

Research on Fund Evaluation Based on the Factor Analysis

Jinru Wei

Department of Science, Chang'an University, Xi'an Shaanxi
Email: weijinru_derek@163.com

Received: Dec. 26th, 2018; accepted: Jan. 9th, 2019; published: Jan. 16th, 2019

Abstract

Factor analysis is one of the classic multivariate statistical analysis methods, which can extract few unobservable factors from the large quantity of observable variables. These factors can reveal the complex structures of the original variables. For the problem of the evaluation of fund products that the ordinary investors faced, such method could obtain explicit factor indices as the reference of evaluation from the complicated financial data. This article first introduces the basic theory of factor analysis, and then presents the general steps to analyze the financial data of fund products. Finally, an empirical analysis was taken on the yearly data of stock funds of China, 2017, aiming at providing a reference of data-based evaluation of fund products.

Keywords

Factor Analysis, Dimensionality Reduction, Fund Evaluation

基于因子分析的基金评价研究

魏劲如

长安大学理学院, 陕西 西安
Email: weijinru_derek@163.com

收稿日期: 2018年12月26日; 录用日期: 2019年1月9日; 发布日期: 2019年1月16日

摘要

因子分析是一种将大量可观测变量抽象为少量客观存在的不可观测因子的经典多元统计分析方法。这些因子能揭示出原观测变量内部复杂的结构关系。对于日常投资者面临的基金产品评价问题而言, 将因子分析应用于此领域, 能从庞杂的基金产品财务指标数据中得出较为明确的因子指标作为基金产品质量的

评价依据。本文首先介绍了因子分析的相关理论,进而给出了分析一般基金财务数据的基本流程。最后对我国2017年度股票型基金的数据进行实证分析。旨在为投资者提供一种从数据出发的基金产品评价参考。

关键词

因子分析, 降维, 基金评价

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人们投资意识的不断提高,金融理财产品已成为我们日常生活中的一大经济支出。相比银行的定期存款而言,以股票为主的理财产品因其相对较高的收益在近年来受到了来自各方投资者的广泛关注,并因此衍生出了各种不同的投资理财产品。然而,高收益的理财产品通常伴随着高风险。作为普通投资者,在日常的情况下不具有进行个人投资交易的能力与专业知识。因此,人们往往会选择购买基金产品作为日常理财的手段,以期借助专业人士的投资操作,获取更大的投资收益。面对品种繁多的基金产品与其庞杂的财务数据,如何选择合适的基金,成为了人们所关注的问题。

对于特定的基金产品而言,我们可以从其披露的财务报表中获取各类与之相关的财务指标。这些财务指标种类多样,信息重叠关系复杂,但却真实地刻画了该产品的财务情况与运营历程。其中又以年报数据最为丰富,能反映较为全面的财务问题[1]。因此,需要采用合理的方法,对这些财务指标数据进行筛选与处理,为普通投资者对该基金产品的投资价值提供更为清晰直观的解释。本文从统计学的角度出发,给出了基于因子分析方法评价基金财务状况的一般步骤,进而对我国2017年度股票型基金财务数据进行了实证分析。为广大投资者提供了一个评价基金财务状况的可行方法,用以在投资行为中选择理想的基金产品。

2. 因子分析基本理论

因子分析,作为多元统计分析方法的一种,最早由 Charles Spearman 在 1904 年提出,并在心理学、社会学等领域被广泛应用。不同于主成分分析直接降维的方法,它的目的在于研究原始变量内部的复杂关系,并将之简化[2]。通过对于研究对象中可观测变量间的依赖关系与基本结构进行抽象,利用少数不可观测的潜在变量来表示其数据内部的基本结构,以达到对数据进行降维的目的。这些少数潜在变量被称为因子,通常能极大程度地揭示大量观测变量所蕴含的统计信息。

常见的对变量 X_i 进行因子分析的模型为 R 型因子分析模型:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{ij}F_j + \cdots + a_{im}F_m + \varepsilon_i, \quad i = 1, 2, \cdots, p; j = 1, 2, \cdots, m \quad (1)$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} = (A_1, A_2, \cdots, A_m), \quad m \leq p \quad (2)$$

