

doi: 10.3969/j.issn.1674-8425(z).2017.05.032

主成分分析和因子分析 在中国股票评价体系中的应用

胡书文¹ 徐建武²

(1. 重庆大学 数学与统计学院, 重庆 401331; 2. 92514 部队, 山东 烟台 264007)

摘 要: 目前在中国股票市场上有近 3 000 只股票, 在对股票进行投资之前, 需要分析各只股票的盈利情况, 从而投资那些盈利能力强、风险小、财务指标好的股票。但股票的评价指标有很多, 如何在一个评价体系里考虑所有指标是一个很现实的问题。运用主成分分析和因子分析对中国股票市场进行分析。主成分分析和因子分析能提取出反映股票各方面能力的成分, 比如股本扩张能力主成分、盈利能力主成分等, 从而可对各股票进行排名, 帮助投资者进行决策。

关 键 词: 股票评价; 主成分分析; 因子分析

中图分类号: O212.4; F830.91 文献标识码: A 文章编号: 1674-8425(2017)05-0192-11

Application of Principal Component Analysis and Factor Analysis in China's Stock Evaluation Market

HU Shu-wen¹, XU Jian-wu²

(1. College of Mathematical and Statistics, Chongqing University, Chongqing 401331, China;
2. The No. 92514th Troop of PLA, Yantai 264007, China)

Abstract: At present, there are nearly 3 000 stocks in the Chinese stock market. Before entering in the stock market, it is necessary to analyze the profitability of each stock and invest in that stocks with strong profitability, low risk and good financial indicators. However, there are many evaluation indicators of the stocks and how to evaluate all the indicators systematically is a very important problem. China's stock market was analyzed by principal component analysis and factor analysis. These two measures can extract the components that reflect the various aspects of the stock, such as the principal component of capital expansion capacity, principal component of profitability, and so on. After getting the principal component scores and factor scores, the stocks whose ranks were relatively high and low were analyzed and it is perfect to meet the actual situation. So the final ranks were given after considering the two cases.

Key words: stock evaluation; principal component analysis; factor analysis

收稿日期: 2016-12-17

基金项目: 国家自然科学基金青年基金资助项目(11001286)

作者简介: 胡书文(1992—), 女, 硕士研究生, 主要从事纵向数据、线性模型方面的研究, E-mail: swhu@cqu.edu.cn。

引用格式: 胡书文, 徐建武. 主成分分析和因子分析在中国股票评价体系中的应用[J]. 重庆理工大学学报(自然科学), 2017(5): 192-202.

Citation format: HU Shu-wen, XU Jian-wu. Application of Principal Component Analysis and Factor Analysis in China's Stock Evaluation Market[J]. Journal of Chongqing University of Technology(Natural Science) 2017(5): 192-202.

随着我国股票市场的不断发展,股票投资已成为投资者的主要途径之一。尽管从理论上讲股票投资能给投资者带来收益,甚至是巨大收益,但同时投资者也面临着巨大风险。所以投资者目前的主要目的是在保值的情况下实现最大的增值,这样就有了股票投资分析。当前股票投资分析方法主要有两种:一是基本分析,二是技术分析。但上述分析方法都是定性分析,通常只做一些描述性的说明且带有很强的主观成分,投资者很难从这些主观性的描述中确定哪些是值得投资的股票。因此,如何合理地建立一个投资评价体系成为亟待解决的问题。

2001年S Gnaesalingnaja, KuldeepKumar^[1]利用因子分析法,通过对1986—1991年美国证券市场中部分公司的数据进行分析得出了因子分析适合于股票市场的结论。张宗强、任敬喜应用因子分析应用于对2002年度汽车类上市公司投资价值分析^[2]。王文哲、李真燕将因子分析应用于我国家电业上市公司竞争力的分析^[3]。顾文炯采用因子分析对农业上市公司进行财务分析^[4]。

综上所述,国内外将主成分分析和因子分析应用到股票价值的评价体系中的研究并不是很多,因此本文结合主成分分析和因子分析来对股票进行定量评价,试图得到进一步的结果。

1 评价体系的建立

1.1 指标的选取原则

在建立评价体系时,指标的选取显得尤为重要。不仅要考虑指标反映投资价值的全面性,还要考虑指标体系是否科学、是否具有可比性、是否具有操作性以及数据能否有效得到等许多问题^[5]。因此,本文主要考虑以下几个指标选取原则:1)全面系统原则;2)恰当适用原则;3)灵活可控原则;4)简明科学原则。

1.2 指标的选取

指标的选取对评价体系有很重要的作用^[6]。

本文选取能反映股票价格的重要财务指标^[7]。见表1。

表1 投资价值评价体系指标分类

指标分类	指标名称	指标代码	指标计算公式
盈利能力	净利润率	X1	净利润/主营业务收入
	毛利率	X2	(主营业务收入 - 主营业务成本)/主营业务收入
偿债能力	流动比率	X3	流动资产/流动负债
	资产负债率	X4	负债总额/资产总额
资产管理能力	存货周转率	X5	销货成本/平均存货余额
	应收账款周转率	X6	销售收入/平均应收账款
	总资产周转率	X7	营业收入净额/平均资产总额
股本扩张能力	每股净资产	X8	股东权益/总股数
	每股公积金	X9	公积金/股票总股数
股本结构	流通股比例	X10	流通股数/总股本数

