

基于分解的库兹涅兹指数的 中国地区间行业收入差距问题研究

赵丹^a, 王涛^b

(山西财经大学 a.公共管理学院; b.统计学院, 太原 030006)

摘要:当我们研究行业收入差距问题时,均通过构建一个指数来综合测度一国或一地区的行业收入差距,比如基尼系数、变异系数等,但指数方法缺乏对各个行业的深入考察。文章对库兹涅兹指数(“库指数”)进行了分解,分解后的“库指数”有效的解决了收入差距测度指数不能分解的问题。基于分解的“库指数”,进一步讨论了中国各地区行业收入差距的特征,并通过因子分析的方法找出造成行业收入差距扩大的主要行业。

关键词:行业收入差距;基尼系数;库兹涅兹指数

中图分类号:C81

文献标识码:A

文章编号:1002-6487(2016)03-0133-04

1 问题的提出

收入差距是学界一直以来关注的主题,随着经济全球化的进一步深入,世界格局的改变,以及经济金融危机的频发,收入分配问题也呈现多元、不确定的变化趋势,甚至在一国内部,不同区域间,收入分配差距的问题也存在多样性。因此,收入分配问题的研究内容、方法也在随之发生改进。

学界关于行业收入差距测度方法较多,比较简单而且易懂的方法主要是涉及少数几个样本比较的指标,比如极差、分位差、高低比等,这些方法有共同的缺点,缺乏对所有行业的考察,也缺乏对行业中所有数据的分析,因此缺乏遍历性,但是作为精度一般的说明性阐述也可满足。相应的,较为精确的具有代表性的测度方法需要考察每个行业 and 每个数据,这样的方法也比较丰富和成熟,较为常见的方法有变异系数、麦克伦指数、收入不良指数、库兹涅兹指数、威廉逊指数、基尼系数。这些方法在测度行业收入差距时存在一些缺点,主要表现在两个方面:其一,大多数国内外文献均采用单个指标或指数来度量差距,简单方法有极差、高低比,比较复杂的方法有变异系数、基尼系数、泰尔指数等。单一指标反应的结论虽然能就总体收入差距问题进行描述,但是,不同的指标的理论出发点不同,表现形式不同,则对同一个问题测度的结果可能存在不一致。其二,上述测度指标和方法,均着眼于一个国家或地区的整体情况,并不关注每个行业在收入不平等中的影响,比如,在行业收入差距的研究中,我们往往只关注总体的不平等程度,却忽视了究竟是哪些行业属于高收入行业,以及他们对收入不平等的影响有多大,同时,各行业间

收入高低的比较,在各个测度指标中均没有反应。基于以上的分析,本文将库兹涅兹指数进行了分解,进而研究行业收入差距问题。

2 库兹涅兹指数的分解和应用

2.1 库兹涅兹指数的介绍

库兹涅兹指数来源于库兹涅兹对经济发展和收入分配的解读,他将二者的关系绘制在平面图中,形成了库兹涅兹曲线,该曲线以人均财富增长为横坐标,以人均财富分配为纵坐标,二者关系遵循倒U型曲线规律。一国或一地区发展初期,经济不发达,那么收入分配相对公平,随着经济的增长,收入分配出现两极分化,收入不平等现象日趋严重,当到经济发展到一定阶段的时候(富裕阶段),收入差距达到了高峰值,也称之为拐点,之后,随着经济的进一步增长,收入差距开始缩小,不平等现象缓解,直到极度发达时,收入差距几乎不存在,这是对库兹涅兹曲线的解释。将该理论投影到社会发展阶段,则有如下的发展变化规律:农业社会时,收入分配相当均等;工业化开始后,农业社会向工业社会转变,经济快速增长,由于劳动力迅速向高收入的生产部门流动,人均财富分配不平衡明显加剧,曲线上扬;一段时间内,工业社会继续发展,曲线相对平缓;工业化完成之后,收入分配不平衡缩小,曲线下降。

库兹涅兹比率库兹涅兹在研究经济发展与收入分配关系时提出的测度收入分配差距的一种方法。计算公式为:

$$R = \sum_{i=1}^n |y_i - p_i| \quad (1)$$

其中: R 为库兹涅兹比率, y_i 、 p_i 分别表示各组的收入比重和人口比重。如果计算得到的结果 R 越大,表示

基金项目:国家社会科学基金资助项目(14ATJ003;13CTJ010);山西省高等学校哲学社会科学研究项目(PSSR)

