

证券投资的主成分因素分析方法

田絮资

(宝鸡文理学院 计算机系, 陕西 宝鸡 721007)

摘要: 为了研究证券投资的规律, 利用主成分因素分析方法对影响证券投资的因素进行了分析。分析结果表明, 经济因素和证券市场技术因素是影响证券投资的主要因素。结果对实际有一定的指导意义, 且为证券市场的研究提供了有效的分析方法和可靠的理论依据。

关键词: 证券投资; 主成分因素; 应用数学

中图分类号: O29 **文献标识码:** A **文章编号:** 1000-274 X (2003)03-0281-04

证券投资者, 无论是个人投资者还是机构投资者, 其根本目的在于获得最大的预期收益。为此, 投资者可以把全部资金投在一种或少数几种收益高的证券上, 以期获得最大限度的收益。然而, 影响证券投资的因素很多, 有经济因素、证券市场技术因素、行业因素和社会心理因素等。与这些因素有关的如: 宏观经济政策与态势、公司实力与业绩、证券市场规律、行业现状和前景、盲目跟风心理、对前景预测的态度等。因素分析是一种统计技术, 广泛应用于心理学、教育学、社会学, 以及经济学等领域。它是从众多的可观测的“变量”中推论出少数的“因素”, 用最少的“因素”来概括和解释大量的观测事实。本文通过证券投资中多方面的统计数据, 利用因素分析法对影响证券投资的因素进行了分析。虽然, 影响证券投资收益的因素很多, 但是起主要作用的只有两种因素。

1 主成分因素分析

主成分分析和因素分析模型(简称为主成分因素模型)是一种多元回归模型, 针对研究对象, 它通过引入一组与之有关的公共因素, 可较好地分析各个因素之间的关系, 以及每种因素对研究对象的影响程度。

若 $X = (x_1, x_2, \dots, x_n)^T$ 为 n 的维随机变量, x_i 的取值是对不同个体的观测值, x_i 的均值和方差分别为 \bar{x}_i 和 σ_i^2 。引入一个新变量 $Z = (z_1, z_2, \dots, z_n)^T$ 为

X 的标准化变量。

$$z_i = \frac{x_i - \bar{x}_i}{\sigma_i}, i = 1, 2, \dots, n, \quad (1)$$

则可建立分析模型

$$\begin{cases} z_1 = p_{11}F_1 + p_{12}F_2 + \dots + p_{1m}F_m + Cc_1\nu_1, \\ z_2 = p_{21}F_1 + p_{22}F_2 + \dots + p_{2m}F_m + Cc_2\nu_2, \\ \dots\dots\dots \\ z_n = p_{n1}F_1 + p_{n2}F_2 + \dots + p_{nm}F_m + Cc_n\nu_n. \end{cases} \quad (2)$$

式中 $F = (F_1, F_2, \dots, F_m)^T$ 为公共因素, 用来表示 $z_i (i \in \{1, 2, \dots, n\})$ 的线性组合。 $V = (\nu_1, \nu_2, \dots, \nu_n)^T$ 为特殊因素, ν_i 只与 x_i 有对应关系 ($i = 1, 2, \dots, n$)。 P 和 C 分别为 F 和 V 的负荷矩阵, 那么

$$Z = PF + CV,$$

矩阵

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nm} \end{bmatrix}, C = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_n \end{bmatrix}.$$

不妨假定, 所有公共因素和特殊因素均相互独立, 且每个因素服从 $N(0, 1)$ 分布。 F, V 和 Z 的相关系数为

$$r(z_i, F_j) = \frac{\text{cov}(z_i, F_j)}{\sqrt{D(z_i)D(F_j)}} = \text{cov}(z_i, F_j) = p_{ij}, \quad (i \in \{1, 2, \dots, n\}; j \in \{1, 2, \dots, m\}). \quad (3)$$

收稿日期: 2003-02-04

基金项目: 陕西省自然科学基金资助项目(2001G11); 陕西省教育厅科学研究基金资助项目(00JK021)

