

评价研究 ·

# 非线性科技评价方法合适性测度研究\*

## Determination of Appropriateness of Science and Technology Evaluation Indicators

俞立平 潘云涛 武夷山

(中国科学技术信息研究所 北京 100038)

**摘要** 为了对非线性科技评价方法的科学合理性进行测度,提出两种方法:一是将各指标作为解释变量,总指标值作为被解释变量,然后进行回归分析,利用拟合优度  $R^2$  值检验指标体系的合适性;二是将各指标作为投入,总指标值作为产出,利用数据包络方法进行分析,然后用效率的集中度来评价指标体系的合适性。在此基础上,比较了两种方法的特点和适用范围,认为两种方法结合运用效果较好,同时可以推广到其他指标体系合适性的评价。

**关键词** 科技评价 指标体系 合适性 回归分析 数据包络分析

中图分类号 G304

文献标识码 A

文章编号 1002-1965(2010)02-0044-03

### 0 引言

评价通常是指根据决策者、资助者以及其他利益群体的需求,确定价值标准,用科学的方法收集和处理相关信息,判断价值实现程度的过程。目前国内、外建立的综合评价方法有数十种之多。大致上分为两大类:第一类是线性评价方法,其基本原理是对评价指标进行主客观赋权,然后将数据标准化后加权汇总,评价结果与指标值之间的关系是线性的。如专家会议法、德尔菲法、层次分析法、熵权法、变异系数法等。第二类是非线性评价方法,主要指运用运筹学、模糊数学、系统工程等领域的方法进行评价,评价值与指标值之间的关系是非线性的。一些系统评价方法无须赋权,如主成分分析法、数据包络分析、TOPSIS 法等;另一些系统评价方法也可以赋权,采用主客观赋权方法均可,但一般以主观赋权为主,如加权 TOPSIS 法、ELECTRE 法、模糊综合评价法、PROMETHEE 等。学术期刊评价方法体系如图 1 所示。

在科技评价领域,从研究方法上看,张克英、黄瑞华等<sup>[1]</sup>应用灰色系统理论、客观赋权与聚类分析相结合进行科技创新评价。赵国杰、邢小强<sup>[2]</sup>利用网络层次分析法(ANP)评价区域科技实力。樊华<sup>[3]</sup>将数据包

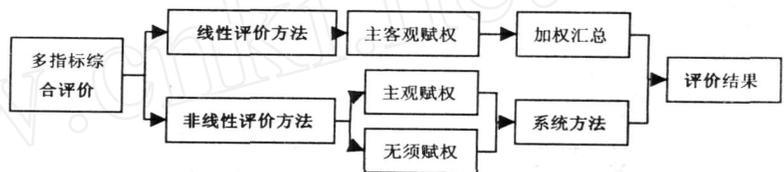


图1 科技评价方法体系

络分析 DEA 与主成分分析(PCA)结合起来,分析了省域高校的科技绩效。王孝宁、黄亚明等<sup>[4]</sup>利用 TOPSIS 法与神经网络法的结合,评价了卫生科技人员科研业绩。张渊、陆玉梅等<sup>[5]</sup>运用多层次灰色评价法,对科技计划项目进行了绩效评估。吴慈生、李兴国等<sup>[6]</sup>运用系统聚类分析的理论和方法对安徽省科技三项费用投入情况进行了分析。朱梅红、李爱华<sup>[7]</sup>运用基于熵值的赋权法对指标赋权,并采用线性加权模型进行加权综合评分,对各省份的综合科技实力以及各层次的科技实力进行比较和分析。吴晓梅、石林芬<sup>[8]</sup>应用因子分析模型对我国中心城市的科技竞争力进行了比较研究。朱干江、王桂芝<sup>[9]</sup>利用秩和比法对江苏各地市科技进步状况进行分析和综合评价研究。此外,还有若干学者用其他方法进行过类似研究。

总之,绝大多数赋权方法在科技评价实务或科技评价研究中都得到了不同程度的应用,并且出现了将多种评价方法结合运用的趋势。但是,除了部分非线

收稿日期:2009-05-04

修回日期:2009-05-27

基金项目:国家十一五支撑计划项目“基于海量信息的科技发展与科技评价的科学计量学研究”(编号:2006BAH03B05);国家自然科学基金资助“强竞争与弱竞争态势下的科技发展与科技评价的科学计量学研究”(编号:70673019)。

