

主成分分析在中国上市公司综合评价中的作用

田波平^{1, 2}, 王 勇¹, 郭文明¹, 葛喜鹏¹

(1. 哈尔滨工业大学数学系, 黑龙江 哈尔滨 150001)

(2. 哈尔滨工业大学管理学院, 黑龙江 哈尔滨 150001)

摘要: 主要介绍了主成分分析在上市公司综合评价中的作用。主成分分析作为一种客观赋权的方法, 权数是能随着宏观经济环境变化而变化的动态权数体系, 因为主成分分析所应用的数据来源于上市公司年度报告所提供的财务指标。它主要对所选择的 40 只绩优股, 进行横向的比较, 并根据第一主成分得分进行排序, 给广大的投资者提供参考。

关键词: 主成分分析; 第一主成分; 财务指标; 综合评价

1 上市公司综合评价方法的来源

随着中国证券市场的不断壮大, 上市公司的数量也不断增加, 证券在社会经济生活中的地位也越来越重要。投资者在进入股市后将面对上千只不同行业, 不同背景的股票。众多的投资者除了进行政策面分析外, 还希望对这些股票进行理性客观的评价, 特别是长线投资者更希望选择那些业绩优良的股票。证券投资是个渐进的过程, 它包括证券投资的分析、决策、操作和管理一系列阶段, 其中证券投资分析是基础性的核心环节。它可以使投资者提高投资收益, 降低风险损失。除了进行宏观经济分析和行业分析外, 公司分析则是证券投资基本分析的最关键的一环。证券投资者能否获得与其所承担的风险相应的收益, 主要取决于上市公司的经营状况和财务状况。要避免投资的盲目性就需要投资者对发行证券的公司了解透彻, 分析全面。我们利用多元统计方法对上市公司的财务状况进行分析会对投资者选择理想的投资对象起到很好的参考作用。

1.1 综合评价方法研究所依托的客体 and 目的

1) 公司综合评价所依托的客体

上市公司综合评价方法研究所依托的客体是沪深两市所有上市公司定期公布的 2002 年年报数据, 这些数据既有总量指标又有比率指标, 既有正指标, 也有逆指标, 又有适度指标, 它们基本上反映了上市公司的经营状况, 尤其是上市公司的盈利能力和发展前景。选择这一客体还有一个原因, 就是这一数据资料取得比较容易, 而且指标口径变动不大, 连贯性强, 用这些指标既可以进行横向分析, 又可以进行纵向分析。

2) 上市公司综合评价方法研究的目的

评价的一个重要功能就是其导向作用。本文旨在设计一种客观赋权的多指标综合评价方法——主成分得分法。该评价方法的权数是能随着宏观经济环境变化而变化的动态权数体系, 主成分得分法是利用上市公司年度报告所提供的财务指标, 因为本方案所运用的权数

体系来源于特定年份的实际数据, 所以它是能与宏观环境相匹配的动态权数体系。本方案还可以根据研究目的的需要, 对上市公司中期或年度报告的财务指标进行组合, 就上市公司做出指标组合中所选定的多指标进行综合评价排名, 进而对行业、区域经济进行横向和纵向的比较。

1.2 国内外有关公司评价方法的现状^[5]

当前国内外对上市公司进行的分析评价多数采用单项指标排名评价的方法。这样做往往会在不同单位指标的评价之间发生相互矛盾的情况, 从而影响评价的整体性, 所以, 如何合理科学公正的评价和分析上市公司经营业绩已经成为当前证券投资理论和实践的一个主要问题。

1) 国外有关公司评价方法的介绍

以美国为代表的西方国家, 很早就展开了对上市公司的评价, 并且在世界上具有较大的影响。以下分别简略介绍目前国外在上市公司评价方面较权威的美国《财富》和《商业周刊》杂志对上市公司的评价方法。

《财富》杂志从 1955 年开始, 在每年的 7 月至 8 月之间, 以上一年度年营业额最大的 500 家上市公司, 即“世界 500 强”; 它另外还设置一些辅助指标, 如利润总额、利润增长额、主营业务利润率、雇员人数、净资产回报率等等, 并分别按照其大小顺序进行多项平行的单指标排名, 列出各项指标最大的前 500 家上市公司。

