

改进熵值法问题的初探

苏洁 沈文成 浙江理工大学经济管理学院 杭州 310018

【内容摘要】

熵值法是一种客观赋权方法,它通过计算指标的信息熵,根据指标的相对变化程度对系统整体的影响来决定指标的权重,相对变化程度大的指标具有较大的权重,此方法现广泛应用于统计学等各个领域,具有较强的研究价值。但随着社会的发展、科学的进步及我们研究问题的复杂性越来越高,传统的熵值法已经不能完全满足研究的需要,这样就有必要对它进行一定的改进。本文就针对此问题进行了初步探索,并对改进的熵值法和其它客观赋权法的进行了比较。

【关键词】

熵值法;改进熵值法;比较

一、绪论

熵原是热力学的一个物理概念。在信息系统中的信息熵是信息:无序度的度量,信息是系统有序程度的度量,两者绝对值相等,符号相反。信息熵越小,信息的无序度越低,其信息的效用值越大,指标的权重也越大;反之,信息熵越大,信息的无序度越高,其信息的效用值越小,指标的权重也越小。据此性质,统计学广泛应用信息熵反映系统信息的有序程度和信息的效用值,进行客观赋权从而作出综合评价。

二、熵值法的主要原理

设有 m 个样本, n 项评价指标,形成原始指标数据矩阵 $X = (x_{ij})_{m \times n}$, 对于某项指标 x_j , 指标值的差距越大,则该指标在综合评价中所起的作用越大,如果某项指标的指标值全部相等,则该指标在综合评

价中不起作用。在信息论中信息熵

$$H(x) = -\sum_{i=1}^n p(x_i) \ln p(x_i)$$

表示系统的有序程度,一个系统的有序程度越高,则信息熵越大,其信息的效用值越小;反之,一个系统的无序程度越高,则信息熵越小,其信息的效用值就越大。所以,可以根据各项指标效用值的差异程度,利用信息熵这个工具,计算出各指标的权重,为多指标综合评价提供基础。

三、用熵值法进行综合评价的步骤

第一,将各指标数据标准化

原始指标 x_{ij} 可以分为正向指标和负向指标,对于正向指标,记 M_j 为其理想值,对于负向指标,记 m_j 为其理想值。理想值的获取可以通过原始数据,把极值作为理想值,即令 $M_j = \max(x_{ij})$, $m_j = \min(x_{ij})$, 定义 x_{ij}^* 为 x_{ij} 对于理想值的接近度,对于正向指标, $x_{ij}^* = x_{ij} / M_j$, 对于负向指标 $x_{ij}^* = m_j / x_{ij}$, 定义其标准化值 $y_{ij} = x_{ij}^* / \sum_{i=1}^m x_{ij}^*$ 。

第二,计算指标信息熵和信息效用值

$$\text{第 } j \text{ 项指标的信息熵为 } e_j = -k \sum_{i=1}^m y_{ij} \ln(y_{ij}),$$

其中, k 为常数,对于一个信息完全无序的系统,其熵值最大,此时 x_{ij} 对于给定的 j 全部相同,那么 $y_{ij} = 1/m$, 此时, e 取极大值。令 $k = 1/\ln m$, 则有 $0 \leq e_j \leq 1$ 。

某项指标的信息效用值: $d_j = 1 - e_j$ 。

第三,计算指标权重和综合评价值

某项指标的信息效用值越高,则对于评价的重要性就越大,则第 j 项指标的权重为: $w_j = d_j / \sum_{j=1}^n d_j$ 。第 i 个样本的综合评价值为: $f_i = \sum_{j=1}^n w_j x_{ij}^*$ 。

