

# 资产组合协方差矩阵的信息结构

罗 英<sup>1</sup>, 蔡玉梅<sup>2</sup>, 崔小梅<sup>2</sup>, 孙坚强<sup>2</sup>

(1. 厦门大学 管理学院, 福建 厦门 361005; 2. 华南理工大学 经济与贸易学院, 广东 广州 510006)

**摘 要:** 本文应用随机矩阵理论及方法解析资产组合协方差矩阵的信息结构。实证发现, 我国股票组合的协方差矩阵存在市场因素、行业因素的信息结构。最大特征值携带着反映市场因素的信息, 影响程度非常高, 是股票资产间相关性的主导因素。最大特征值偏离 RMT 上界的倍数远远高于其他市场。偏离 RMT 上界的特征值, 携带着反映行业因素的信息, 影响同一行业或者主营业务相同的公司, 但随着特征值与 RMT 上界靠近, 信息特征减弱。

**关键词:** 随机矩阵; 投资组合; 协方差矩阵

中图分类号: F832.5 文献标识码: A 文章编号: 1003-5192(2013)04-0026-05

## The Information Structure of the Covariances between Financial Returns

LUO Ying<sup>1</sup>, CAI Yu-mei<sup>2</sup>, CUI Xiao-mei<sup>2</sup>, SUN Jian-qiang<sup>2</sup>

(1. School of Management, Xiamen University, Xiamen 361005, China; 2. School of Economics and Commerce, South China University of Technology, Guangzhou 510006, China)

**Abstract:** We use the methods of random matrix theory (RMT) to investigate the information structure of the covariance matrix between stock returns. The largest eigenvalue are found to represent the market information. Chinese stock portfolio has a particularly high value of largest eigenvalue, which is about 175 times larger than the RMT upper bound. Market information is the dominate factor determining the correlations between stock returns. The other eigenvalues deviating from the RMT upper bound represent information affecting stocks belonging to similar or related industries.

**Key words:** random matrix; portfolio; covariance matrix

## 1 引言

金融资产间的交互相关作用, 表现为资产收益之间的协方差矩阵或者相关矩阵, 这对于资产组合选择和风险管理等重要的金融应用都是决定性的。目前理论研究和业界实践普遍根据历史数据经验地估计该矩阵。近期, Laloux 和 Plerou 等学者<sup>[1, 2]</sup>应用随机矩阵理论 (Random Matrix Theory, RMT) 研究发现, 经验估计的协方差矩阵或者相关矩阵, 具有随机矩阵的特征值, 存在噪音干扰, 经验矩阵携带的信息并非完全是资产交互作用的真实信息。资产之间交互作用的真实信息如何体现在协方差矩阵上, 或者说, 协方差矩阵的真实信息结构是重要的问题。我们将采用并扩展随机矩阵理论的方法, 针对中国股票市场的资产组合, 分析股票资产组合协方差矩阵的真实信息结构。

## 2 文献综述

Laloux 等<sup>[1]</sup>和 Plerou 等<sup>[2]</sup>较早地应用随机矩阵理论分析资产组合的相关矩阵或者协方差矩阵。他们分解矩阵的特征值谱, 并与随机矩阵进行对比。经验矩阵的特征值被分为两组, 一组落在随机矩阵特征值的理论取值范围 (也称 RMT 范围, Random Matrix Theoretical Bounds) 内, 如果这部分特征值具有 RMT 普适性质 (Universal Properties), 则为噪音特征值, 不携带真实信息; 另一组则在理论取值范围外 (也称偏离 RMT 范围, Deviations From the RMT), 部分偏离 RMT 上界, 部分偏离 RMT 下界, 这些特征值则为真实特征值, 可能携带着反映资产间交互作用的真实信息, 这为探析组合协方差矩阵的信息结构提供了线索。

对于落入 RMT 范围的信息成份, 多数文献已

收稿日期: 2013-01-26

基金项目: 国家社会科学基金资助项目 (09CJY013, 11CJY098); 教育部人文社会科学研究资助项目 (08JC790038); 广东省哲学社会科学“十一五”规划青年资助项目 (08YE-02); 中央高校科研业务费资助项目 (x2jmD2117980, x2jmD2118000)