《商业周刊》是美国一家以财经报道为主的, 在这方面具有广泛影响力的综合性的国际出版物。该周刊十多年来一直致力于对上市公司的评价和排名, 它既有单指标排名, 又有多指标综合评价排名。单指标排名根据美元计价的上市公司市价总值大小排出全球市价总值最大的 1000 家上市公司。而综合评价排名是对纳入标准普尔 500 指数中的 500 家上市公司, 根据 8 个指标的综合结果加以排名, 评出前 50 名为“《商业周刊》最佳 50 家上市公司。”其综合评价方法选取的 8 个指标, 分别是: 当年的总收益、三年来的总收益、一年来的销售总额增长率(%)、三年来的销售总额增长率(%)、一年来的利润增长率(%)、三年来的利润增长率(%); 一年来的净利润、一年来的净资产收益率(%). 综合的方法是对这 8 项指标分别进行排名, 取得分值最大的 20% 为 A 级, 其次的 20% 为 B 级, 依次类推, 最后的 20% 为 E 级。各上市公司的总得分是其 8 个指标的评分之和, 然后按总得分排名。

2) 国内有关公司评价方法的介绍

目前, 在国内上市公司业绩评价中享有较高声誉的评估机构是中国诚信证券评估有限公司。自 1996 年来, 该公司与《中国证券报》合作, 每年对上市公司的经营业绩进行评价。其评价方法是: 根据各上市公司公开披露的有关资料, 通过对上市公司的: 净资产收益率; 资产总额增长率; 利润总额增长率; 负债比率; 流动比率; 全部资本化比率共六个指标分别进行考核, 然后按各指标的重要性不同确定各指标在综合评价时的权重 (见表 1)。在这种评价方法下, 各公司的最后得分是在各单项指标考核评分的基础上, 乘以每项指标的权数, 然后相加得到总评。

表 1 中国诚信公司业绩评价财务指标体系及其权重 单位(%)

指标	X1	X2	X3	X4	X5	X6
权重	55	9	13	7	7	9

此外,我国大陆有关上市公司的评价方法主要还有在国内证券界相当有影响力的《上海证券报》和《上市公司》杂志的综合评价方案。《上海证券报》的评价方案是根据上市公司当年的盈利水平、发展速度和财务状况三方面的 6 个指标,再按各指标的重要性不同来确定其权重,各项指标的得分与权重之积即为上市公司的总得分。《上市公司》杂志从 1999 年开始进行了“沪市 50 强上市公司”评选活动,其评价方案采用了四个指标:主营业务收入;净利润;总资产;市值。评价的过程首先是计算沪市全部上市公司各单项得分,即分别按各指标大小顺序进行单项排名,前 50 名的单项名次即为这些上市公司该项指标的得分,例如单项排名第 10 名的单项得分即为 10 分,第 51 名以后的上市公司得分均为 51 分;其次是计算综合得分,4 项指标的单项得分加总即为综合得分;然后按照综合得分从小到大的顺序进行排名,对于综合得分相同的上市公司,再按主营业务收入的大小顺序排名;最后剔除净利润为负值的上市公司后,取前 50 名为“沪市 50 强上市公司”。

2 主成分分析方法^[1, 2, 4, 6]

主成分分析是多元统计分析中的一种常用方法。主成分概念首先由 Karl Parson 在 1901 年引进,不过当时只是对非随机变量来讨论。1933 年 Hotelling 将这个概念推广到随机向量。主成分分析是数学上处理降维的一种方法,它的基本思想就是设法将原来众多具有一定相关性的指标(比如 p 个指标),重新组合成一组新的相互无关的综合指标来代替原来指标,同时根据需要从中选取几个较少的综合指标尽可能多地反映原来指标的信息。

2.1 主成分的数学模型

设有 n 个样品,每个样品观测 p 项指标(变量) X_1, X_2, \dots, X_p , 得到原始数据资料阵:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}$$

其中

$$x_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T, \quad i = 1, 2, \dots, p$$

用数据矩阵 X 的 p 个向量(即 p 个指标向量) x_1, \dots, x_p , 作线性组合(即综合指标向量)为:

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{pi}x_p, \quad i = 1, 2, \dots, p$$

上述方程要求:

$$a_{1i}^2 + a_{2i}^2 + \dots + a_{pi}^2 = 1, \quad i = 1, 2, \dots, p$$

且系数 a_{ij} 由下列原则决定:

F_i 与 F_j ($i \neq j, i, j = 1, \dots, p$) 不相关

F_1 是 x_1, x_2, \dots, x_p 的一切线性组合(系数满足上述方程)中方差最大的, F_2 是与 F_1 不相关的 x_1, x_2, \dots, x_p 一切线性组合中方差最大的,依次类推, F_p 是与 F_1, F_2, \dots, F_{p-1} 都不相关的 x_1, x_2, \dots, x_p 的一切线性组合中方差最大的。

2.2 主成分分析的计算步骤

设有 n 个样本单位, p 项指标, 可得原始数据阵 X :

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \quad (1)$$

将原始数据阵标准化

设 x_{ij} 表示第 i 个样本单位的第 j 项指标值, 对数据进行标准化变化

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{2}$$

式中:

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n - 1}$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

其中, $i = 1, 2, \dots, n; j = 1, 2, \dots, p$

为了书写方便, 不妨设上边矩阵已标准化了.

