# 因子分析中指标数据如何正确预处理

范 坤 冯长焕(教授)

(西华师范大学数学与信息学院 四川南充 637002)

【摘要】本文通过建立合理的指标正向化转换模型证明了曲线模型不适合做指标正向化处理,然后用常用的热平台插补法对样本缺失值进行了填补,同时注意控制样本指标缺失率。最后基于上市电子企业对处理前后的指标数据进行了因子分析,并对排名结果进行了对比分析,证实了对指标数据合理地预处理后其评价结果更合理、客观有效。

【关键词】因子分析 数据处理 指标正向化 缺失值

因子分析是目前国内外进行综合评价的主要方法之一,也是多元统计分析中应用较为广泛的处理数据降维的方法。在实际应用中主要是通过寻找多个指标的少数独立的、专业上有意义的公因子来探索解释原多个指标对个体特征描述的关系。

在做因子分析之前指标数据必须进行样本缺失值、指标正向化和量纲化预处理。当个别样本(公司)缺失某些指标数据,若不对其进行缺失值填补则做因子分析时会被剔除进而不能参与最终的各项排名和综合评价,而一般的用指标均值代替缺失值又不尽如人意,因此有必要选择更合适的填补方法,同时注意控制样本的指标缺失率。本文建立了合理的指标正向化转换模型,基于上市电子企业对处理前后的指标数据进行了因子分析,并对排名结果进行了对比分析。

## 一、缺失值填补

本文以2010年133家上市电子公司的37个财务指标作为因子分析的原始变量为例,用比较常用的热平台插补法进行缺失值填补(数据来源于中国上市公司资讯网)。

37个财务指标分别为:每股收益(扣除)、每股净资产、每股资本公积金、每股未分配利润、每股现金净流量、每股经营活动现金流量、销售毛利率、主营业务利润率、销售净利率、总资产收益率、净资产收益率、加权平均净资产收益率、应收账款周转率、存货周转率、固定资产周转率、股东权益周转率、总资产周转率、流动比率、速动比率、股东权益与固定资产比率、利息保障倍数、资产负债比率、股东权益比率、固定资产比率、主营业务收入增长率、营业利润增长率、净利润增长率、利润总额增长率、净资产增长率、总资产增长率、现金比营运指数、销售现金比率。

首先由软件计算出37个财务指标之间的相关系数矩阵,然后将与含待估缺失值指标相关性最强的指标排序,其次,对待估指标排序并求出排序后的一阶差分,再求出一阶差分的平均值,最后依据其相关性最强的指标的排序和一阶差分平均值填补上缺失值。由于原始数据中一共存在32个缺失值,涉及6个公司的18个财务指标,下面以现金比率指标(含一个缺失值利达光电公司)为例进行填补,其他指标类似。具体填补过程如下:

与现金比率指标相关性较强的有三个指标,依次为速动比率、流动比率和总资产增长率,其相关系数矩阵如表1。因为流动比率和速动比率中也含有缺失值,因而用相关性次之的总资产增长率来估计缺失值。

**表1 总资产增长率、流动比率、速动比率和现金比率的相关系数矩阵**

| | 总资产增长率 | 流动比率 | 速动比率 | 现金比率 |
|---|---|---|---|---|
| 总资产增长率 | 1.000 | 0.731 | 0.740 | 0.758 |
| 流动比率 | 0.731 | 1.000 | 0.998 | 0.986 |
| 速动比率 | 0.740 | 0.998 | 1.000 | 0.990 |
| 现金比率 | 0.758 | 0.986 | 0.990 | 1.000 |

**表2 依据总资产增长率指标的公司排序与对照表**

| 公司 | 中京电子 | 蓉胜超微 | 深桑达A | 利达光电 | 深天马A | 超华科技 | 莱宝高科 |
|---|---|---|---|---|---|---|---|
| 总资产增长率 | 25.46 | 24.85 | 24.28 | 23.78 | 23.09 | 22.41 | 22.28 |
| 排 序 | 72 | 73 | 74 | 75 | 76 | 77 | 78 |
| 现金比率 | 21.48 | 17.59 | 29.63 | 缺 | 40.01 | 46.37 | 330.59 |
| 排 序 | 122 | 124 | 111 | | 96 | 85 | 38 |

