

基于改进因子分析的投资组合问题的研究

赵建喜, 杨永愉, 赵丽娜 *

(北京化工大学 理学院, 北京 100029)

摘 要: 提出了一种新的投资组合优化方案. 由于部分上市公司财务指标数据不全, 采用证券日收益率作为基本的处理数据; 考虑用变异系数法对因子分析法中标准化处理进行适当改进; 将证券看作指标变量, 日收益率看作样本进行分析; 结合 Markowitz M-V 模型, 计算出证券的最优投资比例. 以中国股票市场为例, 确定出了 32 支最有投资价值的股票和它们的最优投资比例.

关键词: 变异系数法; 因子分析; 证券市场; 投资组合; Markowitz M-V 模型

1 引言

我国宏观经济的持续快速增长, 为中国证券市场的发展创造了有利条件. 证券投资基金在资本市场中占有重要地位, 研究中国证券市场投资组合问题有利于促进中国资本市场的繁荣稳定^[1].

目前已有许多关于证券投资基金量化方面的文章. 杨明媚^[2]、饶彬^[3]和郭飞腾^[4]等提出了应用多元统计方法进行股票投资的方案; Xiaohua Ma 等^[5]提出了基于混合差分进化的基数约束下的投资组合优化问题; Bruce I. Jacobs 等^[6]将杠杆规避法应用到投资组合理论中. 投资组合理论往往涉及到多指标处理的问题, 而多元统计分析在多指标处理方面优势明显, 其中具有代表性的方法是因子分析.

2 理论模型

2.1 用变异系数法改进的因子分析

因子分析的过程中需要将数据标准化, 标准化后的数据使得各指标变量的水平相同, 如 z-score 标准化后, 各指标数据均值为 0, 标准差为 1; 极差标准化后, 各指标数据均在 [0,1] 中. 但事实上, 变量水平不一定相同. 指标数据的离散化程度将是影响样品选择的重要因素. 所以, 考虑用变异系数法挖掘各指标的相对重要性大小, 对指标进行加权处理, 再用因子分析法处理加权后的数据^[7].

由于部分上市公司的财务指标数据不全, 而其日收盘价数据全面且容易获取, 但是直接将不同证券的日收盘价进行对比分析, 没有太大价值, 通常的做法是把日收盘价换算成日收益率 (见式 (1)), 但是我们只有收益率这一个指标. 如果将单一指标看作变量, 证券看作样本进行改进因子分析, 显然不合适. 为了解决这个问题, 我们将证券看作变量, 日收益率看作样本

收稿日期: 2014-03-31

资助项目: 国家自然科学基金 (11301021)

* 通讯作者

进行分析. 根据方差贡献率的大致比值, 确定在每个因子对应的证券中挑选几支最有投资价值的证券.

下面是具体的计算步骤:

1) 根据分散投资可以减小风险的原则, 选定几个市场行情看好的证券板块, 确定所要投资的证券总数, 计算出每个板块中大约挑选几支证券.

2) 收集每个板块中各证券一段时期的日收盘价, 根据式 (1) 计算出相应日收益率. 理性投资者都希望你获得收益, 故去除预期收益为负的证券.

$$x_{ij} = \ln \left(\frac{p_{ij}}{p_{(i-1)j}} \right) \quad (1)$$

其中, p_{ij} 是证券 j 的 i 日收盘价, x_{ij} 是证券 j 的 i 日收益率.

将基于变异系数法的改进因子分析分别用于选好的每个证券板块中, 把证券看作指标变量, 日收益率看作样本进行分析, 组成资料矩阵式 (2).

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = (X_1, X_2, \cdots, X_p) \quad (2)$$

其中, n 是天数, p 是证券数.

3) 对各证券日收益率进行极差标准化:

$$y_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}$$

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

4) 计算各证券变异系数:

$$V_j = s_j / \mu_j$$

其中, $\mu_j = \frac{1}{n} \sum_{i=1}^n y_{ij}$, $s_j = \left(\frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \mu_j)^2 \right)^{\frac{1}{2}}$.

5) 求各证券权重:

$$w_j = V_j / \sum_{j=1}^p V_j$$

6) 将权重赋给各证券数据:

$$y'_{ij} = w_j y_{ij}$$

7) 求样本相关系数矩阵 $R = (\gamma_{ij})_{p \times p}$, 其中,

$$\gamma_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}}, \quad s_{ij} = \frac{1}{n} \sum_{\alpha=1}^n (y'_{\alpha i} - \bar{y}_i)(y'_{\alpha j} - \bar{y}_j)$$

8) 求相关系数矩阵 R 的特征值:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$$

以及相应的因子载荷矩阵 A .

9) 选择主要的因子: 当前 k 个因子的方差累计贡献率 P (见式 (3)) 达到大部分时, 说明前 k 个因子包含了主要信息, 就保留前 k 个因子及相应的因子载荷向量. 有些文献 [8] 建议

累计贡献率达到 85% 以上, 其实这只是一个大体的说法, 具体选几个, 还要根据实际情况而定.

$$P = \sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i$$

(3)

10) 对因子载荷矩阵 A 施行方差最大正交旋转.

