

中国股票市场收益率分布特征探索

邝志飞, 唐邵玲

(湖南师范大学 数学与计算机科学学院, 湖南 长沙 410081)

摘要: 介绍了正态分布和分形分布的主要特征, 借助分形理论对中国股票市场的收益序列进行了实证研究, 估计出分形分布的参数, 绘制了分形分布的密度曲线, 并对其进行了检验。实证结果表明, 分形分布能较好地拟合中国股票的收益率曲线。

关键词: 收益率; 分布函数; 分形分布; 正态分布

中图分类号: F830.91

文献标识码: A

文章编号: 1673-9833(2007)04-0100-05

Exploration on Characteristics of Return Distribution in Chinese Stock Market

Kuang Zhifei, Tang Shaoling

(College of Mathematic and Computer Science, Hunan Normal University, Changsha 410081, China)

Abstract: The main characteristics of normal distribution and fractal distribution are introduced. It draws density curve of fractal distribution with the studied return series in Chinese stock market by employing fractal theory and estimating parameter of fractal distribution. The result shows that the return in Chinese stock market is characterized by the fractal distribution rather than normal distribution.

Key words: return; distribution functions; fractal distribution; normal distribution

0 引言

金融资产收益率的分布假设是现代金融理论和金融市场风险分析的重要前提, 通常假设金融资产收益率服从正态分布, 而实际金融数据并非如此, 往往具有尖峰胖尾等特性。在对股市收益率分布函数的分析中, 通常有2种不同的思路: 其一是分析价格的形成机制, 即分析导致股价产生变动的因素, 如剖析信息的到达、交易量与交易行为对价格变动的影响, 然后再寻找一个合适的分布函数来描述经验数据; 另一种方法是通过直接对经验数据进行分析, 如研究收益率数据的基本数字特征(如尖峰、厚尾、非对称性与稳定性等数据的经验规律), 从而根据数据的这些经验特征拟合分布。虽然2种分析思路各有侧重, 但所得的结论在某种程度上是一致的。本文采用第二种方式研究收益率的分布函数。

1 正态分布的特征及应用^[1]

正态分布假设观点始于法国数学家 Bachelier (1900), 他在确定标的资产价格变动规律的过程中, 发现资产价格的无条件分布为正态分布; Kendall (1953) 对英国股市价格数据进行研究, 认为股票价格的变化近似服从正态分布; Osborne (1959) 对美国股市的收益率数据进行了研究, 认为用几何布朗运动来描述股价的变动是合适的。由 Bachelier 提出, 经 Osborne 完善的描述股票收益率行为的正态分布模型的基本理论假设是: 1) 对每一只股票而言, 从一笔交易到另一笔交易, 其价格的变化是独立同分布的随机变量; 2) 交易在时间上是均匀分布的, 并且交易之间的价格变化有有限方差; 3) 在所分析的时间区间上, 发生交易的数量是较大的。然而, 只有在收益率的考察标度区间较大时, 正态分布的假设才成立。当时间

收稿日期: 2007-04-27

作者简介: 邝志飞 (1971-), 女, 湖南永兴人, 湖南师范大学硕士研究生, 主要研究方向为金融统计;

唐邵玲 (1963-), 女, 湖南邵阳人, 湖南师范大学副教授, 主要研究方向为金融统计。

区间很小时, 收益率的分布密度函数在尾部呈厚尾特征。从最早的 B. B. Mandelbrot 和 Fama, 到最近的 Hish 和 Anderson 的研究均表明, 股票市场及其它金融资产收益率表现为非正态分布特性, 用正态分布来描述往往效果很差。虽然基于正态分布假设的有效市场理论, 因其科学、严谨、简洁而最终成为现代金融经济学的支柱理论之一, 但有效市场理论在解释实际金融现象时遇到了很多问题, 如日历效应、规模效应、股权溢价之谜、封闭式基金之谜。^[2]

2 分形分布

2.1 分形分布的特征函数及参数^[3]