表2给出了总资产增长率指标中利达光电公司前后的6家公司排序,以及对照给出了这6家公司在现金比率指标中的排序。然后对现金比率排序并计算出一阶差分平均值为23.8628,这样就可依据利达光电在总资产增长率中的排序相对其在现金比率中的位置进行插补,可用深天马A公司的现金比率值减去一阶差分值即40.01−23.862 8=16.147 2。其他指标中缺失值同样按照上述方法进行填补,其结果见表3:

**表3　　其他指标缺失值对应的公司估计值**

| 公　司 | *ST夏新 | 利达光电 | 安居宝 | *ST博信 | 长白5 | ST三星 |
|---|---|---|---|---|---|---|
| 每股收益扣除元 | | | | | 0.319 466 | |
| 销售毛利率 | 22.044 7 | | | | | |
| 主营业务利润率 | 10.677 1 | | | | | |
| 销售净利率 | 10.929 6 | | | | | |
| 加权平均净资产收益率 | 59.854 5 | | | | −0.274 5 | |
| 应收账款周转率 | 0.675 2 | | | 4.404 1 | | 5.428 8 |
| 存货周转率 | 0.274 1 | 0.527 3 | | 2.092 3 | | |
| 利息保障倍数 | | 1 572.866 5 | | | | |
| 主营业务收入增长率 | | | | −85.04 | | |
| 经营净现金比率 | | 23.683 9 | 35.756 1 | | | |
| 经营现金负债总额比 | | 28.359 3 | 51.879 3 | | | |
| 全部资金现金回收率 | | 15.318 7 | 8.238 7 | | | |
| 净收益营运指数 | | −0.522 | −0.216 3 | | | |
| 现金营运指数 | | 0.796 4 | 0.381 5 | | | |
| 销售现金比率 | −18.494 8 | 36.184 8 | 18.114 8 | | | |
| 流动比率 | | 1.368 1 | | | | |
| 速动比率 | 0.619 7 | 0.930 7 | | 0.489 2 | | |

## 二、指标正向化处理

### (一)合理的指标正向化公式

**1.** 范坤等(2012)给出了判断指标正向化转换模型合理性的定义,并给出了逆指标和适中指标正向化的合理转换模型。指标数据在处理前后的相对位置必须保持不变，否则以适中指标为例，可能会造成转换前离适中值远的数据转换后反而比离适中值近的数据效果好。

定义(保距性):设f为一一变换,$f:x_i\to r_i=f(x_i)$,对任意的$r_i$、$r_j$,且$r_i\leqslant r_j$,令$r_{min}=\min f(x_i)$,$r_{max}=\max f(x_i)$,$\alpha=\min x_i$,$\beta=\max x_i$,适中值为$u_0$(下同),则对于适中指标转换模型:

当$x_i\leqslant u_0$时,有$\frac{r_j-r_i}{r_{max}-r_{min}}=\frac{x_j-x_i}{u_0-\alpha}$;

当$x_i\geqslant u_0$时,有$\frac{r_j-r_i}{r_{max}-r_{min}}=\frac{x_i-x_j}{\beta-u_0}$,则称f是合理的。

对于逆指标转换模型,若$\frac{r_j-r_i}{r_{max}-r_{min}}=\frac{x_i-x_j}{\beta-\alpha}$,则称f是合理的。

**2.** 关于因子分析应用的文章有很多,但很多研究者并没有考虑指标正向化的问题或者是没能合理地对指标正向化处理。下面给出正确处理指标正向化的方法:

逆指标正向化转换模型为:

$$f(x_i)=a-bx_i \qquad (1)$$

其中:$b>0$,a为任意常数,$f(x_i)\in[a-b\beta,a-b\alpha]$。

适中指标正向化转换模型为:

$$f(x_i)=\begin{cases}a+b(x_i-u_0),x_i\leqslant u_0\\ a-b(x_i-u_0),u_0\leqslant x_i\end{cases}$$

或 $f(x_i)=a-b|x_i-u_0|$ (2)

其中$b>0$,a为任意常数,函数在$\mu_0$处取得最大值$\alpha$,其转换后限定区间为:$u_0$的左边$f(x_i)\in[a+b(\alpha-u_0),a]$,右边$f(x_i)\in[a-b(\beta-u_0),a]$。

因此,在式(1)和(2)中,当a和b分别取不同值时都是合理的指标正向化模型,满足定义中的保距性。

例如:逆指标和适中指标的正向化公式可分别取$f=-x$,$f=-|x-u_0|$。

### (二)曲线模型不能进行指标正向化

由式(1)和式(2)可知线性模型是合理的转换模型,而有些学者通常喜欢用$f(x_i)=\frac{1}{x_i}$对逆指标正向化处理,这是不合理的。下面证明所有曲线模型不能合理地对指标进行正向化处理(限于篇幅,只讨论$x_i\leqslant u_0$的情况,$u_0\leqslant x_i$的情况或逆指标的证明类似)。

