

中国股票市场行业收益率序列 动态聚类分析

劳兰珺, 邵玉敏

(复旦大学 管理学院, 上海 200433)

摘要:文章提出对行业指数收益率序列分阶段进行聚类分析的动态分析方法,以考察行业间的相互关系及其演化过程。文章利用深交所的行业指数数据进行实证研究,提出基准类的概念,分析了各行业相对于基准类的紧密性、稳定性,以及各行业间的相似程度,并探讨了有关宏观经济事件对行业间相互关系的影响。由于文章所采用的研究方法并不依赖于行业收益率序列的概率分布,因此其研究结果有助于加深投资者及监管部门对行业间相互关系的了解,对投资决策具有参考价值。

关键词:行业分析;行业指数;聚类分析;宏观经济分析

中图分类号:F830.91 **文献标识码:**A **文章编号:**1001-9952(2004)11-0075-08

一、引言

宏观经济分析和行业分析是金融投资界进行证券分析的两个重要方面。上市公司的业绩无疑受到宏观经济形势的影响,并且与其所处的行业密切相关。在不同的经济时期,各行业由于各自的背景、结构、变化规律、所处生命周期阶段等互不相同,导致其盈利水平的高低、经营状况的稳定程度也有所不同。在不同的宏观经济形势下,各行业的表现如何、相互关系如何是投资者在进行投资决策时所关心的问题。

国外学者对上市公司所属行业与股票回报率之间的关系进行了多方面的研究。Lessard(1974)考察了各国不同的行业结构对股票回报率的影响。Roll(1992)认为指数构造方法、指数行业构成以及汇率表现是三个引起不同国家指数组合回报率差异的主要因素。Mazzucato和Semmler(1999)的研究表明美国自动化行业的生命周期的变化与股票价格的波动性密切相关。

收稿日期:2004-08-30

基金项目:国家自然科学基金项目(70371010)和教育部留学回国基金项目

作者简介:劳兰珺(1965—),女,浙江慈溪人,复旦大学管理学院财务金融系教授;

邵玉敏(1975—),女,河北定州人,复旦大学管理学院财务金融系硕士生。

Schwert(2002)发现自1998年中期以来 Nasdaq 的大公司的非正常波动与这些公司所属行业有关。Moskowitz 和 Grinblatt (1999)认为股票收益率的动量与其所属行业有关,其后也有文章认为行业(或行业效应)不能解释股票收益率的动量(Lee 和 Swaminathan(2000), Grundy 和 Martin (2001))。

国内学者也结合国情开展了许多富有特色的关于行业分析方面的研究工作。有对单个行业进行的分析。如朱武祥、宋勇(2001)对家电行业上市公司的股权结构与企业价值的关系进行了分析;田素华、顾允平(2000)对 WTO 体制下的中国服装业进行了分析。也有关于上市公司所属行业与股票价格变化关系的研究,如范龙振、王海涛(2003)的研究表明,中国股票市场具有明显的行业和地区效应;宋逢明、梁洪昀(2001)研究了取消发行市盈率限制后的 95 只股票的初始回报率水平及影响这一水平的因素,发现二级市场上行业的平均市盈率以及股价总水平等是影响初始回报的主要因素。还有就某一事件或某一模型对各个行业展开的分析,如王幼军(2000)分析了加入 WTO 对我国各个行业的影响;赵宇龙、易琮(1999)利用经典的股票计价模型和资本资产定价模型计算各行业的成长系数,揭示了各行业未来不同的成长能力。

本文旨在研究各行业股票回报率间的相互关系。我们认为行业间的相互关系不是一成不变的,而是随着宏观经济形势的变化和行业本身生命周期的变化而不断演变的。与上述已有的研究不同,本文将采用分阶段聚类分析方法对各行业的回报率进行聚类分析,探讨行业收益率间的相近程度,以及这种相近程度的动态变化。本文还将利用动态聚类分析的结果来分析有关宏观经济事件对行业间相互关系所产生的影响,并且探讨行业收益率聚类分析对投资决策的意义。

二、研究方法、行业划分及数据

1. 研究方法

对行业收益率间相互关系的最直观、简单的分析方式是对各行业的收益率进行排序(Bodie, 2002),并用图表加以表示。行业收益率顺序关系体现了各行业股票回报率的相对优劣关系,排序时使用的是行业平均收益率。平均收益率是一个简单的统计量,在其计算和排序的过程中丢失了许多收益率序列间相互关系的信息。另一个常用的分析相互关系的方法是相关分析,行业收益率序列间的相关系数可用来反映行业收益率间的线性相关性(常用的回归分析本质上也是一种相关分析),但对非线性相关性无能为力。采用平均收益率排序和计算相关系数均不能全面反映行业收益率之间相互关系的综合特征,我们对此举例说明。

