

基于PCA和Logistic回归的A股36家上市券商排名的研究

王宇茜 彭连 夏乙月

四川大学经济学院

DOI:10.32629/ej.v3i3.443

[摘要] 金融业和计算机技术的快速发展,为获取超额或稳定投资收益的量化选股研究在我国兴起创造了条件。本文从wind、csmar数据库中选取A股36家上市券商的2017年第四季度到2019年第三季度的市盈率等25个财务指标数据,以主成分分析的80%累计贡献率为标准构建12个综合指标,采用Logistic回归最终选取9个综合主成分指标,建立量化模型,对A股36家上市券商的投资价值进行了综合排名,最终排名前三依次是中信建投、长城证券和华林证券。

[关键词] PCA; Logistic回归; 券商; 财务指标

1 研究背景和意义

随着我国经济的发展,人们的生活水平逐渐提高,可支配收入不断增加,投资意识慢慢增强,股票在我国居民投资占较大份额,如何选择优质股票是社会大众十分感兴趣的议题。当前,计算机的发展为选股的大数据定量分析奠定了基础,因此基于主成分分析的Logistic回归构建量化模型进行股价预测和投资指导便具有可行性。

与传统的定性投资分析方法相比,量化投资分析方法具有很大的优势,主要为精力无限、概率取胜和纪律性。文章旨在通过建立科学的投资方法帮助投资者进行股票选择,从而有效地分配资本及实现其需要的长期合理投资预期回报。因此,本文的研究具有现实和应用价值。

2 研究过程

2.1 数据选取

本文选取了A股上市的华西证券、海通证券、东方证券等36家券商。通过搜集wind、csmar上36家券商的2017年第四季度始至2019年第三季度的季报财报财务数据为研究样本,并且剔除了一些缺失的无效数据,共161个时间变量。选取25个财务状况指标作为自变量,因变量为二元变量,若2019年第三季度财报发布时间股票相对于对照时间上涨为1,股票价格下跌为0。

选取资产负债率、应收账款周转率、营业利润、现金与利润总额比、营业收入增长率、每股收益、营业总成本、加权平均资产收益率、市盈率、市销率、市净率、PEG、ROE、ROA、每股盈余公积、项目毛利率、流动比率、速动比率、BPS、每股现金流量净额、销售利润率、税项、总资产周转率、资本项目规模维持率、管理费用25个财务指标,依次记为。

每个指标下面均设有7个具有从“2017年第四季度到2019年第三季度”的以季度为时间跨度的小指标:对于 $n=1, 2, 3, \dots, 25$,有 $X_{n-2017-4}$: (2017年第四季度)、 $X_{n-2018-1}$ (2018年第一季度) \dots $X_{n-2019-2}$ (2019年第二季度)、 $X_{n-2019-3}$ (2019年第三季度)。

2.2 方法及模型描述

2.2.1 主成分分析

(1) 数据标准化

本文采用的是Z-score数据标准化方法即对原 p 个具有相关性的随机变量 $x_1, x_2, x_3, \dots, x_p$ 进行数据标准化,即将原始数据:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

标准化为

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

标准化后的指标数据平均值为0,标准差为1,其中。

$$a_{ij} = (x_{ij} - \bar{x}_j) / \sqrt{\frac{(x_{ij} - \bar{x}_j)^2}{n}}, i=1, 2, \dots, n; \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, j=1, 2, \dots, p$$

对各变量标准化后,发现营业总成本和管理费用无法进行标准化,舍去这两个变量,考虑剩下23个变量的贡献程度。

(2) 求标准化后指标数据的相关矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{p3} \end{bmatrix}$$

其中,

$$r_{jk} = \frac{\sum_{i=1}^n (a_{ij} - \bar{a}_j)(a_{ik} - \bar{a}_k)}{\sqrt{\sum_{i=1}^n (a_{ij} - \bar{a}_j)^2 \sum_{i=1}^n (a_{ik} - \bar{a}_k)^2}}, i=1, 2, \dots, n$$

$$j, k = 1, 2, \dots, p, \text{ 而 } \bar{a}_j = \frac{1}{n} \sum_{i=1}^n a_{ij}$$

(3) 求相关矩阵的特征向量和特征值

$$\text{凭据 } |R - \lambda_j E| = 0, j = 1, 2, \dots, p$$

获得特征值 $\lambda_j, j = 1, 2, \dots, p$ 为特征矩阵；凭据

$$(R - \lambda_j E)u_k = 0, j, k = 1, 2, \dots, p$$

获得特征向量 $u_k, k = 1, 2, \dots, p$ 其中

$\lambda_j, j = 1, 2, \dots, p$ 是之前求得特征值。

(4) 得主成分

特征值大小进行排列，

$$\lambda_1 > \lambda_2 > \dots > \lambda_p \text{ 凭据}$$

$$\sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i \geq 85\% \text{ 确定主成分数}$$