作者简介:赵丹(1981—),女,山西阳城人,博士,讲师,研究方向:宏观经济统计。

收入差距就越大,反之,收入差距就越小。该方法从计算角度来讲比较简单,也更容易理解,如果收入比重和人口比重相等,那么绝对值为0,如果这两者差距越大,那么结果也会相应增大。而且直接从该指数的计算公式来看,该指数的适用范围也没有直接明了的说明,对于问题的解释需要进一步的挖掘。

2.2 库兹涅茨指数的分解

本文以库兹涅茨指数为基础,借鉴泰尔指数和威廉逊指数的理论经验,对库兹涅茨指数进行变换和分解,从另一个角度解释该指数,从而为本文随后的研究提供理论支持。

假设 W 表示所有组的收入总数, w_i 表示第 i 组的收入总数, H 表示所有组的人口总数, h_i 表示第 i 组的人口总数, n 表示所有的组数,那么库兹涅茨指数可表示成如下形式:

$$R = \sum_{i=1}^n \left| \frac{w_i}{W} - \frac{h_i}{H} \right| \quad (2)$$

其中 $y_i = \frac{w_i}{W}$ 表示 i 组的收入占总收入的比重, $p_i = \frac{h_i}{H}$ 表示第 i 组的人口占总人口的比重。下面对该式进行进一步的变形和转换:

$$\begin{aligned} R &= \sum_{i=1}^n \left| \frac{w_i}{W} - \frac{h_i}{H} \right| \\ \Leftrightarrow R &= \sum_{i=1}^n \frac{h_i}{H} \left| \frac{w_i}{h_i} \times \frac{H}{W} - 1 \right| \\ \Leftrightarrow R &= \sum_{i=1}^n \frac{h_i}{H} \left| \frac{w_i}{h_i} \times \frac{W}{H} - 1 \right| \\ \Leftrightarrow R &= \frac{H}{W} \sum_{i=1}^n \frac{h_i}{H} \left| \frac{w_i}{h_i} - \frac{W}{H} \right| \\ \Leftrightarrow R &= \frac{1}{\bar{x}} \sum_{i=1}^n p_i |\bar{x}_i - \bar{x}| \end{aligned} \quad (3)$$

其中, $\bar{x} = \frac{W}{H}$ 表示所有组的人均收入, $\bar{x}_i = \frac{w_i}{h_i}$ 表示第 i 组的人均收入,此时的 R 由两部分构成,分别是 $\frac{1}{\bar{x}}$ 和 $\sum_{i=1}^n p_i |\bar{x}_i - \bar{x}|$,其中 $\sum_{i=1}^n p_i |\bar{x}_i - \bar{x}|$ 成为不平等测度的关键, $|\bar{x}_i - \bar{x}|$ 表示第 i 组的人均收入与总的平均收入的绝对差距, p_i 表示第 i 组的人口比重,即第 i 组不平等的权重,这个部分充分解释了收入不平等的成因,收入不平等由各行业平均收入与所有行业平均收入的绝对差距以及该行业的就业比重共同造成。由于各行业收入与所有行业平均收入绝对差具有量纲,因此,总体除以行业平均工资 \bar{x} ,从而消除了量纲的影响,使得库指数具有可比性。又由于库指数的主要部分是求和的形式,求和项数多少依据所分析的问题而定,因此,该指数只适合同类型问题的比较,而且可以比较分组较多情形,比如可以比较中国不同地区行业收入差距问题。

除此之外,变换之后的形式具有比其他测度指标更加

明显的优势,即该指数具有可分解性,可以将每个行业对总体行业收入差距的影响进行单独研究,形成每行业对总体行业收入差距的贡献率。假设 m_i 表示第 i 个行业对总体行业收入差距贡献率,则

$$m_i = p_i |\bar{x}_i - \bar{x}| / \sum_{i=1}^n p_i |\bar{x}_i - \bar{x}|, 0 < m_i < 1, \sum_{i=1}^n m_i = 1 \quad (4)$$