$\forall x_i\in[\alpha,u_0]$,令$A(\alpha,r_{min})$,$B(u_0,r_{max})$此时指标正向化公式为曲线,则存在无数条与割线AB平行的割线,只有满足此条件下即平行于割线AB对应的$x_i$、$x_j\in[\alpha,u_0]$且$x_i<x_j$才能合理地进行转换(满足定义相对距离不变),而对于曲线模型$\forall x_i,x_j\in[\alpha,u_0]$,且$x_i<x_j$,经过一一变换后得到的$r_i$、$r_j$($r_i<r_j$)是无法满足相对距离不变的。因此,曲线模型无法满足任意性,当不平行于割线AB的割线对应的$x_i$,$x_j$经过曲线模型转换得到的$r_i$、$r_j$($r_i<r_j$)将改变了转换前后的相对距离,因此就不满足定义,也就证明了曲线模型不能进行指标正向化处理。

## 三、实证分析

对缺失值填补得到完整数据集后再进行指标正向化和标准化处理,然后做因子分析其结果才是最客观,最有效的。在式(1)和式(2)中,当a和b分别取不同值时都是合理的指标正向化模型,而无论是形式、意义不同还是转换到不同的限定区间上,进行转换后的指标其标准化后的数据是相同的,因此其因子分析结果也是相同的,这个从理论上不难验证,且通过SPSS直接保存标准化后的指标进行对比验证也可得知。

下面基于2010年上市电子企业分别对未处理过的原始数据和预处理过的数据进行因子分析，并建立因子得分排名和综合排名,然后进行对比分析。本文所用样本数据37个财务指标中有5个适中指标和1个逆指标(见表4),其他为正指标。

**表4　　指标目标极性表**

| 指标名称 | 指标计算公式 | 企业适中值 | 目标极性 |
|---|---|---|---|
| 流动比率(倍) | 期末流动资产/期末流动负债 | 2 | 适中 |
| 速动比率(倍) | 期末速动资产/期末流动负债 | 1 | 适中 |
| 资产负债比率(%) | 期末负债总额/期末总资产 | 70% | 适中 |
| 股东权益比率(%) | 股东权益总额/资产总额 | 30% | 适中 |
| 现金比率 | (现金+有价证券)/流动负债 | 60% | 适中 |
| 固定资产比率(%) | 固定资产总额/资产总额 | | 逆指标 |

**1.** 因子分析过程。原始数据存在缺失值，所以在进行因子分析时剔除了含有缺失值数据的公司，从而其没有参与排名，分析结果只包含 127 个公司。原始数据因子分析过程略，本文重点给出数据预处理后的因子分析过程和排名结果。

由于预处理后的数据是完整数据集，所有 133 个公司将进入因子得分排名和综合排名。由表 5 知 KMO 值为 0.687，球形 Bartlett 检验近似卡方值为 6 985.657，显著性概率 P 值为 0.000，远小于 0.05，可以认为适合因子分析。

**表 5　　数据处理后的 KMO and Bartlett´s 检验**

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. KMO值 | | 0.687 |
| 球形Bartlett´s检验 | 近似卡方值 | 6 985.657 |
| | 显著性概率P值 | 0.000 |

确定因子变量的方法有很多，本文选择基于主成分模型的主成分分析法对数据进行分析。根据特征值大于 1 的原则提取的 10 个因子所解释的方差占整个方差的 79.944%，因此能够比较全面地反映原有信息。为了使找到的这 10 个主因子更易于解释，采用最大方差旋转法。

通过旋转后的因子荷载矩阵，我们发现第 1 个公因子与每股净资产（元）、每股现金净流量（元）、每股资本公积金(元)、净资产增长率(%)、总资产增长率(%)、每股收益(扣除)(元)、固定资产比率(%)关系密切；第 2 个公因子与总资产收益率(%)、加权平均净资产收益率(%)、主营业务利润率(%)、销售净利率(%)、销售毛利率(%)、每股未分配利润(元)、现金营运指数关系密切；第 3 个公因子与全部资金现金回收率、净资产收益率(%)、经营现金负债总额比、经营净现金比率关系密切；第 4 个公因子与股东权益周转率(次)、资产负债比率(%)、股东权益比率(%)关系密切；第5 个公因子与利润总额增长率(%)、净收益营运指数、利息保障倍数(倍)关系密切；第 6 个公因子与流动比率(倍)、速动比率(倍)、现金比率关系密切；第 7 个公因子与销售现金比率、每股经营活动现金流量(元)、主营业务收入增长率(%)关系密切；第 8 个公因子与固定资产周转率(次)、股东权益与固定资产比率(倍)关系密切；第 9 个公因子与净利润增长率(%)、营业利润增长率(%)关系密切；第 10 个公因子与应收账款周转率、存货周转率(次)、总资产周转率(次)关系密切。