实际金融数据往往具有尖峰厚尾特性, 而分形分布能很好地刻划数据的尖峰厚尾特征。在经济文献中, 分形分布 (fractal distribution) 又称为 Pareto 分布^[4]、Pareto-levy 分布或 Stable-Paretian 分布。该分布的性质最早是由 Levy (1937) 推导出来的, 而他的工作又是以 Pareto (1897) 有关收入分布的研究工作为基础的。Pareto 发现, 占个人收入 97% 的收入分布接近对数正态分布, 但对于其余 3% 的收入被发现迅速增加, 且服从逆幂规律, 从而产生胖尾特征。产生的原因在

$$\log f(t) = \log E(e^{iXt}) = \begin{cases} i\delta t - \gamma^\alpha |t|^\alpha \left[1 + i\beta \tan \frac{\pi\alpha}{2} (\text{sign}(t)) \right] \left(|\gamma t|^{1-\alpha} - 1 \right), & \alpha \neq 1; \\ i\delta t - \gamma |t| \left[1 + i\beta \frac{2}{\pi} (\text{sign}(t)) \log(\gamma |t|) \right], & \alpha = 1. \end{cases}$$

其中: X 是随机变量; t 是任意实数; i 是虚数单位。

分形分布的特征函数由 4 个参数决定, 即 α , β , δ , γ 。 α , β 是 2 个关键参数, 4 个参数的不同组合产生不同的分形分布形式。各参数的取值范围及意义说明如下:

1) 参数 α 称为分形分布的特征指数, $0 < \alpha \leq 2$, 它不仅刻画了分布的尖峰、肥尾特征, 即分布在尾部极端取值的概率以及密度函数的高度, 同时它也决定该分布的矩、随机变量和的分布及其标准化特征。当 $\alpha = 2$ 时, 分形分布就是正态分布; 当 $0 < \alpha < 2$ 时, 分形分布相对正态分布就有更厚的尾部, 且 α 的值越小, 分布在尾部取极端值的概率越大。同时, 对于该分布只有当 $\alpha = 2$ 时, 才有有限的方差; 只有当 $\alpha > 1$ 时, 均值才存在。

2) 参数 β 称为偏度指数, $-1 \leq \beta \leq 1$ 。当 $\beta = 0$ 时, 分形分布关于 δ 对称; 当 $\beta > 0$ 时, 分布右偏; 当 $\beta < 0$ 时分布左偏。 β 的绝对值越大, 偏斜的程度越大。

3) 参数 δ 是分布的位置参数, $\delta \in (-\infty, +\infty)$ 。当特征参数 $\alpha > 1$ 时, δ 是分布的均值; 当 $\alpha \leq 1$ 时, 分布没有均值。 δ 表示分形分布的中心位置, 当分形分

于财富的杠杆效应。我们之所以称其为分形分布, 是因为它是一个具有统计自相似性的概率密度函数, 也就是说, 在不同的时间增量下, 统计特征保持相同。分形分布得到广泛应用的原因至少有以下几方面:

1) 大量研究表明很多的随机变量服从非正态的分形分布;

2) 广义中心极限定理表明, 对独立同分布随机变量的和进行适当的标准化处理后, 如果极限分布存在, 则此分布一定属于分形分布族;

3) 由于许多大型数据集具有胖尾、尖峰特征, 用正态分布描述效果不好, 而用分形分布去逼近却可以取得比较理想的效果。

分形理论创始人 Mandelbrot 早在 20 世纪 60 年代就研究了棉花价格的变动, 导致经济系统中自相似性的发现和阐述, 价格每一次特定的变化是随机的和不能预言的, 但成串的变化又是与时间标度无关的, 价格的日变化和月变化曲线非常相似, 发现资本市场收益率曲线不符合正态分布, 而是服从分形分布, 这种分布在均值处有高峰, 也有胖尾, 其特点是“方差”无穷大, 倾向于有趋势和循环, 同时也有突然的和不连续的变化。但分形分布的密度函数没有一个统一的解析表达式, 它的特征函数为:

布被标准化时, δ 表示均值且为 0。

4) 模参数 $\gamma \in (0, +\infty)$, 控制分形分布曲线的宽度。当 γ 越大时, 分形分布越宽, 反之, 分形分布则越窄。当 $\alpha = 2$, 分布退化为正态分布时, γ 是正态分布的标准差; 当 $\alpha < 2$ 时, 分形分布的方差不存在, γ 不代表方差。

2.2 分形分布的基本性质^[5]