求指标数据的相关系数阵

$$R = (r_{ij})_{p \times p}, \quad i = 1, 2, \dots, p; j = 1, 2, \dots, p \tag{3}$$

式中, r_{ij} 为指标 i 与指标 j 的相关系数

$$r_{ij} = \frac{1}{n - 1} \sum_{i=1}^n [(x_{ij} - \bar{x}_i)/s_i] [(x_{ij} - \bar{x}_j)/s_j]$$

其中, $i = 1, 2, \dots, p; j = 1, 2, \dots, p$

求 R 的特征根 $\lambda_1, \lambda_2, \dots, \lambda_p$ 及相应的单位特征向量, 并且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 其中

$$\hat{a} = (a_{11}, a_{21}, \dots, a_{p1})^T$$

$$\hat{a} = (a_{12}, a_{22}, \dots, a_{p2})^T$$

$$\dots\dots\dots$$

$$\hat{a} = (a_{1p}, a_{2p}, \dots, a_{pp})^T$$

写出主成分, 并由标准化后的原始数据阵求主成分得分

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{pi}X_p \tag{4}$$

式中, $i = 1, 2, \dots, p$

2.3 贡献率及累计贡献率

在解决实际问题时, 一般不是取 p 个主成分, 如果前 k 个主成分的贡献率已达到 85%, 表明取前 k 个主成分基本包含了全部测量指标所具有的信息, 这样既减少了变量的个数, 又便于对实际问题的分析与研究. 由于

$$\text{Var}(F_i) = \frac{\lambda_i}{p} \sum_{i=1}^p \lambda_i \tag{5}$$

及第一主成分贡献率为 $\frac{\lambda_1}{\sum_{i=1}^p \lambda_i}$, 所以

$$\frac{\lambda_1}{\sum_{i=1}^p \lambda_i} = \frac{\text{Var}(F_1)}{\sum_{i=1}^p \text{Var}(F_i)} \tag{6}$$

因此第一主成分的贡献率就是第一主成分的方差在全部方差 $\sum_{i=1}^p \lambda_i$ 中的比值. 这个值越大,

表明第一主成分综合 x_1, x_2, \dots, x_p 信息的能力越强 前两个主成分的累计贡献率定义为

$$\frac{(\lambda_1 + \lambda_2)}{p}, \text{ 前 } k \text{ 个主成分的累计贡献率定义为}$$

$$\sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i \quad (7)$$

3 主成分分析方法在证券个股评价中的实证研究^[3, 4, 7]

3.1 样本股票的选择

对于样本股票的选择,我们主要是通过金融街(www. jrj com. cn)和中信证券网上股票交易系统网站上市公司统计年报检索系统,从沪深两市中挑选了具有各自代表性的 40 只股票 这里所谓的代表性是指: 行业代表性; 板块代表性; 业绩代表性 这 40 只股票分别代表了“信息产业”、“旅游行业”、“医药行业”、“金融、保险行业”、“采掘业”、“交通运输行业”和“制造行业” 其中 x_1 表示主营业务收入; x_2 表示主营业务利润; x_3 表示利润总额; x_4 表示净利润; x_5 表示总资产; x_6 表示固定资产; x_7 表示净资产收益率; x_8 表示每股净资产; x_9 表示每股资本公积金; x_{10} 表示每股收益 这 10 个指标中既有反映企业投入产出规模的总资产、主营业务收入、净利润指标,也有反映企业投资者投入效率的每股收益、净资产收益率指标,所选样本股票的原始数据见金融街(www. jrj com. cn)和中信证券网上股票交易系统网站

3.2 数据分析与处理

根据上述的原始数据和矩阵标准化公式((1), (2)),用 MATLAB 程序将矩阵标准化 然后,利用多元统计中的主成分分析法,使用 SAS 统计计算软件包,对所选的 40 只股票在 2002 年年报中所参考的财务指标进行分析,可以得到样本相关系数阵((3)),相关系数阵的特征值、特征向量以及各个主成分的贡献率和 k 个主成分的累计贡献率((5)——(7)).