经证实为噪音干扰,包括美国、日本等成熟金融市场的研究<sup>[3-6]</sup>,以及近期对南非、印度和中国等新兴金融市场的研究<sup>[7-9]</sup>。他们对股票市场资产组合的实证发现,根据历史数据估计的协方差矩阵或者相关矩阵,其特征值分布与随机矩阵的理论分布较为吻合,绝大部分的特征值落入 RMT 范围,且落入 RMT 范围内的特征值具有随机矩阵的普适性质。这意味着落入 RMT 范围内的特征值成份为噪音干扰,不携带反映资产交互作用的真实信息。

资产组合协方差或者相关矩阵的真实信息结构,目前文献较多倾向于以市场因素、行业因素结构进行解析。Laloux 和 Plerou 等实证发现<sup>[1,2]</sup>,相关矩阵的最大特征值明显偏离 RMT 上界,大小约为 RMT 上界的 25 倍。偏离程度越高,说明该特征值携带的信息越重要。Plerou 等发现<sup>[3]</sup>,最大特征值的特征向量值分布较为一致,方向相同,因此他们认为,最大特征值携带的信息为市场因素,该特征值量化了市场因素对所有股票的共同影响。Utsugi 等<sup>[5]</sup>在对日本市场的实证分析发现,最大特征值约为 RMT 上界的 52 倍,偏离程度更高。他们通过构建虚构的确定性相关矩阵来与资产组合真实的相关矩阵对比,并利用蒙特卡罗数值模拟进行分析,亦认为最大特征向量代表着市场信息。Wilcox 和 Gebbie<sup>[7]</sup>在对南非股票市场分析中发现最大特征值约为对应的 RMT 理论上界的 9.5 倍,不同的是最大特征值的特征向量对应持有市值大、交易量大的多头策略。Pan 和 Sinha<sup>[8]</sup>对印度市场分析发现最大特征值约为 RMT 理论上界的 28 倍,他们同时检验了较大特征值信息成份的稳定性,发现最大特征值的信息成份最稳定。

Plerou 等<sup>[3]</sup>同时分析了前 10 个偏离 RMT 上界的特征值。在通过单因子定价模型移除最大特征值的贡献后,他们发现,这些特征值的特征向量分量值较大的往往是来自同一行业的公司,或者是主营业务相同。这意味着偏离 RMT 上界的大特征值携带影响同行业公司的信息。如第三大特征值对应着电子和计算机行业。尽管个别特征值对应的公司明显来自不同行业,如第五大特征值,贡献最大的几个公司来自电子通讯、钢铁开采和银行业,但这些公司在拉丁美洲的经营活动都很活跃。Utsugi 等<sup>[5]</sup>则对比日本市场 TSE 股票和美国 S&P500 成份股两组样本,发现结构有一定差异。S&P500 组合相关矩阵中,对应较大特征向量的只有一个特定行业,但 TSE 股票组合相关矩阵的大特征值却会出现若干不同行业的公司。结合数值

的分析,他们认为几个行业的企业可以形成一个较强的行业相关组,信息共同表现在某个大特征值上,这可能是由于 TSE 中有更复杂的市场结构,或是两个国家的经济结构差异所导致。

行业信息结构在其他市场却表现较不明显。Pan 和 Sinha<sup>[8]</sup>对印度股票市场分析,只发现第三大特征值明显对应技术、IT 及通讯行业的公司,其余偏离特征值没有特定的行业特征。他们认为市场、行业信息结构的出现,是该市场是否进入成熟市场的标志性特征。Wilcox 和 Gebbie<sup>[7]</sup>发现南非股票组合协方差矩阵的偏离特征值与行业因素并不存在明显的对应关系。他们同时提出另外一种解析结构,也即最大特征值对应持有市值大、交易量大的多头策略,而次大特征值的特征向量对应持有市值大、交易量大的空头策略。

