

基于多重分形参量 C_q 的股票市场分析

张琛, 郑婷婷, 万涛, 程花花, 章意成

(安徽大学 数学科学学院, 安徽 合肥 230039)

摘要:对沪深300整体指数及其中相关行业的股票进行多重分形分析,发现股票市场具有多重分形性。其中金融业、运输业、电气业的股票指数波动较小,钢铁业、煤炭业和地产业的股票指数波动较大,土木建筑业、医药业的股票指数波动处于中间位置。分析参量 C_q 的图像发现每一支股票都有自身的临界点,并且在由 C_q 构成的三维图像中代表同一行业股票的点比较集中,形成一个簇,从该图中可以看出整个大盘指数与钢铁业、煤炭业和地产业的联系比较紧密。

关键词:多重分形;多重分形谱;广义分形维数 D_q ;参量 C_q

中图分类号:F830.9

文献标识码:A

文章编号:1673-629X(2011)01-0178-03

Stock Market Analysis Based on Parameter C_q of Multifractal

ZHANG Chen, ZHENG Ting-ting, WAN Tao, CHENG Hua-hua, ZHANG Yi-cheng

(School of Mathematical Sciences, Anhui University, Hefei 230039, China)

Abstract:Based on the multifractal analysis of the Shanghai and Shenzhen 300 index and stocks related industries, found the stock market has multifractal nature. Financial sector, transportation, electrical industry have smaller stock index volatility, steel industry, coal industry and real estate have larger stock index volatility, civil construction, pharmaceutical industry in the middle of stock index volatility. By analyzing the parameter C_q find that each stock has its own critical point, and in three-dimensional images constituted by C_q to represent shares of point in the same industry are concentrated, forming a cluster. From the figure can see that the relationship between the whole market index and steel industry, coal industry, real estate sectors are more closely.

Key words:multifractal; multifractal spectrum; generalized fractal dimension D_q ; parameter C_q

0 引言

多重分形分析的方法广泛应用于金融序列中,早期使用的方法是多重分形消除趋势波动分析(multifractal detrended fluctuation analysis),简称MF-DFA方法,它可以描述金融序列在不同时间标度上的统计特征^[1-4],由此确定了广义Hurst指数,发现了金融序列具有多重分形特征。对金融序列进行重排处理和相位随机化处理,将处理后的序列与原序列进行比较,可以得到引起金融序列多重分形特征的因素。除此之外,利用多重分形谱的宽度能表示金融序列的波动情况,多重分形谱的宽度越大,金融序列分布越不均匀,多重分形强度越大。在文中,利用多重分形谱的宽度,广义分形维数 D_q ,以及改进的参量 C_q 等对金融序列进行进一步的分析研究,重点利用由参量 C_q 构成的三维图像来分析沪深300中相关行业股票的联系,以及整个

大盘指数与哪些行业的股票联系最为紧密。

1 多重分形

多重分形也称作多标度分形,是Mandelbrot在1972年研究湍流时首先提出的。多重分形,是定义在分形结构上的由多个标度指数的分形测度组成的无限集合,它刻画了分布在子集上的具有不同标度和标度指数的分形子集的局部标度性。后来被广泛应用于股票市场^[4,5],土壤^[6],生物信息学^[7-9],空气质量^[10]等领域。

1.1 多重分形定义

(1)设研究对象为一长度为 N 的时间序列 $x(t)$, $t = 1 \cdots N$,对其进行归一化处理得到

$$p(t) = \frac{x(t)}{\sum x(t)}, t = 1 \cdots N$$

将归一化后的序列成长度相同的互不重合的 T 个小区间。利用盒计数法,设每个盒子的长度为 ε ,分形体生长界面在该小区域的生长几率为 $P_i(\varepsilon)$,不同小区域生长几率不同,可用不同指数 α 来表征, $P_i(\varepsilon) \propto \varepsilon^\alpha$, $t = 1 \cdots N$

收稿日期:2010-05-16;修回日期:2010-08-21

基金项目:安徽省高等学校省级自然科学研究项目(KJ2007B239)

作者简介:张琛(1986-),女,安徽淮南人,硕士研究生,研究方向为智能计算、分形、金融工程;郑婷婷,博士,副教授,硕士生导师,研究方向为分形、智能计算、金融工程。

