

文章编号: 1005-2542(2002)01-0012-03

# 中国股票市场混沌动力学预测模型

刘文财, 刘 豹, 张 维  
(天津大学 管理学院, 天津 300072)

**【摘要】**运用混沌动力系统理论对中国股票市场的上证综合指数、深证成分指数进行了初步的分析,表明上证综合指数、深证成分指数两个序列是由高维的混沌系统所产生。并对上证综合指数、深证成分指数两序列建模预测,结果表明,新预测模型比均值模型的预测效果好。

**关键词:** 股票; 混沌动力学; 预测

**中图分类号:** F 832.5      **文献标识码:** A

## Forecasting Model of Chaos Dynamics in Stock Market in China

LIU Wen-cai, LIU Bao, ZHANG Wei

(School of Management Tianjin Univ., Tianjin 300072, China)

**【Abstract】**In this paper, we have analyzed the indices of two stock markets based on the chaos dynamics theory. They are Shanghai Composite Index (SCI), Shenzhen Sub-Index (SSI). The results show that the time series of SCI and SSI are generated by a high dimensional chaotic system. We have developed a new forecasting model for SCI and SSI. And our results suggest that the new model outperforms the model of mean.

**Key words:** stock; chaos dynamics; forecasting

对股票市场的准确预测是值得怀疑的,已经提出的许多模型证明了这一点。这些模型似乎为法玛(Fama)的半强有效市场理论(即价格反映了所有公共信息)的成立提供了佐证,许多预测模型都失效了。问题在于为什么仍有众多个人或金融机构继续在做预测,并想以此来击败市场?一个理由可能是人们认为有效市场理论所基于的假设太过理性与完美了,现实中股票市场并非如此。暂且不论市场的完备性如何,一个不争的事实是市场中投资者永远处于信息的极大不对称情况下作决策的,即投资者不可能获得关于股票价值的完全信息与完整知识。且大部分投资者对信息的反应有滞后性的与群集突发性的,并非及时与线性的。股票市场上投资者是以正反馈的方式进行投资的,因此导致了反映市场总况的

指数时间序列呈现了非线性的、长记忆的特征。国外与国内的近期研究成果证明了这一点<sup>[1-4]</sup>。这些研究成果为股票市场的预测带来了新的希望。股票市场并非是复杂得如随机游走一样不可捉摸,其表面上体现出来的随机行为本质上是由两部分构成的,一是内随机性,这是由混沌动力系统所产生,可以把握与预测,至少短期是可以的。二是外随机性,这完全由市场信息的随机达到所决定,是真正不可知的部分。基于以上思想,本文研究了中国股票市场的混沌动力学预测模型。

## 1 方法

### 1.1 选择最佳的模型嵌入维数

从系统论的角度来看,股票市场是一个系统,而股票价格是这个系统的一个重要的状态变量。当然这个状态变量是由系统中的其他变量所决定,因为其他变量与价格的相关关系都隐含在价格的变化过程之中。现在不知道这个系统的动力学行为,因为它们之间的确定函数关系还未知。但可以通过对价格

收稿日期: 2001-07-4 修订日期: 2001-10-15

基金项目: 国家自然科学基金资助项目(79970083)

作者简介: 刘文财(1976-),男,博士生,主要从事金融系统工程的研究。

序列的重构,构造出与原系统“等价”的相空间。设有一时间序列  $\{x_t\}, t=1, \dots, N$ , 通过重构相空间矢量来重构相空间。相空间矢量为

$$V_n = (x_n, x_{n-1}, \dots, x_{n-(d-1)f}) \quad (1)$$

式中  $f$  为延滞时间;  $d$  为嵌入维数;  $n=(d-1)f+1$ 。如果时间序列是由混沌动力系统所产生的,则由嵌入理论可知<sup>[5-6]</sup>,存在一映射  $F: R^d \rightarrow R^d$  使得

$$V_{n+1} = F(V_n) \quad (2)$$

当时间序列的观察函数是光滑的且嵌入维数  $d$  足够大时,式(2)的动力学行为与未知系统的动力学行为是拓扑等价的。在实际应用中,使用一标量方程来代替式(2)的矢量方程,即

