

基于可见图的沪深股市波动分析

张帆, 严广乐

(上海理工大学 管理学院, 上海 200093)

摘要: 利用可见图算法, 考虑沪深股市不同时间频率(日、周、月)的收益率序列, 将其映射成网络, 发现所构造的网络具有分层结构和无标度性. 网络节点度分布幂律指数与 Hurst 指数服从线性规律, 验证了可见图算法计算 Hurst 指数的可靠性. 同时, 截取 2007~2008 年沪深股市大涨大跌收益率序列, 将其分别映射成网络, 得出股市大涨与大跌在网络拓扑结构方面的非平凡性质.

关键词: 可见图算法; 无标度性; Hurst 指数; 分层结构; 复杂网络

中图分类号: N 949

文献标志码: A

Stock Market Volatility Analysis Based on Visibility Graph Algorithm

ZHANG Fan, YAN Guang-le

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: By means of visibility graph algorithm, the indices of monthly, weekly and daily returns of Shanghai and Shenzhen stock markets were investigated. It is found that the series can be converted into a scale-free network with hierarchical structure. The relationship between the scaling exponents of the degree distributions and the Hurst exponents obeys the linear law, which proves the visibility graph can give reliable values of Hurst exponents of the series. Comparing the returns of Bull and Bear markets between 2007 and 2008, it reveals the nontrivial nature of the network topology structure of stock market.

Key words: *visibility graph algorithm; scale-free; Hurst index; hierarchically structure; complex network*

复杂网络是对复杂系统的抽象和描述方式, 它突出强调了系统结构的拓扑特征. 原则上说, 任何包含大量组成单元(或子系统)的复杂系统, 都可以当作复杂网络来研究^[1-2]. 目前, 复杂网络已经成为了

学术界研究的一个热点, 它在工程技术、社会、政治、医药、经济、管理等领域都有着潜在的广泛应用. 证券市场作为一个复杂的经济系统, 可以用复杂网络理论来抽象和描述. 复杂网络理论是研究证券市场

收稿日期: 2013-05-27

基金项目: 上海市一流学科建设资助项目(S1201YLXK)

第一作者: 张帆(1989-), 男, 硕士研究生. 研究方向: 经济金融复杂系统. E-mail: zhangfan8998@126.com

通讯作者: 严广乐(1957-), 男, 教授. 研究方向: 复杂系统理论与方法. E-mail: glyan2003@yahoo.com.cn

内在结构和功能的有力工具,利用复杂网络的理论和研究方法研究股票关联网络的拓扑性质和聚类结构对于理解网络的形成机制、发生在网络上的动力学行为具有重要意义.近年来研究学者将复杂网络理论运用于证券市场分析,大部分学者使用股票波动收益率序列的相关性来构建网络.例如,Boginski等^[3]研究了美国证券市场6546支股票,发现股票相关性呈现无标度性.庄新田等^[4]以上证交易所交易的股票为节点,股票间相关性为边构建无向无权网络,并发现该网络具有小世界效应和无标度特性.此外,一些学者将可见图算法来构建复杂网络.例如,Ni等^[5]考察了分数布朗运动和多重分形随机游走序列的可见图方法,发现这种机制生成的多分形的可见图具有无尺度性质,分数布朗运动的可见图的度分布满足幂律函数;理论推导知,幂律指数 α 是Hurst指数 H 的线性函数: $\alpha = 3 - 2H$.Yang等^[6]首次应用于汇率序列的分析,发现汇率序列的非平凡性质.Shao等^[7]将其应用于心跳信号分析,研究发现,相应的网络都是无标度网络,具有明显的聚类系数、层次结构.目前,运用可见图算法进行证券市场分析还没有深入展开,本文将可见图算法应用于证券市场波动分析,对不同频率及涨跌数据分组,构建无向无权网络,研究证券市场网络拓扑结构,为证券市场波动分析提供一个全新的思路.

1 理论与方法

1.1 可视算法

可见图方法是关于将时间序列映射为相应网络的数学算法^[8],该方法的优点不但保留了原有序列的大部分性质,而且通过研究网络参数可以提供更多关于有关序列的信息.

先假设一标量序列为 $\{x_i | i = 1, 2, \dots, K\}$,其中, K 为记录序列的最大值.若对于该序列任意两节点 $A(a, x_a), B(b, x_b)$ 之间的任意节点 $C(c, x_c)$ 满足条件

$$x_c \leq x_b + (x_a - x_b) \frac{b - c}{b - a} \quad (1)$$

那么,节点 $A(a, x_a), B(b, x_b)$ 相连,或称其可见.如图1所示,有边连接,反之没有.一般情况下,周期序列、随机序列、分形序列最终分别转化为规则网络、随机网络和无标度网络.例如,标量序列 $\{x_i\} = \{0.87, 0.49, 0.36, 0.83, 0.87, 0.49, 0.36, 0.83, 0.87, 0.49, 0.36, 0.83, 0.87, 0.49, 0.36, 0.83\}$.

