

开放式基金投资评判中的聚类分析

宋光辉(博士生导师) 李晓钊

(华南理工大学工商管理学院 广州 510640)

【摘要】随着近年来开放式基金市场风险的日趋增大,研究怎样理性地进行开放式基金投资就成为一个非常具有价值的课题。基于此,本文参考开放式基金的内在考评指标,并借鉴聚类分析的原理和方法,对开放式基金进行比较、分析、甄选并得出结论,以为投资者选择最为适合的开放式基金提供参考。

【关键词】聚类分析 开放式基金 数据标准化

一、问题的提出

开放式基金又称共同基金,它是指基金发行总额不固定,基金单位总数随时增减,可视投资者需求,随时按基金报价在规定的营业场所进行申购或赎回基金份额的一种基金。近年来,开放式基金逐渐替代封闭式基金,成为国际基金市场的主流形式。美国、英国等成熟金融市场有90%以上的基金份额是开放式基金。我国的开放式基金市场起步较晚,但是发展速度十分迅速,到2012年12月1日已增至1 227只。

与外汇、储蓄、股票证券等投资方式相比,开放式基金具有专家理财、收益稳妥的特点。然而随着2008年世界金融危机的爆发,国际资本经济大环境日益衰退,开放式基金市场投资风险日益增大。在这样的背景下怎样进行理性的开放式基金投资?投资什么类型的开放式基金?又是怎么对不同类型开放式基金进行评判?

开放式基金投资价值评判的方法主要有两种:一种是按照某个标准直接进行比较,操作简单,一目了然,但是不能发现数据之间的内在联系;另一种是间接比较,将基金进行归类,并通过同类型的基金进行分析来评定基金业绩的优劣。本研究将采用后者,利用系统聚类分析的方法甄选出业绩突出、收益良好、具有投资潜力的股票,实现理性投资。

二、开放式基金的聚类分析

1. 方法介绍。聚类分析是一种将研究对象分为相对同质群组的统计分析技术。聚类分析也叫分类分析或数值分类,是用数学的方法来研究和处理给定对象的分类,即对同类型对象抽象出其共性,从而形成“类”。本研究采用SPSS中默认的聚类分析方法,即组间平均链锁法,此方法定义了每个样本和类别之间的组间平均链锁距离。聚类分析先将最相似的两个变量聚为一小类,再与最相似的变量或小类合并,如此分层依次进行。假设共有N个样本,第一步,将每个样本看成单独一类,共有N类;第二步,根据样本距离,把样本距离较近的两类样本聚为一类,其他样本仍聚为一类,共有N-1类;

第三步,将距离最近的两类进一步聚为一类,则共有N-2类,依此类推,最终将所有的样本聚为一类。而在该方法中距离的选择上默认使用欧氏距离,衡量单个样本与属于该类别中每个样本的平均值。

此外,为了能够直观地反映系统聚类过程,本文还会进行聚类分析树状图的绘制,丰富研究结果。

值得注意的是,要进行聚类分析首先应建立一个由某些事物属性构成的指标体系,或者说是一个变量组合。入选的每个指标必须能刻画事物属性的某个侧面,所有指标组合起来形成一个完备的指标体系,相互配合共同刻画事物的特征。简单地说,聚类分析的结果取决于变量的选择和变量值获取两个方面。变量选择越准确,测量越可靠,得到的分类结果越能描述事物各类间的本质区别。

2. 研究样本及样本数据。样本选择上首先依据中国基金网现有的1 227只开放式基金排列序号和基金代码的对应顺序进行编号定位,然后采用Excel中的“int(rand()*(b-a))+a”函数对1至1 227进行随机排列,最后选择随机处理后的前30个序号的基金作为本次研究的样本(见表1)。

本文样本数据来源于和讯网、中国基金网、搜狐基金网等大型基金综合类网站。截取时间点为2013年5月20日。累计净值、最新规模、当年收益率均以时间点为准,詹森指数、特雷诺指数、夏普比率均为该基金成立之后的整体表现。

3. 开放式基金的评判指标甄选。充分借鉴具有丰富理论和实践经验的银行证券、海通证券、招商证券、晨星公司等专业基金评级机构的评级标准,并在此基础上从客观性、独立性、全面性、可量化、实用性、适用性原则出发,最终选取基金规模、当年收益率、詹森指数、特雷诺指数、夏普比率共六项评判指标。这六项指标各自从不同角度对基金进行评判,均是当今各大专业机构基金评级时的重要参考依据。