1) 矩特性 胖尾特性产生的结果之一就是——不是所有的矩都存在。在多数统计问题中, 一阶矩 (均值) $E(X)$ 和二阶矩 (方差) $\text{Var}(X) = E(X^2) - (E(X))^2$ 常被用来描述统计分布。然而, 对于胖尾分布来说, 这些不是普遍有用, 因为这些表达式是发散的。当 $0 < \alpha < 2$ 时, 对于任意 $p \in [0, \alpha]$, $E|X|^p$ 是有限的, 但是对于 $p \geq \alpha$, $E|X|^p = +\infty$ 。因此, 当 $0 < \alpha < 2$ 时, $E|X|^p = +\infty$, 这时分形分布就没有有限的二阶矩或方差; 当 $1 < \alpha \leq 2$, $E|X| < +\infty$,

并且 $EX = \delta_1 = \delta_0 - \beta \gamma_0 \tan \frac{\pi\alpha}{2}$ 。

2) 稳定性 Levy (1937) 指出, 如果对于所有的 $b_1, b_2 > 0$ 时, $F(x)$ 满足下列关系式:

$$F(x/b_1)F(x/b_2) = F(x/b),$$

或者分布函数的特征函数有类似的关系式:

$$f(b_1u)f(b_2u)=f(bu),$$

则 $F(x)$ 是稳定的。这表明 $F(x/b_1)$ 、 $F(x/b_2)$ 、 $F(x/b)$ 具有相同的形式, 只是尺度不同而已。应用这一关系式到分形分布, 就有关系式: $b_1^\alpha + b_2^\alpha = b^\alpha$, 这表明具有特征指数 α 的 2 个分形分布是稳定的, 即它们的和也是具有特征指数 α 的分形分布, 并且是稳定的。因此, 具有相同 α 、 β 的分形分布的随机变量相加, 所得到的随机变量也是具有相同 α 、 β 的分形分布。尽管 δ 、 γ 可能发生变化。

3) 尖峰胖尾性 分形分布具有尖峰、胖尾特性, 特征指数 α 决定了这一性质。在同等级离散程度下, 特征指数 α 越小, 分形分布的中心越尖, 尾部越胖。

4) 非连续性 分形分布可能存在无限均值与无限方差, 这是由于价格发生不连续跳跃所导致的。随着样本观测值的增加, 样本均值与方差在变化过程中会发生跳跃, 即使样本观测值无限增加, 样本均值与方差也不会收敛到一个有限数。因为总体均值与方差可能不存在, 分形分布过程的大变化不是大量小变化的和, 而是小量的大变化, 这些大变化是非连续的。当市场的分形结构崩溃时, 市场的大变化就会发生。正态分布注重的是大量连续的小变化, 而分形分布注重的是小量不连续的大变化。

5) 自相似性 无论怎样调整规模参数 γ 的值, 分形分布的特征指数 α 和偏度指数 β 总是保持不变, 即分形分布的形状保持不变。因此, 在不同的规模调整尺度下, 分形分布具有统计自相似的结构。如果日收益的均值是 m , 特征指数为 α , 则 5 日收益的均值

是 $5m$, 特征指数保持不变。当调整时间标度时, 收益序列的分布函数形状保持不变, 该序列就是标度不变序列。分形分布还有一个重要性质是它们的自变量较大时的幂关系^[2], 常被称为“帕累托尾部”, 即有

$$p_\alpha(x) = p(X > x) = \frac{C}{|x|^{1+\alpha}}, \quad (X \text{ 是随机变量}), \text{ 当}$$

$|x| \rightarrow \infty$ 。

式中: $0 < \alpha < 2$ 是一个指数;

C 为常数。

3 实证研究

3.1 样本与数据说明

我国从 1990 年 12 月 19 日上海证券交易所成立, 开始股票的集中交易至今, 也就 10 多年的历史, 期间交易与管理度经多次调整, 特别是 1996 年 12 月 16 日, 沪深股市实行涨跌停板制度, 对股票收益有较大的影响, 徐绪松、陈彦斌 (2001) 认为这之后中国股市更趋于理性。因此, 本文选取的数据样本取自 1997 年 1 月 2 日~2006 年 12 月 31 日的上证指数和深证成指的日收盘价, 共计 2 401 个数据, 分别为上证指数、深证成指周收盘价以及月收盘价。在计算过程中, 股票市场每日收益率以相邻营业日股指收盘价对数的一阶差分