2 相关统计方法

2.1 主成分分析

主成分分析法就是一种既能保留原有信息又能简化指标变量的多元统计方法,它的目的是降维,设法将原始指标进行综合,求出一组原始指标的线性组合,这些线性组合即为主成分。在这个数学变换中,要求总的方差不变,然后使第一变量的方差最大,即为第一主成分,方差仅次于第一变量且与第一变量独立的第二变量,即为第二主成分,依此类推。在实际应用中,一般提取前几个方差较大的主成分,使其尽可能多地反映原始信息。另外,当主成分个数确定后,可以以各主成分的方差贡献率为权数,构造综合主成分。通过综合主成分的得分得到股票的排名进而构建评价体系。

主成分分析的基本步骤如下^[8]:

1) 确认待分析的原有若干变量是否适合做主成分分析,即确定变量之间是否具有较强的相关关系。一般可以做Bartlett球度检验,它是以变量的相关系数矩阵为出发点,零假设是 H_0 :相关

系数矩阵是一个单位阵。统计量根据相关系数矩阵的行列式计算得到,如果该统计量值比较大,且其对应的概率值小于某个给定的显著性水平,则应拒绝 H_0 ,认为相关系数矩阵不太可能是单位阵,适合做主成分分析;反之,则不适合。

2) 对原始指标数据标准化处理,消除计量单位不同量纲不一等因素。设有 p 个随机变量 X_1, \dots, X_p , 它们在第 i 次试验中的取值分别为 $x_{i1}, x_{i2}, \dots, x_{ip}$, 其中 $i = 1, \dots, n$ 。写成矩阵形式有

$$X = (x_1, x_2, \dots, x_p) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

对样本矩阵标准: $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$, $i = 1, 2, \dots, n; j = 1, 2, \dots, p$, 其中 \bar{x}_j, s_j 分别为 x_j 的样本均值和方差。得到标准化阵 $Z = (z_{ij})$ 。

3) 求出 Z 的相关系数矩阵 R :

$$R = \frac{Z^T Z}{n-1}$$

4) 求 R 的特征值和其对应的特征向量:

解相关系数矩阵 R 的特征方程 $|\lambda E - R| = 0$, 得到 p 个特征根,按大小排序为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, 然后由 $(\lambda E - R)X = 0$ 分别求出特征根 $\lambda_j (j = 1, 2, \dots, p)$ 所对应的 p 个特征向量: $\eta_1, \eta_2, \dots, \eta_p$ 。

5) 确定主成分个数

结合具体情况,一般按照累积贡献率 $\geq 70\%$

来确定主成分个数 m , 即 $\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.7$ 。 m 个主成分

分为 $U_i = X\eta_i (i = 1, 2, \dots, m)$ 。

6) 计算综合主成分得分

对 m 个主成分加权求和,就可以得出综合主成分得分,各主成分的权重通过主成分的方差贡献率来确定。

2.2 因子分析

因子分析是主成分分析的推广,它也是从研

究相关系数矩阵内部的关系出发,把一些具有相关关系的变量归结为少数几个综合因子的一种多元统计分析方法。其基本思想是:根据相关性大小将变量分组,使得同组内的变量之间相关性较高,但不同组的变量相关性较低。每组变量代表一个基本结构,这个基本结构称为公共因子。对于所研究的问题就希望用最少数个数的不可测的公共因子的线性组合与特殊因子之和来描述原来观测的每一分量。在实际问题中,描述一种现象的指标很多,从一些有关联的现象中找出少数几个主要因子,每一个主要因子就代表一种解释,抓住这些主要因子就可以帮助我们对复杂的问题进行分析和解释。

设 X 为 $p \times 1$ 随机向量,其均值为 μ ,协方差阵为 $\Sigma = (\sigma_{ij})$,若 X 能表示为 $X = \mu + Af + u$, 其中 Σ 是 $p \times k$ 未知常数阵, f 是 $k \times 1$ 随机变量, μ 是 $p \times 1$ 随机向量,且

$$\begin{cases} E(f) = 0, \text{Var}(f) = I \\ E(u) = 0, \text{Var}(u) = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \\ \text{Cov}(f, u) = 0 \end{cases}$$

则 $X = \mu + Af + u$ 称为因子分析模型, f 称为公共因子, u 称为特殊因子。

因子分析的基本步骤如下^[9]:

1) 确认待分析的原有若干变量是否适合做因子分析,即确定变量之间是否具有较强的相关关系。一般可以做 Bartlett 球度检验,具体步骤和上文主成分分析类似。

2) 构造因子变量和计算因子载荷矩阵,方法包括主成分法、主因子法和极大似然法等。其中,比较常用的是主成分法,主因子法和主成分分析的方法类似,故省略。唯一区别是因子分析要求因子载荷矩阵 $A = (\sqrt{\lambda_1}\eta_1, \sqrt{\lambda_2}\eta_2, \dots, \sqrt{\lambda_m}\eta_m)$,再计算各因子得分。