目 k ，特征值的排列顺序与特征向量 u_1, u_2, \dots, u_k 相对应，再凭据之前所

确定的 k 的值，确定特征向量，将原 p 个标准化后的指标变量数据和特征向量相乘得到 k 个综合的主成分指标数据，方程和实证结果如下：

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

$$\times \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_k \end{bmatrix}$$

(5) 求综合得分

将 k 个综合的主成分指标数据与上面求出的对应主成分贡献率相乘再求和即为综合得分。通过主成分分析选出了 12 个累计贡献度到 80% 的主成分以及各成分的得分。

综合主成分计算公式如下：

$$\text{Comp1} = 0.0055\text{std}x_{120174} + 0.0027\text{std}x_{20181} + \dots + 0.0424\text{std}x_{2420184} + 0.0682\text{std}x_{20192}$$

$$\text{Comp2} = -0.1177\text{std}x_{120174} - 0.1316\text{std}x_{20181} + \dots - 0.0490\text{std}x_{2420184} - 0.0238\text{std}x_{20192}$$

$$\dots$$

$$\text{Comp12} = 0.1493\text{std}x_{120174} + 0.1484\text{std}x_{20181} + \dots + 0.0545\text{std}x_{2420184} + 0.0418\text{std}x_{20192}$$

2.2.2 Logistic回归

本文探究的上市券商是否值得投资的问题，研究变量的取值只有两个，股票价格上涨(为1)和下跌(为0)，对于因变量值的选取，考虑了两个时间点，2019年9月相对于2017年12月的股票价格的涨跌，其中7个证券公司在2017年12月之后才上市，2017年第四季度的股票价格缺失，使用其他日期的收盘价替代。华西证券用2018年2月的收盘价、长城证券用2018年10月、华林证券用2019年1月、中信建投用2018年6月、天风证券用2018年10月、南京证券用2018年6月、红塔证券用2019年7月的收盘价替代。

通过Logistic回归拟合出如下方程：

$$\text{logit}(p) = \ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$$

其中 $p(Y=1)$ 是在因子 X_1, X_2, \dots, X_m 下券商财务指标不良的概率，则 $1-p(Y=1)$ 是财务指标良好的概率， X_1, X_2, \dots, X_m 是由主成分分析得到的贡献率超过80%的因子， $\beta_0, \beta_1, \dots, \beta_m$ 是待估参数。

以获得的主成分作为自变量，2019年第三季度上市券商的股票涨跌为因变量，对所有指标进行Logistic回归，并选择通过了显著性检验的自变量建立通过检验的回归方程：

$$\text{logit}(p) = \ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad n=1, \dots, 12$$

$X_i (i=1, \dots, 12)$ 为通过主成分分析得到的衡量上市券商投资价值的主成分。通过Logistic回归分析，发现有一些变量的 p 值并不显著，于是删去检验结果

不显著的 $p > 0.5$ 的变量，再进行一次 Logistic 回归。

2.2.3 综合评分模型构建

将A股上市的36家券商样本的25项大类评价指标通过标准化进行无量纲化处理后，通过PCA和Logistic回归最终选取出 Comp. 1, Comp. 3, Comp. 4, Comp. 6, Comp. 7, Comp. 8, Comp. 10, Comp. 11, Comp. 12 的值，运用综合主成分得分法计算各公司的综合主成分得分，并进行排序，按照排名从中选择出具有投资价值的券商。其中，综合评分模型如下所示：

$$F = -2.916 + 0.210 \cdot \text{Comp.1} + 0.282 \cdot \text{Comp.3} - 0.203 \cdot \text{Comp.4} + 0.502 \cdot \text{Comp.6} + 0.384 \cdot \text{Comp.7} + 0.429 \cdot \text{Comp.8} - 0.395 \cdot \text{Comp.10} - 1.292 \cdot \text{Comp.11} + 0.903 \cdot \text{Comp.12}$$

2.3 实证结果

通过综合评分模型计算出券商投资价值得分及排名，排名前十的证券如下表所示：

表1 前十券商得分

名称	股票代码	得分	名次
中信建投	601066	7.01639	1
长城证券	002939	5.672076	2
华林证券	002945	5.371765	3
海通证券	600837	4.509564	4
申万宏源	000166	4.217663	5
天风证券	601162	4.205426	6
中信证券	600030	4.132206	7
红塔证券	601236	3.807987	8
国海证券	000750	3.590505	9
华泰证券	601688	3.380736	10

通过对36家A股市场券商投资价值进行了综合分析实证检验，得出了36家公司的投资价值的综合得分排名，其中中信建投、长城证券、华林证券这三家券商相对投资价值更高，建议投资者关注这些公司的股票，中长期持有或者可以适量增仓。