$$x_{n+1} = f(V_n) \quad (3)$$

从一个时间序列中确定嵌入维数  $d$  已有很多方法。如 Grassberger & Procaccia<sup>[7]</sup>通过计算吸引子的关联维数与 Lyapunov 指数来确定嵌入维数  $d$ ; Broomhead & King<sup>[8]</sup>, Vautard & Ghil<sup>[9]</sup>的奇异值分解法; Kennel 等<sup>[10]</sup>的虚假邻域法。Cao<sup>[11]</sup>指出了这些方法在选择嵌入维时都包含主观参数或主观判断。如在国内股票市场混沌判断研究上,文献[3]利用了 Grassberger & Procaccia 算法,通过试选取嵌入维数的方法来观察上证指数对数收益率序列混沌吸引子的关联维数与 Lyapunov 指数随嵌入维数增加的稳定性情况。在确定 Lyapunov 指数时,文献[3]给出了当  $d$  增加到 25 时, Lyapunov 指数才基本稳定。然而在观察关联维数时,当  $d$  增加到 10 时,关联维数就趋于稳定。两个不同的  $d$  值暴露了 G-P 算法的不足之处。Cao 等<sup>[12-13]</sup>给出一种不依赖主观参数的零阶近似法。本文就采用这种方法来确定嵌入

维。以下对此方法作一简单的叙述。

对一给定的嵌入维  $d$ , 通过式(1)可得到一系列滞后向量  $V_n$ , 对每个  $V_n$  都能找到最近的一点  $V_{j(n)}$ , 使得

$$V_{j(n)} = \arg \min \{ \|V_n - V_j\| : j = (d-1)f+1, \dots, N, j \neq n \} \quad (4)$$

即

$$\|V_{j(n)} - V_n\| = \min \{ \|V_n - V_j\| : j = (d-1)f+1, \dots, N, j \neq n \}$$

其中范数

$$\|V_n - V_j\| = \|(x_n, x_{n-1}, \dots, x_{n-(d-1)f}) - (x_j, x_{j-1}, \dots, x_{j-(d-1)f})\| = \left[ \sum_{i=0}^{d-1} (x_{n-if} - x_{j-if})^2 \right]^{1/2}$$

在此基础上,定义:

$$E(d) = \frac{1}{N - J_0} \sum_{n=J_0}^{N-1} |x_{n+1} - x_{j(n)+1}| \quad (5)$$

$$J_0 = (d-1)f+1$$

$E$  的值依赖于嵌入维  $d$ 。实际上,对于一给定的  $d$ ,  $E(d)$  是零阶近似预测因子绝对预测误差的平均值。零阶近似预测因子  $f$  为:  $x_{n+1} = f(V_n)$ , 且  $x_{n+1} = x_{j(n)+1}$ 。所以选择嵌入维  $d$  使得  $E(d)$  最小。即

$$d_e = \arg \min \{ E(d) : d \in \mathbb{Z} \text{ 且 } d \geq 1 \} \quad (6)$$

在实际中,在  $d$  的所有正整数中取即可,即

$$d_e = \arg \min \{ E(d) : 1 \leq d \leq D_{\max} \} \quad (7)$$

式中  $D_{\max}$  是最大嵌入维数。取  $D_{\max} = 60$  (取  $D_{\max} = 60$ , 是根据对 2 个指数程序运算后确定的)。

图 1(a~e) 分别给出了低维无噪声、高维无噪声、低维噪声、高维噪声及完全随机时间序列的  $E(d)$  典型的图形。

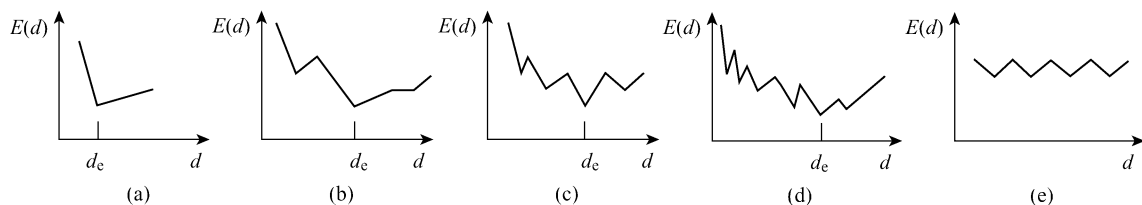


图 1 五种系统的典型  $E(d)$  图

## 1.2 基于混沌动力学的局部线性预测模型

当向量  $V_n$  的嵌入维与延滞时间确定之后,就可以用式(3)进行预测了。剩下的问题是如何去估计函数  $f$ 。假设现在用  $N_f$  个样本值去拟合函数  $f$ , 亦即有  $x^1, x^2, \dots, x^{N_f}$ 。这样就可构造延滞向量  $V_n, n=J_0, J_0+1, \dots, N_f$  且  $J_0=(d-1)f+1$ 。局部线性近似预测的步骤如下:

(1) 找出距  $V_n$  最近的  $l$  个向量, 用  $V_{j_1}, V_{j_2}, \dots,$

$V_{j_l}$  表示,  $J_0 \leq j_k < N_f$  ( $k=1, 2, \dots, l$ )。这样对任意  $k$  都有,  $\|V_{j_k} - V_{N_f}\| \leq \|V_n - V_{N_f}\|$  ( $J_0 \leq n < N_f$  且  $n \neq j_k, k=1, 2, \dots, l$ )。