(1)累计净值。基金累计净值是指基金最新净值与成立以来的分红业绩之和,体现了基金从成立以来所取得的累计收

益,可以比较直观和全面地反映基金在运行期间的历史表现。一般来说,累计净值越高,基金业绩越好。

(2)规模。基金规模是基金绩效的直观影响因素。一般认为,中小规模的基金相较于规模大的基金而言,换仓灵活,能较好地追逐市场热点,换股冲击成本也小。而规模过大的基金在投资上灵活性不足,处于市场震荡阶段或震荡盘升阶段时,在建仓和调仓方面劣势明显;同时在缺乏足够多的大市值及优质新上市公司股票时,可选择的股票难度较大。

(3)收益率。基金收益率最重要的指标是基金投资收益率,即基金证券投资实际收益与投资成本的比率。投资收益率的值越高,则基金证券的收益能力越强。

(4)詹森指数。詹森指数用于测定证券组合经营绩效,是证券组合的实际期望收益率与位于证券市场中的证券组合的期望收益率之差。它能评估基金业绩优于基准的程度,通

过比较考察期基金收益率与由定价模型 CAPM 得出的预期收益率之差(即基金的实际收益超过它所承受风险对应的预期收益的部分)来评价基金,此差额部分就是与基金经理业绩直接相关的收益。值越大说明基金经理的能力越强。

(5)特雷诺指数。特雷诺指数是每单位风险获得的风险溢价。特雷诺指数越大,单位风险溢价越高,开放式基金的绩效越好,说明基金管理者在管理的过程中所冒风险有利于投资者获利。相反特雷诺指数越小,单位风险溢价越低,开放式基金的绩效越差,基金管理者在管理的过程中所冒风险不利于投资者获利。

(6)夏普比率。如果夏普比率为正值,说明在衡量期内基金的平均净值增长率超过了无风险利率,在以同期银行存款利率作为无风险利率的情况下,说明投资基金比银行存款要好。夏普比率越大,说明基金单位风险所获得的风险回报越高。