### 3 协方差矩阵的估计与特征值分解

考虑  $N$  个资产,  $P_i(t)$  为资产  $S_i (i = 1, 2, \dots, N)$  在第  $t$  天的价格,资产的日收益率定义为

$$r_i(t) = \ln P_i(t+1) - \ln P_i(t) \quad (1)$$

资产组合协方差矩阵  $\Sigma$  的经验估计是

$$COV_{ij} = \frac{1}{L} \sum_{t=1}^L (r_i(t) - \bar{r}_i)(r_j(t) - \bar{r}_j), \quad i, j = 1, 2, \dots, N \quad (2)$$

其中  $\bar{r}_i$  为收益率的样本均值,  $L$  为日收益率时间序列的长度。

我们选取上海证券交易所和深圳证券交易所 2011 年市值最大的 1000 只股票,时间范围为 2001 年 1 月 1 日至 2011 年 12 月 31 日,共 2662 个交易日。剔除缺失交易日超过 242 天的股票(2001 ~ 2011 年年平均交易日为 242 天),保留样本股票为 445 只。所选股票 2011 年成交金额占整个市场成交金额的 50.67%,2011 年市值占整个市场的 32.44%。数据来源于国泰安 CSMAR 中国股票市场交易数据库。对缺失交易数据采用线性插值<sup>[7]</sup>方法进行处理。

分解协方差矩阵得到矩阵的特征值  $\lambda_i (i = 1, 2, \dots, N)$  及其对应的特征向量  $V_i (i = 1, 2, \dots, N)$ 。这里特征值大小顺序排序,  $\lambda_1 > \lambda_2 > \dots > \lambda_N$ 。矩阵特征值的密度函数为  $\rho(\lambda)$

$$\rho(\lambda) = \frac{1}{N} \frac{dn(\lambda)}{d\lambda} \quad (3)$$

其中  $\lambda$  为矩阵特征值,  $n(\lambda)$  表示特征值小于  $\lambda$  的个数。

相应地,考虑  $N$  维的随机矩阵  $H, H = (1/L)AA^T$ 。其中矩阵  $A$  为  $N \times L$  矩阵,矩阵  $A$  的元素为独立同分

布的随机变量 具有零均值和有限方差  $\sigma^2$ 。根据随机矩阵中心极限定理,当  $N, L \rightarrow \infty$  且  $Q = L/N \geq 1$  固定时,随机矩阵特征值的密度函数为

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \lambda \in [\lambda_{\min}, \lambda_{\max}] \quad (4)$$

$$\lambda_{\min}^{\max} = \sigma^2 (1 + 1/Q \pm 2\sqrt{1/Q}) \quad (5)$$

这是随机矩阵特征值的理论分布,  $[\lambda_{\min}, \lambda_{\max}]$  为随机矩阵特征值的取值范围,也称为 RMT 范围,  $\lambda_{\min}$  和  $\lambda_{\max}$  分别为 RMT 下界和 RMT 上界。

分解样本股票组合的协方差矩阵,对于  $N = 445$  只股票,  $L = 2661$  天,则  $Q = 5.9798$ 。简单计

表 1 股票组合协方差矩阵最大的 20 个特征值

序号	特征值	序号	特征值	序号	特征值	序号	特征值
1	191.54	6	2.9096	11	2.1079	16	1.7273
2	7.7498	7	2.7164	12	2.0797	17	1.6883
3	5.2311	8	2.5073	13	1.9350	18	1.6686
4	4.9034	9	2.3126	14	1.9014	19	1.6133
5	3.7479	10	2.1803	15	1.8538	20	1.5576

#### 4 协方差矩阵的信息结构解析

落在 RMT 范围内的特征值被多数学者证实为噪音特征值,不携带信息。那么 RMT 范围外的特征值,则可能携带反映资产之间交互作用的真实信息。相对而言,偏离 RMT 范围的特征值更有意义,因为他们反映了资产之间产生作用的内在原因。

##### 4.1 市场因素成份与最大特征值

首先分析最大特征值  $\lambda_1$ ,这是偏离 RMT 上界最远的特征值。偏离 RMT 上界越远,该特征值携带的“信息”越多、也越重要。表 1 显示,最大特征值为 191.54,是调整后 RMT 上界的 174.75 倍。这一倍数远远高于其他股票市场,美国市场约为 25 倍<sup>[3]</sup>,日本市场约为 52.87 倍<sup>[5]</sup>,南非市场约为 9.5 倍<sup>[7]</sup>。