$0.87, 0.49, \dots\}$.

由式(1)得,任意节点可见则表示有边连接,如图1所示,反之没有.

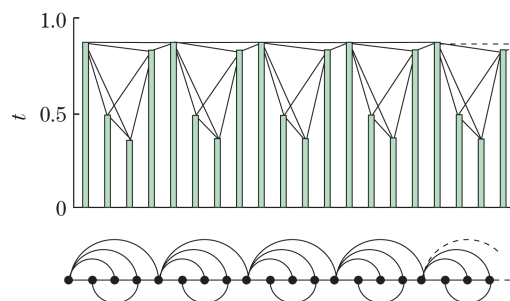


图1 时间序列转换网络规则图

Fig.1 Rule diagram of transformation of time series to network

1.2 度分布

度分布可用于衡量网络的结构性质,节点的度就是与其相连节点的个数.大量现实世界的复杂网络都具有无标度性.在无标度网络里,度分布服从幂律分布函数, $p(k) \propto k^{-r}$.此外,作为另一种重要类型,指数分布也广泛存在于现实世界中,即 $p(k) \propto \exp(-\lambda k)$.Lacasa等研究提出了 $p(k)$ 尾部行为的分析预测.根据式(1),时间序列中极值对应节点的度比较大,从而产生“厚尾现象”.对于极值对应节点 y_e ,其度为 k 的概率为 $P_\theta(k)$.对分形布朗运动的Hurst指数,度分布服从幂律函数,即 $P_\theta(k) \propto k^{-\beta} \sim k^{2H-3}$,其中, β 是标度指数的预测值.因此可以估计Hurst指数, $H = (3 - \beta)/2$.对一个分形布朗运动来说,可见图被证明是Hurst指数估计的一种可靠方法.

1.3 平均最短路径长度

在包含 N 个节点的无向网络中,平均路径长度表示为

$$l = \frac{1}{2N(N-1)} \sum_{i \neq j} d_{ij} \quad (2)$$

式中, d_{ij} 为节点 i 和节点 j 的最短距离.考虑到每个节点到其自身的距离为0,无关联节点的距离为无穷大,修改该定义式,得“调和平均”最短路径长度 $l^{-1} = \frac{1}{2N(N-1)} \sum_{i \neq j} d_{ij}^{-1}$.平均最短路径长度反映网络的大小和分离程度,最短路径的最大值称为网络直径.

1.4 聚集系数

考虑节点 i ,它通过 k_i 条边和其它 k_i 个网络节点相连接,则它们之间最多有 $(k_i - 1)k_i/2$ 条边连

接,但 k_i 个节点实际有 E_i 条边,所以节点的聚集系数定义为 $c_i = 2E_i / (k_i - 1)k_i$,网络的平均聚集系数为 $c = \sum c_i / N$,其中, N 为节点度.聚集系数是用来刻画网络的小集团形态,说明邻近集团在相关性意义上的凝聚程度.若网络的聚集系数与节点度表现出幂律分布特征,则此时的网络具有层次结构.

1.5 R/S 方法

R/S 方法,即重标极差分析(rescaled range analysis)^[9]是一种分析内蕴长期记忆和非周期性循环时间序列的分析方法.设一时间序列 $\{x_t\}$,观测次数为 M ,R/S 分析的基本思想是:将长度为 M 的时间序列 $\{x_t\}$ 分成 A 个长度为 K ($2 \leq K \leq L$) 的相邻子区间, L 是最大子区间的长度, $AK = M$,每个子区间分别记为 I_a ,这里 $a = 1, 2, \dots, A$;每个 I_a 上的 x_t 记为 $x_{k,a}$,这里 $k = 1, 2, \dots, K$.设 I_a 上的 $\{x_t\}$ 的均值为 e_a ,令

$$X_{k,a} = \sum_{i=1}^k (x_{i,a} - e_a), \quad k = 1, 2, \dots, K \quad (3)$$

$$R_{I_a} = \max_{1 \leq k \leq K} \{X_{k,a}\} - \min_{1 \leq k \leq K} \{X_{k,a}\} \quad (4)$$

$$S_{I_a} = \left[\frac{1}{K} \sum_{k=1}^K (x_{i,a} - e_a)^2 \right]^{\frac{1}{2}} \quad (5)$$

式中, K 为子区间的长度; $X_{k,a}$ 为子区间 I_a 上的累积离差; S_{I_a} 为子区间 I_a 上的标准差; R_{I_a} 为子区间 I_a 上的极差,这里 R_{I_a} 随 K 的增大而增大.

