如何正确运用因子分析法进行综合评价

●游家兴

(厦门大学 计划统计系,福建 厦门 361005)

摘要:将因子分析方法运用于综合评价方法,克服了传统评价方法在处理指标高度相关和权重设定上的缺陷,但所构造的因子得分模型仅适用于对评价对象的静态比较,并不适用于动态比较,本文探讨了运用因子分析法进行综合评价应注意的一些问题。

关键词:综合评价;因子分析法

中图分类号:F224 文献标识码:B 文章编号:1005-5762(2003)05-0010-02

在多指标综合评价方法中,传统方法对于权重的设置往往带有一定的主观随意性,将多元统计引入综合评价方法,如因子分析法,可以克服人为确定权数的缺陷,使得综合评价结果唯一,而且客观合理。许多学者将因子分析应用到对上市公司并购前后经营业绩的综合评价上,但在因子分析方法的运用上存在着一些问题,削弱了实证分析研究的解释力和信服力。本文试从如何正确运用因子分析法进行综合评价作一些探讨。

一、因子分析方法的基本思想和运用

因子分析法是把一些具有错综复杂关系的变量归结为少数几个无关的新的综合因子的一种多变量统计分析方法。其基本思想是根据相关性大小对变量进行分组,使得同组内的变量之间相关性较高,不同组的变量相关性较低。每组变量代表一个基本结构,因子分析中将之称为公共因子。

假设观测系统(即评价总体)有 k 个评价指标,n 个观测单位,因子分析的数学模型就是把 n 个观测单位分别表示为 p <k 个公共因子和一个独特因子的线性加权和,即:

$$\chi_i = \alpha_{i1}F_1 + \alpha_{i2}F_2 + \ldots \alpha_{ij}F_j \ldots + \alpha_{ip}F_p + \varepsilon_i \quad (i = 1, 2, \ldots, n) \quad (1)$$

其中:$F_1, F_2, \ldots, F_p$ 为公共因子,它是各个指标中共同出现的因子,因子之间通常是彼此独立的;ε_i 是各对应变量 χ_i 所特有的因子,称为特殊因子,通常假定 $\varepsilon_i \sim N(0, \sigma_i^2)$;系数 α_{ij} 是第 i 个变量在第 j 个公共因子上的系数,称为因子负荷量,它揭示了第 i 个变量在第 j 个公共因子上的相对重要性。

因此,通过因子模型建立综合评价函数的步骤如下:

(1)根据原始变量矩阵估计因子载荷矩阵。因子载荷阵的估计方法有很多,主成分法是其中最为普遍的方法:

设原始变量 $X=(X_1, X_2, \ldots, X_k)'$ 的协方差阵为 Σ,$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k > 0$ 为 Σ 的特征根,λ_i 代表第 i 个主成分的方差,且总方差 $\sum_{i=1}^{k}\sigma_{ii} = \sum_{i=1}^{k}\lambda_i$,$e_1, e_2, \ldots, e_k$ 为对应的标准正交化特征向量,根据线性代数知识 Σ 可分解为:

$$\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \ldots + \lambda_k e_k e_k'$$

$$= (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \cdots \sqrt{\lambda_k}e_k)\begin{bmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \ldots \\ \sqrt{\lambda_k}e_k' \end{bmatrix}$$

上式的分解是公共因子与变量个数一样多的因子模型的协方差阵结构。采用因子分析方法总是希望公共因子的个数小于变量的个数即 m <k。当最后 k-m 个特征根较小时,通常略去最后 k-m 项 $\sum = \lambda_{m+1}e_{m+1}e_{m+1}' + \cdots + \lambda_k e_k e_k'$ 对 Σ 的贡献,从而得到:

$$\Sigma \approx (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \ldots \sqrt{\lambda_j}e_j \ldots, \sqrt{\lambda_m}e_m)\begin{bmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \ldots \\ \sqrt{\lambda_m}e_m' \end{bmatrix}$$

其中:$\sqrt{\lambda_j}e_j$ 为第 j 个公共因子的因子载荷。

(2)将公共因子表示为变量的线性组合,得到评价对象在各个公共因子的得分。由于因子得分函数中方程的个数 m 小于变量个数 p,因此不能精确计算出因子得分,通过最小二乘法或极大似然法可以对因子得分进行估计:

收稿日期:2003-01-05

作者简介:游家兴(1978-)男,厦门大学计划统计系研究生

$$\hat{F}_{ij} = \beta_{i0} + \beta_{i1}\chi_1 + \beta_{i2}\chi_2 + ...\beta_{ij}\chi_j ... + \beta_{ik}\chi_k \quad (2)$$

(3) 以各公共因子的方差贡献率占公共因子总方差贡献率的比重作为权重进行加权汇总，建立因子综合得分函数：

$$Y_j = \gamma_1\hat{F}_{1j} + \gamma_2\hat{F}_{2j} + ... + \gamma_i\hat{F}_{ij} + ... + \gamma_p\hat{F}_{pj} \quad (i = 1, 2, ... p) \quad (3)$$

其中：Y_j 是第 j 个评价对象的综合得分，$\hat{F}_{ij}$ 为第 j 个评价对象在第 i 个公共因子的得分，γ_i 为第 i 个公共因子方差贡献率占公共因子总方差贡献率的比重，即 $\gamma_i = \lambda_i / \sum_{l=1}^{m} \lambda_l$。