四、用功效系数法进行变换

取第 J 项指标值中最好值为 Xh , 最差值为 Xl , 用下列公式进行变换:

$$X_{ij} = \frac{X_{ij} - Xl}{Xh - Xl} \times \partial + (1 - \partial)$$

为避免变换后的数据出现零, ∂ 的范围应取 $(0, 1)$ 。

在用此公式进行变换时,实际上加入了评价者的主观因素,因为 ∂ 的选取是由评价者决定的。如果评价者要加大该指标的权重,可将 ∂ 取大一些,这时数据范围大,用熵值法计算的权重就大;同理,如果要减小该指标的权重,可将 ∂ 取小一些,这时数据范围小,用熵值法计算的权重就小。从这个意义上说,用功效系数法对数据变换后的熵值法不是严格的客观赋权法,而是一种主、客观结合赋权法。

用功效系数法变换后,对极端值作一定的处理,消除了指标值中负值的问题,然后按前面的步骤进行评价,但取不同的 ∂ , 可能会出现不同的评价结果

五、用标准化法进行变换

$$Z_{ij} = (X_{ij} - x_j) / S_j$$

其中 x_j 为第 J 项指标值的均值, S_j 为第 J 项指标值的标准差。然后用 Z 代替前面的步骤重的 X_{ij} 进行评价。

用标准化法进行变换与用功效系数法进行变换的区别是:

- 1) 用标准化法变换不需要加入任何主观信息,是一种完全意义的客观赋权法。
- 2) 用功效系数法变换 ∂ 的选取不同使得评价结果可能是不唯一的,而用标准化法进行变换评价结果是唯一的。
- 3) 标准化法有利于缩小极端值对综合评价的影响。

六、改进的熵值法和其它客观赋权法的比较

客观赋权法从实质上来说可以分为以下几类:

1) 消除指标间的相关性确定权数

综合评价是通过多项指标进行的,如果指标间具有一定的相关关系,说明它们

》转 187 页

于机会不公平而引起的分配关系不和谐的现象,政府应该而且必须有所作为。

第三,完善相关制度,保证规则公平。规则是一种行为规范,是人们一种共有的理念,是大家共同制定的要求大家共同遵守的规定。规则公平意味着不允许有不正当竞争获取私利。市场的自由竞争虽然为机会公平和规则公平提供了一个平台,但是要以一定的条件作为前提。这就是市场必须是完善的,法律法规必须是健全的。市场不完善、非均衡,交换就不可能是等价的。在利益的驱动下,交换双方就有可能产生机会主义行为,交换关系就是不和谐的。市场不完善、非均衡,生产资料所有者能够占有劳动剩余产品,贫富差距扩大,分配关系也是不和谐的。虽然市场可以不断完善,但是均衡却是相对的,当前我国贫富悬殊和分配不公根本上讲不是由于市场引起的,而是由于市场不健全、规则不公平,由于权力寻租、行业行政垄断以及偷税漏税等非市场因素所致,这使得有些人和组织可以在规则之外寻求利益的实现。所有这些因素为人民群众所深恶痛绝,也是导致社会不安定的重要因素。差距本身并不可怕,可怕的是造成分化的原因主要不是客观原因而是主观因素造成的。

第四,完善税收政策和社会保障制度,保证分配结果的公平。分配结果是最关系切身利益的,强调分配结果的公平不等于搞平均主义,而是指适当缩小贫富差距,兼顾全体公民的利益。这不仅不会伤害到效率的提高,相反还会进一步提高效率。即使从公平本身来看,由于造成分配结果不公的原因不仅是人的资源、能力、机遇和先天条件,而且竞争规则和制度的不完善也是造成差距过大的重要原因。因此通过财政、税收等公共政策的实施对市场竞争中所出现的市场失灵和政府失灵进行匡正。这种匡正本身同样体现了公平原则。

“社会主义的本质,是解放生产力,发展生产力,消灭剥削,消除两极分化,最终达到共同富裕。”“社会主义的优越性归根到底要体现在它的生产力比资本主义发展得更快一些、更高一些,并且在发展生产力的基础上不断改善人民的物质文化生活。”“社会主义的目的就是全国人民共同富裕,不是两极分化。如果我们的政策导致两极分化,我们就失败了;如果产生了什么新的资产阶级,那我们就真是走了邪路了。”邓小平的高瞻远瞩,需要我们不断温习和把握。在当前公平严重缺失的条件下,更加关注公平,通过制度公平促进效率的提升,通过制度公平的建设加快和谐社会的建设成为政府急需解决的问题。