表示, 即 $s_t = \ln \frac{p_t}{p_{t-1}}$, 其中: p_t 为 t 时的股票价格; s_t 为 t 时股票的对数收益率。

3.2 样本的描述性统计量

沪、深股指收益率的基本统计详见表 1。

表 1 沪、深股指收益率的基本统计量

Table 1 Elementary statistics on return in Shanahai and Shenzhen stock market

收益率序列	均值	标准差	偏度	峰度	J-B 统计量	概率
日收益率	-0.000 450	0.015 103	-0.008 990	8.705 917	3 255.781	0.000 000
上证指数						
周收益率	-0.002 197	0.030 880	-0.282 879	4.542 267	55.772 49	0.000 000
月收益率	-0.008 572	0.070 254	-0.776 083	4.574 981	24.245 19	0.000 000
日收益率	-0.000 295	0.016 567	-0.075 176	7.973 795	2 474.061	0.000 000
深证成指						
周收益率	-0.001 461	0.034 981	-0.327 868	4.523 673	56.865 74	0.000 000
月收益率	-0.004 937	0.080 459	-0.961 898	5.442 397	47.928 71	0.0000 00

说明: 1. J-B 统计量在 5% 和 1% 显著性水平下的临界值分别为 5.991 5 和 9.210 3; 2. 全部数据来源于搜狐财经。

3.3 实证研究结果

本人利用 eviews、matlab spss 软件得出下列结果。

1) 正态性检验。用频率直方图、Jarque-Bera 检验、 $Q-Q$ 图 3 种方法对中国股市收益率作正态性检验。从表 1 的结果可以看出, 上证指数与深证成指收

益率的偏度值、峰度值都显著异于正态分布, 因为在正态分布中, 两者的值分别为 0 和 3; 另外, 收益率的 J-B 统计量也远大于任意合理水平下的临界值, 而且, 收益率序列的 $Q-Q$ 正态图也不在直线上 (图 1、2) (本文只列出沪市、深市的日收益率 $Q-Q$ 图)。这

些事实表明, 股票的收益率分布不是正态分布, 而是具有尖峰、厚尾的特征 (图 3、4), 这一结论与国内其它研究结果一致。

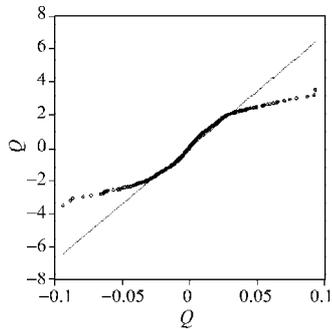


图 1 沪市日收益率序列 $Q-Q$ 正态图
Fig. 1 Normality Quantile-Quantile on daily return series in Shanghai stock market

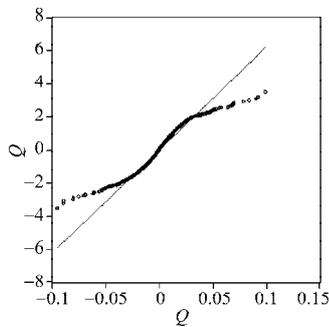


图 2 深市日收益率序列 $Q-Q$ 正态图
Fig. 2 Normality Quantile-Quantile on daily return series in Shanghai stock market

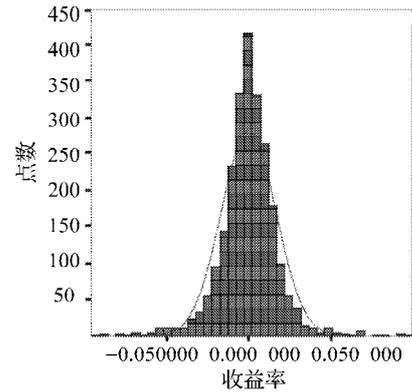


图 3 深市日收益率序列
Fig. 3 Daily return series in Shenzhen stock market

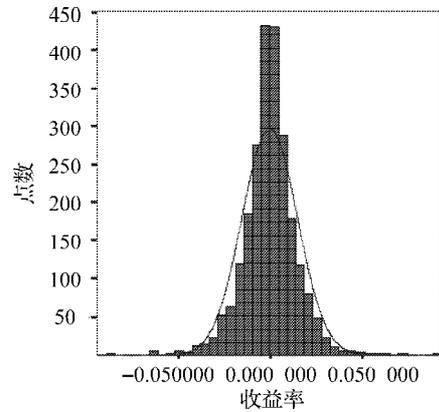


图 4 沪市日收益率序列
Fig. 4 Daily return series in Shanghai stock market

2) 分形分布的参数估计^[4, 6-8]