二、应用因子分析法进行综合评价应注意的问题

1. 原始指标是否需要转换处理

若原始指标的量纲或经济意义不同，将原始指标直接求得综合得分，将很难给予一个合理的经济解释; 若原始指标变量数量级差异较大，则变量值大的对综合指标 (公共因子) 的影响也大。例如, 同样是反映生产能力的产值指标，采用以元为单位和采用以万元为单位，其方差显然是完全不同的。经济意义不变，但以元为单位的产值指标不仅会增加评价指标体系中变量的总方差，也会增加该指标在总方差中的比重，从而增大它在评价指标体系中的作用。因此，在运用因子分析法时，通常需要对原始指标进行无量纲化处理。

对原始指标进行无量纲化处理的方法有很多种，如标准化、均值化或极差正规化。由于标准化处理会保持原始指标数值的相对稳定性，在进行因子运算时会带来许多便捷，如原始指标经过标准化后协方差阵 $\sum$ 就是相关系数矩阵，并且在因子得分和综合得分上意义明确、容易理解，因此是最普遍的做法。标准化的公式为：

$$\chi_{ij}' = \frac{\chi_{ij} - \bar{\chi}_i}{\sqrt{var(\chi_j)}} \quad (i = 1, 2, ..., n; j = 1, 2, ... k)$$

其中 $\bar{\chi}_j$ 和 $\sqrt{var(\chi_j)}$ 分别表示第 j 个变量的平均值和标准差，则 $E(\chi_j') = 0$，$var(\chi_j') = 1$。对于含有 k 个指标的评价体系中，总方差 $\sum_{i=1}^{k} \sigma_{ii} = \sum_{i=1}^{k} \lambda_i = k$，数学上可以证明，经过标准化处理后，每个公共因子得分均值为 0，方差为 1; 综合得分的均值为 0，方差为 $\sum \alpha_i^2$。

2. 什么评价指标适合运用因子分析方法

因子分析方法在多元统计中属于降维思想中的一种，其目的在于简化数据，通过较少的公共因子反映复杂现象的基本结构。原始评价指标少，意义明确，能较好地反映评价对象，这时，不一定要使用因子分析。如果强行运用，不仅会加大计算量，而且意义不大。

此处，使用因子分析法进行综合评价目的之一是为了避免评价指标之间的相关性所引起权重的偏倚，因此其中一个前提条件是评价指标之间应该有较强的相关关系。如果指标之间的相关程度很小，指标不可能共享公共因子，公共因子对于指标的综合能力就偏低。一般来说，可以通过对指标的相关矩阵进行检验，如果相关矩阵的大部分系数都小于 0.3，则不适合做因子分析。

3. 因子模型应选取几个因子进行分析

因子分析的目的是寻求用少数的几个公共因子解释协方差结构的因子模型。选取的因子过多，应用因子分析方法就失去原有的意义; 但选取的因子过少，又可能造成原始信息量的大量损失。通常有以下三种准则：

(1) 以主成分的特征值为标准选取公共因子。原始评价指标标准化后，由于每个指标的方差为 1，假如主成分所对应的特征值小于 1，意味着该主成分连一个指标的方差都无法解释，所以应选取特征值大于或接近于 1 的主成分作为公共因子，舍弃特征值远小于 1 的其它主成分。

(2) 以主成分的方差累计贡献率为标准来选取公共因子。方差累积贡献率反映了主成分保留原始信息量的多少。一般而言，主成分累积贡献率达到 85% 以上就可以很好地说明和解释问题，因此可以以此为标准选取累积贡献率达到 85% 以上的那些主成分作为公共因子。

(3) 根据分析问题的需要或具体问题的专业理论来选取公共因子。在多维数据中，当维数大于 3 时便不能画出几何图形，但通过因子分析法选取主要的两个公共因子，画出正交因子得分图，以反映评价对象在二维平面上的分布情况，从而直观地找出各评价对象在公共因子中的地位，进而还可以对评价对象进行分类处理。

4. 初始公共因子是否需要旋转

建立因子分析模型的目的不仅是要找出主因子，更重要的是要知道每个主因子的意义，以便对实际问题进行分析。通过式 (1)、式 (2)，只是确立初始公共因子, 这些初始因子是否具有明确意义，需要进一步分析因子载荷阵才能得出。如果从每个初始因子能较好地找出所代表的原始指标，我们就可以直接赋予这些因子合理的经济解释，进行下一步的分析研究。但如果因子载荷量较为平均，难以判别哪些指标与哪个因子联系较为密切，无法从原始指标中寻求评价对象在各个因子上得分差异的原因，这时就需要进行因子旋转。

因子旋转的直观意义是经过旋转后，公共因子的贡献越分散越好，使指标仅在一个公共因子上有较大的载荷，而在其余公共因子上的载荷比较小。因子旋转的方法很多，如正交旋转、斜交旋转等，正交旋转又包括方差最大化旋转、四次方最大化旋转等，但基本思路就是在寻求极值的前提下，用一个正交阵 (对正交旋转) 或非正交阵 (对斜交旋转) 右乘因子载荷阵，达到简化因子载荷阵结构的目的。由于因子旋转涉及到十分复杂的矩阵运算，一般统计软件如 SPSS、SAS 都可以按照研究方法和研究目的的需要，直接得出, 本文不做详细推导，感兴趣的读者可参阅多元统计知识的相关介绍。

参考文献:

[1] 罗积玉, 邢英. 经济统计分析方法及预测[M]. 北京: 清华大学出版社, 1987.