(2) 定义一个配分函数 $\chi_q(\varepsilon)$, 对概率 $P_i(\varepsilon)$ 用 q 次方进行加权求和, 其数学表达式为:

$$\chi_q(\varepsilon) = \sum P_i(\varepsilon)^q = \varepsilon^{\tau(q)}$$

如果上式后面的等式成立, 即配分函数和 ε 有幂函数关系, 则可以从 $\ln \chi_q \sim \ln \varepsilon$ 曲线斜率得到

$$\tau(q) = \frac{\ln \chi_q(\varepsilon)}{\ln \varepsilon} \quad (\varepsilon \rightarrow 0)$$

一般把 $\tau(q)$ 称为质量指数。

(3) 由 $\tau(q)$ 可以得到广义分形维数 D_q :

$$\text{当 } q \neq 1 \text{ 时 } D_q = \frac{\tau(q)}{q-1} = \frac{\ln \chi_q(\varepsilon)}{(q-1) \ln \varepsilon} \quad (\varepsilon \rightarrow 0)$$

$$\text{当 } q = 1 \text{ 时 } D_1 = \frac{\sum P_i \ln P_i}{\ln \varepsilon} \quad (\varepsilon \rightarrow 0)$$

其中 D_1 称为信息维数, D_2 称为相关维数, 在以 q 作为横坐标, D_q 作为纵坐标的直角坐标系上, 如果 D_q 是一条接近于 1 的平滑直线, 即 D_q 与 q 不相关, 则可以说时间序列服从随机游走, 序列的波动较小, 反之则拒绝随机游走假设; 若 D_q 与 1 偏离越大, 随机游走假设就越弱, 序列的波动越大。

(4) 由 Legendre 变换得到

$$\alpha = \frac{d\tau(q)}{dq}, f(\alpha) = \alpha q - \tau(q)$$

α 是奇异指数, 用来描述复杂体系中的各个区间不同的奇异性, α 越小, 奇异性越大, α 越大, 奇异性越小。 $f(\alpha)$ 称为多重分形谱。多重分形谱的宽度为 $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$, 表征了最大概率和最小概率之间的差别, $\Delta\alpha$ 越大表示时间序列分布越不均匀, 多重分形强度越大^[11]。

(5) 利用 $\tau(q)$ 还能得到类似热能函数的量 C_q :

$$C_q = -\frac{\partial^2 \tau(q)}{\partial q^2} \approx \tau(q+1) - 2\tau(q) + \tau(q-1)$$

利用 C_q 的图像得到时间序列走势的临界点, 在该点系统将发生突变, Canessa 在文献[12]中利用 C_q 的图像发现了美国股市走势的临界点, 例如: 1987 年的股灾“黑色星期一”。

1.2 实验结果与分析

文中的数据主要来自沪深 300 中从 2005 年 1 月 1 日—2009 年 12 月 31 日的股票日收盘价。所选取的股票反映了不同行业的景气状况及其股价整体变动状况, 市值覆盖率高, 与现有市场指数相关性高, 样本股集中了市场中大量优质股票, 具有良好的市场代表性。

根据文献[11], 计算发现股票市场具有多重分形的特征。对金融业, 运输业, 电气业, 土木建筑业, 医药业, 钢铁业, 煤炭业, 地产业这八个行业内的所有股票, 以及沪深 300 整体指数的收盘价进行多重分形分析, 为了便于对结果进行直观的分析, 文中仅呈现出每个