其中 F 称为公共因子, ε 称为特殊因子。 a_{ij} 称为第 i 个变量在第 j 个因子上的载荷, 亦表示 m 维空间中一点 X_i 在坐标轴 F_j 上的投影, 故 A 称为因子载荷矩阵。因子载荷 a_{ij} 表示了变量 X_i 对公共因子 F_j 的依赖程度与相对重要性, 对因子含义的理解具有很重要的作用。通常情况下, 因子载荷矩阵 A 还具有如下统计意义:

$$h_i^2 = \sum_{j=1}^m a_{ij}^2, \quad i=1,2,\dots,p \quad (3)$$

$$g_j^2 = \sum_{i=1}^p a_{ij}^2, \quad j=1,2,\dots,m \quad (4)$$

其中 h_i^2 为 A 中第 i 行元素平方和, 称为变量 X_i 的共同度, 表示了全部公共因子对变量 X_i 的总方差所作的贡献。 g_j^2 为 A 中第 j 列元素平方和, 表示同一公共因子 F_j 对各变量所提供的方差的总和, 是衡量每一个公共因子相对重要性的尺度。实际应用中, 建立因子分析的具体模型, 就是根据样本数据通过“主轴因子法”对因子载荷矩阵进行求解的过程。

为了寻求抽象因子与研究变量间的关系, 往往需要使用因子旋转的方法, 使每个变量仅在一个因子上具有较大的载荷, 以突出其与因子之间的联系。常用的方法有最大方差旋转法, 即对因子载荷矩阵 A 进行正交变换, 使得公共因子向量的各个分量也成为互不相关的公共因子[2]。

在因子分析中, 不考虑特殊因子 ε , 当 $m=p$, 且因子载荷矩阵 A 可逆时, 可以得到每个样本在因子 F 上的相应取值 $F=A^{-1}X$ 。即该样本在因子 F 上的因子得分。然而因子分析模型实际应用中更多情况下要求 $m < p$, 故仅能通过对因子得分的估计进行计算。常用方法有回归得分, Bartlett 得分, Anderson-Rubin 得分等[3]。以 Thompson 回归得分为例:

$$\hat{F} = A^T R^{-1} X \quad (5)$$

其中 \hat{F} 为估计的因子得分, R 是 X 的相关系数矩阵。通过因子得分可以建立因子与样本间的关系, 以构造对样本的评价度量。

3. 基于因子分析的基金投资评价的一般步骤

因子分析作为一种有效的数据降维方法在不同的应用背景中具有一定的限制性。为此我们给出利用因子分析处理基金投资评价问题的一般步骤:

首先, 对基金项目的所有财务指标进行筛选, 去除数据缺失等无效指标。为保证因子得分结果有一定的趋势性, 需采用负向指标正向化转换模型对指标进行合理的正向化处理[4]。对剩余指标进行相关性检验, 将与其他变量皆具有高度相关关系的变量进行筛选剔除。因为此类变量在计算因子载荷过程中容易出现特征值为负值的情形, 在样本量有限的情况下, 影响因子分析效果。

其次, 对数据进行 KMO 检验和 Bartlett 球形检验, 以判断该问题是否适合于因子分析方法。其中 KMO 检验统计量主要判定变量间相关系数平方和和偏相关系数的平方和间的差异。KMO 值越接近于 1, 证明变量间相关性越强, 越适合进行因子分析。若 KMO 值小于 0.5, 则不适合进行因子分析。Bartlett 球形检验用于检验变量相关矩阵是否为单位阵, 即变量间是否独立。当且仅当 Bartlett 检验统计量 p 值小于 0.05 时, 方能证明变量适合做因子分析。在此基础上对样本进行因子分析, 分析因子的方差贡献率信息以及旋转后的因子载荷分布, 以确认变量与因子之间的关系。依照基金领域相关概念对所提取的公共因子赋予语义上的解释, 以总结其作用与影响。

最后, 通过估计的因子得分情况进行因子综合得分的计算, 给出直观的量化指标来确定样本的综合排序。为此我们给出因子综合得分的一般定义:

设对 p 个变量 X_1, X_2, \dots, X_p ，有 m 个公共因子 F_1, F_2, \dots, F_m 。第 j 个因子对于方差的贡献率为 V_j ，由方差贡献率决定的因子间的权重为 W_j ，第 i 个变量关于第 j 个因子的得分是 S_{ij} 。则有以下等式成立：

$$W_j = \frac{V_j}{V_1 + V_2 + \dots + V_m} \quad (6)$$

$$T_k = \sum_{j=1}^m W_j \cdot S_{ij} \quad (7)$$

其中 $i=1, 2, \dots, p; j=1, 2, \dots, m$ ， T_k 即为第 k 个样本在 m 个公共因子上的综合得分。

根据上述等式，投资者可计算因子得分数据并依照对单一因子作用的偏好进行排序，亦可根据计算得到的综合得分数据对所有样本整体进行直观的量化排序。

4. 实证研究

利用上述评价方法对 2017 年度股票型基金财务数据进行基于因子分析的实证研究，数据来源于 CCER (China Center for Economic Research) 中国经济金融数据库。

首先从股票型基金数据库中随机抽取 150 条基金产品财务数据，删除无效及不相关指标，并对负向指标进行正向化。进行相关性检验，并删去具有广泛相关性的指标，最终筛选至 20 个财务指标，分别为：最新基金份额(X_1)，期末单位净值(X_2)，净值增长率(X_3)，加权平均单位基金净收益(X_4)，佣金(X_5)，股票交易额(X_6)，债券交易额(X_7)，股票投资市值(X_8)，债券投资市值(X_9)，银行存款(X_{10})，实收基金(X_{11})，基金资产估计(X_{12})，未分配收益(X_{13})，股票利息收入(X_{14})，债券利息收入(X_{15})，存款利息收入(X_{16})，收入合计(X_{17})，股票投资成本(X_{18})，债券投资成本(X_{19})，费用合计(X_{20})。基于 148 条有效样本数据进行因子分析，得到如下结果：

Table 1. KMO and Bartlett's test
表 1. KMO 和 Bartlett 特检验

KMO 检验	Bartlett 球形检验		
	近似卡方	自由度	显著性
0.797	7023.787	190	0.000

表 1 为 KMO 检验统计量和 Bartlett 球形检验的结果。其中 KMO 检验统计量的值为 0.797，大于 0.7，表示该数据很大程度上适用于因子分析的方法。同时 Bartlett 球形检验统计量的显著性值为 0.000，小于 0.05，说明相关系数矩阵非单位阵，亦表明因子分析在该问题上是有意义的。

进而根据表 2 得到各个因子对于总方差的解释程度。

其中因子分析一共提取到 5 个公共因子，其累计方差贡献率为 87.638%，超过 85%。亦即利用因子分析的方法将原问题中 20 个财务指标变量通过 5 个公共因子代替，对样本数据做到了较大程度的降维。进而可以通过表 3 中旋转后的因子载荷量，分析变量与公共因子间的关系。

可以看到，第一个因子在变量 $X_1, X_8, X_{11}, X_{12}, X_{14}, X_{16}, X_{17}, X_{18}, X_{20}$ 上有较大载荷，这些变量皆与股票型基金的基本业务有关，可将第一个因子命名为主营业务因子。第二个因子在变量 X_7, X_9, X_{15}, X_{19} 上有较大载荷，这些变量皆与债券投资业务有关，可将第二个因子命名为副营业务因子。第三个因子在 X_3, X_4 上有较大载荷，这两个指标体现了股票型基金产品的净收益率与净收益增长率，可将第三个因子命名为净收益能力因子。第四个因子在 X_2, X_{10}, X_{13} 上有较大载荷，这三个指标反应了股票型基金产品的期末资金结余情况，可将第四个因子命名为期末盈余因子。第五个因子在 X_5, X_6 上有较大载荷，这两个指标反应了股票型基金产品的主要收入来源，可将第五个因子命名为盈利能力因子。

Table 2. Total variance explained
表 2. 特征根与方差贡献率表

因子	初始特征值			旋转后载荷平方和		
	特征值	方差贡献率(%)	累积贡献率(%)	特征值	方差贡献率(%)	累积贡献率(%)
1	10.045	50.224	50.224	7.295	36.476	36.476
2	3.038	15.190	65.414	3.596	17.978	54.454
3	2.365	11.827	77.241	2.378	11.892	66.346
4	1.167	5.837	83.078	2.249	11.247	77.592
5	0.912	4.560	87.638	2.009	10.046	87.638
6	0.796	3.982	91.620			
7	0.492	2.459	94.079			
8	0.366	1.831	95.910			
9	0.266	1.329	97.239			
10	0.219	1.095	98.334			
11	0.091	0.457	98.791			
12	0.077	0.387	99.178			
13	0.056	0.278	99.456			
14	0.041	0.205	99.661			
15	0.039	0.195	99.856			
16	0.022	0.109	99.965			
17	0.005	0.025	99.989			
18	0.002	0.010	99.999			
19	9.928E-5	0.000	100.000			
20	1.563E-5	7.816E-5	100.000			