[参考文献]

- 1、马克思.《马克思恩格斯全集》.[M]第23卷第59页.北京:人民出版社,1972年。
- 2、斯蒂格利茨等.《政府为什么干预经济——政府在市场经济中的角色》.[M].郑秉文,译.第69页.北京:中国物资出版社,1998.
- 3、[美]阿瑟·奥肯.《平等与效率》[A].梁小民,姚开建.《西方经济学必读手册》[C].北京:中国物资出版社,1999.
- 4、邓小平.《邓小平文选》(第三卷).[C].北京:人民出版社,1993,第373页。

- 5、邓小平.《邓小平文选》(第三卷).[C].北京:人民出版社,1993,第63页。
- 6、邓小平.《邓小平文选》(第三卷).[C].北京:人民出版社,1993,第110-111页。

[作者简介]

- 1、王俊奇(1968年3月生),男,副教授,硕士学位,主要从事马克思主义政治经济学和邓小平理论的研究。
- 2、康小莉(1966年5月生)女,教授,硕士学位,主要从事毛泽东思想和邓小平理论的研究。

》接188页

反映被评对象的信息有一定的重复 相关指标如不作变换处理而直接合成,合成的结果必定要包含重复的信息.这样就会歪曲对被评对象的评价结果。

于是,有一类客观赋权法的思路是:对原来相关的各原始变量作数学变换.使之成为相互独立的分量,这样就消除了指标间相关对被评对象的重复信息。这种变换一方面要消除相关性,另一方面对原变量进行简化,使少数几个综合因子能尽可能地反映原来变量的信息量 这类方法主要是多元分析中的主成分分析法和因子分析法

2)根据指标间的重复信息量确定权数

对于具有重复信息的指标来说,也可以按照另一种思路来确定权数 即某评价指标如果和其它评价指标重复信息越多,说明该指标的变化越能被其它指标的变化所解释,因而该指标在综合评价中所起的作用较小,所以应赋以其较小的权重;反之,若某项指标和其它指标的重复信息少,则该指标应赋以较大的权数。根据指标间的重复信息量来确定权数主要有复相关系数法。

3)根据指标的变异信息量确定权数

前两类赋权方法的共同之处是二者都是从指标间的重复信息量出发赋权的.或是消除重复信息后赋权,或是直接根据重复信息赋权。而熵值法是根据指标的变异信息量确定权数,这是该方法与前面两类客观赋权法的根本的区别。

按这种思路确定权数的方法还有变异系数法熵值法与因子分析法(包括主成分分析法)的区别还有:因子分析法消除了变量相关对综合评价的影响,熵值法不具

备消除这种影响的作用;因于分析法从众多的变量中分解出少数几个公共因子,减少了评价指标维数,而熵值法不能减少评价指标维数;因子分析法中因子与变量是线性函数关系,而熵值法的评价结果与变量不是线性函数关系;因子分析法计算过程比较复杂,一般必须用计算机完啦.熵值法计算过程相对简单

[参考文献]

- 1、郭显光.一种新的综合评价方法 组合评价法.统计研究,1995(5):56-59
- 2、情春花,蓝盛芳.环境价值的多角度评估,农业环境与发展,1999,Vo1.2 No.2
- 3、郝晓辉. Ecc0 模型:持续发展的全新定量分析方法,中国人口、资源与环境,1995,(1):43-46
- 4、张世秋.可持续发展环境指标体系的初步探讨,世界环境,1996(3):8-9
- 5、张成科.基于熵的水质模糊评价模型及应用[J].系统工程理论与实践,1998(6):80-85.
- 6、王靖,张金锁.综合评价中确定权重向量的几种方法比较[J].河北工业大学学报,2001,30(2):52-57

[作者简介]

- 1、苏洁(1981—),浙江理工大学经济管理学院管理科学与工程研究生。
- 2、沈文成(1983—),浙江理工大学经济管理学院管理科学与工程研究生。