作者简介:俞立平(1967-)男,博士,博士后,教授,研究方向为信息经济、科学计量;潘云涛(1967-)女,研究员,研究方向为科技评价;武夷山(1958-)男,研究员,研究方向为情报学、科技管理、科学计量学。

性评价方法内含着对自身合适性的检验(如层次分析法的总排序一致性检验)以外,至今没有一种通用方法对所有非线性评价方法的合适性进行评价。于是,当人们对同一问题采用不同的赋权方法进行评价时,就无法对这些方法的合适性直接进行比较。不同学者采用不同赋权方法所得出的评价结论可能差异很大,结果指标体系的权威性难以得到公认。

考虑到回归分析法可以定量评价样本的总体拟合优度,而数据包络分析法也可以根据效率水平来定量分析样本点的分布,本文决定利用这两种方法对非线性科技评价方法的合适性进行测度,并比较二者的差别与适用范围,从而试图为非线性科技评价方法本身的评价提供一种行之有效的方法,同时也为进一步优化指标体系提供客观的评价依据。

## 1 采用回归分析法对科技评价指标体系的合适性进行评价

回归分析本来是研究一个变量关于另一个(些)变量的具体依赖关系的计算方法和理论。它关心的是根据解释变量的已知或给定值,考察被解释变量的总体均值,即当解释变量取某个确定值时,与之统计相关的被解释变量所有可能出现的对应值的平均值。

对科技评价指标体系而言,可以假设所有的单个指标为自变量,总指标值为被解释变量,建立如下模型:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 \dots b_n x_n \quad (1)$$

其中,  $x_1, x_2, x_3, \dots, x_n$  为指标值,  $b_1, b_2, b_3, \dots, b_n$  为权重。

不管什么赋权方法,其基本的前提是必须能“自圆其说”。尽管研究背景、思路、研究原则不同,但指标体系必须能说明总指标值是肯定的。为考察指标体系的合适性,可以根据公式(1)按统计指标数据残差平方和最小的原则来估计总体回归模型中的参数,即应用普通最小二乘法(OLS)进行估计,同时得出方程对统计样本的拟合优度即  $R^2$  值。

$R^2$  值就是回归平方和与总离差平方和的比值,  $R^2$  值越高,残差平方和越小,即样本点更加接近拟合曲线,于是  $R^2$  值可以用来作为判断指标体系合适性的指标。一套合适的科技评价指标体系由于指标选用科学合理,权重分配较为科学,其  $R^2$  值必定较高,而赋权不合理的指标体系,  $R^2$  值必然较低。一般认为,  $R^2 < 0.5$  属于低相关,  $0.5 < R^2 < 0.7$  属于中相关,  $R^2 > 0.7$  属于高相关。就科技评价指标体系而言,一般  $R^2$  值大于 0.7 就可以了。

## 2 采用数据包络分析(DEA)对科技评价指标体系的合适性进行评价

2.1 数据包络分析简介 DEA 是一种测算具有相同类型投入和产出的若干系统或部门(简称决策单元, DMU)相对效率的有效方法。其实质是根据一组关于输入输出的观察值,采用数学规划模型,来估计有效生产的前沿面,再将各 DMU 与此前沿面比较,进而衡量效率。凡是处在前沿面上的 DMU, DEA 认定其投入产出组合最有效率,将其效率指标定为 1;不在前沿面上的 DMU 则被认定为无效率,同时以效率前沿面之有效点为基准,给予一个相对的效率指标(大于 0, 小于 1)。此外, DEA 还可以判断各个 DMU 的投入规模的适合程度,给出各 DMU 调整其投入规模的方向和程度。