Table 3. Rotated Factor Matrix
表 3. 旋转后的因子载荷阵

	因子				
	1	2	3	4	5
最新基金份额(X_1)	0.856	0.036	-0.008	-0.251	0.271
期末单位净值(X_2)	-0.059	0.149	0.495	0.624	0.128
净值增长率(X_3)	0.089	-0.001	0.950	0.065	-0.015
加权平均单位基金净收益(X_4)	0.061	0.095	0.950	0.212	-0.009
佣金(X_5)	0.504	0.156	0.011	0.149	0.823
股票交易额(X_6)	0.497	0.153	0.006	0.129	0.831
债券交易额(X_7)	0.202	0.598	0.060	0.028	0.080
股票投资市值(X_8)	0.864	0.175	0.092	0.343	0.230
债券投资市值(X_9)	0.118	0.964	0.055	0.124	0.082
银行存款(X_{10})	0.604	-0.023	0.104	0.670	0.033
实收基金(X_{11})	0.954	0.051	-0.004	0.037	0.104
基金资产合计(X_{12})	0.867	0.185	0.105	0.381	0.152

Continued

未分配收益(X_{13})	0.323	0.329	0.239	0.780	0.163
股票利息收入(X_{14})	0.843	0.250	0.112	-0.033	0.211
债券利息收入(X_{15})	0.067	0.961	0.017	0.068	0.046
存款利息收入(X_{16})	0.691	0.061	-0.100	0.414	0.267
收入合计(X_{17})	0.733	0.268	0.454	0.097	0.044
股票投资成本(X_{18})	0.861	0.166	0.043	0.341	0.230
债券投资成本(X_{19})	0.117	0.964	0.054	0.123	0.081
费用合计(X_{20})	0.791	0.185	-0.005	0.231	0.494

上述五个因子体现了股票型基金产品的不同表现。可以为投资者的选择提供一定的参考。可根据表 4 中的因子得分情况结合因子的方差贡献率则可对样本数据进行综合得分的计算。

Table 4. Factor score coefficient matrix

表 4. 因子得分系数矩阵

	因子				
	1	2	3	4	5
最新基金份额(X_1)	0.199	-0.015	0.053	-0.316	0.016
期末单位净值(X_2)	-0.153	-0.023	0.130	0.329	0.118
净值增长率(X_3)	0.001	-0.042	0.471	-0.164	0.039
加权平均单位基金净收益(X_4)	-0.031	-0.020	0.437	-0.062	0.033
佣金(X_5)	-0.126	-0.032	0.032	-0.046	0.602
股票交易额(X_6)	-0.128	-0.032	0.034	-0.058	0.613
债券交易额(X_7)	0.010	0.185	0.006	-0.071	-0.024
股票投资市值(X_8)	0.130	-0.014	-0.030	0.081	-0.078
债券投资市值(X_9)	-0.039	0.305	-0.029	-0.018	-0.031
银行存款(X_{10})	0.077	-0.085	-0.098	0.380	-0.200
实收基金(X_{11})	0.240	-0.024	-0.023	-0.110	-0.197
基金资产合计(X_{12})	0.149	-0.009	-0.039	0.112	-0.155
未分配收益(X_{13})	-0.072	0.021	-0.046	0.428	-0.014
股票利息收入(X_{14})	0.175	0.041	0.056	-0.194	-0.059
债券利息收入(X_{15})	-0.035	0.313	-0.038	-0.041	-0.048
存款利息收入(X_{16})	0.067	-0.048	-0.132	0.201	-0.006
收入合计(X_{17})	0.169	0.040	0.191	-0.146	-0.166
股票投资成本(X_{18})	0.131	-0.016	-0.054	0.091	-0.080
债券投资成本(X_{19})	-0.039	0.305	-0.029	-0.019	-0.031
费用合计(X_{20})	0.046	-0.015	-0.031	0.006	0.194