虽然贡献率 m_i 的计算给出了各行业对总体行业收入差距的影响,但由于公式中绝对值得存在,忽略了每个行业对行业收入差距正负影响,即高收入行业 and 低收入行业在此式中没有能够体现,因此,可以将上式进行简单变换,得到每个行业对总体收入差距的相对贡献率。

$$h_i = p_i (\bar{x}_i - \bar{x}) / \sum_{i=1}^n p_i (\bar{x}_i - \bar{x}), \sum_{i=1}^n h_i = 0 \quad (5)$$

该贡献主要特点在于不仅区分贡献率的绝对值大小,而且可以区分贡献的方向,即高收入拉动还是低收入拉动。 h_i 表示第 i 个行业对整个地区行业收入差距的相对贡献率, h_i 的绝对值即为 m_i 。下面笔者运用分解库兹涅茨指数对我国各地区的行业收入差距进行研究。

3 中国各区域行业收入差距特征分析

表1 中国各地区行业收入差距的库指数分析表

地区	总指数	高收入拉动最大行业	相对贡献率(%)	低收入拉动最大行业	相对贡献率(%)
山西	0.310765	采矿业	40.8	公共管理	-11.6
浙江	0.300802	金融业	15.1	建筑业	-24.3
北京	0.277943	金融业	22.1	制造业	-13.2
上海	0.265571	金融业	24.3	制造业	-25.8
海南	0.255941	教育	14.5	农业	-24.9
江苏	0.234681	公共管理	13.7	制造业	-29.4
广东	0.224343	金融业	16.5	制造业	-31.8
宁夏	0.221569	采矿业	26.7	制造业	-13.0
黑龙江	0.21555	采矿业	14.7	农业	-39.3
新疆	0.203698	采矿业	22.6	农业	-39.5
天津	0.202913	金融业	9.4	制造业	-15.9
内蒙古	0.19765	教育	12.7	农业	-20.3
重庆	0.188406	金融业	19.1	建筑业	-25.5
辽宁	0.186197	金融业	13.2	农业	-16.1
四川	0.181449	金融业	12.9	建筑业	-24.5
河北	0.178042	采矿业	15.7	建筑业	-14.2
山东	0.172653	采矿业	15.7	制造业	-25.1
陕西	0.166093	采矿业	14.6	制造业	-15.6
福建	0.161058	金融业	16.2	制造业	-37.2
云南	0.157108	金融业	21.7	建筑业	-26.4
河南	0.150894	采矿业	23.1	制造业	-15.3
青海	0.149985	科学研究	9.4	建筑业	-14.6
吉林	0.138413	金融业	15.4	农业	-17.9
安徽	0.136804	采矿业	33.8	建筑业	-12.2
广西	0.132746	金融业	22.5	制造业	-12.8
甘肃	0.127721	采矿业	20.8	公共管理	-12.2
湖南	0.12722	金融业	19.5	建筑业	-21.1
贵州	0.121576	金融业	22.2	制造业	-17.5
江西	0.099024	金融业	11.8	建筑业	-19.9
湖北	0.090371	金融业	17.0	批发和零售业	-11.7

3.1 各省收入差距总指数和贡献率的结果分析

根据2013年国家统计年鉴,我国各地区的行业门类包括:农林牧渔业、采矿业、制造业、电力燃气及水的生产和供应业等19类,行业数据统计对象是城镇单位,按照前面提到的分解计算方法,分别结算相对贡献率和总库兹涅兹指数,鉴于所计算的结果数据较多,可以得到 30×19 的数据表,笔者摘取各地区(西藏除外)相应的总指数即具有代表性的行业进行说明计算的结果如表1所示。

由表1可以得出如下各区域行业收入差距特征:

(1)金融业在16个省份均是最高收入拉动收入差距的行业,超过了全国的一半,有些地区金融业对行业收入差距的贡献率几乎接近四分之一,比如上海和北京,这说明这两个地区的金融行业较发达,不仅平均收入高,而且吸引了大量的金融从业人员,这与北京和上海的金融中心地位相符。其次,一些不发达省份比如云南、广西、甘肃和贵州的金融行业的相对贡献均超过了五分之一,究其原因在于,这些地区金融行业的平均工资远高出行业平均工资,几乎是当地平均工资的两倍。