特征向量的 445 个分量——对应着各个资产,分量值的大小则反映了该特征值的信息对相应股票的影响程度,或股票在该特征值信息成份的贡献大小。对最大特征值对应的特征向量进行分析发现,最大特征向量的分量均为正数,大小较为一致,取值范围为 0.0287 至 0.0556,平均值为 0.0472,标准差仅为 0.0048(均值的 10.08%)。这意味着所有股票方向一致,程度较为一致地受到该特征值所携带信息的影响。自然地,我们将这一信息成份解释为影响所有股票的市场因素。

为了验证上述解释,我们以最大特征向量的分

算,随机矩阵特征值的理论上下界分别为  $\lambda_{\min} = 0.3494$ ,  $\lambda_{\max} = 1.9851$ 。表 1 给出经验矩阵最大的 20 个特征值,注意到有 12 个大于 RMT 上界,  $\lambda > \lambda_{\max}$ 。这些特征值明显偏离 RMT 上界,所携带的信息并非随机噪音。参考 Laloux 等<sup>[1]</sup>的方法,移除最大两个特征值在总方差  $\sigma^2$  中的贡献,调整后的方差为  $[\sigma_{ad}^2 = \sigma^2 (1 - (\lambda_1 + \lambda_2)/N)]$ 。相应地,调整后, RMT 下界为  $\lambda_{\min} = 0.1929$ , RMT 上界为  $\lambda_{\max} = 1.0961$ 。对比发现,有 38 个特征值偏离 RMT 上界, 62 个特征值偏离 RMT 下界,共 100 个特征值偏离 RMT 范围。

量值作为权重,构建如下虚拟市场指数

$$M(t) = \sum_{j=1}^N V_{1j} r_j(t) \quad (6)$$

其中  $V_1$  为最大特征值对应的特征向量,  $V_{1j} (j = 1, 2, \dots, N)$  为向量的分量。图 1 描绘了该虚拟指数在 2001 年 1 月 1 日至 2011 年 12 月 31 日期间的日收益。我们同时选择上证综合指数同时期的序列作为对比,如图 2。对比发现,虚拟市场指数与上证综合指数的变化高度一致,简单计算,两个指数的相关系数高达 0.9274。因此,将最大特征值携带的信息成份解释为市场因素信息是合理的。