(2) 构造一局部线性预测模型, 把  $V_{j_k}$  看成邻域内的一个点, 而  $x_{j_k+1}$  是相应点。这样, 对  $l$  对  $(V_{j_k}, x_{j_k+1})$  ( $k=1, 2, \dots, l$ ), 可以用最小二乘法找到这个线

性函数, 用  $\hat{F}$  表示, 使得  $\sum_{k=1}^l |x_{j_k+1} - \hat{F}(V_{j_k})|$  最小。

(3)  $x_{N_f+1}$  的预测值  $\hat{x}_{N_f+1} = F(V_{N_f})$

$l$  一般要大于或等于局部线性法中参数的数量,但要比总延滞向量数小,即  $N_f - J_0 > \geq d + 1$

2.3 预测精确性的评价

因为预测的对象是股票指数的收益率,对这个有较长数据的时间序列,曾尝试用 ARIMA 模型作为混沌动力学模型的比较对象。但用 SAS 软件来建模并拟合,最后得到是出人意料的结果,即,对这两个时间序列来说,最好的模型不是 ARIMA 模型而是均值模型。这可能是由于股指的收益率序列确有其特殊性的原因,故本文就把局部线性预测模型的预测精度与均值预测模型比较。

均值预测模型所指的是,预测值  $\hat{x}_{i+1} = \bar{x}$   $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  是序列的均值。对这两个预测模型,用下面统计量来比较它们的预测精度:

$$RMSE = \left[ \sum_{i=p}^{\infty} (\hat{x}_i - x_i)^2 \right]^{1/2} \bigg/ \left[ \sum_{i=p}^{\infty} (\bar{x} - x_i)^2 \right]^{1/2}$$

(8)

式中:  $p$  表示外推的数据量;  $\hat{x}_i$  为  $x_i$  的预测值。如果  $RMSE < 1$ , 则表明局部线性模型比均值模型的预测效果好; 如果  $RMSE \geq 1$ , 则表明它并不比均值模型的效果好。

2 数据及结果

选取从 1996-12-16~2001-05-31(选取样本数据从 1996-12-16 开始,是因为从这天开始中国股票市场实行了涨跌停板制造。这种交易制度的变更深刻地影响着股票市场的性质。在此以后的股票市场是可以在统计意义上作研究的) 的上证综合指数(000001)、深圳成分指数(9901)的对数收益率数据(对数收益率,即  $r_t = \ln(p_{t+1}/p_t)$  其中  $p_t$  为第  $t$  天的收盘价)。样本数量为 1 069 个。把它们分为 3 个部分,第 1~960 个观察值来确定预测模型,第 961~1 000 个观察值来优化模型的参数值,第 1 000~1 069 个观察值来检验模型的预测效果。仅用最新的数据作一步预测。

在确定两指数的嵌入维时,用第一部分与第二部分的样本值,即 1~1 000 个观察值。按照 1.1 节给出的算法,选取延滞时间  $\tau = 1$ , 计算出各个股票的  $E(d)$  值。图 2 列出了两指数的  $E(d)$  与  $d$  之间的关系图。由图 2 可见,上证综合指数(000001)与深证成分指数(9901)的  $E(d)$  值随着  $d$  的增加有较明显的衰减趋势。比较图 1 可知,000001 与 9901 的高维混沌动力学特征是明显的。

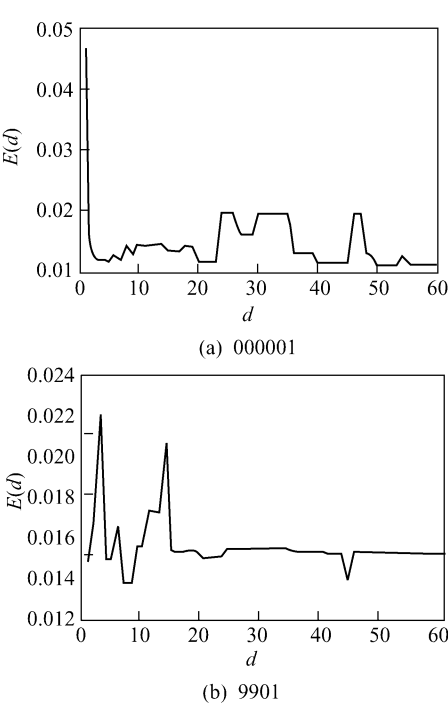


图 2 两指数的  $E(d)$  与  $d$  关系图

由上所述,用 1.2 中所述的方法对 000001 与 9901 进行建模与预测。表 1 给出了它们的嵌入维,最近点数及 RMSE 值。两个指数的 RMSE 都比 1 小,这表明使用基于混沌动力学的局部线性预测模型比均值预测模型好。但并不好的太多,这由于为去掉原始序列趋势而取对数后,致使序列中的内随机性减弱,外随机性信号相对加强造成的干扰结果。因此,如果能找到一个较好的去掉趋势而不削弱内随机性影响的方法,则能提高该模型的预测效果。