所以我们可以将这 10 个因子根据其财务指标的构成内容分别命名为盈利能力因子、销售能力因子、变现能力因子、风险控制因子、营运能力因子、偿债能力因子、现金周转因子、固定资产生产能力因子、利润增长因子、资产管理因子，然后可以建立因子得分排名和综合排名。

我们以旋转后 10 个因子的方差贡献率为权重，建立综合函数，公式如下：

$$Zf=18.844\%\times F_1+13.342\%\times F_2+9.276\%\times F_3+7.41\%\times F_4+7.409\%\times F_5+7.314\%\times F_6+4.992\%\times F_7+3.862\%\times F_8+3.757\%\times F_9+3.739\%\times F_{10}$$

**2.** 结果分析。表 6 和表 7 分别是数据处理前后因子分析各因子和综合排名结果前 15 名。

**表 6　　原始数据排名结果**

| 公　司 | 偿债能力 | 销售能力 | 现金周转 | 风险控制 | 利润增长 | 盈利能力 | 营运能力 | 资产管理 | 变现能力 | 固资生产 | 综合排名 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 国民技术 | 1 | 14 | 10 | 14 | 8 | 6 | 14 | 10 | 10 | 2 | 1 |
| 中瑞思创 | 2 | 13 | 7 | 10 | 12 | 11 | 5 | 3 | 7 | 12 | 2 |
| 乾照光电 | 4 | 10 | 6 | 5 | 4 | 12 | 2 | 14 | 1 | 15 | 3 |
| 北京君正 | 13 | 1 | 1 | 12 | 15 | 14 | 11 | 8 | 15 | 1 | 4 |
| 信维通信 | 5 | 7 | 4 | 7 | 13 | 15 | 4 | 9 | 2 | 11 | 5 |
| 英飞拓 | 3 | 5 | 14 | 6 | 10 | 13 | 13 | 6 | 9 | 14 | 6 |
| 海康威视 | 10 | 2 | 11 | 9 | 14 | 3 | 12 | 12 | 11 | 6 | 7 |
| 莱宝高科 | 15 | 3 | 3 | 1 | 3 | 8 | 9 | 7 | 4 | 13 | 8 |
| 长信科技 | 12 | 6 | 9 | 2 | 11 | 10 | 6 | 4 | 3 | 9 | 9 |
| 兴森科技 | 11 | 11 | 8 | 8 | 7 | 2 | 8 | 2 | 8 | 10 | 10 |
| 瑞凌股份 | 7 | 4 | 13 | 13 | 9 | 7 | 15 | 5 | 14 | 3 | 11 |
| 台基股份 | 9 | 8 | 12 | 3 | 1 | 9 | 7 | 11 | 13 | 8 | 12 |
| 福星晓程 | 6 | 9 | 15 | 4 | 5 | 4 | 10 | 13 | 12 | 5 | 13 |
| 七星电子 | 8 | 15 | 2 | 15 | 2 | 1 | 1 | 15 | 6 | 4 | 14 |
| 华映科技 | 14 | 12 | 5 | 11 | 6 | 5 | 3 | 1 | 5 | 7 | 15 |

