

改进的熵值法及其在经济效益评价中的应用

郭显光

(安徽财贸学院经济信息管理系, 安徽 蚌埠 233041)

摘要 运用熵值法进行经济效益评价时常会遇到某些指标的极端值无法处理的困难, 本文通过对熵值法的改进, 有效地解决了这个问题。

关键词 熵值法 客观赋权 经济效益 评价

Application of Improved Entropy Method in Evaluation of Economic Result

Guo Xiangang

Dept. of Economic Information Management,
Anhui Institute of Finance and Trade, Bengbu 233041

Abstract It's hard to evaluate the economic result with entropy method when confronted with some special data. In this paper, the author improves entropy method and solves the problem effectively.

Keywords entropy method; objective weighing; economic result; evaluation

在多指标综合评价中, 确定指标权重的方法主要有主观赋权法和客观赋权法。主观赋权法是一类根据评价者主观上对各指标的重视程度来决定权重的方法; 客观赋权法所依据的赋权原始信息来源于客观环境, 它根据各指标的联系程度或各指标所提供的信息量来决定指标的权重。客观赋权法有熵值法、主成分分析法、因子分析法、复相关系数法等等, 本文针对熵值法在应用中遇到的一些问题对熵值法进行改进。

1 熵值法的原理

设有 m 个待评方案, 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 x_{ij} 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 $H(x) = - \sum_{i=1}^n p(x_i) \ln p(x_i)$ 是系统无序程度的度量, 信息是系统有序程度的度量, 二者绝对值相等, 符号相反。某项指标的指标值变异程度越大, 信息熵越小, 该指标提供的信息量越大, 该指标的权重也应越大; 反之, 某项指标的指标值变异程度越小, 信息熵越大, 该指标提供的信息量越小, 该指标的权重也越小。所以, 可以根据各项指标指标值的变异程度, 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依据。

用熵值法进行综合评价的步骤是:

1) 将各指标同度量化, 计算第 j 项指标下第 i 方案指标值的比重 p_{ij}

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (1)$$

本文于 1997 年 6 月 26 日收到

2) 计算第 j 项指标的熵值 e_j

$$e_j = -k \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (2)$$

其中 $k > 0$, \ln 为自然对数, $e_j \in [0, 1]$ 。如果 x_{ij} 对于给定的 j 全部相等, 那么

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} = \frac{1}{m}$$

此时 e_j 取极大值, 即

$$e_j = -k \sum_{i=1}^m \frac{1}{m} \ln \frac{1}{m} = k \ln m$$

若设 $k = \frac{1}{\ln m}$, 于是有 $0 \leq e_j \leq 1$

3) 计算第 j 项指标的差异性系数 g_j

对于给定的 j , x_{ij} 的差异性越小, 则 e_j 越大; 当 x_{ij} 全部相等时, $e_j = e_{\max} = 1$, 此时对于方案的比较, 指标 x_j 毫无作用; 当各方案的指标值相差越大时, e_j 越小, 该项指标对于方案比较所起的作用越大。定义差异性系数

$$g_j = 1 - e_j \quad (3)$$

则当 g_j 越大时, 指标越重要。

4) 定义权重

$$a_j = \frac{g_j}{\sum_{j=1}^n g_j} \quad (4)$$

5) 计算综合经济效益系数 v_i

$$v_i = \sum_{j=1}^n a_j p_{ij} \quad (5)$$

v_i 为第 i 个方案的综合评价值。

2 改进的熵值法

在用熵值进行企业经济效益评价时, 时常会遇到一些极端值, 例如 1996 年各地区工业效益指标中, 选取了产品销售率、资金利税率、成本费用利润率、劳动生产率、流动资金周转次数、增加值率等六项指标, 资金利润率指标中有 27 个地区在 2.42%~10.75% 之间, 但云南为 28.30%, 青海为 -0.88%, 这是两个极端值(资料来源:《中国统计摘要 1997》)。指标值为负时不能直接计算比重, 也不能取对数, 而为保证数据的完整性这两个值又不能删去, 因此需要对该项指标数据进行变换。

2.1 用功效系数法进行变换

取第 j 项指标值中最好值为 $x_j^{(h)}$, 最差值为 $x_j^{(s)}$, 用下列公式进行变换:

$$x_{ij} = \frac{x_{ij} - x_j^{(s)}}{x_j^{(h)} - x_j^{(s)}} \times \alpha + (1 - \alpha) \quad (6)$$

为避免变换后的数据出现零, α 的范围应取 $(0, 1)$ 。

在用公式(1)进行变换时, 实际上加入了评价者的主观因素, 因为 α 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 α 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 α 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, 然后可按公式(1)~(5)的步骤进行评价, 但取不同的 α 可能会出现不同的评价结果。

笔者取 $\alpha = 0.6$, 对资金利税率和成本费用利润率两项指标进行了变换。综合评价的结果见表 1 中熵值

法 2 的数据。

2.2 用标准化法进行变换

$$x_{ij} = (x_{ij} - \bar{x}_j) / s_j \quad (7)$$

其中 \bar{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。

一般地, x_{ij} 的范围在 -5 到 5 之间, 为消除负值, 可将坐标平移, 令

$$Z_{ij} = 5 + x_{ij} \quad (8)$$

然后用 Z_{ij} 代替 x_{ij} 按公式 (1)~(5) 的步骤进行评价, 评价结果见表 1 中熵值法 1 的数据。

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) 用功效系数法变换因 α 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。