3) 因子变量的命名解释

因子变量的命名解释是因子分析中的重要问题,也是区别于主成分分析的地方。对于上面计算得到的因子载荷矩阵 A 进行观察,一般会发现

这样的现象: A 可能在某一行的许多列都有较大的取值, 或 A 可能在某一列的许多行上都有较大取值。这说明: 某个原有变量可能同时与几个因子都有比较大的相关关系。因此, 可以通过因子旋转的方法使每个变量在尽可能少的因子上有比较高的载荷, 即在理想状态下, 让某个变量在某个因子上的载荷趋于 1, 而在其他因子上的载荷趋于 0。这样, 一个因子就能够成为某个变量的典型代表, 实际含义更加清楚。

因子旋转的方法包括正交旋转法、斜交旋转法等, 本文分析采用方差极大法。即选取方差最大的正交旋转, 将某个因子旋转到某个位置, 使每个变量在旋转后的因子轴上的投影向最大、最小两级分化, 从而使每个因子中的高载荷只出现在少数的变量上。最后得到的旋转因子载荷矩阵中, 每列元素除几个值外, 其余的均接近于 0。

4) 计算因子得分

得到公共因子和因子载荷后, 需要反过来考察每个样本的得分情况, 从而对样本进行评价和分类。估计因子得分的方法有很多, 如汤普森 (Thompson) 1939 年提出的回归法, 也叫汤普森法, 约内斯克 (Joreskog) 和罗莱 (Lawley) 在 1967 年提出了一种较为实用的迭代法, R 中的因子分析函数 `factanal` 就采用了这种算法。

2.3 回归分析

线性回归分析最早由 19 世纪末期高尔顿提出。线性回归常表示成如下形式:

$$\begin{cases} Y = X\beta + \varepsilon \\ E\varepsilon = 0, \text{cov}(\varepsilon) = \sigma^2 I_n \end{cases}$$

其中 X 是一个纯量矩阵, 称为设计矩阵或结构矩阵, 在回归分析中一般假设 X 为列满秩, 即 $\text{rank}(X) = p + 1$; 然后求得 β 的最小二乘估计: $\hat{\beta} = (X^T X)^{-1} X^T Y$ 。从而得到 Y 的估计值为: $\hat{Y} = X \hat{\beta} = X(X^T X)^{-1} X^T Y$ 。

至于回归分析的显著性检验, 常见的有 F 检验和 t 检验。 F 检验是对模型的显著性检验, t 检验是对系数的显著性检验。

3 评价体系在我国股票市场中的应用

3.1 数据的来源

本文数据来源于 2015 年 10 月 31 日沪深的全部股票, 通过西南证券金点财富管理软件获得 2 780 只股票, 其中沪市 1 062 只, 深市 1 718 只。

3.2 数据的处理

3.2.1 数据的标准化处理

在做数据分析时, 引入的各变量常有不同的单位和不同的变异程度。例如: 本文中引入的变量净利润率与每股净资产, 不在一个数量级上。如果不对数据进行处理, 直接引入, 会造成每股净资产这个变量对评价系统的影响高于净利润率, 与实际经济情况不符。所以, 为了消除量纲影响, 需将数据进行标准化处理。数据标准化处理的方法有很多种, 常用的有“最小—最大化标准化”、“Z-score 标准化”和“按小数定标标准化”等。统计中常采用 Z-score 标准化。即: $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$, 其中 \bar{x}_j , s_j 分别为 x_j 的样本均值和方差。

3.3.2 数据的缺失值处理

在实际问题中, 有一些数据往往无法获取, 进而对整体的数据分析产生影响。所以有必要采取一些方法对缺失值进行处理。常见的方法有: 个案剔除法、均值替换法、回归替换法和多重替换法。本文采用剔除法和均值替换法。最后的股票数为 2 685 只。

3.3 主成分分析

1) 依据上文主成分分析的步骤, 首先判断 10 个变量之间的相关性, 相关系数如表 2。从表 2 中可看出: 各变量之间的相关系数不是很大。之后做 Bartlett 球度检验, p 值很小, 接近 0, 则应拒绝 H_0 , 认为相关系数矩阵不太可能是单位阵, 适合做主成分分析。

2) 用 R 中的函数 `scale()` 对原始数据进行标准化^[10]