11) 挑选影响因子的主要证券: 根据 k 个因子的方差贡献率的大致比值和在该板块要选多少支证券两个条件, 来确定每个因子中挑选几支证券 (比如 m 支). 选出每个因子载荷向量中元素值最大的 m 个元素所对应的证券.

所有板块选出的证券合在一起, 即是我们所要投资的最有价值的证券.

2.2 Markowitz M-V 模型

Markowitz M-V 投资组合模型简述如下 [4]:

目标函数: $\min \sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{cov}(r_i, r_j)$

约束条件:
$$\begin{cases} \sum_{i=1}^n w_i = 1, & w_i \geq 0 \\ \bar{r}_p = \sum_{i=1}^n w_i \bar{r}_i \end{cases}$$

其表明在约束条件下, 各证券投资比例 w_i 为多少时, 可以使组合风险 σ_p^2 达到最小值 [9]. 该模型为非线性规划, 可用 matlab 金融工具箱计算.

3 实证分析

本文以股票市场为例. 选取近来, 市场行情看好的股票板块: 180 资源、180 金融、380 医药、服装及其他、建筑指数和交通设施. 我们共从中选取 32 支股票, 平均每个板块约挑选 5 支股票. 用大智慧软件分别下载各板块每支股票 2012 年 5 月 3 日及之前近半年的收盘价 (每支股票 125 个数据), 去掉数据个数不足的股票, 再算出日收益率均值 (预期收益). 理性投资者都希望投资能获得收益, 故删除收益率为负的股票.

分析 180 资源板块. 把该板块的股票看作指标变量, 日收益率看作样本进行分析. 用变异系数法计算出各股票权重 (见表 1).

表 1 180 资源板块各股票权重

股票	权重	股票	权重
DR 包钢稀	0.0541	厦门钨业	0.0522
驰宏锌锗	0.0535	阳泉煤业	0.0581
广晟有色	0.0584	中国神华	0.0685
吉恩镍业	0.0554	中国石化	0.0575
江西铜业	0.0582	中国石油	0.0427
兰花科创	0.0659	中金黄金	0.0554
潞安环能	0.0537	中煤能源	0.0600
南山铝业	0.0674	紫金矿业	0.0614
盘江股份	0.0775		

从表 1 可以看出, 兰花科创、南山铝业、盘江股份、中国神华、中煤能源和紫金矿业的权重稍微大些, 说明其数据离散程度相对较大, 对 180 资源板块的影响大些. 将表 1 权重乘以相应的股票数据, 继续第七步及以后步骤, 求得相关系数矩阵的特征值、方差贡献率、累计贡献率 (表 2).

表 2 180 资源板块特征值和贡献率

序号	特征值	贡献率	累计贡献率
1	6.174	0.3632	0.3632
2	2.896	0.1703	0.5335
3	1.722	0.1013	0.6348
4	1.126	0.0662	0.7010
5	0.982	0.0578	0.7588

从表 2 可以看出, 前 3 个特征值的累计贡献率是 63.48%, 已占了大部分比例. 虽然和 85% 还有一定差距, 但是从第四个特征值开始, 每个特征值的贡献率都很小. 由于我们在每个板块平均提取约 5 支股票, 因此不值得在因子 4 及以后因子中提取股票. 故我们只考虑前三个因子.

观察表 3, 对因子 1 正面影响较大的股票是驰宏锌锗、潞安环能、盘江股份和厦门钨业等, 在东方财富网查询知, 这三家公司的营业收入相对较低, 故命名低营业收入因子; 对因子 2 正面影响较大的股票是南山铝业、中煤能源和紫金矿业等, 这几家公司的营业收入中等, 故命名中等营业收入因子; 而对因子 3 影响较大的公司既有营业收入很高的 (中国石化和江西铜业) 也有中等的 (阳泉煤业), 故命名中高等营业收入因子. 由前三个因子方差贡献率的大小 (因子 1 方差贡献率是因子 3 的 3 倍多, 是因子 2 的 2 倍多, 我们也近似看作 3 倍, 即比值约为 3: 1: 1) 和在每个板块约挑选约 5 支股票两个条件, 我们决定在因子 1 中挑选 3 支股票, 因子 2 和因子 3 中各挑选 1 支股票.