考虑 A、B、C、D、E、F 六个行业,各行业的日收益率序列如表 1 所示。我们的问题是:如何判断行业间的相近程度或相似性?若按平均收益率排序,由

于 A 与 C、E 与 F 两两之间的平均收益率相同，结果认为行业 A 和行业 C 无差异、行业 E 和行业 F 无差异。事实上，前五日行业 A 的日收益率呈上升趋势，而行业 C 的日收益率呈下降趋势，两个行业的表现相差甚远。若进行相关分析，将行业收益率序列两两配对计算相关系数（采用夹角余弦等其他相似系数，结果类似），则 A 与 D 的相关系数为 1，E 与 F 的相关系数为 -1，其余的介于 -1 与 1 之间。结果会将 A 与 D 两个收益率相差 15 倍的行业看成最为相近，而将 E 与 F 两个收益率在零附近徘徊相差无几的行业看成差别最大。产生上述似是而非的分析结果的原因在于无论是平均收益率的大小顺序，还是收益率相关系数都只体现了行业收益率序列间的某一方面的关系，要全面综合反映收益率序列间的相互关系尚需寻找更为恰当的分析方法。

表 1 日收益率 单位：万分之一

行业	日收益率序列										日平均收益率
A	1.00	2.00	3.00	4.00	5.00	5.00	4.00	3.00	2.00	1.00	3.00
B	1.10	2.00	3.10	4.00	5.10	5.00	4.10	3.00	2.10	1.00	3.05
C	5.00	4.00	3.00	2.00	1.00	5.00	4.00	3.00	2.00	1.00	3.00
D	15.0	30.0	45.0	60.0	75.0	75.0	60.0	45.0	30.0	15.0	45.0
E	0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	-0.01	0.01	-0.01	0.00
F	-0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	0.01	-0.01	0.01	0.00

我们认为可以采用聚类分析方法来探讨行业回报率间的相互关系。聚类分析通过对样品的各个指标（每个行业为一个样品，每个日收益率为一个指标）进行比较来分析行业间的相似程度，使用的信息量大，更能综合反映行业之间的关系。使用欧氏距离进行聚类分析，结果表明 A 与 B 属同一类，E 与 F 属同一类，C、D 各属一类，与人们的直观认识相一致。

行业收益率序列聚类分析方法综合考虑了行业的收益、风险等因素，能对不同行业在相同经济环境下的表现进行有效的分类。当经济发展处于不同阶段时，不同行业之间可能会表现出相互各异的业绩；宏观经济事件的发生也会对各个行业产生不同的影响。将时间划分为多个阶段，分阶段进行聚类分析，从中观察行业聚类关系的变化，有助于分析各个行业对经济环境变化的敏感度，以及行业间相互关系的演变。进一步再将分阶段聚类分析方法与排序方法相结合，既能反映各个经济发展阶段中各行业间业绩表现的相近程度，又能大致体现各阶段行业间平均回报率的高低，可以较全面地刻画行业回报率间的相互关系，从而为投资决策提供有价值的参考信息。

谱系聚类法是应用较为广泛的一种聚类方法。根据谱系聚类法，首先是各样品自成一类，然后把最相似（指标间距离最小）的样品聚为小类，再将已聚合的小类按其相似性（用类间距离度量）进行进一步聚合，如此逐步扩大，直到所有的样品都聚集完毕，形成一个表示亲疏关系的谱系图。最后，根据谱系图按照某种标准对样品进行分类。

本文采用谱系聚类法对深市行业指数进行聚类。把行业指数作为样品,日收益率作为指标,并对数据进行标准化,使各指标具有相同的均值和标准差,其分辨能力完全被同化。样品之间的距离采用欧氏距离:

$$d_{ij}^2 = (x_i - x_j)^T (x_i - x_j)$$

其中, x_i 和 x_j 为样品 i 和样品 j 的各指标的多元观测数据。类间距离采用离差平方和距离:

$$D_{pq}^2 = \frac{n_p n_q}{n_p + n_q} (\bar{x}_p - \bar{x}_q)^T (\bar{x}_p - \bar{x}_q)$$