表 1 原始数据和进行标准化处理后的结果

基金代码	基金名称	累计净值(元)	最新规模(亿元)	当年收益率(%)	詹森指数	特雷诺指数	夏普比率	累计净值(元)	最新规模(亿元)	当年收益率(%)	詹森指数	特雷诺指数	夏普比率
270010	广发沪深300指数	1.38	24.01	7.61	0.001 1	0.001	0.04	0.02	0.27	1.34	0.32	-0.02	-0.42
050006	博时稳定债券B	1.26	3.14	2.64	0.000 8	0.002 4	0.09	-0.13	-0.62	-0.29	0.02	0.00	-0.17
270030	广发聚财信用债B	1.04	17.87	1.28	0.000 3	-0.030 6	0.18	-0.43	0.00	-0.74	-0.47	-0.36	0.28
161909	万家添利分级债券A	1.07	29.02	1.13	0.000 3	0.087 9	0.89	-0.38	0.48	-0.79	-0.47	0.92	3.82
398041	中海量化策略股票	0.94	4.24	4.15	0.000 9	0.000 9	0.03	-0.56	-0.58	0.20	0.12	-0.02	-0.47
481001	工银核心价值股票	3.69	77.05	9.20	0.000 2	0.000 2	0.01	3.10	2.53	1.86	-0.57	-0.02	-0.57
070099	嘉实优质企业股票	1.79	81.74	3.00	0.001 8	0.002 1	0.07	0.57	2.73	-0.17	1.01	0.00	-0.27
159910	嘉实深圳基本面120	0.81	4.72	5.95	-0.000 1	-0.000 1	-0.01	-0.74	-0.56	0.80	-0.87	-0.03	-0.66
165516	信诚周期轮动股票	1.01	1.49	2.53	0.003 0	0.000 1	0.01	-0.46	-0.69	-0.33	2.20	-0.03	-0.57
165705	诺德双翼分级债券	1.07	3.12	1.53	0.000 7	-0.267 8	0.36	-0.38	-0.62	-0.66	-0.08	-2.91	1.18
253061	国联安信心增长B	1.08	2.57	1.72	0.001 1	-0.158 7	0.33	-0.38	-0.65	-0.59	0.32	-1.73	1.03
160809	长盛同辉深证100	0.94	16.33	3.14	0.000 0	0.002 3	0.08	-0.56	-0.06	-0.13	-0.77	0.00	-0.22
163805	中银动态策略	1.30	13.13	3.43	0.001 7	0.001 8	0.06	-0.08	-0.20	-0.03	0.91	-0.01	-0.32
160512	博时卓越品牌	1.01	2.83	12.84	0.001 5	0.001 9	0.07	-0.47	-0.64	3.06	0.72	-0.01	-0.27
519692	交银成长	2.83	65.04	5.57	0.002 3	0.002 3	0.08	1.95	2.02	0.67	1.51	0.00	-0.22
070021	嘉实主题新动力	0.80	34.91	-0.25	0.000 3	0.000 3	0.01	-0.75	0.73	-1.24	-0.47	-0.02	-0.57
530015	建信基本面60联接	0.88	3.10	6.41	0.000 1	0.000 1	0.00	-0.65	-0.63	0.95	-0.67	-0.03	-0.62
270002	广发稳健增长	3.04	64.33	1.74	0.000 5	0.000 7	0.02	2.24	1.99	-0.59	-0.27	-0.02	-0.52
310379	申万菱信添益宝	1.15	3.43	1.71	0.000 8	0.011 2	0.24	-0.28	-0.61	-0.60	0.02	0.09	0.58
450004	富兰克林国海价值	1.59	17.07	5.44	0.003 0	0.000 4	0.13	0.31	-0.03	0.63	2.20	-0.02	0.03
590006	中邮核心中小盘	0.80	6.01	5.66	-0.000 8	-0.000 8	-0.03	-0.74	-0.50	0.70	-1.56	-0.04	-0.76
550008	信诚优胜精选	0.80	12.18	-0.25	0.000 9	0.000 8	0.03	-0.74	-0.24	-1.24	0.12	-0.02	-0.47
163110	申万菱信量化小盘	0.82	1.86	-0.49	0.000 8	0.000 7	0.03	-0.72	-0.68	-1.32	0.02	-0.02	-0.47
213001	宝盈鸿利收益	2.30	3.90	3.97	0.000 7	0.000 8	0.03	1.25	-0.59	0.15	-0.08	-0.02	-0.47
253020	国联安德盛增利A	1.27	10.05	1.77	0.001 3	0.034 4	0.49	-0.12	-0.33	-0.58	0.52	0.34	1.83
420102	天弘永利债券B	1.21	14.90	1.61	0.000 9	0.379	0.44	-0.20	-0.12	-0.63	0.12	4.05	1.58
620007	金元惠礼保本混合	1.05	1.02	0.88	0.000 1	0.002 1	0.05	-0.41	-0.71	-0.87	-0.67	0.00	-0.37
570007	诺德优选30	0.72	4.90	0.84	-0.002 1	-0.002 1	-0.08	-0.85	-0.55	-0.88	-2.85	-0.05	-1.01
161207	瑞和300	0.75	7.80	7.2	0.000 9	0.000 8	0.03	-0.82	-0.43	1.21	0.12	-0.02	-0.47
217002	招商安泰平衡	2.42	1.04	3.88	0.000 3	0.000 5	0.02	1.41	-0.71	0.12	-0.47	-0.02	-0.52

标准化处理后的结果

4. 数据标准化处理。由于不同变量间存在较大的数量级的差别,因此对数据变量采用Z得分值标准化的方法进行标准化,这种方法给予原始数据的均值和标准差进行数据的标准化,普遍适用于最大值和最小值未知的情况。具体计算公式如下所示:

$$\hat{Y} = \frac{(Y - \bar{Y})}{S.D.}$$

式中: \hat{Y} 为标准化处理值; Y 代表原数据; \bar{Y} 代表均值; $S.D.$ 代表标准值。

根据此公式,对原始数据进行标准化处理,详见表1。

5. 聚类结果分析。表2显示的是系统聚类分析的类成员聚类表,从表2中可以知道类别从3到7时的个案所属的类别。当类别为5时,万家添利和国联安德是一类,博时卓越是一类,其他样本分属其他三类。

表3显示的是系统聚类的凝聚状态。表中,第一列表示聚类分析的第几步;第二、三列表示本步聚类中哪两个个案聚

成一类;第四列是个案距离;第五、六列表示本步骤类中参与聚类的个案是个案还是小类,0表示个案,非0表示由第几步聚类生成的小类参与本步骤类;第七列表示本步骤的结果将在以下第几步中用到。

在聚类分析的第一步中,个案序号为8(嘉是深圳)和17(建信基本)据成一小类,这个小类将在第5步用到。同理,聚类分析的第二步,序号为24的个案(宝盈鸿利)和30的个案(招商安泰)又进行聚类,并且将在第20步用到。最终,30个个案聚成了一个大类。