行业中的一支股票的实验结果, 另外, 文中选取的 q 的范围均是从 -10 到 10, 中间间隔 1 个单位。如图 1 所示。

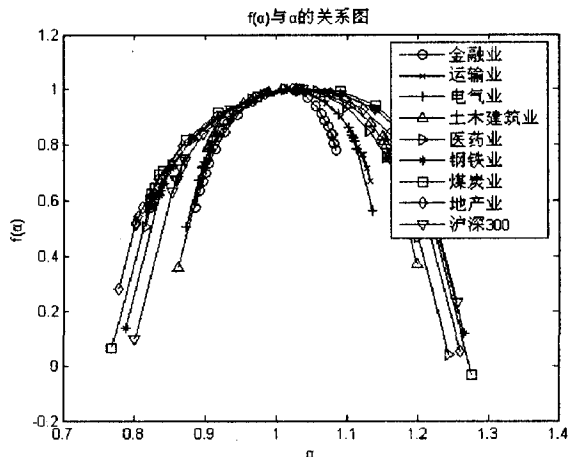


图1 各行业股票的多重分形谱函数 $f(\alpha)$

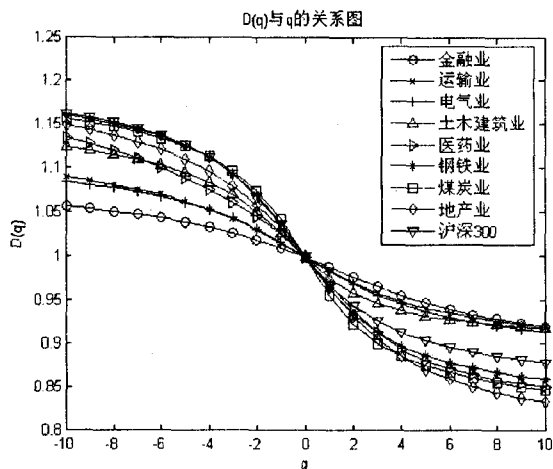
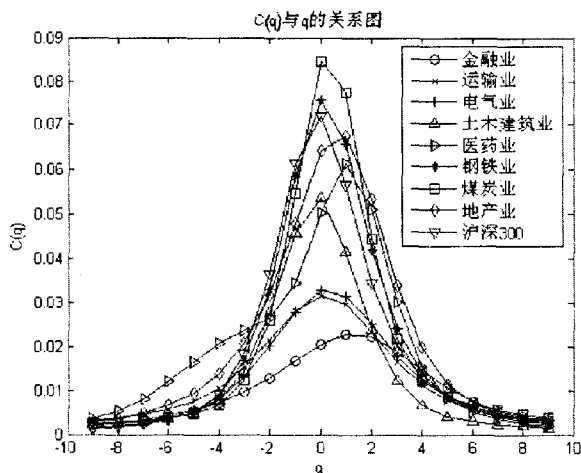
多重分形谱的宽度 $\Delta\alpha$ 反映了序列的涨落程度。 $\Delta\alpha$ 越大表示金融序列分布越不均匀, 多重分形强度越大, 即股票指数波动越激烈, 反之, 表示股票指数波动越小; 当 $\Delta\alpha = 0$ 时, 表示完全均匀分布的情况。结合图 1 与表 1, 发现金融业, 运输业, 电气业的多重分形谱宽度 $\Delta\alpha$ 的平均值较小, 所以股票指数的波动较小, 钢铁业, 煤炭业和地产业的多重分形谱宽度 $\Delta\alpha$ 的平均值较大, 股票指数波动较大, 而土木建筑业, 医药业的股票指数波动处于中间位置。

表1 每个行业的多重分形谱宽度 $\Delta\alpha$ 的平均值以及沪深 300 整体指数的多重分形谱宽度

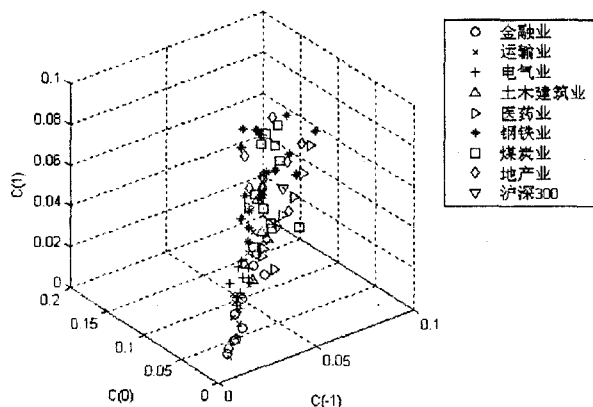
行业	金融	运输	电气	土木建筑	医药
$\Delta\alpha$	0.2275	0.2337	0.2644	0.3417	0.3997
行业	钢铁业	煤炭	地产	沪深 300	
$\Delta\alpha$	0.4555	0.4562	0.4657	0.4556	

这点也可以图 2 看出, 金融业, 运输业, 电气业的广义分形维数 D_q 与 1 偏离较小, 故股票指数波动较小, 钢铁业, 煤炭业和地产业的分形维数与 1 偏离较大, 故股票指数波动较大, 而土木建筑业, 医药业的 D_q 曲线处在中间位置。