一般若前 k 个主成分的累计贡献率达到 85% 以上,就表明取前 k 个主成分基本包含了全部测量指标所具有的信息,因为前三个主成分的累计贡献率 $p = 93\%$,已经大于 85%,则取前 3 个主成分变量代替原来 10 个标量,并根据特征向量来求主成分得分,提出使用第一主成分来评价个股业绩,利用第一主成分来对样本股票进行排序,得到所选样本股票 2002 年的经营状况,给广大投资者一个理性的参考作用 前 3 个主成分对应的特征向量列于表 2

3.3 计算主成分得分并对个股进行评价

利用(4)中的计算公式:

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{pi}x_p, \quad i = 1, 2, 3; p = 1, 2, \dots, 10$$

可以分别计算各主成分的值

表 2 特征向量

主成分	1	2	3
x_1	0.401048	0.094703	0.089112
x_2	0.403139	0.090295	0.080681
x_3	0.414576	0.085301	0.039037
x_4	0.406954	0.095834	0.067097
x_5	0.345556	0.021592	0.005027
x_6	0.408464	0.045950	0.060557
x_7	0.024043	0.393296	-0.646526
x_8	-0.160771	0.468642	0.473934
x_9	-0.151813	0.490794	0.447068
x_{10}	-0.058561	0.590469	-0.365173

第一主成分计算公式:

$$F_1 = 0.401048x_1 + 0.403139x_2 + 0.414576x_3 + 0.406954x_4 + 0.345556x_5 \\ + 0.408464x_6 + 0.024043x_7 - 0.160771x_8 - 0.151813x_9 - 0.058561x_{10}$$

第二主成分计算公式:

$$F_2 = 0.094703x_1 + 0.090295x_2 + 0.085301x_3 + 0.095834x_4 + 0.021592x_5 \\ + 0.045950x_6 + 0.393236x_7 + 0.468642x_8 + 0.490794x_9 + 0.590469x_{10}$$

计算前 2 个主成分得分, 并按第一主成分之值按降序排列

表 3 按第一主成分之值降序排列

第一主成分名次	个股名称	第一主成分	第二主成分	第三主成分	第一主成分名次	个股名称	第一主成分	第二主成分	第三主成分
1	中国石化	13.4017	1.2511	1.1297	21	四川长虹	-0.5419	-0.2892	0.8742
2	中国联通	3.1959	-0.9146	0.9783	22	新农开发	-0.5923	-0.9564	-0.9792
3	招商银行	2.6160	-0.0311	-0.2109	23	首旅股份	-0.5938	-0.3389	-1.1511
4	宝钢股份	2.2731	0.0271	0.3667	24	亚星客车	-0.6366	-0.7988	-0.2368
5	深发展 A	0.6350	-0.8643	0.2277	25	南京新百	-0.6685	-0.2314	-0.1528
6	东方航空	0.2594	-1.9742	-0.3807	26	澳柯玛	-0.6685	-0.6592	-0.3892
7	TCL 通讯	-0.2631	3.7395	0.0729	27	上菱电器	-0.6700	0.8461	0.0472
8	白云山 A	-0.2639	-1.3660	1.5709	28	万科 A	-0.6952	1.0604	0.5275
9	哈空调	-0.3192	-2.0780	0.1971	29	东软股份	-0.7185	-0.0452	-0.3955
10	长江投资	-0.3237	-1.6834	-0.9318	30	古井贡 A	-0.7917	-0.1552	0.0537
11	五粮液	-0.3390	-0.2908	1.0363	31	中兴通讯	-0.8109	1.0072	-5.6019
12	工大首创	-0.3497	-1.5703	-0.7419	32	燕京啤酒	-0.8390	0.5939	-0.2033
13	一汽轿车	-0.3628	-0.9179	2.6450	33	金地集团	-0.8670	0.9162	0.2771
14	上海机场	-0.3871	0.0854	0.2315	34	春兰股份	-0.8869	0.3322	-0.2729
15	国旅联合	-0.4326	-1.5952	1.2882	35	新华医疗	-0.8870	0.4139	-0.6256
16	鲁能泰山	-0.4416	-1.0236	2.0592	36	海王生物	-0.9252	0.1794	-0.4305
17	东阿阿胶	-0.4911	-1.2382	0.6094	37	新大陆	-1.0070	0.7514	-0.7509
18	万杰高科	-0.4985	-0.8176	-0.2349	38	鄂尔多斯	-1.0222	0.5215	-0.0613
19	丽珠集团	-0.5140	-0.7549	0.8344	39	波导股份	-1.0543	3.5375	0.2003
20	哈药股份	-0.5334	0.5633	-0.3279	40	用友软件	-1.9846	4.7682	-1.1483