作者简介: 田絮资(1954-), 女, 湖南株州人, 宝鸡文理学院副教授, 从事应用数学研究。

$$r(z_i, v_j) = \begin{cases} c_i (i \neq j) \\ 0 (i = j) \end{cases},$$

$$(i \in \{1, 2, \dots, n\}; j \in \{1, 2, \dots, m\}),$$

$$r(z_i, z_j) = \sum_{k=1}^m p_{ik} p_{jk}, i \neq j,$$

$$r(z_i, z_i) = \sum_{k=1}^m p_{ik}^2 + c_i^2 = h_i^2 + c_i^2,$$

$$i = 1, 2, \dots, n. \quad (4)$$

式中 h_i^2 称为 z_i 的公共因素方差,它反映了全部公共因素对 z_i 的贡献。 h_i^2 越接近于 1,则公共因素越能代表 z_i ,从而越能精确地成为 z_i 的线性组合。当 $h_i^2 = 1$ 时, z_i 完全可由公共因素线性组合表示。

用 R 表示 Z 中各变量间的相关矩阵,则 $R = PP^T + CC^T$ 。

若将 F_j 对所有变量的方差贡献定义为 $U_j (j = 1, 2, \dots, m)$, 则

$$U_j = \sum_{i=1}^n p_{ij}^2, j = 1, 2, \dots, m.$$

再由此定义方差贡献率 T_j

$$T_j = \frac{U_j}{\sum_{k=1}^m U_k}.$$

那么, T_j 可作为衡量 F_j 重要性的指标。将公共因素中方差贡献率较大的因子定义为“主成分因素”。于是,可以根据贡献率的大小选择主成分因素,并且用它来对变量 X 进行线性组合。这就是主成分因素模型的原理。

以“方差贡献率最大”作为主成分因素的选择和确定负荷矩阵 P 的原则。第一主成分因素 F_1 应是对各变量的方差贡献最大者,方差贡献 U_1 即相关矩阵 R 中的最大特征值, F_1 上的负荷 $P_1 = (p_{11}, p_{12}, \dots, p_{1n})$, 由非线性规划

$$\begin{cases} \max_{a_{1i}} U_1 = \sum_{i=1}^n p_{1i}^2, \\ s. t. \quad r_{ij} = \sum_{k=1}^m p_{ik} p_{jk}, \end{cases}$$

$(i \in \{1, 2, \dots, n\}; j \in \{1, 2, \dots, m\})$ 。

r_{ij} 为相关矩阵 R 的元素。

利用 Lagrange 乘子法,令

$$L = U_1 - \sum_{i=1}^n \sum_{j=1}^m s_{ij} (\sum_{k=1}^m p_{ik} p_{jk} - r_{ij}).$$

式中 s_{ij} 是 Lagrange 乘子 ($s_{ij} = s_{ji}$)。 L 存在极值的必要条件为 $(R - U_1 E)P_1 = 0$, 其中 E 为 n 阶单位矩阵, $P_1 = (p_{11}, p_{12}, \dots, p_{1n})^T$ 。记 R 的最大特征值为 λ_1 , 相应的特征向量为 $\alpha_1 = (\alpha_{11}, \alpha_{21}, \dots, \alpha_{n1})^T$ 。利用正交

变换,可将实对称矩阵 R 对角化,即

$$\alpha^T R \alpha = \Lambda, \quad R = \alpha \Lambda \alpha^T.$$

Λ 为对角矩阵。若将 R 的特征值 $\lambda_i (i = 1, 2, \dots, n)$ 由大到小排列,则 $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 在不考虑特殊因素 V 时, $R = PP^T = \alpha \Lambda \alpha^T$, 因此 $P = \alpha \sqrt{\Lambda}$, 显然 $p_{ij} = \alpha_{ij} \sqrt{\lambda_j}$ 。式中

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} \cdots & \alpha_{nn} \end{bmatrix},$$

可得 F_1 的负荷 $P_1 = \alpha_1 \sqrt{\lambda_1}$ 。

对剩余相关矩阵 $R_{(1)} = R - P_1 P_1^T$, 用同样的方法抽取第二主成分因素 F_2 , 可得相应的负荷 P_2 。如此,依次可求出 F_3, \dots, F_n 的负荷 P_3, \dots, P_n 。这样,回归方程可表示为 $Z = PF$ 。

为了确定从 F 中选择多少个分量去线性表示 Z , 给出统计意义下的定量标准。定义 $\sum_{i=1}^j T_i$ 为前 j 个主成分因子的方差累积贡献率。若 $\sum_{i=1}^j T_i \geq 85\%$, 则确定 j 为所选 F 的分量个数。