(2)对于一些矿产资源较发达省份,特别是中西部地区,采矿业的相对贡献较高,比如山西的采矿业的高收入拉动行业不平等达到了40%,这是所有地区相对贡献率最高的,这也是造成山西是行业收入差距最大的主要原因。其他中西部地区,比如安徽、甘肃、新疆河北、陕西等,采矿业的相对贡献均是最高。

(3)对于最低收入拉动不平等的行业主要是制造业,江苏、福建、广东、上海等沿海制造业大省,均存在同样的现象,主要原因在于这些省份的低端制造加工业较繁荣,吸引着大量的内地流动务工人员,又因为低端制造业工资较低,两者共同影响使得这些地区的制造业低收入拉动效果突出。

(4)除了制造业之外,低收入拉动行业还有建筑业和农业,浙江、重庆、四川、湖南等南部省份的建筑业均是最低拉动,原因类似于制造业,其一,从业人员较多,其二,建筑业工资较低。农业的不平等现象主要集中在东北三省、内蒙古和新疆等农业大省,农业的平均工资在很多省份均是最低的,再加之这三个省份的农业从业人员较多,最后导致农业的低收入拉动最明显。

3.2 各地区行业不平等的特征研究

通过计算总指数与相对贡献,我们对中国各地区的大致情况有了初步的认识,但是缺乏规律性或集中性的认识,在此采用多元统计方法,希望得出相关结论:其一,行业收入不平等现象不可避免,那么全国30个地区的不平等现象是否存在相似性,是否某些省份的不平等现象相似,而另外一些省份的又具有相同的特征,此分析有利于清晰的认识区域特征;其二,对于19个行业来讲,各个行业在不同省份对行业收入差距的影响是不同的,那么哪些行业对整个行业收入差距的影响是明显的。

3.2.1 中国各地区行业收入差距的聚类分析

聚类分析是多元统计分析中常用的方法,其目的是对

变量或样本进行分类,使得分类之后的数据具有更加明显的特征。全国包含30个地区,通过聚类的方法,可以将行业收入差距类似的省份归为一类,这样分析使的全国各地区的行业收入差距特征呈现空间相似性。通过前面相对贡献率的计算,可以得到 30×19 原始数据表,笔者以19个行业为变量,30个省份为样本,采取Q性聚类方法,对30个省份进行聚类(聚类过程略),聚类的结果如表2所示。

表2 30个地区行业收入差距聚类分析

类别	地区
1	北京、湖北、湖南、广西、贵州
2	上海、江苏、福建、广东、浙江
3	山西、安徽、河南、宁夏
4	黑龙江、海南、新疆
5	天津、河北、内蒙古、辽宁、吉林、江西、山东、重庆、四川、云南、陕西、青海、
6	甘肃

聚类采取K-均值聚类方法,共得到6个分类,每一类中的地区均具有相似的特征,有些特征比较明显可以直接描述,比如第2类中有五个省份,这五个省份均是沿海制造业大省,金融业也相对发达,而且地理区域上相邻,使得这四个地区的收入不平等具有相似性。第3类中有四个省份,这四个省份均是地处中西部的资源大省,而且这些地区资源行业的收入不平等贡献率最高。其他类别中的地区,也呈现相同的特征,但是这种相似性不是非常的明显。

3.2.2 影响收入差距的主要行业分析

在对全国30个地区进行聚类之后,笔者将对19个行业进行深入分析,19个行业的收入变化在30个地区的不同表现使得30个地区的行业收入差距呈现不同的特征,探究哪些行业成为影响收入差距如此大的原因,笔者采用的方法是多元统计中的因子分析。

因子分析是一种典型的降维和简化数据的计数,通过研究众多变量之间的内部依赖关系,探求观测数据中的基本结构,并用少数几个“抽象”的因子变量来表示其基本数据结构,其中,因子变量便是能够反映众多变量的主要信息。

在做因子分析前,首先对数据进行KMO和Bartlett检验,KMO的结果越接近于1,说明因子分析的结果会很好,Bartlett的结果中P值越接近于0,说明原始数据越适合做因子分析。检验结果如表3所示。