从表 3 可以认为第一主成分代表总的业绩水平, 亦即综合财务业绩。从第一主成分的计算公式可以看到第一主成分较大的系数是在 x_1 、 x_2 、 x_3 、 x_4 、 x_5 及 x_6 上, 也就是说这几个指标最能代表综合财务业绩。由此, 我们按第一主成分之值排序, 得到了表 3, 它是按第一主成分的数值由大到小, 即综合财务业绩由强到弱而列出的。此结果显示: 中国石化 ($F_1 = 13.4017$) 明显居于首位, 中国联通 ($F_1 = 3.1959$)、招商银行 ($F_1 = 2.6160$)、宝钢股份 ($F_1 = 2.2731$) 次之。可以看出, 综合财务业绩的排序与 x_1 、 x_2 、 x_3 、 x_4 、 x_5 及 x_6 的总的顺序大体保持一致, 但与其中每一个并不总保持完全一致。这说明了综合财务业绩并不能用任何一个指标完全代替, 即使是采用最有代表性的主营业务收入 x_1 也不行, 而是通过考虑了多个指标的综合情况来得到的。从上面第二主成分的计算公式可以看出, 它在变量 x_7 、 x_8 、 x_9 和 x_{10} 上有较大的正值系数, 其含义是有较多的净资产收益率、每股净资产、每股资本公积和每股收益将获得较大的数值。用友软件 ($F_2 = 4.7682$) 名列第一, TCL 通讯 ($F_2 = 3.7395$)、波导股份 ($F_2 = 3.5375$)、中国石化 ($F_2 = 1.2511$) 次之。而哈空调 ($F_2 = -2.0780$) 排在最后。第二主成分是衡量各股稳健发展的象征, 从第二主成分的得分可以

看出,名列前茅的多是科技含量较高、资本雄厚、经营比较稳健的股票。它们一般属于信息技术产业和钢铁石油产业。对于此类股票中长线投资价值较大。总之,面对上市公司财务报表中的众多指标,使用多元统计分析中的主成分分析法,可以用很少的综合指标代替原来众多的原始指标。以第一主成分的得分作为个股综合业绩的度量,能够真实的反映原始变量的主要信息,结果可靠。再加上第二主成分,第三主成分,广大的投资者就可以比较清楚的了解上市公司的财务状况了。

4 结 论

本文首先通过对沪深两市所选 40 样本股票的主成分得分的实证研究,应用一种客观赋权的动态评价方法,此方法包含了多指标的综合性 and 权数的动态可变性。利用上市公司年度报告所提供的财务指标,应用多元统计中的主成分分析方法对多指标表现进行综合。此外,本方法所运用的权数体系来源于特定年份的实际数据,所以它是能与宏观环境相匹配的动态权数体系,而在已介绍的国内外评价方法的主要特点就是单指标或固定权数综合。然后对所选样本股票来自 2002 年年报的财务指标,利用 SAS 统计软件进行处理,算出各只股票的第一主成分得分,并利用得分的大小进行降序排名,从而给广大的投资者一个显性的参考。

参考文献

- [1] 张尧庭,方开泰.多元统计分析引论[M].科学出版社,1987.
- [2] 于秀林.多元统计分析及程序[M].中国统计出版社,1993.
- [3] 江冬明.主成分分析在证券市场个股评析中的作用[J].数理统计与管理,2001,(3).
- [4] 潘琰,程小可.上市公司经营业绩的主成分评价方法[J].会计研究,2000,(1).
- [5] 董逢谷.上市公司综合评价[M].上海财经大学出版社,2002.
- [6] Anderson T W. An introduction to multivariate statistical analysis[J]. Wiley, 1984.
- [7] 李艳双.主成分分析法在多指标综合评价方法中的应用[J].河北工业大学学报,1999,(1).

Role of Principal Component Analysis in the Valuation of Chinese listed Companies on Stock Market

TIAN Bo-ping^{1,2}, WANG Yong¹, GUO Wen-ming¹, GE Xi-peng¹

(1. Math Dept of Harbin Institute of Technology, Harbin Heilongjiang 150001, China)

(2. Management Institute of Harbin Institute of Technology, Harbin Heilongjiang 150001, China)

Abstract Introduce principal component analysis when appearing on the stock market role in company's comprehensive evaluation. Principal component analysis is the method of a kind of objective weight defining. The weight is a dynamic weight system that can change as macro-economics environment changes frequently. Because the financial index and that data source are offered on the market company's financial report that principal component analysis applies. We mainly carry out more lateral comparison for the 40 blue chips selected, and they have a rank according to the scores of the first principal component to offer reference to masses of investor.

Keywords principal component analysis; first principal component; financial index; comprehensive evaluation