在图 3 中, 发现每条曲线都有一个峰值, 即在该点系统将会发生突变。表现在股票市场上, 即股票指数在该点将会发生较大的波动。从该图中发现金融业, 运输业, 电气业的曲线处于较低的位置, 指数波动较小, 钢铁业, 煤炭业和地产业的曲线处于较高的位置, 指数波动较大。土木建筑业, 医药业的曲线位于中间的位置。这个性质与上面利用多重分形谱宽度和广义分形维数 D_q 分析得到的结果一致。并且只有在 D_q 曲线较为光滑时, 研究 C_q 曲线才是有意义的。

图2 各行业股票的广义分形维数 $D(q)$ 图3 各行业股票的类似热函数 $C(q)$ 的图像

对每支股票分别取 $q = -1, 0, 1$ 时的 C_q 值, 构成一个三维向量 $\{C_{-1}, C_0, C_1\}$, 每支股票都有一个这样的三维向量并且分别对应于图 5 中的一个点。

图4 各行业股票关于三维向量 $\{C_{-1}, C_0, C_1\}$ 的点分布图

从图 4 可以发现, 代表同一行业股票的点相对而言比较集中, 不同行业股票的点则比较分散, 距离较近的点所代表的股票具有一定的相似性。金融业, 运输业, 电气业的点聚成一簇, 土木建筑业, 医药业的点聚

成一簇, 钢铁业, 煤炭业, 地产业的点聚成一簇, 并且从图中发现沪深 300 整体指数的点落在工业, 煤炭业和地产业所形成的区域中, 可以认为沪深 300 整体指数与钢铁业, 煤炭业和地产业联系较大, 即整个大盘指数与钢铁业, 煤炭业和地产业的联系最为紧密。

2 结束语

在利用多重分形分析股票市场时, 可以较好反映股票市场内部的波动情况, 是迄今为止比较全面的描述金融序列波动情况的模型。在文中, 计算出各个行业的股票波动情况, 并分析出同一行业股票的波动具有一定的相似性, 沪深 300 大盘指数与钢铁业, 煤炭业和地产业的联系最为紧密。在以后的研究中可以进一步研究 C_q 的图像与股票波动的联系, 例如可以利用这些多重分形方法得到的数据, 结合智能计算方法进行聚类, 这些都有待于日后的研究。

参考文献:

- [1] 苑莹, 庄新田. 国际汇率的多重分形消除趋势波动分析[J]. 管理科学, 2007, 20(4): 80-85.
- [2] 施锡铨, 艾克凤. 股票市场风险的多重分形分析[J]. 统计研究, 2004(9): 33-36.
- [3] 卢方元. 中国股市收益率的多重分形分析[J]. 系统工程理论与实践, 2004(6): 50-54.
- [4] 曹广喜, 史安娜. 上海股市收益的多重分形分析——滑动窗 MFDDFA 方法的应用[J]. 数理统计与管理, 2007, 26(5): 875-880.
- [5] 魏宇, 黄登仕. 基于多标度分形理论的金融风险测度指标研究[J]. 管理科学学报, 2005(8): 50-59.
- [6] 白晓宇, 袁峰, 周涛发, 等. 多重分形方法识别铜陵矿区土壤中 Cd 的地球化学异常[J]. 矿物岩石地球化学通报, 2008(3): 702-704.
- [7] 郑婷婷, 毛军军. 蛋白质序列中的多重分形分析[J]. 计算机工程与应用, 2007, 43(20): 46-48.
- [8] Zhou Li-Qian. A fractal method to distinguish coding and non-coding sequences in a complete genome based on a number sequence representation[J]. Journal of Theoretical Biology, 2005, 232: 559 - 567.
- [9] Zu-Guo Yua, Vo Anh, Ka-Sing Lau. Chaos game representation of protein sequences based on the detailed HP model and their multifractal and correlation analyses[J]. Journal of Theoretical Biology, 2004, 226: 341 - 348.
- [10] Anh V V, Leung Y, Lam K C, et al. Multifractal characterization of Hong Kong air quality data [J]. Environmetrics, 2005, 16: 1-12.
- [11] 孙霞, 吴自勤, 黄均. 分形原理及其应用[M]. 合肥: 中国科学技术大学出版社, 2003.
- [12] Canessa E. Multifractality in time series [J]. J. Phys. A: Math. Gen, 2000, 33: 3637 - 3651.