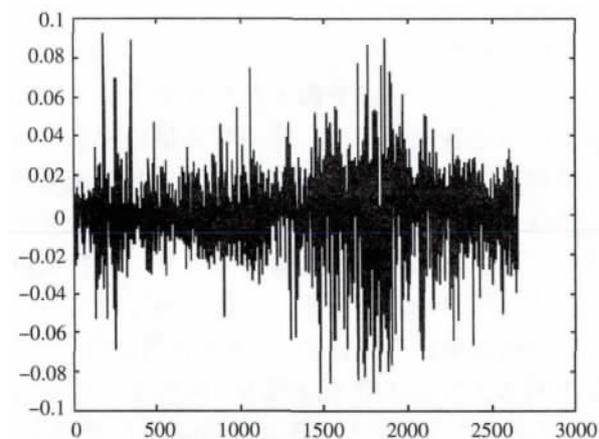


图 1 虚拟市场指数日收益序列

对于中国股票资产组合,协方差矩阵最大特征值远远偏离 RMT 上界,对这一“高倍数”现象,一个可能的解释是,作为新兴市场的中国股票市场,

股票价格变动较大程度上受市场共同因素的影响。事实上,中国股票市场出现绝大多数股票同涨同跌的现象并不少见。

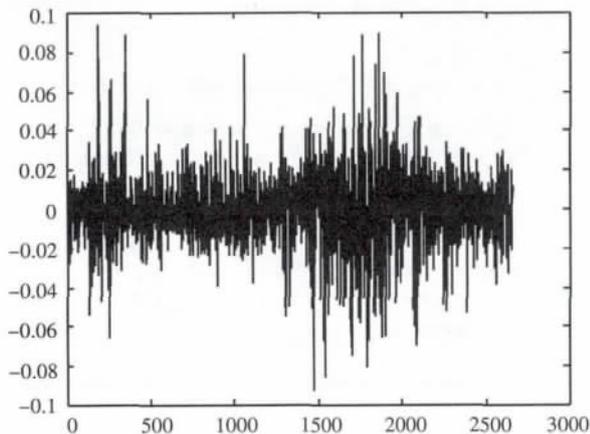


图2 上证综合指数的日收益序列

进一步,我们采用单因子模式提取协方差矩阵最大特征值的信息。单因子基本模型如下

$$r_i(t) = \alpha_i + \beta_i M(t) + \varepsilon_i(t), \quad t = 1, 2, \dots, L \quad (7)$$

其中参数  $\alpha$  和  $\beta$  根据最小二乘方法获得。将各股票收益率序列  $r_i(t)$  的市场影响  $M(t)$  移除,以剩余的残差  $\varepsilon_i(t)$  重新构建协方差矩阵。进行特征值分解后,发现移除市场因素后协方差矩阵的最大特征值显著下降为 13.57,偏离 RMT 上界倍数亦显著下降为 12.37 倍。并且最大特征值的特征向量各分量不再表现一致性。此外,移除市场因素前,股票之间平均相关程度为 0.4262(相关矩阵上三角元素的均值),移除市场因素后,股票之间的平均

相关程度显著下降为 0.0001。这些不仅证实最大特征值携带市场因素信息,同时反映出对于中国市场的股票组合,资产间的相关性主要由市场因素主导。

#### 4.2 行业因素与偏离 RMT 上界的特征值

最后,我们分析其他偏离 RMT 上界的特征值。对于每个特征值,选取特征向量正负分量最大的各 10 只股票。分析发现,这些受到该特征值影响程度较高的股票往往来自同一行业板块或者有着相同的主营业务。从第二至第十特征值对应的行业、影响方向、代表性股票及其行业分类号如表 2 所示。对于第二特征值  $\lambda_2$ ,正分量有 9 个股票来自医药板块,包括医药制造业、药品及医疗器械批发业和生物制品业;对于第三特征值  $\lambda_3$ ,所有负分量股票均来自房地产开发与经营业,其中中天城投 2011 年 94.6% 的主营业务收入来自房地产开发与经营;第四特征值  $\lambda_4$ ,正向影响房地产开发与经营,负向则影响黑色金属冶炼及压延加工业,这个特征值携带的板块因素最为清晰;第五特征值  $\lambda_5$ ,正向影响有色金属冶炼及压延加工业,负向影响黑色金属冶炼及压延加工业;第六和第七特征值没有明显的版块特征值;第八特征值同样影响医药板块;第九、十特征值分别对应煤炭采选业、黑色金属冶炼及压延加工业和交通运输设备制造业。前十大特征值的信息特征较为明显。但随着离 RMT 上界越近,信息特征越来越弱,第十一以上的特征值,没有发现清晰的信息特征。综合来看,偏离 RMT 上界的特征值,携带影响某个行业板块的行业信息。

表 2 偏离 RMT 上界的特征值的信息成份

特征值	行业	影响方向	代表性股票及行业分类号
第二特征值	医药制造业、生物制品业、药品及医疗器械批发业	正向	哈药股份 C81 长春高新 C85 华东医药 H0160 昆明制药 C81 海南海药 C81 中新药业 C81 华润三九 C81 国药一致 H0160
第三特征值	房地产开发与经营业	负向	中天城投 M ( 主营业务房地产开发与经营) 上实发展 J01 中弘股份 J01 天房发展 J01 泛海建设 J01 首开股份 J01 深振业 A J01 名流置业 J01 中粮地产 J01 万业企业 J01
第四特征值	黑色金属冶炼及压延加工业	负向	凌钢股份 C65 酒钢宏兴 C65 包钢股份 C65 西宁特钢 C65 韶钢松山 C65 马钢股份 C65 南钢股份 C65
	房地产开发与经营业	正向	中华企业 J01 首开股份 J01 泛海建设 J01 金融街 J01 招商地产 J01 金地集团 J01 万科 AJ01
第五特征值	黑色金属冶炼及压延加工业	负向	马钢股份 C65 南钢股份 C65 韶钢松山 C65 宝钢股份 C65 包钢股份 C65 武钢股份 C65
	有色金属冶炼及压延加工业	正向	包钢稀土 C67 焦作万方 C67 铜陵有色 C67 云铝股份 C67 中金岭南 C67 锡业股份 C67 云南铜业 C67