3) 求标准化后数据的特征值

2 524.126 4 2 335.578 3 1 804.966 5 827.995 4。

10 个特征值依此为: 5 158.753 3 3 333.053 2 ,

4) 直接使用 R 提供的主成分函数 `princomp()` 进

2 786.389 5 2 734.092 8 2 694.469 0 2 640.575 7 , 行主成分分析 结果分别见图 1 , 2 , 3。

表 2 相关系数矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.000	-0.029	0.010	-0.063	0.000	-0.008	0.017	0.041	0.006	-0.022
x_2		1.000	0.018	-0.085	0.009	-0.003	-0.009	-0.026	0.008	-0.044
x_3			1.000	-0.238	-0.001	-0.006	-0.048	0.033	0.045	-0.074
x_4				1.000	-0.012	0.015	0.039	-0.182	-0.167	0.263
x_5					1.000	-0.002	-0.023	-0.006	0.007	0.032
x_6						1.000	-0.010	0.010	-0.031	0.032
x_7							1.000	0.073	-0.027	0.011
x_8								1.000	0.675	-0.150
x_9									1.000	-0.223
x_{10}										1.000

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.3861181	1.1141636	1.0187058	1.0091006	1.0017618	0.99169281
Proportion of Variance	0.1922039	0.1241823	0.1038148	0.1018663	0.1003901	0.09838211
Cumulative Proportion	0.1922039	0.3163862	0.4202010	0.5220674	0.6224574	0.72083955
	Comp.7	Comp.8	Comp.9	Comp.10		
Standard deviation	0.96957946	0.93266369	0.81990290	0.55531807		
Proportion of Variance	0.09404346	0.08701857	0.06724912	0.03084931		
Cumulative Proportion	0.81488301	0.90190157	0.96915069	1.00000000		

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
净利润率..			0.651	-0.429	-0.123	0.291	0.498	-0.153	0.117	
毛利率..		0.299	-0.309	0.526	-0.131	0.503	0.368	-0.340	0.116	
流动比率	0.180	0.534		-0.184	0.139	-0.100	-0.482	-0.442	0.429	
资产负债率..	-0.381	-0.487	-0.100					-0.121	0.755	
存货周转率			-0.407	-0.499	-0.582	0.387	-0.222	0.204		
应收账款周转率			-0.240	-0.293	0.778	0.473		0.145		
总资产周转率		-0.289	0.449	0.356		0.512	-0.555			-0.101
每股净资产	0.576	-0.371						-0.199		0.694
每股公积	0.591	-0.307	-0.158					-0.145		-0.700
流通股比例	-0.366	-0.259	-0.101	-0.201			-0.113	-0.724	-0.449	

图 1 主成分分析结果

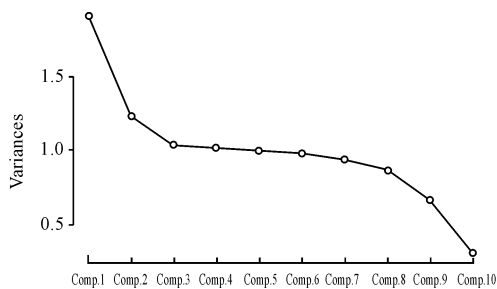


图 2 主成分碎石图

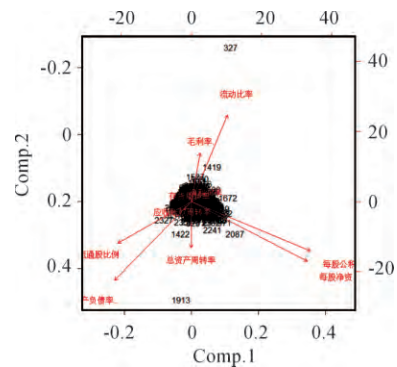


图 3 主成分分析散点图

由图 1 知: 前 6 个主成分的方差之和占全部方差之和的 72% , 所以按照本文标准选 6 个主成分为宜。这样, 原来的 10 个指标转化成 6 个, 起到了降维的作用。且由主成分碎石图可以看出: 前 3 个主成分解释的原始信息较多, 第 5 主成分到第 6 主成分没有明显的下降趋势, 但前 5 个主成分的方差之和只占全部方差之和的 62% , 所以还是选前 6 个主成分。

第 1 主成分 $Y_1 = 0.18x_3 - 0.38x_4 + 0.58x_8 + 0.59x_9 - 0.37x_{10}$

第 2 主成分 $Y_2 = 0.3x_2 + 0.53x_3 - 0.49x_4 - 0.29x_7 - 0.37x_8 - 0.31x_9 - 0.26x_{10}$

第 3 主成分 $Y_3 = 0.65x_1 - 0.31x_2 - 0.1x_4 - 0.41x_5 - 0.24x_6 + 0.45x_7 - 0.16x_9 - 0.1x_{10}$

第 4 主成分 $Y_4 = -0.43x_1 + 0.53x_2 - 0.18x_3 - 0.5x_5 - 0.29x_6 + 0.36x_7 - 0.2x_{10}$