其中, n_p 为类 G_p 中包含的样品的个数, $\bar{x}_p = \frac{1}{n_p} \sum_{i=1}^{n_p} x_i$ 为类 G_p 中所有样品的均值。

类间距离的递推公式为:

$$D_{rk}^2 = \frac{n_p + n_k}{n_r + n_k} D_{pk}^2 + \frac{n_q + n_k}{n_r + n_k} D_{qk}^2 - \frac{n_k}{n_r + n_k} D_{pq}^2, \quad k \neq p, q$$

其中, D_{ij} 为类 G_i 与类 G_j 的距离, n_i 为类 G_i 中包含的样品个数, 类 G_r 由类 G_p, G_q 合并所得, $n_r = n_p + n_q$ 。关于聚类分析方法可参见范金城、梅长林的《数据分析》(2002)。

2. 行业分类及有关数据

中国证监会 2001 年 4 月发布了《上市公司行业分类指引》(下文简称《指引》),它是目前用于中国上市公司行业分类的官方标准。《指引》将上市公司的经济活动分为门类、大类两级,中类作为支持性分类参考。深交所依据《指引》中的门类划分,编制了 13 个行业指数,行业指数名称和行业代码(见表 2)。

表 2 上市公司的行业划分

行业代码	行业指数	行业代码	行业指数	行业代码	行业指数
A	农林牧渔指数	F	运输仓储指数	K	社会服务指数
B	采掘业指数	G	信息技术指数	L	传播文化指数
C	制造业指数	H	批发零售指数	M	综合类指数
D	水电煤气指数	I	金融保险指数		
E	建筑业指数	J	房地产业指数		

深交所从 2001 年 7 月 2 日起开始编制并公布这些行业指数数据。本文采用深交所自 2001 年 7 月~2003 年 9 月公布的行业指数数据,对其日收益率($r_t = \log(p_t/p_{t-1})$, p_t 代表当日收盘价)进行了分阶段聚类分析,每阶段为一个季度,以研究各行业间的相互关系及其变化过程。

三、基于深市的行业相互关系的实证研究

各行业的表现受宏观经济形势的影响。一般地,经济繁荣时,各行业的业绩呈上升趋势;经济衰退时,各行业的表现也相应低迷。因而以市场主体为参照来研究各行业业绩的关系比孤立地观察各行业更能揭示行业关系的本质特

性。为此，本文引入基准类概念（基准类的表现最接近整个市场的表现），并以基准类为参照来研究行业间的关系。

本节首先考察了 13 个门类的相互关系及其变化过程，然后分析行业间的关系及其演化过程对投资决策的意义。

1. 门类行业指数的聚类结果及其分析

我们对各行业按季度进行分阶段聚类分析，结果列于表 3 之中，其中各类的平均收益率按降序排列。在 13 个门类中，制造业指数(C)的成分股最多，为 330 个，其他指数的成分股个数之和为 217。从聚类结果可看出，制造业指数所在类包含的行业门类最多，占整个市场的份额也最大，其表现最接近整个市场的表现。因而把制造业指数所在类作为基准类。

首先，我们考察各行业关于市场的紧密性，亦即关于基准类的密切程度。通过对聚类结果的分析，我们发现批发零售指数(H)、社会服务指数(K)和综合类指数(M)与市场主流的密切程度最高，亦即它们与制造业指数(C)表现接近，共同形成市场的主流；信息技术指数(G)、金融保险指数(I)、传播文化指数(L)与市场主流的密切程度最低。其他指数与市场的密切程度居中。

其次，考察基准类外行业间的相似性。建筑业指数(E)和房地产业指数(J)与市场的密切程度不高，但两者间的相似性强，表现接近。在九个阶段的聚类中，只有在 2003 年第二季度中两者不属于同一类，其余八个阶段中两者均属于同一类。

再次，分析各行业与基准类相对关系的稳定性。通过对聚类结果的观察，我们发现金融保险指数(I)日收益率最不稳定，时而高于基准类，时而低于基准类，变化最为频繁，而且与基准类的差别很大，相对于基准类来说稳定性最差。原因在于金融保险指数(I)的成分股少，且金融保险业本身就具有较高的风险特性。建筑业指数(E)和房地产业指数(J)在基准类附近来回波动，与基准类的相对关系不稳定，但比金融保险指数(I)的稳定性好。