为了比较不同类型的时间序列,赫斯特用观测值的标准差去除极差得下列关系式

$$(R/S)_K = \frac{1}{A} \sum_{a=1}^A (R_{I_a} / S_{I_a}) = (aN)^K \quad (6)$$

式中, a 为常数; H 为 Hurst 指数,且 $0 \leq H \leq 1$, $0 \leq H \leq 1$, $2 \leq K \leq L$.

对等式(4)两端取对数得

$$\log((R/S)_K) = H \log(K) + \log a \quad (7)$$

按式(7)将 $\log((R/S)_K)$ 对 $\log(K)$ 做最小二乘法(ordinary least square, OLS)估计,即可得到 H 指数.

2 实证研究

2.1 数据来源

采用“锐思数据库”提供的上海和深圳证券交易所所得上证综合指数和深成指数的日、周、月收盘价,采样跨度为1990年至今(见表1).此处采用指数的对数收益率.

表1 股票指数样本表

Tab.1 Sample of stock index

指数	样本空间	日观察个数	周观察个数	月观察个数
上证综指	1990-12-19 至 2012-09-28	5 337	1 100	266
深证成指	1991-04-03 至 2012-09-28	5 295	1 084	262

2.2 网络拓扑特性

沪深股市月数据、周数据、日数据网络拓扑结构特征各异:

首先,沪深股市3个不同周期数据所构建的网络聚集系数平均分别为0.758 7,0.756 7,聚集系数较高,说明网络构建聚合度较好.同时,日数据网络边数达到上万级别,结构庞大,属于大型复杂网络.

其次,网络直径和网络平均路径长度随周期减小而增大;单就深成指网络而言,深证成指明显地高于上证综指,尤其是日数据网络平均路径长度达到了5.576 2,明显高于上综指的4.801 8,说明深市网络更为复杂,波动更大,涉及更为广泛.

表2 沪深指数网络拓扑结构特征表

Tab.2 Network topology characteristics of Shanghai & Shenzhen stock index

数据	时间单位	节点数	边数	网络直径	平均路径长度	聚集系数
上证指数	月数据	265	1 108	6	3.392 0	0.756 1
	周数据	1 099	4 808	8	4.027 1	0.761 5
	日数据	5 336	23 234	10	4.801 8	0.758 6
深证成指	月数据	261	1 062	7	3.733 5	0.755 1
	周数据	1 083	4 729	8	4.315 6	0.758 7
	日数据	5 294	22 539	10	5.576 2	0.756 2

2.3 度分布

实验研究发现,大量真实网络的节点度服从幂律分布.计算上证和深证各周期的网络节点的度分

布情况,如图2所示.图2(a),2(b)中各个数据点散落在双对数平面上的一条直线附近,表明各周期的网络节点度服从幂律分布.因此,上海和深圳证券市

场网络具有无标度特性.同时,节点度分布产生了“厚尾”现象,而且随着周期减小,变得越为明显.网络的幂指数 r 随着周期的增大而减小.