第 5 主成分 $Y_5 = -0.12x_1 - 0.13x_2 + 0.13x_3 - 0.58x_5 + 0.78x_6$

第 6 主成分 $Y_6 = 0.29x_1 + 0.5x_2 - 0.1x_3 + 0.39x_5 + 0.47x_6 + 0.51x_7$

第 1 主成分方差贡献率为 19.22% , 主要与每股净资产和每股公积有关, 这 2 个指标能衡量股本的扩张能力, 称为股本扩张能力主成分; 第 2 主成分方差贡献率为 12.42% , 主要与流动比率和资产负债率有关, 而这 2 个指标可衡量偿债能力, 称为偿债能力主成分; 第三主成分方差贡献率为 10.38% , 主要与净利润率有关, 但其他各方面也都涵盖, 把它称为综合主成分; 第 4 主成分方差贡献率为 10.19% , 主要与净利润率和毛利率有关, 这 2 个指标能衡量盈利能力, 称为盈利能力主成分; 第 5 主成分方差贡献率为 10.04% , 主要与存货周转率和总应收账款周转率有关, 可衡量资产管理能力, 称为资产管理能力主成分; 第 6 主成分方差贡献率为 9.8% , 主要与毛利率和总资产周转率有关, 叫做盈利能力和资产管理能力主成分^[11]。

5) 令第 1 主成分 Y_1 , 第 2 主成分 Y_2 , 第 3 主成分 Y_3 , 第 4 主成分 Y_4 , 第 5 主成分 Y_5 , 第 6 主成分

Y_6 为自变量, 现价 Y 为因变量, 做回归分析。其中 $R^2 = 0.3012$, 调整的 $R^2 = 0.2996$, 说明拟合优度不是很好, 但 p 值很小, 说明模型通过了检验。 $Y = 20.76 + 6.51Y_1 + 0.03Y_2 + 0.1Y_3 + 0.84Y_4 - 0.27Y_5 + 1.08Y_6$, 而 Y_2, Y_3 和 Y_5 的系数没有通过检验, 猜测可能是自变量之间存在着严重的相关性, 因为 Y_3 被命名为综合主成分, 所以和其他主成分之间存在交叉。

然后采用逐步回归法, 利用 R 中的 stepAIC 函数, 该函数以 AIC 信息量为准则, 默认的是向后法, 从所有变量开始, 逐步通过选择最小的 AIC 信息量达到增删变量的目的。结果最后保留的变量是第 1 主成分 Y_1 , 第 4 主成分 Y_4 , 第 6 主成分 Y_6 , 回归方程为 $Y = 20.76 + 6.51Y_1 + 0.84Y_4 + 1.08Y_6$, 且系数也都通过了显著性检验。说明股价主要和第 1 主成分股本扩张能力主成分、第 4 主成分盈利能力主成分、第 6 主成分盈利能力和资产管理能力主成分有关, 这也符合实际情况。

6) 求出各主成分的得分 $f_1, f_2, f_3, f_4, f_5, f_6$, 以各主成分的方差贡献率 α_i 作为权重, 构造综合评价函数 $f = \sum_{i=1}^6 \alpha_i f_i$, 就可以得到主成分的综合得分 f 并进行排序, 主成分得分排名见表 3。

限于篇幅, 这里只给出了排名前 20 名和后 20 名的股票。本文的评价系统对每个股票的评价具有现实的参考价值。

3.4 因子分析

1) 按照上述因子分析的步骤, 首先确认原有若干变量是否适合做因子分析, 即是否具有相关性, 根据主成分分析的结果, 答案是肯定的。

2) 构造因子变量和计算因子载荷矩阵

因子分析和主成分分析有一个不同之处在于主成分分析是根据算出来的累积贡献率来确定主成分个数, 但是因子分析是在之前就预判出要提取的因子个数。此处采用 R 中的 fa.parallel 函数, 结果建议选择 3 个因子, 如图 4。但 3 个因子总的方差贡献率仅为 42% , 为防止遗漏重要信息, 选取 5 个因子。5 个因子的方差贡献率为 62.2% 。

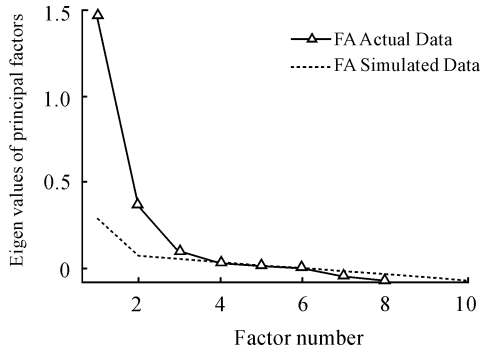


图4 因子选取

3) 因子变量的命名解释,因子载荷矩阵见

表4。

由表4可以看出: 尽管一个因子可以解释许多变量信息,但它却只能解释某个变量的少部分信息,不是任何一个变量的典型代表。这样的情况必然使得因子变量的实际含义模糊不清。因此,可以通过因子旋转的方法使每个变量在尽可能少的因子上有比较高的载荷。这样,一个因子就能够成为某个变量的典型代表,实际含义更加清楚。这里采用方差极大法进行因子旋转。旋转后的因子载荷矩阵见表5。