2 证券投资的主成分因素分析

为了应用主成分因素分析法对证券投资的因素进行分析,我们对 10 组(共 100 名) 证券投资者进行了与收益相关的 5 项测验:① 经济政策调控;② 行业动态;③ 盲目跟风心理;④ 公司实力和业绩;⑤ 股市操作规律,得到原始数据矩阵,如表 1。

表 1 5 项测验的原始数据

变量	1	2	3	4	5
投资者					
A 组	15.32	10.90	17.36	14.58	10.82
B 组	10.12	11.56	10.08	6.91	12.28
C 组	8.18	6.31	8.07	9.15	11.71
D 组	10.18	6.67	11.59	10.20	8.56
E 组	15.74	20.03	23.74	24.37	7.91
F 组	13.51	16.89	16.40	19.50	5.55
G 组	18.55	10.70	9.64	17.02	4.44
H 组	23.29	16.46	16.22	22.10	4.85
I 组	11.15	9.30	11.05	17.25	6.87
J 组	14.52	7.62	12.83	16.82	4.36

根据表 1, 利用公式(1)将数据标准化得到矩阵 Z , 如表 2。

表2 数据标准化矩阵
Tab.2 Matrix of standardized data

变量 投资者	1	2	3	4	5
A组	0.295 3	-0.167	0.816 8	-0.225 6	1.071 4
B组	-0.917 3	-0.020	-0.806 4	-1.656 5	1.579 4
C组	-1.371 1	-1.200	-1.253 7	-1.237 6	1.382 3
D组	-0.904 2	-1.120	-0.469 9	-1.042 1	0.285 9
E组	0.391 6	1.887	2.237 5	1.599 9	0.062 7
F组	-0.126 5	1.181	0.602 2	0.692 7	-0.758 3
G组	1.048 3	-0.212	-0.904 9	0.229 1	-1.145 3
H组	2.152 7	1.085	0.562 4	1.176 9	-1.003 4
I组	-0.677 8	-0.527	-0.590 4	0.272 0	-0.301 4
J组	0.109 0	-0.906	-0.193 6	0.191 4	-1.173 4

再根据表2,计算出5项测验之间的相关系数 矩阵 R ,如表3。

表3 相关系数矩阵
Tab.3 Matrix of correlation coefficient

变 量	1	2	3	4	5
1	1.000 0	0.558 9	0.456 6	0.724 1	-0.645 0
2	0.558 9	1.000 0	0.813 2	0.747 1	-0.259 1
3	0.456 6	0.813 2	1.000 0	0.744 6	-0.170 5
4	0.724 1	0.747 1	0.744 6	1.000 0	-0.704 1
5	-0.645 0	-0.259 1	-0.170 5	-0.704 1	1.000 0
Σ	2.094 6	2.860 2	2.843 9	2.511 8	-0.778 7

对表3中的数据,利用公式(3)和(4)及特征值 因素 F_1, F_2, F_3, F_4 及 F_5 ,得到因素负荷矩阵 P ,如表方法^[1],进行因素分析、因素抽取,回归计算出公共 4。

表4 未经旋转的因素负荷矩阵
Tab.4 Irrotated matrix of the load on common factors

因素 变量	F_1	F_2	F_3	F_4	F_5	$h^2(2F)$	$h^2(5F)$
1	0.805 5	-0.311 1	0.465 6	0.094 4	0.000 5	0.774 28	0.999 96
2	0.888 6	0.427 0	0.065 2	-0.325 1	-0.028 6	0.889 28	1.000 00
3	0.815 9	0.534 4	-0.117 3	0.242 5	-0.096 5	0.918 15	1.000 00
4	0.946 2	-0.067 2	-0.197 0	0.043 2	0.178 2	0.927 56	1.000 00
5	-0.619 6	0.704 6	0.236 3	0.058 9	0.107 7	0.929 10	0.999 99
特征值	3.372 8	1.065 6	0.329 4	0.178 7	0.053 5		
贡献率	0.674 6	0.887 7	0.953 6	0.989 3	1.000 0	(累积方差比例)	

在表4的最后一行,给出了依重要性排列的各因素的累积方差比例。

为了能用主成分因素来概括问题,从 F_1, F_2, F_3, F_4 及 F_5 中,选择可以线性表示出 Z 的分量,根据表4中累积方差贡献率 $\sum_{i=1}^j T_i$ 的大小来确定。若 $\sum_{i=1}^j T_i \geq 85\%$,则确定 j 为所选 F 的分量个数。由表4可知,因素 F_1 和 F_2 的方差累积贡献率为88.77%,而其他因素对提高累积方差的贡献很小,因此, F_1