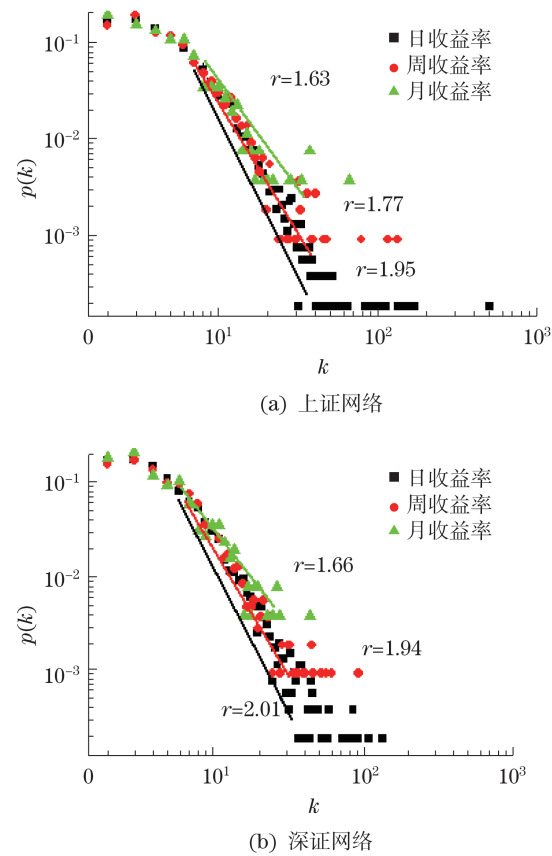


图2 网络节点度分布
Fig.2 Degree distribution of networks

实际网络的幂指数一般在 $2\sim3$ 之间^[10],股票网络具有比实际网络较小的幂指数,但这也是有其实际意义的,这是因为证券市场股票网络本身就是一个虚拟的网络,网络中节点的连接不会受到现实中技术资源的影响和限制,不需要高昂的构造费用^[11].

上证和深证网络节点度分布幂指数 r 相对较小,是因为由对数收益率构建的网络中的 Hub 节点较多,Hub 节点对整个网络的影响较大.根据可见图原理,在收益率波动过程中,指数在某段时期的较大波动(大涨或大跌)对此后指数波动具有一定的影响力.

表3是沪深各周期Hurst指数和对应网络幂指数.其中 H 表示通过 R/S 分析法计算得出的Hurst值, H_0 表示通过 $H = (3 - r)/2$ 关系式计算的估计值.可以发现,估计值 H_0 与计算值 H 相差不大,这也验证了可见图被证明是分形布朗运动Hurst指数估计的一种可靠方法.值得注意的是深证成指日数

据 H_0 预测偏差稍大,这说明深证成指不是简单的分形布朗运动,其波动更为复杂.

表3 沪深指数Hurst指数和幂指数
Tab.3 Hurst index and power-law exponent of Shanghai & Shenzhen stock indices

数据	时间单位	H	H_0	r
上证指数	月数据	0.67	0.68	1.63
	周数据	0.63	0.61	1.77
	日数据	0.59	0.52	1.95
深证成指	月数据	0.62	0.67	1.66
	周数据	0.56	0.53	1.94
	日数据	0.62	0.50	2.01

同时,截取沪指从2005年07月11日至2008年11月04日、深指从2005年11月15日至2008年11月04日为原始序列,同时这两段原始序列分别以2007年10月16日、2007年10月31日分界,从而区分出原始序列、上涨序列和下跌序列.应用可见图算法,计算其网络特性.如图3所示,三序列对应的网络度分布服从幂律形式,且幂律指数明显不同.通过对比,可见图算法揭示出沪深指数非平凡的性质.

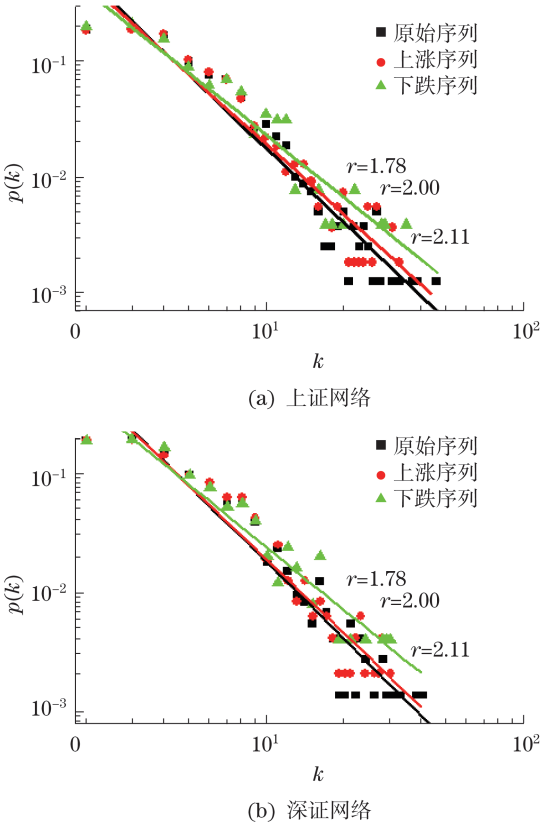
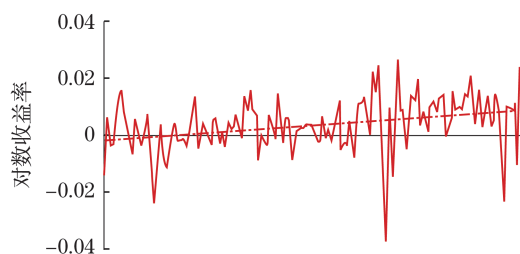