3 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; 因子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, 而熵值法的综合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成, 而熵值法计算过程相对简单。

为了比较两种改进的熵值法与其它客观赋权法评价结果的差异, 笔者另外用因子分析法和主成分分析法对表 1 的数据进行了综合评价, 评价结果见表 1。

表 1 几种客观赋权法的评价结果

地区	因子分析	名次	主成分	名次	熵值法 1	名次	熵值法 2	名次
北京	0.2525	8	0.84744	6	0.03694	6	0.04003	3
天津	0.2370	10	0.47300	10	0.03565	13	0.03511	13
河北	0.1832	11	1.16535	5	0.03787	4	0.03524	11
山西	- 0.3620	19	0.84503	7	0.03625	10	0.03154	20
内蒙古	- 0.4814	24	- 0.29006	16	0.03326	17	0.03039	23
辽宁	- 0.4051	22	- 0.82016	20	0.03212	22	0.02956	25
吉林	- 0.6130	26	- 1.49248	26	0.03042	25	0.02911	26
黑龙江	0.2778	9	1.73513	3	0.03768	5	0.03690	9
上海	1.1812	2	2.26665	2	0.04017	2	0.04672	2
江苏	0.6029	4	0.26347	12	0.03666	9	0.03952	5
浙江	0.5274	5	0.27530	11	0.03590	11	0.03807	7
安徽	0.0965	13	0.10204	15	0.03507	15	0.03414	15
福建	0.3189	7	0.64995	9	0.03686	7	0.03710	8
江西	- 0.6759	27	- 0.88165	21	0.03251	19	0.02698	28
山东	0.6309	3	1.45545	4	0.03895	3	0.03958	4
河南	0.0786	14	0.7388	8	0.03671	8	0.03434	14
湖北	0.1485	12	0.18020	14	0.03545	14	0.03573	10
湖南	- 0.3762	21	- 0.80217	19	0.03307	18	0.02977	24
广东	0.4899	6	0.24668	13	0.03582	12	0.03932	6
广西	- 0.0944	15	- 0.57668	17	0.03373	16	0.03513	12
海南	- 0.1288	16	- 2.49866	28	0.02801	28	0.03203	18
四川	- 0.3467	18	- 0.89332	22	0.03233	21	0.03073	21
贵州	- 0.4070	23	- 1.20513	24	0.03022	26	0.03174	19
云南	2.2380	1	7.73393	1	0.05085	1	0.05363	1
陕西	- 0.7730	28	- 1.70838	27	0.02988	27	0.02742	27
甘肃	- 0.3358	17	- 0.73788	18	0.03235	20	0.03217	17
青海	- 1.3345	29	- 4.67674	29	0.02252	29	0.02456	29
宁夏	- 0.3630	20	- 1.02794	23	0.03126	24	0.03043	22
新疆	- 0.5665	25	- 1.36722	25	0.03150	23	0.03301	16
西藏								

从表 1 中可以看出,因选用不同的评价方法评价结果也出现了一定的差异。对于不同的评价结果如何确定一种合理的评价结果,一般有两种方法:一种方法是对各种不同的结果进行组合,笔者已专门作过介绍(见文献[1]);另一种方法是从多种评价结果中选取一种较合理的评价结果。

为评价某种评价结果的合理性,可以用一种非参数统计方法——等级相关系数法来度量该方法与其它方法评价结果的密切程度。

用斯皮尔曼等级相关系数公式



$$r = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n-1)} \quad (9)$$

来计算不同方法排序结果的等级相关系数。

如果某种评价方法和其它方法评价结果的等级相关系数较大, 则该种方法的评价结果较为合理。

根据表 1 中的数据, 笔者计算了熵值法 1 和因子分析法、熵值法 1 和主成分分析法、熵值法 2 和因子分析法、熵值法 2 和主成分分析法的等级相关系数, 分别用 r_{31} , r_{32} , r_{41} , r_{42} 表示。

$$\begin{aligned} r_{31} &= 0.9837 & r_{32} &= 0.8429 \\ r_{41} &= 0.9505 & r_{42} &= 0.8120 \end{aligned}$$

由等级相关系数可见, 用标准化法对数据处理后的熵值法更具有合理性。

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

(上接第 97 页)

此外, 需要说明的是本文的结论并不是试图否认理性假设的重要性, 更谈不上对现有理论体系的否定或挑战, 而是为了强调实际应用中问题界定的重要性。

参 考 文 献

- 1 Anand P. The Philosophy of Intransitive Preference. The Economic Journal, 1993, 103: 337~ 346
- 2 Anand P. Are the Preferences Really Rational. Theory and Decision, 1987, 23: 189~ 214
- 3 Sen A. K., Social Choice Theory, in Kenneth J. Arrow and Michael Intriligator, eds, Handbook of Mathematical Economics, Vol III, Amsterdam: North-Holland, 1986, 1078~ 1181
- 4 Suzumura K. Rational Choice, Collective Decisions and Social Welfare. Cambridge: Cambridge University Press, 1983
- 5 Sugden R. Why be Consistent? A Critical Analysis of Consistency Requirements in Choice Theory. Economica, 1985, 52: 167~ 183
- 6 罗云峰. 社会选择的若干理论问题研究. 华中理工大学博士后研究报告, 1997