表3 主成分得分排名

排名	名称	主成分得分	排名	名称	主成分得分
1	五矿稀土	4.609 155 226	2 666	云维股份	-0.695 550 024
2	栋梁新材	1.487 688 630	2 667	珠江控股	-0.698 723 188
3	赛升药业	1.445 858 497	2 668	四川金顶	-0.706 306 943
4	北京君正	1.417 241 033	2 669	实达集团	-0.714 643 598
5	牧原股份	1.236 005 664	2 670	* ST 云网	-0.722 465 161
6	上海钢联	1.161 548 029	2 671	ST 景谷	-0.760 577 622
7	贵州茅台	1.087 673 995	2 672	* ST 中昌	-0.769 336 890
8	裕兴股份	1.018 246 629	2 673	黑化股份	-0.785 238 197
9	欧亚集团	1.016 268 515	2 674	南华生物	-0.792 270 948
10	天孚通信	0.983 915 319	2 675	* ST 南化	-0.831 107 438
11	永兴特钢	0.976 677 562	2 676	汇通能源	-0.844 506 191
12	日机密封	0.937 451 165	2 677	* ST 美利	-0.887 193 266
13	德尔股份	0.936 809 348	2 678	* ST 新都	-1.071 560 834
14	国光股份	0.914 920 289	2 679	* ST 京蓝	-1.184 656 737
15	恒锋工具	0.913 896 283	2 680	* ST 川化	-1.318 994 857
16	爱普股份	0.911 922 996	2 681	生意宝	-3.277 438 450
17	广宇发展	0.902 325 686	2 682	天润控股	-3.383 379 400
18	金雷风电	0.896 838 121	2 683	黔源电力	-3.564 502 740
19	嘉宝集团	0.865 043 887	2 684	ST 宜纸	-3.796 176 598
20	焦点科技	0.854 236 866	2 685	西水股份	-5.629 082 895

表4 因子载荷矩阵

	Factor1	Factor2	Factor3	Factor4	Factor5
x_1	-0.092 223	0.052 205	0.663 188	-0.432 854	-0.123 081
x_2	-0.060 547	0.333 529 3	-0.315 288	0.531 206	-0.130 932
x_3	-0.249 921	0.595 163 7	0.083 510 7	-0.185 519	0.139 094
x_4	0.528 782	-0.543 162	-0.102 102	0.026 055	-0.019 629
x_5	0.009 283	0.0317 860	-0.414 769	-0.503 705	-0.582 843
x_6	0.048 761	-0.056 923	-0.244 291	-0.295 853	0.779 213
x_7	0.000 810	-0.321 772	0.457 164 9	0.359 131	-0.060 639
x_8	-0.798 370	-0.412 984	-0.065 328	-0.044 983	0.026 073
x_9	-0.818 847	-0.342 272	-0.161 387	-0.022 094	-0.023 309
x_{10}	0.507 370	-0.288 298	-0.102 457	-0.203 191	-0.009 741

表5 旋转后的因子载荷矩阵

	Factor1	Factor2	Factor3	Factor4	Factor5
x_1		0.173	0.776		-0.144
x_2		0.242	-0.620		-0.249
x_3		0.665			0.131
x_4	0.190	-0.739			
x_5				-0.862	-0.134
x_6				0.105	0.865
x_7		-0.253	0.160	0.478	-0.346
x_8	-0.900				
x_9	-0.897				
x_{10}	0.288	-0.495		-0.178	0.159

每股净资产 x_8 , 每股公积金 x_9 在 Factor1 上有较大载荷, 称为股本扩张能力因子; 流动比率 x_3 , 资产负债率 x_4 , 流通股比例 x_{10} 在 Factor2 上有较大载荷, 称为偿债能力因子和股本结构因子; 净利润率 x_1 , 毛利率为 x_2 在 Factor3 上有较大载荷, 称为盈利能力因子; 存货周转率 x_5 , 总资产周转率 x_7 在 Factor4 上有较大载荷, 称为资产管理能力因子; 应收账款周转率 x_6 在 Factor5 上有较大载荷, 称为资产管理能力因子。

4) 计算因子得分

用 R 中的因子分析函数 `factanal` 计算因子得分, 然后以各因子的方差贡献率为权数, 算出总的

因子得分并进行排名。见表6。

与前面的主成分得分排名进行对比发现: 主成分得分排名第1的五矿稀土在因子得分中排名倒数第7, 而查看五矿稀土的具体情况不难看出其综合主成分得分为4.61, 明显高于后面股票的综合主成分, 原因在于它的第2主成分显著较高, 而第2主成分为偿债能力主成分, 并非越高越好, 而是需要在一个合适的范围。且因子得分中五矿稀土的得分比较低是因为第5因子资产管理能力因子特别低, 即应收账款周转率特别低, 说明企业的资金周转不好, 所以综合来看, 虽然五矿稀土的主成分得分排名第1, 但是不推荐。

因此, 考虑把主成分得分和因子得分综合起来考虑, 取平均值, 得到的最终排名如表7(排除五矿稀土和西水股份, 因为其主成分得分明显异常), 明显看出排名较差的20名中有好多是ST板块的, 也比较符合实际情况。搜索了一下排名靠前的股票, 例如上海钢联、贵州茅台和赛升药业等, 也都是利好消息, 可以投资。例如上海钢联排名第一, 查阅相关信息可知^[12]: 上海钢联在国内钢铁工业陷入严冬之际, 主营线上钢市信息和钢材交易, 在过去一段时间里实现了股价十几倍的暴涨。这家公司在2011年登陆创业板, 此前主营业务是钢铁行业的咨询信息服务。在传统钢贸业企业纷纷倒塌的时候, 该公司开始向后端的线上