和 F_2 可以近似线性表示,即 F_1 和 F_2 可作为主成分因素。

然而,现在的问题是: F_1 和 F_2 的含义以及这两个因素与各变量之间的关系还不清楚。通过旋转变换的方法,可以解决这一问题。旋转变换的方法有多种,我们这里使用“最大方差旋转”法^[1],计算出 F_1 和 F_2 经过旋转后的负荷矩阵,如表5。

从表5中可以明显看出,变量1,4主要由因素 F_1 决定,负荷分别是0.992 78和0.986 39,变量5主

表 5 两个因素经过旋转的负荷矩阵

Tab. 5 Rotated matrix of the load two factors

	F_1	F_2	$h^2(2F)$
1	0.992 78	0.119 95	0.999 99
2	0.580 44	0.814 20	0.999 84
3	0.477 26	0.878 65	0.999 81
4	0.986 39	0.399 17	0.999 10
5	0.266 14	0.963 82	0.999 78
方差比例	0.674 56	0.213 12	\
累积方差比例	0.674 56	0.887 68	\

要由因素 F_2 决定, 负荷是 0.963 82, 而变量 1, 4 的含义分别是“经济政策调控”和“公司实力和业绩”, 都属于经济因素, 变量 5 的含义是“股市操作规律”, 属于证券市场技术因素, 因此, 可以将 F_1 和 F_2 解释为“经济因素”和“证券市场技术因素”。同时, F_1 和 F_2 这两个因素在各变量上的方差和 $h^2(2F)$ 均接近于 1, 即 Z 主要由这两个因素决定, 所以, 对证券投资的收益来说, 经济因素和证券市场技术因素起着

主要作用。

3 结 论

中国证券市场从建立到发展仅有短短数年, 同国外证券市场相比, 发展还很不成熟。对于证券市场的研究都还是刚刚起步。对影响证券投资的因素分析, 特别是对影响证券投资收益的因素的量化分析尤为重要。本文利用主成分因素分析法对影响证券投资的诸因素进行了量化分析, 其结果表明经济因素和证券市场技术因素是影响证券投资的主要因素。

参考文献:

- [1] 谢小庆, 王 丽. 因素分析[M]. 北京: 中国社会科学出版社, 1989.
- [2] 吴晓求. 证券投资学[M]. 北京: 北京理工大学出版社, 1993.
- [3] 乌 杰. 系统科学理论与应用[M]. 成都: 四川大学出版社, 1996. 635-640. (编辑 姚 远)

Method of main factor analysis of the securities investment

TIAN Xu-zi

(Department of Computer, Baoji College of Arts & Science, Baoji 721007, China)

Abstract: The securities investment is affected by many factors. In order to research the law of the securities investment, the factors which will affect securities investment are analyzed by the method of main factor analysis. The analyzed results showed that the economic factor and the market technology factor are the main factors affecting securities investment.

Key words: securities investment; main factors; factors analysis; applied mathematics

· 学术动态 ·

我校将在郭杜工业园区建设西北大学科技园

2003年4月10日, 受西安市政府委托, 由西安市发展计划委员会主持, 邀请陕西省教育厅、西安市科技局、规划局、国土局、建委、环保局、长安区等有关部门负责人, 就组建西北大学科技园进行协调。我校朱恪孝副校长主持会议, 科研、财务、产业及生命科学院等部门参加了会议。

与会领导听取了科研处处长郭鹏江教授所作的西北大学科技园项目情况汇报, 并就各自关心的问题进行了质询。会议认为: 兴办大学科技园是促进高等学校丰富的智力资源与其他社会资源的优化组合, 推动高新技术成果转化的有效途径。这既是党中央提出的重要任务, 也是推动科技成果转化的高新技术产业化、提高我国技术创新能力和国际竞争力的客观要求。西北大学科技园的建立必将为推动科技成果转化和高新技术产业产业化、提高技术创新能力、推进西部大开发和地方经济建设做出积极贡献。

西北大学科技园位于郭杜工业园区、西北大学郭杜新校区以北, 占地约 60 hm²。科技园依托学校学科、人才、成果等优势, 以生物医药产业为发展龙头, 重点发展生物芯片、生物化工、计算机软件、电子信息技术、新材料技术等高新技术, 建成后将成为科技人才创新、创业平台和高新技术企业的孵化和培育基地。

(王方元)