在 2010 年和 2012 年比交通通信支出的极化低,但在 2014 年和 2016 年高于交通通信支出。每个维度极化的变动较为频繁,但总体上变动幅度不是太大,其中 2016 年非匿名的财产和受教育年限极化分别比 2010 年上升 7%、5%,收入和交通通信支出极化分别下降 3%、12%。

表 1 中国城乡居民之间的四个维度极化

方法	年份	绝对数				相对数			
		收入	财产	交通通信	教育	收入	财产	交通通信	教育
匿名	2010	0.0285	0.0365	0.0308	0.0125	100	100	100	100
	2012	0.0289	0.0371	0.0295	0.0111	101	102	96	89
	2014	0.0288	0.0357	0.0281	0.0123	101	98	91	98
	2016	0.0289	0.0370	0.0279	0.0127	102	101	91	102
非匿名	2010	0.0338	0.0433	0.0367	0.0140	100	100	100	100
	2012	0.0325	0.0428	0.0340	0.0134	96	99	93	96
	2014	0.0323	0.0426	0.0321	0.0143	95	98	87	102
	2016	0.0327	0.0462	0.0323	0.0148	97	107	88	105
长期	2010	0.0338	0.0433	0.0367	0.0140	100	100	100	100
	2 年之和	0.0305	0.0418	0.0326	0.0094	90	97	89	67
	3 年之和	0.0290	0.0408	0.0299	0.0078	86	94	82	55
	4 年之和	0.0285	0.0417	0.0282	0.0069	84	96	77	49

就长期极化而言,四个维度的极化都呈现下降趋势,且随着观察时期的延长,下降幅度增大。与2010年相比,极化程度最高的财产在整个考察期内仅降低4%,下降幅度最小;之后依次为收入极化下降16%,交通通信支出极化下降23%。受教育年限极化下降了51%。可以看出,2010-2016年城乡之间的财产极化仅得到少许缓解,而受益于义务教育普及、高等教育扩招等政策,城乡在受教育水平上的极化得到了很大改善。

## 2. 城乡多维极化。

测度多维极化时涉及多维联合密度的求解,虽然关于居民收入分布函数的估计方法比较多,但其他维度分布函数的估计非常少,且多维分布函数将更加复杂,因此本文直接采用离散数据进行测度。依据式(9),本文分别按照熵权法和等权法计算了城乡之间的多维极化,具体测算结果见表2,其中熵权法的权重是由四个年份的长期数据计算得到。由于熵权法和等权法的权重差别不大,因而两种权重测算结果的差异很小。

首先,匿名性和非匿名性的测算结果虽然存在一定差异,但是两者变动趋势近乎一致。就四个年份的变动方向来看,两者都经历了“降→升→降”的

过程,只是变动幅度稍有差异。其中,匿名性多维极化最低的年份为2012年,而非匿名性多维极化最低的年份为2016年。总的来看,2016年城乡匿名性多维化相比2010年下降了8.8%,非匿名性多维极化下降了10.5%,城乡多维极化出现了一定程度的下降。

其次,随着观察期的延长,长期多维极化呈现出稳定的下降趋势。将四个年份加总起来的四年期长期极化与2010年相比,熵权法和等权法分别下降了11.6%和11.4%,比单一年份2016年相比2010年下降的幅度更大。这说明我国城乡居民之间的多维极化有了一定程度的改善,且这种改善真实地反映了农村居民相比城镇居民在多维极化上的逐渐缩小。

最后,与四个维度分别计算的极化相比,多维极化高于受教育年限的极化,但低于收入、财产和交通通信支出三个维度的极化,且四年期多维极化的下降幅度大于长期财产极化,低于其他三个维度的长期极化。因此,当前如果仅仅关注收入或财产等某个维度极化的变动而忽略多维极化,可能会对城乡极化做出不客观的判断,进而不能全面评价相关政策的实施效果。

表2 中国城乡居民之间的多维极化

方法	年份	绝对数		相对数	
		熵权法	等权法	熵权法	等权法
匿名	2010	0.0162	0.0163	100	100
	2012	0.0147	0.0148	90.73	90.79
	2014	0.0152	0.0153	93.78	93.82
	2016	0.0148	0.0148	91.20	91.21
非匿名	2010	0.0198	0.0198	100	100
	2012	0.0179	0.0180	90.61	90.63
	2014	0.0180	0.0181	90.97	91
	2016	0.0177	0.0178	89.47	89.47
长期	2010	0.0198	0.0198	100	100
	2年之和	0.0183	0.0184	92.74	92.87
	3年之和	0.0179	0.0180	90.43	90.63
	4年之和	0.0175	0.0176	88.36	88.59