交易、金融等板块延伸,意在打造垂直于钢铁全产业链的电商生态系统。被贴上互联网、金融、电商标签的上海钢联迅速成为资本市场的宠儿。从2013年的5月到2015年的3月,股价从8元左右飙升至143元左右,涨超16倍。2015年9月到12月,该股的关注度高于行业内的其他113家公司,排名第41。共有5家机构发布了18篇关于该股的研究报告,该股综合评级为增持,维持前期评级。虽然截止目前(2016年1月8日)的消息^[13]是:上海钢联跌停,报于36.38元,但其实是跟这

些天股市的不稳定形态有关^[14],在熔断机制运行的4个交易日中已有1月4日和1月7日两天被触发熔断,同时均伴随市场暴跌。1月4日下午,A股大跌连续两次触发熔断并暂停交易;3天后,沪深300指数在开盘不到半小时内,再度两次触发熔断。实施熔断机制连续4天以来,上证指数跌幅累计达11.96%。所以,这不影响上海钢联的内在投资价值。排名最后的ST宜纸本身就处于ST板块,且最新的消息是“ST宜纸‘易主+重组’方案遭上交所问询”,也已停牌多日,显然不适合投资。

表6 因子得分排名

排名	名称	因子得分	排名	名称	因子得分
1	上海钢联	1.759 426 476	2 666	戴维医疗	-0.465 863 483
2	栋梁新材	1.306 157 208	2 667	京天利	-0.465 925 677
3	大商股份	1.255 492 333	2 668	江苏三友	-0.469 662 958
4	贵州茅台	1.110 839 018	2 669	继峰股份	-0.478 779 140
5	西泵股份	1.085 226 025	2 670	印纪传媒	-0.481 079 363
6	王府井	1.059 879 497	2 671	海思科	-0.486 103 036
7	天齐锂业	1.059 486 878	2 672	利君股份	-0.487 028 209
8	上海物贸	1.053 679 349	2 673	青海春天	-0.487 089 571
9	蓝星新材	0.960 785 193	2 674	海洋王	-0.488 233 460
10	国药一致	0.926 566 433	2 675	浙江美大	-0.489 214 521
11	联创股份	0.922 428 137	2 676	龙津药业	-0.491 008 707
12	西水股份	0.920 416 651	2 677	建新矿业	-0.495 939 377
13	人人乐	0.913 141 578	2 678	大康牧业	-0.499 124 970
14	设计股份	0.898 011 009	2 679	五矿稀土	-0.499 439 146
15	力生制药	0.887 962 971	2 680	凯美特气	-0.504 804 546
16	爱施德	0.878 655 517	2 681	大豪科技	-0.510 249 584
17	宇顺电子	0.859 281 430	2 682	海南矿业	-0.517 255 781
18	亚玛顿	0.851 734 551	2 683	保千里	-0.529 772 365
19	重庆百货	0.847 732 564	2 684	神州高铁	-0.534 967 941
20	云南铜业	0.828 440 223	2 685	中毅达	-0.676 069 598

表 7 综合排名

排名	名称	f	排名	名称	f
1	上海钢联	1.458 687 761	2 666	* ST 中昌	-0.321 780 206
2	栋梁新材	1.395 495 597	2 667	武昌鱼	-0.321 938 403
3	贵州茅台	1.098 971 322	2 668	四川金顶	-0.327 901 174
4	赛升药业	1.005 774 046	2 669	珠江控股	-0.332 955 478
5	西泵股份	0.931 219 315	2 670	* ST 云网	-0.335 332 481
6	大商股份	0.909 762 488	2 671	天津磁卡	-0.336 521 329
7	欧亚集团	0.864 796 652	2 672	* ST 美利	-0.339 425 411
8	德尔股份	0.811 784 628	2 673	嘉凯城	-0.340 469 857
9	焦点科技	0.802 108 407	2 674	华塑控股	-0.343 921 504
10	天齐锂业	0.769 045 651	2 675	万方发展	-0.344 253 350
11	力生制药	0.761 655 706	2 676	实达集团	-0.345 368 964
12	蓝星新材	0.760 586 528	2 677	汇通能源	-0.360 953 127
13	上海物贸	0.756 276 244	2 678	南华生物	-0.383 726 161
14	联创股份	0.749 849 472	2 679	* ST 新都	-0.420 662 867
15	王府井	0.749 706 115	2 680	* ST 川化	-0.429 269 923
16	国药一致	0.724 375 028	2 681	* ST 京蓝	-0.544 891 272
17	金雷风电	0.723 827 872	2 682	黔源电力	-1.609 929 598
18	日机密封	0.701 850 937	2 683	生意宝	-1.828 241 497
19	设计股份	0.699 940 400	2 684	天润控股	-1.847 819 969
20	老凤祥	0.694 222 309	2 685	ST 宜纸	-2.090 403 260