表 2 列出了上证指数、深证成指的日收益、周收益的分形分布的参数估计值。

表 2 上证指数、深证成指的日收益、周收益的分形分布的参数估计

Table 2 Parameter estimate on fractal distribution

序列	分布	α	β	γ	δ
上证指数日收益	分形分布	1.500 359	-0.069 363	0.757 171e-02	-0.184 355e-03
	正态分布	2.000 0	0.000 0	0.106 793e-02	-0.445 053e-03
上证指数周收益	分形分布	1.797 995	0.274 747	0.188 214e-01	-0.251 140e-02
	正态分布	2.000 0	0.000 0	0.218 355e-01	-0.219 742e-02
深证成指日收益	分形分布	1.558 416	-0.120 405	0.848 357e-02	0.223 688e-03
	正态分布	2.000 0	0.000 0	0.117 149e-01	-0.294 742e-03
深证成指周收益	分形分布	1.548 882	-0.152 897	0.201 464e-01	0.889 905e-03
	正态分布	2.000 0	0.000 0	0.247 350e-01	-0.146 132e-02

由表 2 可以看出上证指数、深证成指的日收益率与周收益率的特征指数均在 1.0~2.0 之间, 表明上证指数、深证成指收益分布比正态分布更尖峰和更胖尾, 分布的均值存在, 但是分布的二阶矩却不存在或无穷。上证日收益, 深证日、周收益的偏斜指数 $\beta < 0$, 表明上证日收益, 深证日、周收益序列均为右偏; 上证周收益序列左偏。深证日收益、周收益的特征指数相差不大, 说明两者在一定程度上相似。上证指数的

日收益比周收益更具尖峰胖尾性, 从而具有更强的长期记忆性。

3) 分形分布、正态分布的拟合及检验。笔者利用 matlab 编写程序, 绘制了上证指数、深证成指的日收益的实际分布图、拟合的正态密度曲线和分形分布密度曲线来进行对比, 如图 5、6 所示。从图中可清楚看到, 分形分布密度曲线能较好地拟合实际股票收益率序列, 比正态分布密度曲线的拟合效果好。