### (三) 城乡多维极化的分解

#### 1. 静态分解结果。

依据式(13), 本文测算了城乡多维极化的静态分解结果, 由于熵权法和等权法的结果比较接近, 因而表3只展示了熵权法的分解结果。可以看出, 无论匿名性还是非匿名性多维极化, 每个维度贡献率的排序没有变化, 从大到小依次为财产、交通通信支出、收入和受教育年限, 其中前三个维度贡献率的差距不大, 大致处于25%~31%, 教育的贡献率相对较小, 约为15%。此外, 四个维度在每个年份贡献率的变动幅度都较小, 最大变动幅度不超过3%, 说明四个维度的贡献率较为稳定。

就长期多维极化的贡献率而言, 随着时间的延长, 长期财产极化的贡献率呈现稳步上升, 四年财产极化的贡献率上升到35.25%; 长期教育极化的贡献率逐步下降, 由2010年的14.02%逐步下降到四年的8.39%, 而长期收入和长期交通通信支出的贡献率变化不大, 基本稳定在27%和30%。

#### 2. 动态分解结果。

根据式(15), 本文计算了城乡多维极化的动态分解, 熵权法的分解结果见表4。从匿名性和非匿名性的动态分解结果来看, 相关效应除了2012-2014年外均小于0, 也即多数年份相关效应起着缩

小多维极化的作用。2010-2016年相关效应对匿名性和非匿名性多维极化下降的贡献率分别达到87%、92% (由表中数据除以多维极化的变动量得到, 下同), 是促使多维极化下降的最重要原因。这其中收入和财产的相关效应最大, 两者对非匿名性多维极化下降的贡献率分别为34%和47%, 匿名性的贡献率分别是38%和36%。维度极化效应的贡献率相对较小, 这也从一个侧面说明城乡各个维度极化在各年的变动不大。不过值得注意的是, 2010-2016年交通通信支出的极化效应有利于多维极化的下降, 匿名性和非匿名性的贡献率分别达到39%、43%; 而财产的极化效应起着促进多维极化上升的作用, 匿名性和非匿名性两种情形下对多维极化下降的贡献率分别约为-20%、-30%。

与两个年份简单对比不同的是, 相关效应均促使长期多维极化的上升, 极化效应均有利于长期多维极化的下降。相关效应对两年期、三年期和四年期多维极化下降的贡献率分别为-43%、-55%和-40%, 这其中交通通信支出的负贡献率最多, 分别达到-22%、-28%和-27%; 其他三个维度的相关效应也都促使多维极化的上升, 只是数值相对较小, 对四年期多维极化下降的贡献率分别为-3%、-4%和-6%。对应地, 极化效应是促使长期

表3 中国城乡居民之间多维极化的静态分解 (%)

方法	年份	收入	财产	交通通信支出	教育
匿名	2010	26.88	28.39	29.34	15.39
	2012	27.07	29.76	29.37	13.80
	2014	25.89	30.14	28.43	15.53
	2016	25.81	29.58	28.02	16.58
非匿名	2010	27.03	29.33	29.62	14.02
	2012	26.83	30.15	29.16	13.86
	2014	25.34	31.45	28.16	15.05
	2016	25.83	30.85	27.61	15.71
长期	2年之和	27.10	32.09	30.38	10.43
	3年之和	26.64	34.09	30.15	9.12
	4年之和	27.08	35.25	29.28	8.39

表 4 中国城乡居民之间多维极化的动态分解

方法	年份	维度相关效应(%)					维度极化效应(%)				
		收入	财产	交通通信	教育	总和	收入	财产	交通通信	教育	总和
匿名	2010 - 2012	-0.36	-0.26	-0.22	-0.19	-1.04	-0.01	0.04	-0.22	-0.27	-0.46
	2012 - 2014	-0.06	0.15	0.22	0.06	0.37	0.02	0.05	-0.22	0.27	0.12
	2014 - 2016	-0.11	-0.40	-0.06	-0.01	-0.59	-0.01	0.19	-0.12	0.10	0.17
	2010 - 2016	-0.54	-0.51	-0.05	-0.13	-1.23	0.00	0.28	-0.56	0.09	-0.19
非匿名	2010 - 2012	-0.39	-0.52	-0.23	-0.20	-1.35	-0.14	0.13	-0.40	-0.09	-0.51
	2012 - 2014	-0.02	0.14	0.23	0.04	0.39	-0.23	0.11	-0.38	0.18	-0.32
	2014 - 2016	-0.27	-0.61	-0.08	-0.01	-0.97	0.28	0.41	-0.10	0.08	0.68
	2010 - 2016	-0.71	-0.97	-0.09	-0.16	-1.92	-0.07	0.63	-0.89	0.17	-0.16
长期	2010 与 2 年长期	0.11	0.10	0.31	0.08	0.61	-0.49	-0.02	-0.60	-0.94	-2.04
	2010 与 3 年长期	0.15	0.23	0.54	0.13	1.04	-0.73	0.07	-1.00	-1.27	-2.94
	2010 与 4 年长期	0.07	0.09	0.61	0.14	0.91	-0.68	0.27	-1.35	-1.45	-3.21