4 结论与展望

4.1 主要结论

本文分析了影响股票的一些重要财务指标，利用主成分分析和因子分析建立了股票的评价体系，并利用相关数据分析得出影响股价最重要的因素有股本扩张能力、盈利能力和资产管理能力，给出了一个最终的股票排名。表 7 是根据这个评价体系给出的股票投资的参考。

4.2 展望

本文所提出的投资评价体系仍然存在不足之处。例如指标的选取不全，没有考虑到影响股票的其他宏观因素、数据的选取问题等。总之，要更

加透彻地研究股票评价模型，不仅需要进一步深化学习，还应该努力追踪国内外最新研究进展，通过对股票影响因素的深入挖掘，结合我国股票变化特点，形成一套行之有效的理论。

参考文献：

[1] GAMESALINGAM S ,KUMAR K. Detection of financial distress via multivariate statistical analysis[J]. Managerial Finance 2001 27(4) :45 - 55.

[2] 张宗强 ,任敬喜. 对 2002 年度汽车类上市公司投资价值的因子分析[J]. 价值工程 2004(5) : 109 - 112.

[3] 王文哲 ,李真燕. 我国家电上市公司竞争力的因子分析[J]. 价值工程 2005(3) : 105 - 108.

- [4] 顾文炯. 用因子分析对农业上市公司进行财务评价[J]. 安徽大学学报 2005 29(3): 136-139.
- [5] 杜茜 张雪梅. 清洁能源行业上市公司投资价值评估研究[D]. 北京: 中国地质大学 2013.
- [6] 任福匀. 因子分析法在我国股票市场行业投资价值评价中的应用[D]. 长沙: 中南大学 2005.
- [7] 姜爱宇. 部分线性模型在股票价格预测中的应用研究[D]. 大连: 辽宁师范大学 2012.
- [8] 王学民. 应用多元分析[M]. 上海: 上海财经大学出版社 2009.
- [9] 杨虎. 金融大数据统计方法与实证[M]. 北京: 科学出版社 2016.
- [10] 汤银才. R 语言与统计分析[M]. 北京: 高等教育出版社 2008: 316-324.
- [11] 张甜 牛明飞. 基于主成分分析的股价因素分析[D]. 兰州: 兰州大学 2013.
- [12] 新京报. 上海钢联: 从 8 元涨到 143 元的“电商故事”[EB/OL]. [2015-04-13]. <http://tech.163.com/15/0413/02/AN22CJ0700094ODV.html>.
- [13] 金融界网站. 快讯: 上海钢联跌停 报于 36.38 元[EB/OL]. [2016-01-08]. <http://stock.jrj.com.cn/hotstock/2016/01/08095720376470.shtml>.
- [14] 东方网. 熔断重启时间未定证监会称将“不断完善相关机制”[EB/OL]. [2016-01-08]. <http://finance.sina.com.cn/roll/2016-01-08/doc-ifxnkeru4776113.shtml>.

(责任编辑 陈 艳)

(上接第 168 页)

- [8] 余素红 张世英. SV 和 GARCH 模型拟合优度比较的似然比检验[J]. 系统工程学报 2005 19(6): 625-629.
- [9] JACQUIER E POISON N G ROSSI P E. Bayesian Analysis of Stochastic Volatility Models[J]. Journal of Business & Economic statistics 1994(12): 371-389.
- [10] 朱慧明 李峰. 基于 MCMC 模拟的贝叶斯厚尾金融随机波动模型分析[J]. 运筹与管理 2007 16(4): 111-115.
- [11] 苏理云 彭相武 王杰. 基于状态空间 SV-T-MN 模型的股指波动率预测[J]. 数理统计与管理 2016(5).
- [12] LONGIN F M. From value at risk to stress testing: The extreme value approach[J]. Journal of Banking & Finance 2000 24(7): 1097-1130.
- [13] 周炳均 王沁 郑兴. 基于两种分布下的 SV 模型与 GARCH 模型的 VaR 比较[J]. 重庆文理学院学报(社会科学版) 2016(5): 133-137.
- [14] 胡晓馨. 基于极值理论的黄金期货市场风险度量研究[D]. 杭州: 浙江大学 2014.
- [15] MCNEIL A J. Calculating quantile risk measure for financial time series using extreme value theory[J]. ASTIN Bulletin 2000 27(1): 117-137.
- [16] 桂文林 韩兆洲. POT 模型中 GPD“厚尾”性及金融风险测度[J]. 数量经济技术经济研究 2010(1): 107-118.
- [17] 耿贵珍 王慧彦. 基于 POT-GPD 模型的地震巨灾损失分布研究[J]. 自然灾害学报 2016(3): 153-158.
- [18] PAUL H K. Techniques for Verifying the Accuracy of Risk Measurement Models[J]. The Journal of Derivatives 1995 3(2): 73-84.

(责任编辑 何杰玲)