表 3 180 资源板块旋转后的因子载荷矩阵

股票	因子 1	因子 2	因子 3	股票	因子 1	因子 2	因子 3
DR 包钢稀	-0.1927	0.5310	0.2131	厦门钨业	0.7390	0.1129	0.2911
驰宏锌锗	0.8554	0.1046	0.2237	阳泉煤业	0.4614	0.1554	0.6606
广晟有色	0.2347	0.3058	-0.0071	中国神华	0.8606	-0.0544	0.1849
吉恩镍业	0.3039	0.4081	-0.0875	中国石化	0.2871	-0.1896	0.6556
江西铜业	0.5498	0.0269	0.7389	中国石油	0.6717	-0.1520	0.0823
兰花科创	-0.0450	0.3090	-0.1692	中金黄金	-0.0788	0.1541	-0.0516
潞安环能	0.9043	0.0128	0.2519	中煤能源	0.1642	0.7510	0.0085
南山铝业	0.1687	0.8578	0.0187	紫金矿业	-0.0938	0.6038	0.3788
盘江股份	0.8831	0.1140	0.2205				

由表 3 知, 因子 1 中三个值最大的载荷值所对应的股票是潞安环能、盘江股份和中国神华; 因子 2 中最大的载荷值所对应的股票是南山铝业; 因子 3 中最大的载荷值所对应的股票

是江西铜业. 所以我们在 180 资源板块中选出了以上 5 支股票.

其他板块的分析过程与 180 资源板块相似.

把选出的 32 支股票放入股票池, 用马克维茨 M-V 模型进行分析. 在给定股票组合预期收益 \bar{r}_p 的前提下, 求出最优投资比例.

给定 $\bar{r}_p = 0.0008$, 用 matlab 金融工具箱 quadprog 算出最优投资比例 w_i (见表 4). 求得的 w_i 值多数为零, 此属禁止卖空条件下的正常现象, 表示有许多股票不能被投资或投资机会不能为投资者所用.

表 4 挑选出的 32 支股票的最优投资比例

股票	权重	股票	权重	股票	权重	股票	权重
潞安环能	0	华泰证券	0	雅戈尔	0	东南网架	0.0517
盘江股份	0	中国银行	0.1661	朗姿股份	0.0386	金螳螂	0
中国神华	0	交通银行	0.2151	金飞达	0.0542	厦门空港	0
南山铝业	0	金宇集团	0	浪莎股份	0	唐山港	0.0576
江西铜业	0	中牧股份	0	美尔雅	0	上海机场	0
南京银行	0.0326	现代制药	0	深天地 A	0	皖通高速	0.0830
浦发银行	0	益佰制药	0.0266	深天健	0	山东高速	0
招商银行	0	七匹狼	0.0175	围海股份	0	深基地 B	0.2570

4 结 论

为减小投资风险, 本文采取分散投资的方法进行投资组合. 在此遇到两个问题: 一是如何选择投资对象, 二是如何确定最优投资比例. 为解决第一个问题, 我们用基于变异系数法的改进因子分析挑选最优投资证券, 该法有效解决了证券水平相同的问题. 部分上市公司的财务指标数据不全, 为避免这一现象, 我们利用收益率作为基本数据进行分析. 而且我们将证券作为指标变量, 日收益率作为样本进行分析. 为解决第二个问题, 我们利用 Markowitz M-V 模型确定最优投资比例. 解决这两个问题的过程蕴含了分层优化的思想.

本文从理论和实证两个方面分析了如何选取投资对象和确定投资比例的问题.

参考文献

[1] 王兴运. 中国证券市场投资组合的构建及其风险管理 [D]. 天津: 南开大学硕士学位论文, 2004: 1.
[2] 杨明媚, 李华林. 主成分分析在证券组合投资中的应用 [J]. 统计与信息论坛, 2004, 19(1): 74-77.
[3] 饶彬, 易东云. 主成分分析法在股票评估中的应用 [J]. 统计与决策, 2006(2): 140-142.
[4] 郭飞腾. 基于投资组合理论的股票投资分析 [D]. 上海: 同济大学硕士学位论文, 2008: 1-64.
[5] Ma X H, Gao Y L, Wang B. Portfolio optimization with cardinality constraints based on hybrid differential evolution [J]. AASRI Procedia, 2012(1): 311-317.
[6] Jacobs B, Levy K. Introducing leverage aversion into portfolio theory and practice[J]. The Journal of Portfolio Management, 2013, 39(2): 1-2.
[7] 杨迎辉, 吴昊, 周航. 基于因子分析法的房地产行业可持续发展程度评估 [J]. 数学的实践与认识, 2012, 24(15): 92-99.

- [8] 于秀林, 任雪松. 多元统计分析 [M]. 北京: 中国统计出版社, 2010: 154-166.
- [9] Steinbach M C. Markowitz revisited: mean-variance models in financial portfolio analysis[J]. SIAM Review, 2001, 43(1): 31-85.

Study on the Portfolio Problem Based on an Improved Factor Analysis

ZHAO Jian-xi, YANG Yong-yu, ZHAO Li-na

(School of Science, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: This paper proposes a new portfolio optimization scheme. Because some listed companies have incomplete data of financial indicators, daily return rates of securities are regarded as the basic processing data. This paper considers to use the variation coefficient method to improve the standardization of factor analysis properly. Securities are regarded as indicator variables and daily return rates are regarded as samples. Coupled with Markowitz M-V model, optimal investment proportion of securities can be calculated. This paper takes Chinese stock market an example. The scheme determines 32 most investment valuable stocks and their optimal investment proportion.

Keywords: variation coefficient method; factor analysis; securities market; portfolio; Markowitz M - V model