表 1 两指数的嵌入维及 RMSE

股指代码	嵌入维	最近点数	RMSE
000001	59	64	0.955
9901	44	48	0.973

4 结 语

运用混沌动力系统理论对中国股票市场的两个指数作了初步分析。表明上证综合指数、深证成分指数两个序列是由高维的混沌系统所产生。嵌入维数分别为 59 与 44。并对上证综合指数、深证成分指数两序列建模预测。结果表明,基于混沌动力学的局部线性预测模型比均值预测模型的预测效果好。但并不好的太多,这由于为去掉原始序列趋势而取对数后,致使序列中的内随机性减弱,外随机性信号相对加强造成的干扰结果。因此,如果能找到一个较好的去掉趋势而不削弱内随机性影响的方法,则能提高该模型的预测效果。

(下转第 31 页)

- Res, 1978, 12: 241– 246.
- [ 4 ] Fisk C S, Boyce D E. Alternative variational inequality formulations of the network equilibrium-travel choice problem [J]. Transportation Science, 1983, (4): 454– 463.
- [ 5 ] Lam William H K, Hai-Jun Huang. A combined trip distribution and assignment model for multiple user classes[J]. Transpn Res, 1992, 26B(4): 275– 287.
- [ 6 ] Abrahamsson T, Lundqvist L. Formulation and estimation of combined network equilibrium models with applications to stockholm [J]. Transportation Science, 1999, 33(1): 80– 100.
- [ 7 ] Daganzo C, Sheffi Y. On stochastic models of traffic assignment [J]. Transportation Science, 1977, 11(3): 253– 274.
- [ 8 ] Fisk C. Some developments in equilibrium traffic assignment[J]. Transportation Research, 1980, B-14: 243– 255.
- [ 9 ] Sheffi Y, Powell W. Comparison of stochastic and deterministic traffic assignment over congested network [J]. Transportation Research, 1981, B-15: 53 – 64.
- [10] Boyce D E, Leblanc L, Chon K. Network equilibrium models of urban location and travel choices [J]. Journal of Regional Science, 1988, 28.
- [11] Oppenheim N. Equilibrium trip distribution assignment with variable destination costs[J]. Transportation Research, 1993, B-27(3): 207– 217.

(上接第 14 页)

#### 参考文献:

- [1] Edgar E Peters. Chaos and order in the capital markets. 资本市场的混沌与秩序 (中译本第 2 版) [M]. 北京: 经济科学出版社, 1999.
- [2] 张 维, 黄 兴. 沪深股市的 R/S 实证分析 [J]. 系统工程, 2001, 19(1): 1– 5.
- [3] 高红兵, 等. 我国证券市场的混沌判据 [J]. 系统工程, 2000, 18(6): 28– 32.
- [4] 樊 智. 分形市场理论及其应用 [D]. 天津大学硕士学位论文, 2000.
- [5] Takens F. Detecting strange attractors in fluid turbulence [A]. Dynamical Systems and Turbulence, Lecture Notes in Mathematics [C]. Vol. 898, Berlin: Springer-Verlag, 1981.
- [6] Sauer T, Yorke T A, Casdagli M. Embedology [J]. Journal of Statistical Physics, 1991, (65): 579– 616.
- [7] Grassberger P, Procaccia I. Measuring the strangeness or strange attractors [J]. Physica, 1983, D9, 189– 208.
- [8] Broomhead D S, King G P. Extracting qualitative dynamics from experimental data [J]. Physica, 1986, D20: 217– 236.
- [9] Vautard R, Ghil M. Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series [J]. Physica, 1989, D35: 395– 424.
- [10] Kennel M, Brown R, Abarbanel H. Determining embedding dimension for phase-space reconstruction using a geometrical construction [J]. Physical Review, 1992, A45: 3403– 3411.
- [11] Cao L. Practical method for determining the minimum embedding dimension of a scalar time series [J]. Physica, 1997, D110: 43– 50.
- [12] Cao L, Mees A, Judd K. Dynamics from multivariate time series [J]. Physica, 1998, D110: 43– 50.
- [13] Cao Liangyue, Soofi A S. Nonlinear deterministic forecasting of daily dollar exchange rates [J]. International Journal of Forecasting, 1999, (15): 421– 430.