**表 7　　133 个公司数据预处理后排名结果**

| 公　司 | 盈利能力 | 销售能力 | 变现能力 | 风险控制 | 营运能力 | 偿债能力 | 现金周转 | 固资生产 | 利润增长 | 资产管理 | 综合排名 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 国民技术 | 1 | 13 | 6 | 5 | 9 | 12 | 4 | 5 | 4 | 10 | 1 |
| 中瑞思创 | 9 | 6 | 4 | 1 | 1 | 3 | 3 | 14 | 14 | 14 | 2 |
| 乾照光电 | 10 | 4 | 5 | 12 | 10 | 6 | 5 | 10 | 11 | 9 | 3 |
| 北京君正 | 13 | 2 | 8 | 2 | 2 | 5 | 1 | 15 | 15 | 15 | 4 |
| 信维通信 | 15 | 1 | 1 | 14 | 3 | 14 | 2 | 1 | 12 | 1 | 5 |
| 英飞拓 | 4 | 14 | 14 | 8 | 11 | 1 | 15 | 3 | 1 | 8 | 6 |
| 海康威视 | 12 | 5 | 7 | 4 | 15 | 2 | 9 | 8 | 9 | 12 | 7 |
| 莱宝高科 | 7 | 10 | 12 | 9 | 13 | 11 | 10 | 4 | 7 | 2 | 8 |
| 长信科技 | 5 | 8 | 11 | 11 | 6 | 13 | 6 | 9 | 10 | 11 | 9 |
| 兴森科技 | 2 | 11 | 2 | 7 | 5 | 15 | 8 | 12 | 6 | 5 | 10 |
| 瑞凌股份 | 3 | 9 | 15 | 13 | 14 | 9 | 12 | 11 | 13 | 13 | 11 |
| 台基股份 | 14 | 3 | 9 | 3 | 12 | 10 | 7 | 2 | 8 | 3 | 12 |
| 福星晓程 | 6 | 15 | 13 | 6 | 7 | 4 | 14 | 6 | 2 | 6 | 13 |
| 七星电子 | 11 | 7 | 3 | 15 | 4 | 7 | 11 | 13 | 5 | 4 | 14 |
| 华映科技 | 8 | 12 | 10 | 10 | 8 | 8 | 13 | 7 | 3 | 7 | 15 |

两表对比可知，原始数据处理前后因子分析其排名结果发生了较大的改变。七星电子和海康威视排名都分别上升至第二、三位，且新进入前 15 名的有 7 家公司，值得注意的是，之前有缺失值而未能参与排名的两家公司利达光电和安居宝在经过缺失值填补后又重新回到了排名中，且综合排名分别为第四和第九名。值得注意的是利达光电、*ST 夏新和安居宝在填补之前的缺失指标分别有 11 个、8 个和 6 个，分别约

占总指标数的29.73%、21.62%和16.22%，特别是利达光电公司的指标缺失值相对较多，接近总指标数的三分之一，此时如果缺失值填补方法不够合理的话，对其较多的指标缺失值填补后将与公司真实的绩效和财务状况不相符合，进而参与的各项排名与综合评价也将失去意义。因此，对于指标缺失值较多的公司是否适合填补并参与因子分析还需进一步研究。

下面尝试剔除指标缺失率29.73%的利达光电，其KMO值为0.698，再剔除*ST夏新后KMO值为0.734，如果将缺失率16.22%的安居宝也剔除则KMO值为0.725。可见对本文所用的缺失数据集，将样本公司的指标缺失率控制在20%以内是合适的，此时KMO值最高为0.734，更适合因子分析，其结果将会更合理、符合实际。下面剔除利达光电和*ST夏新公司得到新的样本数据131个公司，然后对其预处理后进行因子分析，由于因子分析过程类似，故只给出最后的排名结果见表8，注意此时公因子有9个。

**表8　　131个公司数据预处理后排名结果**

| 公　司 | 盈利能力 | 销售能力 | 现金周转 | 风险控制 | 营运能力 | 偿债能力 | 固资生产 | 资产管理 | 利润增长 | 综合排名 |
|---|---|---|---|---|---|---|---|---|---|---|
| 国民技术 | 1 | 15 | 9 | 5 | 11 | 12 | 5 | 6 | 4 | 1 |
| 七星电子 | 8 | 13 | 2 | 2 | 1 | 3 | 15 | 14 | 14 | 2 |
| 北京君正 | 15 | 1 | 1 | 13 | 9 | 14 | 1 | 8 | 15 | 3 |
| 海康威视 | 9 | 3 | 5 | 12 | 13 | 4 | 8 | 9 | 13 | 4 |
| 中瑞思创 | 2 | 14 | 3 | 7 | 12 | 15 | 13 | 2 | 10 | 5 |
| 瑞凌股份 | 6 | 11 | 12 | 10 | 14 | 9 | 3 | 5 | 3 | 6 |
| 东方日升 | 5 | 7 | 15 | 9 | 4 | 1 | 7 | 4 | 1 | 7 |
| 大华股份 | 11 | 4 | 11 | 6 | 15 | 2 | 9 | 13 | 8 | 8 |
| 安居宝 | 4 | 12 | 10 | 11 | 2 | 13 | 6 | 11 | 6 | 9 |
| 鸿利光电 | 13 | 5 | 4 | 3 | 3 | 10 | 4 | 7 | 2 | 10 |
| 兴森科技 | 10 | 9 | 7 | 15 | 10 | 6 | 14 | 1 | 9 | 11 |
| 金运激光 | 14 | 2 | 8 | 4 | 6 | 8 | 2 | 12 | 7 | 12 |
| 正海磁材 | 12 | 6 | 6 | 1 | 8 | 11 | 10 | 10 | 11 | 13 |
| 福星晓程 | 3 | 8 | 13 | 14 | 7 | 7 | 12 | 15 | 12 | 14 |
| 超日太阳 | 7 | 10 | 14 | 8 | 5 | 5 | 11 | 3 | 5 | 15 |