续表 2

特征值	行业	影响方向	代表性股票及行业分类号
第八特征值	医药制造业	正向	云南白药 C81 同仁堂 C81 海正药业 C81 中新药业 C81 华东医药 H0160
		负向	辽宁成大 H21 (2011 年生物制药占主营业务收入 11.3%)
第九特征值	煤炭采选业	正向	兖州煤业 B01 煤气化 B01 冀中能源 B01 西山煤电 B01 盘江股份 B01 伊泰 B 股 B01
第十特征值	黑色金属冶炼及压延加工业	负向	韶钢松山 C65 鞍钢股份 C65 宝钢股份 C65 马钢股份 C65
	交通运输设备制造业	正向	江铃汽车 C75 长安汽车 C75 一汽轿车 C75 一汽夏利 C75 上汽集团 C75 福田汽车 C75 洪都航空 C75 中联重科 C73 (2011 年工程机械占主营业务收入 95.7%)

## 5 结论

本文解析我国股票市场资产组合协方差矩阵的信息结构,主要的实证结论包括:

第一,对于我国股票组合的协方差矩阵,存在市场因素、行业因素的结构特征。第二,协方差矩阵的最大特征值携带着反映市场因素的信息,其影响程度非常高,最大特征值偏离 RMT 上界的倍数远远高于其他市场。在移除市场因素后,最大特征值偏离倍数显著下降,市场因素是股票组合资产间相关性的主导因素。第三,偏离 RMT 上界的较大特征值,对应行业因素,受其携带的信息影响程度较高的公司往往来自同一行业,或者具有相同的主营业务。前十大特征值的信息特征较为明显,但随着靠近 RMT 上界,信息特征减弱,第十一以上的特征值,没有发现清晰的信息特征。

目前解析资产组合协方差矩阵的信息结构,主要根据特征值及对应的特征值向量判断,高分量值对应的公司,受其信息影响程度高,这些公司的营业特征提供了信息结构的线索。但对于像我国股票市场和日本股票市场,市场因素占主导地位,剩余的信息结构则表现不清晰,另一重要指标反向参与比可能比特征向量本身更有助于辨析信息结构,这是值得后续分析的问题。

## 参 考 文 献:

- [1] Laloux L, Cizeau P, Bouchaud J, et al. . Noise dressing of financial correlation matrices [J]. Physical Review Letters, 1999, 83(7): 1467-1470.
- [2] Plerou V, Gopikrishnan P, Rosenow B, et al. . Universal and nonuniversal properties of cross correlations in financial time series [J]. Physical Review Letters, 1999, 83(7): 1471-1474.
- [3] Plerou V, Gopikrishnan P, Rosenow B, et al. . Random matrix approach to cross correlations in financial data [J]. Physical Review E, 2002, 65(6): 066126-1-066126-48.
- [4] Pafka S, Potters M, Kondor I. Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization [EB/OL]. <http://arxiv.org/abs/cond-mat/0402573>, 2004-02-24.
- [5] Utsugi A, Ino K, Oshikawa M. Random matrix theory analysis of cross correlations in financial markets [J]. Physical Review E, 2004, 70(2): 026110-1-026110-17.
- [6] Daly J, Crane M, Ruskin H J. Random matrix theory filters in portfolio optimization: a stability and risk assessment [J]. Physica A: Statistical Mechanics and Its Applications, 2008, 387(16): 4248-4260.
- [7] Wilcox D, Gebbie T. An analysis of cross-correlations in an emerging market [J]. Physica A: Statistical Mechanics and Its Applications, 2007, 375(2): 584-598.
- [8] Pan R K, Sinha S. Collective behavior of stock price movements in an emerging market [J]. Physical Review E, 2007, 76(4): 046116-1-046116-10.
- [9] 罗英. 股票市场资产交叉相关的随机信息与真实信息研究 [D]. 厦门: 厦门大学, 2010.