下页树形图显示了在系统聚类的过程中,从每个个体为单独的一类逐步合并,一直到全部合并成一大类,整个过程都在树形图(基金序号同表2所示)中得到了体现。

三、中国开放式基金的投资评判

根据样本数量,将样本聚为三类。表4为各分类样本指标的均值描述。基于该表,我们可以得出相应的结论。

表2 样本分类表

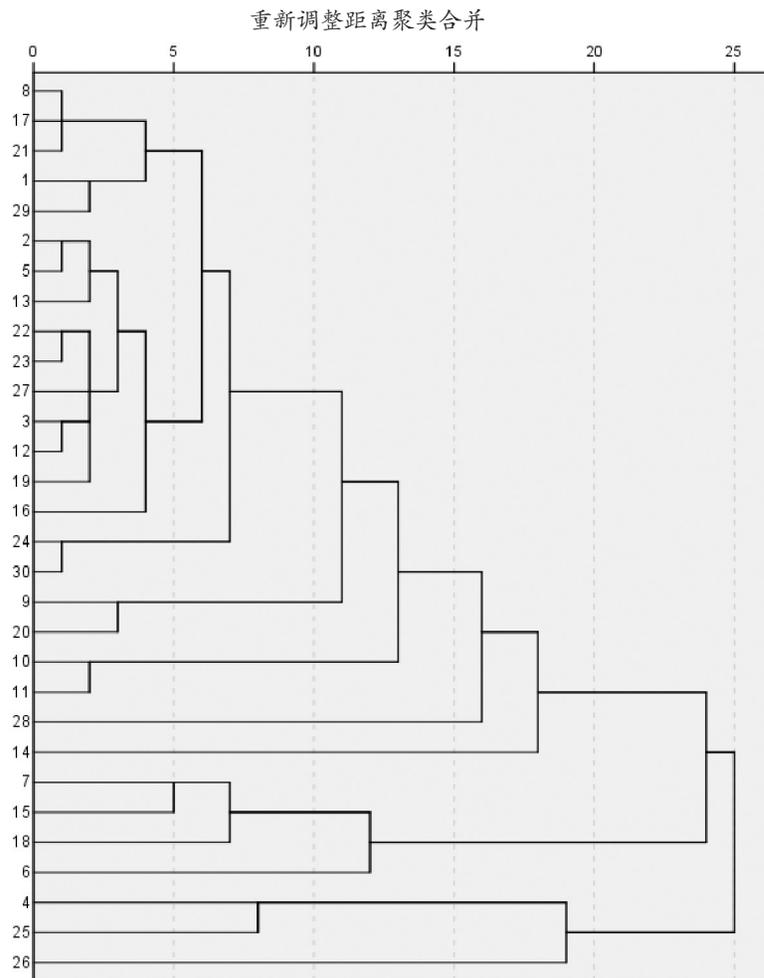
案例	7 群集	6 群集	5 群集	4 群集	3 群集
1:广发沪深	1	1	1	1	1
2:博时稳定	1	1	1	1	1
3:广发聚财	1	1	1	1	1
4:万家添利	2	2	2	2	2
5:中海量化	1	1	1	1	1
6:工银核心	3	3	3	3	3
7:嘉实优质	3	3	3	3	3
8:嘉是深圳	1	1	1	1	1
9:信诚周期	1	1	1	1	1
10:诺德双翼	4	1	1	1	1
11:国联安信	4	1	1	1	1
12:长盛同辉	1	1	1	1	1
13:中银动态	1	1	1	1	1
14:博时卓越	5	4	4	1	1
15:交银成长	3	3	3	3	3
16:嘉实主题	1	1	1	1	1
17:建信基本	1	1	1	1	1
18:广发稳健	3	3	3	3	3
19:申万菱信	1	1	1	1	1
20:富兰克林	1	1	1	1	1
21:中邮核心	1	1	1	1	1
22:信诚优胜	1	1	1	1	1
23:申万菱信	1	1	1	1	1
24:宝盈鸿利	1	1	1	1	1
25:国联安德	2	2	2	2	2
26:天弘永利	6	5	5	4	2
27:金元惠礼	1	1	1	1	1
28:诺德优选	7	6	1	1	1
29:瑞和300	1	1	1	1	1
30:招商安泰	1	1	1	1	1