表3 聚类分析的KMO和Bartlett'检验

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.371
Bartlett's Test of Sphericity	Approx. Chi-Square	557.085
	df	171
	Sig.	0

其中,Bartlett值接近于0,说明数据适合做因子分析,而KMO值比较小,说明如果做因子分析,那么结果可能不会很理想,笔者结合原始数据,将19个变量进行降维,确实存在一些未知的因素,但是还是可以从结果找到一些规律。因子分析的结果如表4、表5所示。

表4 因子分析贡献率

	特征值	方差贡献率	累计贡献率
1	4.7630291	25.068574	25.068574
2	2.4504388	12.897046	37.965621
3	2.1826938	11.487862	49.453483
4	1.9945989	10.497889	59.951372
5	1.7718772	9.3256694	69.277041
6	1.2810618	6.7424305	76.019472
7	1.0233469	5.3860364	81.405508

表5 旋转后因子载荷矩阵

	1	2	3	4	5	6
农业	0.202	0.574	-0.388	-0.014	-0.263	0.498
采矿业	-0.807	0.030	-0.466	-0.106	0.197	-0.162
制造业	-0.167	-0.882	-0.191	0.071	0.024	0.004
电力、热力	-0.046	-0.040	-0.032	-0.072	0.144	0.023
建筑业	-0.195	-0.028	-0.071	0.112	-0.043	-0.898
批发和零售业	0.363	0.436	0.123	0.095	0.375	-0.285
交通运输	0.002	-0.732	0.297	-0.398	-0.039	0.080
住宿和餐饮业	-0.528	-0.012	-0.042	-0.371	0.496	0.100
信息传输、	0.398	0.300	0.081	0.716	-0.209	-0.037
金融业	0.875	0.155	-0.160	0.252	0.074	0.092
房地产业	0.025	0.154	0.241	-0.862	-0.050	0.153
租赁和商务	-0.157	0.166	0.278	0.556	-0.225	0.421
科学研究	-0.008	-0.057	0.063	0.185	-0.705	0.271
水利、环境	-0.079	0.817	0.307	-0.047	-0.130	0.141
居民服务	-0.061	-0.171	-0.038	0.102	0.830	0.169
教育	-0.027	-0.010	0.887	-0.018	-0.031	0.102
卫生	0.849	0.046	0.055	-0.314	-0.001	0.120
文化、体育	0.331	0.334	0.263	0.353	0.049	0.171
公共管理	0.241	0.345	0.809	-0.129	-0.065	-0.099

由上面的因子分析结果不难看到,每个因子的方差贡献率并不是很大,即使是第一个因子,方差贡献率也只有25%,而其他因子方差贡献率只有10%左右,这可能主要

因为变量较多,而且城乡多元变化的现象,但即使如此,还是可以发现一些特征:

(1)第一因子的方差贡献率为25%,也就是说明第一因子所包含的信息能够解释行业收入差距变动的25%,说明第一因子的影响还是比较明显的。而随后的几个因子的方差贡献率迅速缩小,说明这几个因子相对于第一因子的影响比较弱。

(2)观察旋转后的因子载荷矩阵,第一因子在采矿和金融业的载荷较大,这与我们之前的分析相同,这两个行业对收入差距的影响较大,只是一个呈现低收入拉动,一个呈现高收入拉动。

(3)第二个因子在农业、制造业、交通运输和水利环境的载荷较大,这些行业都是与民众生活息息相关的行业。农业和制造业在多个地区对收入差距的拉动比较明显,而交通运输和水利等行业的影响需要进一步的研究。

参考文献:

- [1]王志江,胡日东.修正加权变异系数:度量收入分配平等程度的有用指标[J].数量经济技术经济研究,2006,(6).
- [2]章洪海,苟娟娟.也谈中国的两极分化现状[J].经济问题,2005,(6).
- [3]程永宏.改革以来全国总体基尼系数的演变及其城乡分解[J].中国社会科学,2007,(4).
- [4]洪兴建.基于S基尼系数的中国行业工资差距分析[J].统计研究,2010,(5).
- [5]洪兴建.一个新的基尼系数子群分解公式——兼论中国总体基尼系数的城乡分解[J].经济学(季刊),2008,(8).

(责任编辑/浩 天)