由表6与表8对比可知，排名结果发生了较大的改变。七星电子、北京君正和海康威视分别上升至第二、三、四名，新进入前15名的有7家公司，其中之前有缺失值的安居宝经过填补后排在第9名。

2010年是“十一五”的最后一年，宏观经济在经历了2009年的动荡形势后，2010年市场逐步回暖，在国家振兴装备业、发展新能源产业等系列鼓励政策的推动下，上市电子企业相关业务领域进入新一轮增长周期。

以数据预处理后排名变动较大的三家公司为例，综合排名上升至第2名的七星电子经营业绩在2010年度实现了快速增长，其实现营业总收入81 025.06万元，比上年同期增长37.94%；其中主营业务完成79 316.76万元，比上年同期增长36.58%；实现利润总额11 754.97万元，比上年同期增长61.89%；归属于股东的净利润为8 109.40万元，比上年同期增长63.47%。另外，对比表6与表8，七星电子经过合理的指标正向化后，其风险控制因子由原来的15名上升至第2名，偿债能力因子也由原来第8名上升至第3名，固资生产因子由第4名降到第15名，而这些因子中包含所有的适中指标和逆指标。

新进入前15名综合排名第7的是东方日升。金融危机过后，管理层面对全球市场发展的变化情况，紧抓国际光伏市场的发展契机，公司的研发、生产、市场开拓能力稳步增强。2010年度，公司实现营业收入237 485.72万元，比上年同期增长182.13%；营业利润31 179.53万元，比上年同期增长144.94%，归属于母公司净利润27 512.85万元，比上年同期增长136.92%。公司业绩的大幅增长主要得益于先期使用自有资金预先投入的部分募投项目的达产、光伏行业的复苏以及欧洲各国逐步下调补贴政策引发的装机热。由表8可知，东方日升利润增长因子和偿债能力因子都上升至第1名，盈利能力因子也上升至第5名，公司未来将会继续保持稳定增长趋势。

填补缺失值后新进入前15名排名第9的安居宝2010年营业收入243 378 065.39元，比2009年增长12.24%，利润总额70 153 184.25元，比2009年增长36.37%。公司资产总额1 091 641 255.85元，比2009年增长435.28%，归属于上市公司股东的所有者权益1 015 251 358.69元，比2009年增长646.31%，归属于上市公司股东的每股净资产14.10元，比2009年增长459.52%。由表8可知，其盈利能力因子和利润增长因子也分别上升至第4和第6名。

可见，在进行因子分析前，对指标数据进行合理的样本缺失值填补、指标正向化和量纲化预处理，才能使综合评价结果趋于合理，更客观有效、符合实际。

## 四、结论

本文基于实证分析重点介绍了合理的样本缺失值填补方法和正确的指标正向化公式，并通过数据预处理前后的因子分析排名结果的对比分析，证实了对指标数据合理地预处理后其评价结果更合理、客观有效。

本文也存在一定的局限性：一是以37个财务指标作为业绩评价指标体系，没考虑非财务指标，在指标体系构建方法上还欠缺考虑。二是部分上市公司的财务数据未能全面反映公司的实际运营情况，这将直接影响业绩评价结果，而且股市受市场外部因素影响比较大。

【注】本文系教育部人文社会科学研究规划及专项任务项目“证券市场动态相关性测度的拓展及应用研究”（编号：11XJC910001）的研究成果。

### 主要参考文献

**1.** 范坤，冯长焕.基于因子分析的目标极性模型研究——针对上市电子企业的实证分析. 井冈山大学学报（自然科学版），2012；6

**2.** 钟齐. 基于因子分析的我国上市银行经营效率比较分析.统计与信息论坛，2012；27