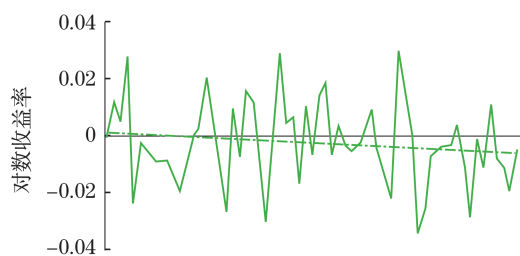
图3 沪深原始、上涨和下跌序列网络度分布
Fig.3 Degree distribution of original, rise and fall time series networks

特别注意到,图 3(a),(b)中上涨和下跌序列对应的网络节点度分布在双对数坐标下所拟合的两条直线相交于一点,该点之前,上涨序列的网络节点度分布概率 $p(k)$ 曲线(红色圆点)在下跌序列(绿色三角形)之上,也就是在同一度的情况下,上涨序列的概率 $p(k)$ 大于下跌序列;该点之后,情况正好相反:上涨序列的概率 $p(k)$ 小于下跌序列,同时形成了“重尾”现象.这说明,该交点在幂率曲线的下降趋势中具有一定特异性质,推测可能与股市网络内在性质有关.

对上证指数该交点进行研究,两直线交点横坐标(平均度 k)在 4.5~4.7 之间,也即交点落在度为 4 和度为 5 之间.对上涨和下降序列分别找出度为 4 和度为 5 对应所有的节点,取出对应的对数收益率,得到图 4.观察可以得到,在交点附近度为 4,5 对应对数收益率,上涨序列和下降序列出现反转,图 4(a)发现对数收益率呈现明显上升的趋势,且图 4(b)对数收益率呈现明显下降趋势,从而说明在该交点附近(度为 4 和度为 5)两侧上涨和下跌序列趋势发生明显反转.



(a) 上涨序列



(b) 下跌序列

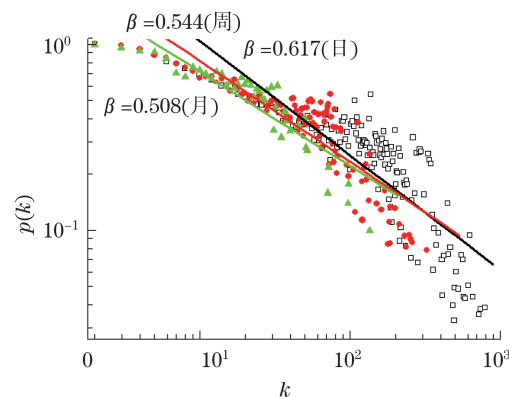
图 4 度为 4,5 对应节点的对数收益率

Fig.4 Logarithmic yield of nodes corresponding to degree of 4,5

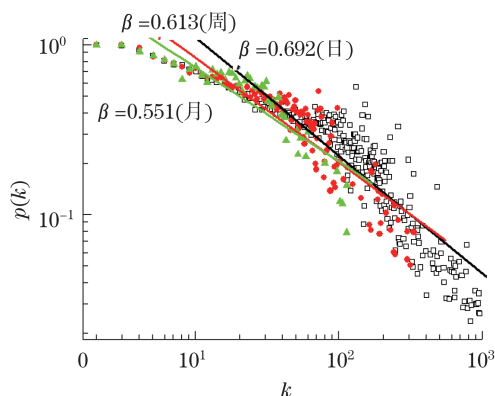
2.4 分层结构

考虑网络节点度与网络聚集系数的关系,图 5 显示网络度与聚集系数在双对数坐标下有幂律分布特征,可知由可见图算法构造的网络具有等级结构

特征.同时,幂律指数 β 都随着周期的增加而减小,也就是说随着周期的增加,由可见图算法构造的网络层级结构特征更为明显.



(a) 上证网络



(b) 深证网络

图 5 沪深股市聚集系数与节点度关系图

Fig.5 Relation between clustering coefficient and degree of Shanghai & Shenzhen stock markets

3 结论

应用可见图算法能有效的表示序列的变化趋势,反映序列的一些非线性特性,如周期性、混沌、分形;将沪深股市序列转换成无标度网络,且幂律指数 r 与 Hurst 指数服从 $H = (3 - r)/2$ 关系式,验证了可见图算法计算 Hurst 指数的可靠性.同时,对沪深股市上涨和下跌序列的分析,发现其网络节点分布也服从幂律指数;同时,发现存在一个交点,对网络节点度分布下降趋势有明显划分,对数收益率明显反转,说明了股市网络性质特异,具有研究价值.由此,可见图算法揭示了沪深股市收益率序列非平凡的特性.

(下转第 263 页)