传统的统计方法是根据大量样本数据推断出样本集合整体的一般情况,其本质是平均性, DEA 则是根据样本数据来判断样本集合中哪些样本个体是属于相对有效的,其本质是最优性。DEA 在测定若干个决策单元的相对效率时注重的是对每一个决策单元进行优化,所得出的相对效率是其最大值,是最有利于该决策单元的相对效率。DEA 方法与其他多目标评价方法相比,具有独特的优势:在对 DMU 进行评价时,不必考虑指标的量纲,可以避免由于指标量纲不同而不得不寻求相同度量因素所带来的许多困难;不需要事先确定指标的相对权重,也不必确定决策单元的各输入输出之间的显式关系,这就排除了许多主观因素,不仅增强了评价结果的客观性,而且还会使问题得到简化。

2.2 数据包络分析在科技评价指标体系评价中的应用 在科技评价中,可以将所有指标的原始数据视为“投入”,将指标总分值视为“产出”,因为所有的单个指标都是总指标的输入数据,从而可以巧妙地借助 DEA 方法进行科技指标体系合适性的测度。需要说明的是,就多数指标体系(任一指标项都不为负)而言,输入指标值小,输出总指标值必然小,输入指标值大,输出指标值必然也大,这和效率分析中投入小、产出大的选优原则是不一样的。

如图 2 所示,为了简化起见,首先假设只有一种投入 X 和一种产出 Y 的情况,图中各点表示样本点。在一定的投入下,产出最大点总是位于最上方,将最上方的点连成一条曲线就是数据包络曲线,斜线是拟合线,表示投入与产出之间的函数关系,实际上就是输入指标与总指标值之间的关系。

根据 DEA 方法的原理,凡是以较小投入获得最大产出的点总是位于数据包络曲线上。假设投入为  $X_0$ ,

有 3 种产出结果,分别为 A、B、C 三点所示, A 点产出  $Y_3$  最大,位于最上方的数据包络曲线上; B 点产出为  $Y_2$ ,处于中等水平,位于拟合曲线上; C 点产出  $Y_1$  最低。从效率角度看, A 点效率最高, B 点效率一般, C 点效率最低。但是,从样本点的拟合情况看, B 点的拟合度最好,而 A 点、C 点的拟合程度较差。

由以上分析可以看出,单纯依靠平均效率无法确定样本点的拟合度,必须采用新的指标进行衡量,本文引入效率集中度指标的概念:

$$\text{效率集中度} = 1 - \text{效率标准差} / \text{效率平均值} \quad (2)$$

效率集中度越大,说明样本点越靠近拟合曲线,指标体系越合适。这一点与效率分析中提倡的效率越高则越优的原则是不一样的。如果从产出投入比的角度分析(这是 DEA 方法得以产生的前提和动力),当然是效率越高越好,但是从提高拟合优度的角度出发,过高效率点与过低效率点必然离拟合曲线较远,平均效率过高与过低都说明拟合度较差,只有在平均效率附近的样本点才拥有较好地拟合优度,这样的指标体系更为科学合理。

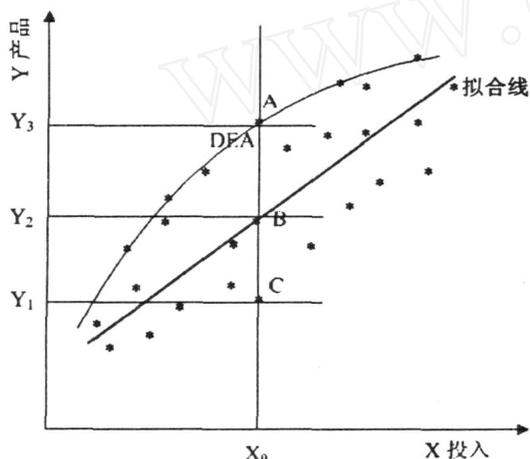


图 2 DEA 用于指标体系评价的原理

### 3 两种方法的比较

那么两种评价科技指标的方法哪种更好呢?任何方法的应用都是有一定的前提条件的,就回归分析法而言,有若干假设和基本前提,如解释变量互不相关,这一点在实际应用中很难做到,许多指标之间都是相关的,完全不相关几乎是做不到的。又如,同方差性的要求,即要求随机解释变量对所有样本点都具有相同方差,不允许有异方差的存在,这一点也很难做到。再如所有评价指标必须服从正态分布,在大样本情况下是可能的,但在小样本情况下,比如数据较少的情况下,这一点也是做不到的。此外,某些输入指标可能是某种排序,而非连续变量,对于这种情况,也不宜简单