(1) 中信建投，从数据中可以看到2017年—2019年间盈利和成长能力很高；在偿债能力和运营管理能力方面，排名也较高；虽然目前净资产并不占优势，但是如果进一步补充净资产，将进一步提高投资价值。

(2) 长城证券，通过数据可以看到在

数字化财会专业人才培养模式探究

罗娟

云南经济管理学院

DOI:10.32629/ej.v3i3.467

[摘要] 随着数字化的不断发展,很多高职院校在对大学生进行财会专业知识教学中,为了提高学生的财会专业成绩以及为学生今后的实际工作中能够提高自身的工作效率,对财会专业人才培养模式进行了改进。制定教学目标的目的是为了给社会培养更多的专业人才,对教学目标进行不断的改进可以为社会培养适应社会发展的专业人才,这也是我国当前的发展形势。为了更好的培养数字化财会专业人才,本文希望通过研究找到更加适合财会专业人才培养模式,为各大院校培养财会专业人才提供一定的借鉴意义。

[关键词] 数字化; 财会专业人才; 培养模式

1 数字化时代下改进财会专业人才培养模式的重要性

数字化时代的到来为财会专业带来了一定机遇的同时也带来了新的挑战。数字化时代下,不断提高高职院校财会专业学生就业情况已经成为了院校考核的主要指标。高职院校要想做好财会专业人才培养,要注意以下几方面:首先,各大企业对财会专业人才的需求量日益增大,每年我国的新增企业达到30万,且如果每个企业需求财会专业人才的数量为1个,那每年就有30万学生有就业机会,但是并不是每一个企业都需要财会人员,

因此,这对财会专业人才的就业率产生了一定的影响。其次,要更多地关注实际技能,随着数字化时代的到来,各企业对财会专业人才的要求不断增高,企业主要关注的重点为人才、技能,而不是求职者本人,随着很多新的会计软件的诞生,很多企业通常将学生对是否会应用新的软件作为衡量学生功能能力的标准。因此,在数字化时代下改进财会专业人才培养模式是非常重要的。

2 财会职业教育的现状

随着高等职业教育的发展,财会职业教育不单单包含于教育的范畴中,同

时在经济领域的人才培养过程中也发挥着重要作用。财会属于一门专业性与应用性较强的专业,人才培养的目标不仅是培养会计职业素养,更需要了解和掌握会计信息系统、财务报表分析、会计业务流程知识等,更为重要的是培养学生较强的金融管理的方法技巧以及具备扎实的财务金融分析能力。但是现阶段高校财会职业教育中依旧存在很多问题,如,财会职业教育脱离实际,毕业生实践能力差,特别是上岗前实际操作能力差,满足不了实际工作的需要。造成这些问题的重要原因是:市场需求不清晰,定位

抗风险能力方面,排名靠前,并且各项指标各方面相对均衡。

(3)华林证券,通过数据可以看到盈利和成长能力仅次于中信建投,但是偿债能力和运营管理能力方面为负数,如果运营管理能力得到改善,将进一步提高投资价值;在抗风险能力方面,表现相对较好。

3 结论与建议

本文从理论和实证相结合的角度对证券选择进行了研究。通过将证券分析指标和行业分析指标相结合,建立了由25个可量化指标出发的研究证券投资价值的综合评判体系,形成了对于证券投资价值分析的基本分析模型。最终对A

股36家证券进行了综合排名,并从理论角度对排名进行了解释,符合实际结果。

综上,建议投资者在投资券商股时,要充分考虑到证券行业的特点,在结合财务报表数据进行投资时,要采用科学的方法进行系统分析,切不可只关注其中几个具体的指标。

[参考文献]

- [1]兰强太.基于主成分分析和BP神经网络算法的综合选股实证研究[D].暨南大学,2017.
- [2]邹亚宝,梁红漫.基于logistic回归的我国上市公司信用评级模型研究[J].西南金融,2013(3):14-19.
- [3]于立勇,詹捷辉.基于Logistic回

归分析的违约概率预测研究[J].财经研究,2004(09):15-23.

[4]齐治平,余妙志.Logistic模型在上市公司财务状况评价中的运用[J].东北财经大学学报,2002(01):60-63.

[5]李靖华,郭耀煌.主成分分析用于多指标评价的方法研究——主成分评价[J].管理工程学报,2002(01):39-43.

作者简介:

王宇茜(1998--),女,汉族,河北保定人,本科,研究方向:金融工程。

彭连(1999--),女,汉族,贵州遵义人,本科,研究方向:财政学。

夏乙月(1999--),女,汉族,四川仁寿人,本科,研究方向:金融工程。