另外，行业和基准类的关系与行业本身的特点有关，且受外部因素的影响。观察行业聚类关系的变化有助于我们分析某些事件所产生的影响。例如，2002 年 11 月 7 日中国证监会和央行联合颁发《合格境外证券投资管理暂行办法》(简称 QFII 事件)后，各行业收益率的分类关系发生了明显变化。对聚类结果进行分析，可发现 QFII 事件后，建筑业指数(E)和房地产业指数(J)与基准类的紧密性加强了；水电煤气指数(D)和运输仓储指数(F)与基准类的密切程度明显降低。水电煤气指数(D)、运输仓储指数(F)、采掘业指数(B)、信息技术指数(G)和传播文化指数(L)的收益率相对于基准类来说有所提高。

2. 研究行业收益率间相近关系对投资决策的意义

对行业收益率序列进行聚类分析的本质是将收益率序列相近的行业归于同一类之中，与相关分析的分析角度完全不同，同一类中的行业收益率序列间

表3 13个门类的行业指数的聚类结果

时段	第一类	第二类	第三类	第四类	第五类
2001年第三季度	L (-0.00091)	I (-0.00128)	ACDEFHJKM (-0.00157)	B (-0.00174)	G (-0.00186)
2001年第四季度	I (-0.00045)	B (-0.00054)	CDFHKM (-0.00056)	AEGJ (-0.00089)	L (-0.00142)
2002年第一季度	EJ (0.00072)	CDFHKM (0.00017)	ABG (0.00016)	L (-0.00058)	I (-0.00072)
2002年第二季度	I (0.00202)	EJ (0.00070)	G (0.00067)	L (0.00046)	ABCDFHKM (0.00045)
2002年第三季度	B (-0.00045)	I (-0.00047)	ACDFHKM (-0.00051)	EJ (-0.00089)	GL (-0.00110)
2002年第四季度	L (-0.00117)	BCDFHKM (-0.00125)	EJ (-0.00137)	AG (-0.00139)	I (-0.00200)
2003年第一季度	I (0.00115)	L (0.00103)	DF (0.00087)	B (0.00077)	ACEGHJKM (0.00060)
2003年第二季度	BD (0.00030)	GL (-0.00020)	J (-0.00042)	I (-0.00061)	ACEFHKM (-0.00063)
2003年第三季度	G (-0.00037)	BF (-0.00039)	L (-0.00056)	ACDEHJKM (-0.00062)	I (-0.00116)

可以是线性无关、弱线性相关,也可以是强线性正相关或强线性负相关。将行业收益率序列的聚类分析、相关分析和平均收益率排序相结合,可为投资决策提供如下的参考信息(以表1中的数据为例):

(1)若属于同一类的行业中存在两个行业的收益率序列高度负相关,则该类中行业的收益率序列波动性不大,投资该类中的行业风险较小(如E与F)。

(2)某一类中若存在收益率序列波动大的行业,则该类中行业的收益率序列间高度正相关(如A与B)。

(3)同一类中行业的平均收益率相近(如A与B),但具有相近平均收益率的行业不一定属于同一类(如B与C)。投资于同一类中的不同行业,收益率序列表现接近(如A与B)。

(4)属于不同类行业的收益率序列间可高度正相关。属于不同类且收益率序列间高度正相关的行业间平均收益率差别大(如A与D,B与D)。

(5)若有两类的平均收益率接近,则这两类间正相关性不强,即分别取自两类的两个行业的收益率序列间正相关性不强(如{A,B}与{C})。

(6)对于积极型投资管理来讲,从不同类中动态选择行业进行资产组合比从同一类中动态选择行业更能体现择时能力的价值。例如,将A与C进行动态组合,假设投资者具有强预测能力,从而具有择时投资的能力,则可在第一天和第二天将全部资产投资于行业C,再从第三天开始将全部资产转移到行业A,则组合后的平均收益率明显大于任何A与C的消极组合。由于同一类中行业收益率序列间的高度相似性,投资于同一类中的行业的差别不大,从而在属于同一类的行业间通过择时进行动态资产配置意义较小。例如,A与B之间无论怎样择时进行动态资产配置,与单独投资于A、B两者之一,或两

者的某种固定组合的差别都不大(不考虑交易费)。

投资哪些行业以及资产如何在这些行业间分配是投资管理者必须考虑的问题。投资管理者在进行投资决策时,首先,要对行业的发展前景和收益率的高低进行预测分析。聚类分析有助于发现那些与市场表现接近的行业,我们可以借助于经济形势的分析来估计这些行业的发展趋势。同时,我们还观察到一些行业的收益率相对于市场(基准类)表现比较稳定,投资于这些行业可获得相对较稳定的平均收益。其次,投资管理者要在行业间配置资产,并且需要根据行业间的收益率关系的变化动态地调整其在行业间的资产配置。本文提出的分阶段聚类分析有助于投资管理者了解宏观经济事件对行业的前景及行业间关系的影响,为投资管理者有效地选择行业并在行业间合理地配置资产与适时调整资产提供参考信息。