地直接进行最小二乘估计。

在回归分析中,必然会碰到异方差、多重共线性问题,导致解释变量的统计检验值不显著、系数符号不正确等问题,但只要待评价的指标的选择基本科学合理,我们仍不妨采用拟合优度对指标体系进行评价。

效率分析对输入数据量纲没有要求,对数据的分布规律要求也不高,也不需要确定输入指标与总指标值之间的显式函数关系,但对异常点非常敏感,因此对数据的精度要求较高。

回归分析与数据包络分析各有其特点。实际工作中,在样本较少的情况下,只能采用数据包络分析进行评估,当样本较多时,可结合两种方法同时进行分析,以取长补短,在此基础上进一步优化科技评价指标体系。

需要说明的是,这两种方法均不适用于线性评价方法的比较,因为在这种情况下,所有样本点都在回归曲线上,拟合优度  $R^2$  为 1,当然所有的效率值也为 1。

### 4 结论

科技评价既可以进行主观评价,也可以进行客观评价。在指标体系已经确定的情况下,必须从评价方法系统外寻找新的方法对科技评价指标体系的合适性进行评估,在此基础上进一步优化各指标的权重,然后再进行评估,直到指标体系合适性指标达到一个较满意值为止,回归拟合优度和效率集中度正是这种思想的体现。

采用回归拟合优度和效率集中度除了可以对不同非线性科技评价方法比较外,还可以寻找到异常样本点。在大样本的情况下,同时采用两种方法进行分析可以找到两种方法共同的异常样本点,从而为指标权重修正提供了可行的路径。

本文介绍的方法具有一定的通用性,也可供评价其它指标体系时进行参考。在评价指标已确定的情况下,上述方法有助于进一步选择较合适的评价方法。

#### 参考文献

- [1] 张克英,黄瑞华等. 多种综合评价方法在科技创新评价指标分析中的应用[J]. 科技管理研究, 2005(12): 65 - 67
- [2] 赵国杰,邢小强. ANP 法评价区域科技实力的理论与实证分析[J]. 系统工程理论与实践, 2004(5): 41 - 44
- [3] 樊华. DEA/PCA 模型在科技绩效评价中的应用[J]. 淮海工学院学报(自然科学版), 2005(6): 81 - 84
- [4] 王孝宁,黄亚明等. TOPSIS 法与神经网络在卫生科技人员科研业绩评价中的应用[J]. 科技进步与对策, 2005(8): 34 - 36
- [5] 张渊,陆玉梅等. 多层次灰色评价法在科技计划项目绩效评

(下转第 65 页)

### 3 结束语

我国农业科技信息共享目前正处于建设项目深入推进阶段,主要的项目评价手段还只是国家对各项科技信息共享工程项目的年度验收审查,其中对 IQ 的评价要求还难以做到统一、规范、科学及量化评测。而各个信息共享建设单位内部的 IQ 评价也存在评价体系不规范、方法单一等问题,难以满足不断发展的用户需求和项目建设管理的需求,为此,本文尝试从信息质量评价框架设计、评价方法和评价指标选择三方面构建农业科技信息共享中信息质量评价体系。

但也看到,本文中观点是结合信息质量评价领域现有研究成果和农业科技信息的特点综合分析研究的基础上进行的阶段性研究探索,相信随着我国农业科技信息共享工作的推进,IQ 评价必将会在不断的理论探索、实践验证的过程中逐渐健全和成熟起来。本文作为农业科技信息共享领域的 IQ 评价的探索性认识,以期引入一种 IQ 评价方法思路,以对具体领域评价指标体系的建立提供参考,希望能为后续研究起参考和借鉴作用。