表3 系统聚类的凝聚状态

阶	群集组合		距离	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	8	17	0.077	0	0	5
2	24	30	0.195	0	0	20
3	22	23	0.210	0	0	7
4	2	5	0.527	0	0	8
5	8	21	0.700	1	0	15
6	3	12	0.862	0	0	10
7	22	27	0.944	3	0	12
8	2	13	1.068	4	0	13
9	1	29	1.255	0	0	15
10	3	19	1.410	6	0	12
11	10	11	1.581	0	0	24
12	3	22	1.693	10	7	13
13	2	3	2.018	8	12	16
14	9	20	2.310	0	0	22
15	1	8	2.644	9	5	18
16	2	16	2.913	13	0	18
17	7	15	3.367	0	0	19
18	1	2	4.576	15	16	20
19	7	18	5.073	17	0	23
20	1	24	5.082	18	2	22
21	4	25	6.044	0	0	27
22	1	9	8.398	20	14	24
23	6	7	9.241	0	19	28
24	1	10	10.346	22	11	25
25	1	28	12.495	24	0	26
26	1	14	13.965	25	0	28
27	4	26	14.810	21	0	29
28	1	6	18.961	26	23	29
29	1	4	20.475	28	27	0

表4 各分类样本指标均值描述

均值	N	累计净值(元)	最新规模(亿元)	当年收益率%	詹森指数	特雷诺指数	夏普比率
第一类	23	1.13	8.29	3.57	0.000 7	-0.018 8	0.074 0
第二类	3	1.18	17.99	1.50	0.000 8	0.167 1	0.606 7
第三类	4	2.84	72.04	4.88	0.001 2	0.001 3	0.045 0



使用平均联接(组间)的树状图

第一类基金:今年整体收益较高。但是历史詹森指数、特雷诺指数和夏普比率均为最低,整体素质较差,说明今年的良好业绩可能是该板块的股票规模较小导致。今年整体股票市场震荡很大,规模较小的基金换仓灵活,能较好地追逐市场热点获利。该类基金未来的走势尚不明朗,最好在获益后逐步退出。

第二类基金:收益较差,规模适中。该类基金的历史詹森指数、特雷诺指数和夏普比率非常高。仔细对数据进行剖析可以发现,特雷诺指数和夏普比率远远超过其他两类基金。前者表示单位风险溢价很高,后者表示基金单位风险所获得的风险回报越高。虽然这两个指标表示风险对应的回报很高,但是风险也会逐渐增大。当这三只基金一直持有高风险类股票时,在今年证券市场整体走弱的情况下,收益率低也是在所难免。但是未来如果证券市场看好,该类基金的收益

将会大幅提升,所以可以在大盘上升趋势明朗时谨慎介入。

第三类基金:收益率最高,累计净值非常高,规模也较大,说明是已经成熟运作很长时间的基金种类。詹森指数、特雷诺指数、夏普比率都表现良好,尤其是詹森指数非常高,远远超过其他两类基金,说明基金经理的能力非常强。这类基金相当于股票市场中的绩优股,市场表现远远好于其他类型基金,建议长期持有。

四、讨论

近年来,基金投资存在着较高的风险性,一旦不慎就会令投资者损失惨重。本研究从基金的内在价值出发,重点从基金规模、当年收益率、詹森指数、特雷诺指数、夏普比率六个可以整体反映基金业绩表现的指标对于基金进行盘点,并进一步运用系统聚类分析方法,对这几个指标进行评判,从而将基金进行类别划分,对应到不同等级的投资价值中去,以指导股票投资。

该方法对于投资者来说,不但有利于缩小投资选择范围,确定投资价值,减少投资者的投资风险,进一步拓展到所有开放式基金的投资价值分析中去。而且对于上市公司的经营管理者来说,也可以借此认清和企业发展现状行业对照,对于薄弱环节分析问题,采取措施以提高企业竞争力。

经过实际操作,可以认为聚类分析对于开放式基金的投资分析有一定的指导作用,对于未来的研究还可以考虑和AHP层次分析法结合起来,定性定量相结合。运用AHP层次分析法对评判基金指标进行两两重要性比较,得

出指标的权重和权值后,赋予每只基金的投资价值得分,再应用聚类分析法进行层次划分,可能划分结果相较于仅聚类分析更具有针对性和精确性。

主要参考文献

1. 张庆利.SPSS宝典(第二版).北京:电子工业出版社,2011
2. 王敏,王颖.基于主成分分析的基金绩效评价模型研究.大连理工大学学报,2005;2
3. Jensen Michael. The performance of funds in the period 1945~1964. Journal of Finance,1968;2
4. Treynor J.. How to rate management of investment funds. Harvard Business Review,1965;43
5. Sharpe, W.F. Mutual fund performance. Journal of Business, 1966;39
6. 高慧璇.应用多元统计分析.北京:北京大学出版社,2005