得到旋转后的因子得分表达式，以主营业务因子 F_1 为例：

$$\begin{aligned}
 F_1 = & 0.199X_1 - 0.153X_2 + 0.001X_3 - 0.031X_4 - 0.126X_5 - 0.128X_6 \\
 & + 0.01X_7 + 0.13X_8 - 0.039X_9 + 0.077X_{10} + 0.24X_{11} + 0.149X_{12} \\
 & - 0.072X_{13} + 0.175X_{14} - 0.035X_{15} + 0.067X_{16} + 0.169X_{17} \\
 & + 0.131X_{18} - 0.039X_{19} + 0.046X_{20}
 \end{aligned} \tag{8}$$

则可以通过因子得分表达式计算样本在单一因子上的表现效果。以因子得分结果作为替代指标近似值,以因子方差贡献率和累计方差贡献率的比值作为替代指标的权重,根据(6)(7)式可以计算得到每个样本的综合得分作为评价基金产品的依据。表5仅展示综合得分排名前10位以及后10位的基金样本情况:

Table 5. Total score of fund products

表 5. 基金产品综合得分

排名前 10 基金产品				排名后 10 基金产品			
基金代码	年度	基金名称	综合得分	基金代码	年度	基金名称	综合得分
161026	2017	富国企改革	3.31	003646	2017	创金中证 1000A	-0.19
540006	2017	汇丰大盘 A	2.75	002236	2017	大成中证 360 互联网	-0.19
000457	2017	上投摩根	2.16	290006	2017	泰信蓝筹	-0.19
050002	2017	博时沪深 300 指数 A	1.94	001611	2017	天弘中证休闲娱乐	-0.20
202001	2017	南方稳健	1.51	003017	2017	广发中证军工 ETF 联	-0.21
000979	2017	景顺长城沪港深精选	1.46	290011	2017	泰信中小盘	-0.21
001542	2017	国泰互联网+股票	1.02	320016	2017	诺安策略	-0.21
001048	2017	富国新兴产业	0.92	161030	2017	体育分级	-0.22
110020	2017	易方达 300	0.88	590007	2017	中邮上证 380	-0.25
161022	2017	富国创业	0.79	360001	2017	保德信量化	-0.30

由表5可以得到在5个因子上综合得分表现较好的十个基金产品与综合表现较弱的十个基金产品。结合因子分析提取出的五个公共因子,还可以从主营业务,副营业务,净收益能力,期末盈余和盈利能力多个角度进行基金产品的具体对比。从而对复杂的财务数据指标,进行了简化解释与量化表达。

5. 结论与展望

本文将因子分析在基金产品财务数据上应用,从统计学的角度,以实际数据为出发点,为分析复杂专业的财务指标数据提供了一种简化方法。从基于2017年度股票型基金实际数据的实证分析结果可以看到,经过筛选得出的财务指标数据,在因子分析过程中有着较为高质量的表现。在由20维数据降为5维数据的过程中,保留了87.638%的完整信息含义,达到了数据在降维过程中的精度要求。并且经过因子旋转的变量使其高载荷的因子具有一定的语义解释功能。进而通过因子得分数据,可以对基金产品样本进行基于单一因子或整体综合的排序,以引导出量化的基金评价依据。

然而,在研究的过程中仍存在一些数据上的限制与可以深究的问题。其中在变量筛选的过程中,并不能做到完全的由数据驱动的变量选择。需要借助一定的主观分析与专业知识指导,方能正确区分有效指标与无效指标。根据因子分析的结果,可以得出初始因子与旋转后的因子并没有语义上的关联。因其计量值不同,不能混淆使用。同时因子分析对于旋转后因子的命名与解释仍旧是一个没有精确定义领域[5]。另外,基于因子分析的理论,存在根据小部分样本的因子得分,向未知样本得分实现推广的可能。然而由于没有参考数据,无法构造损失函数进行优化处理。因此,仍需要从理论层面上对因子分析的相关指标进行更深入的研究与探索。

参考文献

- [1] 王玉春. 简论证券投资基金投资价值的财务分析[J]. 南京财经大学学报, 2008(2): 29-31.
- [2] 张润楚. 多元统计分析[M]. 北京: 科学出版社, 2006: 190-195.

-
- [3] Distefano, C., Zhu, M. and Mindrila, D. (2009) Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment Research & Evaluation*, **14**, 11.
- [4] 范坤. 因子分析中指标数据如何正确预处理[J]. 财会月刊, 2013(6): 85-88.
- [5] 林海明. 因子分析应用中一些常见问题的解析[J]. 统计与决策, 2012(15): 65-69.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: sa@hanspub.org