#### 参考文献

- [1] 孟宪学. 国家农业科学数据中心的设计与建设研究[J]. 农业图书情报学刊, 2004, 16(12): 5 - 8
- [2] 曹孟谊. 国外信息质量评估指标体系研究[J]. 军事运筹与系统工程, 2004, 18(4): 55 - 58
- [3] 曹瑞昌, 吴建明. 信息质量及其评价指标体系[J]. 情报探索, 2002, 84(4): 6 - 8
- [4] 鞠英杰. 网络信息质量评估[J]. 现代情报, 2005(3): 61 - 63
- [5] 刘雁书, 方平. 网络信息质量评价指标体系及可获取性研究[J]. 情报杂志, 2002(6): 10 - 12
- [6] 苏强, 梁冰. 信息质量及其评价指标[J]. 计算机系统应用, 2000(7): 63 - 65
- [7] 马小闯, 龚国伟. 信息质量评估研究[J]. 情报杂志, 2006(5): 19 - 21
- [8] 杨青云. 数据质量评估方法研究[J]. 计算机工程与应用, 2004(9): 3 - 4
- [9] 中国科学院. 《数据质量评价过程》、《数据质量研究报告》[R]. [2005 - 12 - 04], www.sdb.ac.cn
- [10] Eppler M J. Managing Information Quality [M]. Springer, 2006
- [11] Knight Shirlee - ann, Burn Janice. Developing a Framework for Assessing Information Quality on the World Wide Web[J]. Informing Science Journal, 2005(8)
- [12] M B Parker, V moleshe. An Evaluation of Information Quality Framework for the World Wide Web [C]. MIT International Conference on Information Quality, 2006
- [13] 金越. 网络信息资源的评价指标研究[J]. 情报科学, 2004(1): 64 - 66
- [14] 沈洁, 朱庆华. 国内外网络信息资源评价指标研究述评[J]. 情报科学, 2005, 23(7): 1104 - 1109
- [15] 索传军, 吴启琳. 国内外网络信息资源评价研究进展[J]. 现代图书情报技术, 2006(8): 55 - 59
- [16] 袁静. 网络信息资源评价指标研究的回顾及相关问题的思考[J]. 图书馆论坛, 2006, 26(5): 280 - 282
- [17] 张东华. 网络信息资源评价方法的研究[J]. 科技情报开发与经济, 2007, 17(1): 41
- [18] 张家武, 郑德俊. 网络信息本地化转换中的信息质量评价[J]. 图书情报知识, 2004(6): 77 - 79
- [19] 赵俊玲, 陈兰杰. 国外网络信息资源评价研究综述[J]. 图书馆工作与研究, 2004(3): 24 - 26
- [20] 宋敏, 覃正. 国外数据质量管理研究综述[J]. 情报杂志, 2007(2): 7 - 9
- [21] 宋立荣. 基于网络共享的农业科技信息质量管理研究[D]. 北京: 中国农业科学院研究生院, 2008
- [22] 胡耀华, 方佳. 农业科技管理教程[M]. 北京: 中国科学技术出版社, 2003
- [23] 霍佳震. 企业评价创新 - 集成化供应链绩效及其评价[M]. 石家庄: 河北人民出版社, 2001
- [24] 冯维扬. 竞争信息质量评价体系研究[J]. 图书情报工作, 2007, 51(2): 30 - 33

(责编:王平军)

(上接第 46 页)

- 估中的应用[J]. 科技进步与对策, 2006(6): 120 - 123
- [6] 吴慈生, 李兴国等. 基于聚类分析的科技三项费用投入评价与分析[J]. 合肥工业大学学报(自然科学版), 1999(6): 52 - 56
- [7] 朱梅红, 李爱华. 基于熵权的中国西部各省份科技实力综合评价[J]. 数学的实践与认识, 2006(12): 120 - 125
- [8] 吴晓梅, 石林芬. 基于因子分析的中心城市科技竞争力评价研究[J]. 科技管理研究, 2005(1): 46 - 50
- [9] 朱干江, 王桂芝等. 科技进步综合评价中秩和比法的运用[J]. 科技进步与对策, 2007(6): 151 - 154
- [10] Charnes A, Cooper W W, Rhodes E. Measuring the Efficiency of Decision Making Units[J]. European Journal of Operational Research, 1978(2): 429 - 444
- [11] Banker R D, Charnes A, Cooper W W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis[J]. Management Science, 1984, 30(9): 1078 - 1092

(责编:刘影梅)