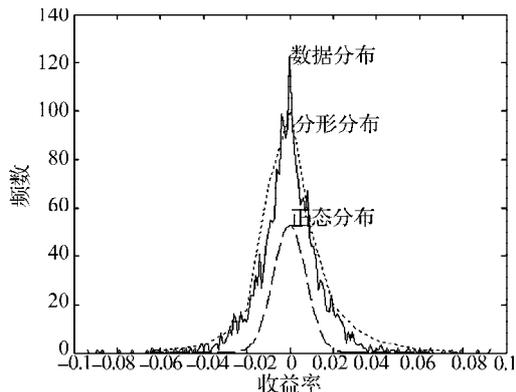


图5 上证指数日收益率

Fig. 5 Daily return on Shanghai stock market

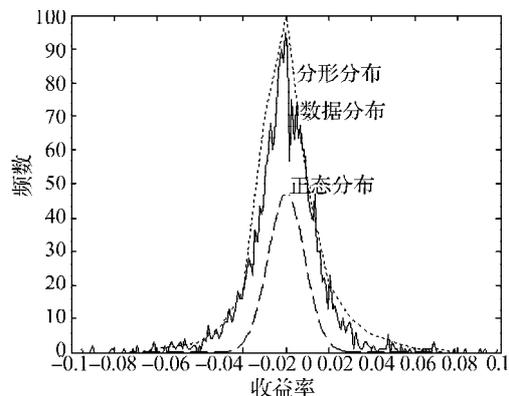


图6 深圳成指日收益率

Fig. 6 Daily return on Shenzhen

利用 pp 图方法, 对分形分布和正态分布的拟合效果进行检验, 检验结果如图 7、8 所示。

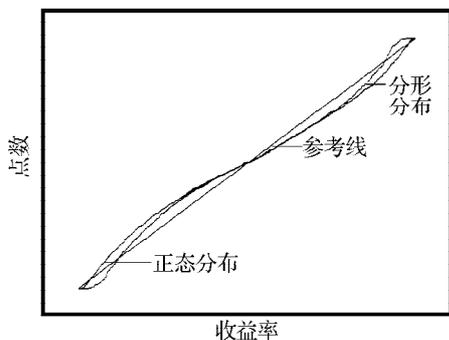


图7 深证成指日收益率检验结果

Fig. 7 The test result of daily return on Shenzhen stock market

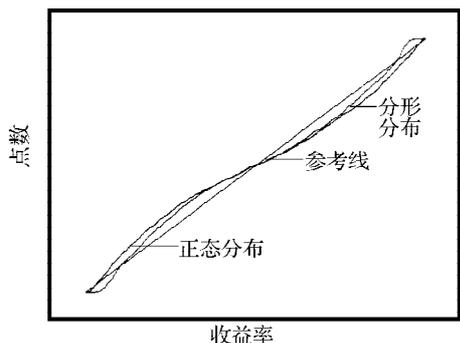


图8 上证指数日收益率检验结果

Fig. 8 The test result of daily return on Shanghai stock market

4 结果分析

1) 基本统计量分析。上证指数和深证成指的 3 种收益率均具有负的偏度, 呈左偏态势。 $\alpha < 2$, 表明在均值附近的频数比正态分布多, 有较厚的尾部。收益率的偏度都显著小于 0, 表明沪市和深市股票出现大幅下跌的概率大于大幅上升的概率。同正态分布相比, 收益率的峰度远远大于 3, 呈现出尖峰、厚尾特征。股票收益率不能用正态分布来进行解释。

2) 由表 1 可知, 上证指数日收益率的峰度值大于周收益率、月收益率的峰度值, 日收益率 $J-B$ 统计量大于周收益率、月收益率 $J-B$ 统计量, 深证成指也有同样的结果。这表明随着时间间隔期的延长, 收益率分布的尖峰、厚尾特征开始弱化。

3) 深证成指收益率均值的绝对值小于沪市上证指数收益率均值的绝对值, 深市的标准差大于沪市的标准差, 说明深市的波动性大于沪市, 即深市的风险高于沪市风险。

4) 分形分布能较好地拟合上证指数和深证成指的日收益率序列。无论从分形分布和正态分布拟合的密度曲线, 还是从对分形分布拟合的 pp 检验结果, 可以看出, 分形分布对实际股票收益率序列的拟合效果令人满意, 而正态分布的拟合效果却不尽如意。人们以非线性方式对信息作出反映, 市场遵循有偏的随机游走, 即时间序列是分形的。因此, 沪深股市收益率分布不是正态分布, 而是更符合分形分布。

参考文献:

- [1] 封建强. 股价行为理论与实证分析[M]. 北京: 经济科学出版社, 2004.
- [2] Peters E E. Fractal market Analysis Applying Chaos Theory to Investment and Economics[M]. New York: John Wiley & Sans Inc, 1994.
- [3] 徐龙炳. 中国股票市场收益稳态特性的实证研究[J]. 金融研究, 2001 (6): 36-43.
- [4] Mandelbrot B B. The Pareto-levy law and the distribution of income[M]. International Economic Review, 1960 (1): 79-106.
- [5] 黄诒蓉. 中国股市收益分形分布的实证研究[J]. 南方经济, 2006 (2): 99-106.
- [6] Nolan J P. Numerical computation of Stable Densities and Distribution Functions[J]. Comm.in Stat-Stochastic Models., 1997, 13: 759-774.
- [7] Nolan J P. Maximum Likelihood Estimation and Diagnostic for Stable Distribution[M]. Washington : Department of Mathematics and Statistics American University, 1999.
- [8] Mantegna R N. Fast accurate algorithm for numerical simulation of levy stable stochastic processes[J]. Physical Teview, 1994, E495: 4677-4683.