多维极化下降的唯一原因,其中长期教育的极化效应对多维极化下降的影响最大,两年期、三年期和四年期贡献率分别达到 66%、67% 和 63%。受教育年限的长期极化效应一直在抑制长期多维极化的增加,与两个年份直接比较的极化效应存在明显差异,特别是 2016 年相比 2010 年的受教育水平极化效应促进了多维极化的增加。由于个人的受教育水平不会下降,因此城乡之间长期受教育水平极化的逐步下降及其对长期多维极化较大的抑制效应,极有可能是部分农村居民在 2010 - 2016 年之间提高了自身受教育水平<sup>⑤</sup>。此外,长期交通通信支出的贡献率排在第二,分别为 42%、53% 和 59%,长期收入对三个时期多维极化下降的贡献率分别约为 34%、38% 和 30%;而长期财产极化效应的贡献率分别为 1%、-4% 和 -12%,是三年期和四年期多维极化增加的因素之一。不同于收入、支出等流量指标,作为存量的财产极化具有一定的累积效应,需要引起警惕。

### 五、结论

在多维极化测度中,测度方法的选择、不同维度及指标的确定、数据的时期设定等都会对结果产生影响,因此寻求相对科学的测度方法是十分必要

的。鉴于现有多维极化测度方法没有准确考虑维度之间的相关性,也没有考虑冲突主要来源于被剥夺的一方,本文引入了余弦相似度作为不同维度相关性的度量指标,并根据相对剥夺的概念构建疏远函数,从而提出了一个新的多维极化测度方法。该测度方法还可以进行静态和动态分解,能够分析各个维度的相关效应和极化效应分别对多维极化的影响。此外,由于每个观察单位在不同年份收入分布、财产分布等分布中的位置可能发生变动,简单将两个年份对比分析的结论可能不准确,需要根据一段时期的长期收入、长期财产等测度长期多维极化。

运用新的多维极化测度方法。本文依据 CFPS 数据实证分析了 2010 年、2012 年、2014 年和 2016 年我国城镇居民家庭和农村居民家庭之间的多维极化,结论主要有四点。第一,虽然 2010 - 2016 年城乡收入、财产、交通通信支出和受教育年限的极化变动方向不一致,但四个维度的长期极化都呈现明显的下降趋势。第二,多维极化表现出不断下降之势,且四年期多维极化下降的幅度远高于 2016 年相比 2010 年的下降幅度。第三,在每个年份的多维极化中,贡献率从大到小依次为财产、交通通



信支出、收入和受教育年限。第四,相关效应是2016年相比2010年多维极化下降的主要原因,但却是长期多维极化上升的因素,而极化效应为长期多维极化下降的唯一原因,其中教育极化效应对长期多维极化下降的贡献率最大,其次是交通通信支出和收入,而财产则是三年期和四年期长期多维极化上升的因素。本文实证结果的一个政策含义是,今后需要将城乡共享发展的重心由收入逐渐转向财产、交通通信支出等方面,要提升城乡之间各个要素的双向流动,避免资源过度集中在城市,稳步缩小城乡在经济、社会、文化等各个方面的多维极化,从而真正实现城乡融合发展。

注释:

①之所以称为拟极化,是因为测量某个维度的极化时,涉及到的密度函数应该是该维度分布的密度函数,而拟极化是所有维度联合分布的密度函数,且本文还包括了相关性指标  $\cos\theta_{i,j}$ 。这个说法类似于基尼系数分解中的拟基尼系数,由 Fei 等(1978)等首先提出。

②Lerman 和 Yitzhaki(1985)在研究基尼系数按收入来源的分解方法中,提出了基尼相关系数(Gini Correlation Coefficient)的概念,受此启发本文提出了极化相关系数。

③中国家庭追踪调查(China Family Panel Studies)来自北京大学“985”项目资助,由北京大学中国社会科学调查中心实施。

④交通和通信产品的升级换代是社会经济发展的重要体现,交通和通信支出也是反映城乡居民生活质量的重要指标,并且在当前消费拉动经济增长中具有不可替代的作用,因此本文选择交通通信支出作为刻画城乡极化的一个维度。

⑤《中国人口和就业统计年鉴》数据显示:2010年我国农村居民的高中、大专及以上人口占比分别为7.72%和2.06%,2016年分别上升到10.39%和3.11%。

参考文献:

民收入两极分化[J]. 经济研究,2007,42(11):139-153.