四、结束语

本文提出对行业指数收益率序列分阶段进行聚类分析的动态分析方法,以考察行业间的相互关系及其演化过程。所提出的研究方法不依赖于行业收益率序列的概率分布特性。本文利用深交所的行业指数数据进行了实证分析,探讨了各行业相对于基准类(市场主流)的紧密性、稳定性,以及各行业间的相似程度,并且尝试考察了 QFII 事件前后行业指数收益率聚类关系的变化,对该事件在股票场所产生的影响进行了一定探讨。

本文的研究结果有助于加深投资管理者及监管部门对市场中各行业间相互关系的了解,并且可为投资决策提供参考信息。但也存在一些值得进一步探讨的问题,比如,是否可考虑收益率以外更多的行业聚类指标(如市盈率、每股净资产、每股净利润等),以更好地刻画各行业的特点。另外,本文提出的动态聚类分析方法还可用来对各类证券和国际金融市场等进行分析,对投资管理者进行资产配置具有参考价值。

参考文献:

- [1]Bodie Z, A Kane, A J Markus. Investments[M]. US, McGraw-Hill Irwin, 2002.
- [2]Grundy B F, S R Martin. Understanding the nature of the risks and the source of the rewards to momentum investing[J]. Review of Financial Studies, 2001, (14): 29~79.
- [3]Lee C, B Swaminathan. Price momentum and trading volume[J]. Journal of Finance, 2000, (55): 2017~2069.
- [4]Lessard, Donald. World, national, and industry factors in equity returns[J]. Journal of Finance, 1974, (24): 379~91.
- [5]Mazzucato M, W Semmler. Market share instability and stock price volatility during the US auto industry life cycle[J]. Journal of Evolutionary Economics, 1999, (9): 67~96.
- [6]Moskowitz T, M Grinblatt. Do industries explain momentum? [J]. Journal of Finance,

- 1999, (54): 1249~1290.
- [7] Roll Richard. Industrial structure and the comparative behavior of international stock indices[J]. Journal of Finance, 1992, (47): 3~41.
- [8] Schwert G W. Stock market volatility in the new millennium: How wacky is Nasdaq? [J]. Journal of Monetary Economics, 2002, (49): 3~26.
- [9] 范金城, 梅长林. 数据分析[M]. 北京: 科学出版社, 2002.
- [10] 范龙振, 王海涛. 上海股票市场行业与地区效应分析[J]. 系统工程学报, 2003, (2).
- [11] 何涛, 陈晓. 现金股利能否提高企业的市场价值——1997~1999年上市公司会计年度报告期间的实证分析[J]. 金融研究, 2002, (8).
- [12] 宋逢明, 梁洪昀. 发行市盈率放开后的A股市场初始回报率研究[J]. 金融研究, 2001, (2).
- [13] 田素华, 顾允平. WTO体制下中国服装业国际比较优势研究[J], 复旦学报(社会科学版), 2000, (2).
- [14] 王幼军. 加入WTO对我国各行业的影响[J]. 财经科学, 2000年增刊.
- [15] 赵宇龙, 易琮. 对我国各行业未来成长能力的实证考察: 一种市场视角[J]. 经济研究, 1999, (6).

Dynamic Clustering Analysis of Return Series of Industrial Indexes in Chinese Stock Market

LAO Lan-jun, SHAO Yu-min

(School of Management, Fudan university, Shanghai 200433, China)

Abstract: The paper suggests clustering the return series of industrial indexes stage by stage to explore the relations among industries and their evolution course. The paper makes an empirical study by using the industrial index data from Shenzhen Stock Exchange. It puts forward the concept of basic(bench mark) class, analyzing the closeness, steadiness of different industries with regard to the basic(bench mark)class, and the similarity of different industries. It also attempts to analyze the influence of some macroeconomic events on the mutual relations among industries. Our research methodology does not depend on the probability distribution of the return series and our research results can help the investors and supervisors to further understand the mutual relations among industries. Thus it has great reference value to investment decisions.

Key word: industry analysis; industry index; clustering analysis; macroeconomic analysis