[2]洪兴建. 居民收入分配失衡的测度方法研究[M]. 北京:经济科学出版社,2010.

[3]罗楚亮. 我国居民收入分布与财产分布的极化[J]. 统计研究,2018,35(11):82-92.

[4]王朝明,李梦凡. 极化效应下我国中等收入者群体的发展问题[J]. 数量经济技术经济研究,2013,30(6):51-64.

[5]Anderson G. Polarization of the Poor - Multivariate Relative Poverty Measurement Sans Frontiers[J]. Review of Income and Wealth, 2010, 56(1):84-101.

[6]Araar A. On the Decomposition of Polarization Indices: Illustrations with Chinese and Nigerian Household Surveys [R]. Cahier de recherche/Working Paper 08-06, 2008.

[7]Chakravarty S R. Inequality, Polarization and Poverty: Advances in Distributional Analysis [M]. New York: Springer, 2009.

[8]Duclos J Y, Esteban J, Ray D. Polarization: Concepts, Measurement, Estimation[J]. Econometrica, 2004, 72(6):1737-1772.

[9]Esteban J M, Ray D. On the Measurement of Polarization[J]. Econometrica, 1994, 62(4):819-851.

[10]Fei J C H, Rainis G, Kuo S W Y. Growth and the Family Distribution of Income by Factor Components [J]. Quarterly Journal of Economics, 1978, 92(1):17-53.

[11]Foster J E, Wolfson M C. Polarization and the Decline of the Middle Class: Canada and the U. S. [R]. Vanderbilt University Working Paper, 1992.

[12]Gigliarano C, Mosler K. Constructing Indices of Multivariate Polarization [J]. Journal of Economic Inequality, 2009, 7(4):435-460.

[13]Lambert P J, Aronson J R. Inequality Decomposition Analysis and the Gini Coefficient Revisited [J]. Economic Journal, 1993, 103(420):1221-1227.

[14]Lasso D L V, Urrutia M A. An Alternative Formulation of the Esteban Gradin - Ray Extended Measure of Bipolarization [J]. Journal of Income Distribution, 2006(15):42-54.

[1]洪兴建,李金昌. 两极分化测度方法述评与中国居

[15] Lerman R I, Yitzhaki S. Income Inequality Effects by Income Source: A New Approach and Applications to the United States[J]. *Review of Economics and Statistics*, 1985, 67 (1): 151 – 156.

[16] Merz J, Scherg B. Polarization of Time and Income—A Multidimensional Analysis for Germany, Economic Well – Being and Inequality: Papers from the Fifth ECINEQ Meeting (Research on Economic Inequality, Vol. 22), Emerald Group Publishing Limited, 2014, pp. 273 – 321.

[17] Runciman, W. Relative Deprivation and Social Justice: A Study of Attitudes to Social Inequality in Twentieth – Century England[M]. Berkeley: University of California Press, 1966.

[18] Srisuma S, Vesperoni A. An Index of Multivariate

Polarization[R]. Working Paper, 2015.

[19] Tewodaj M. A Two – Dimensional Measure of Polarization[R]. IFPRI Discussion Paper 00837, 2008.

[20] Tsui K Y. Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation[J]. *Social Choice & Welfare*, 1999, 16 (1): 145 – 157.

[21] Wang J, Caminada K, Wang C. Measuring Income Polarization for Twenty European Countries, 2004 – 2013: A Shapley Growth Redistribution Decomposition[J]. *Eastern European Economics*, 2017, 55(6): 1 – 23.

[22] Yitzhaki S. Relative Deprivation and the Gini Coefficient[J]. *Quarterly Journal of Economics*, 1979, 93 (2): 321 – 324.

## Multidimensional Polarization: New Measurement and Its Application

Hong Xingjian      Dong Jun

**Abstract:** As the existing measurements of multidimensional polarization fail to reflect the correlation between dimensions, this paper introduces cosine similarity, and proposes a new multidimensional polarization measurement based on the relative deprivation theory. The method can also be decomposed statically and dynamically, and can analyze the contribution rate of each dimension and its change to multidimensional polarization. Using the data from China Family Panel Studies (CFPS), this paper empirically analyzes the multidimensional polarization between urban and rural households in China from four dimensions: income, properties, expenditure on transportation and communications, and educational level. The empirical results show that the long – term multidimensional polarization shows a significant downward trend, and the polarization effect of education, expenditure on transportation and communications, and income is the main reason, but the polarization effect of properties promotes multidimensional polarization.

**Key words:** multidimensional polarization; measurement; dimensional decomposition; urban and rural multidimensional polarization