第20卷 第8期
2010年8月
计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT
Vol.20 No.8
Aug. 2010

基于相空间重构的股价时间序列相关性分析

查春生,倪志伟,倪丽萍,公维峰
(合肥工业大学 管理学院,安徽 合肥 230009;
合肥工业大学 过程优化与智能决策教育部重点实验室,安徽 合肥 230009)

摘 要:股指的相关性研究对于衍生产品定价、风险管理、套期保值和最优投资组合选择等都具有重要的意义。股指相关性研究大都是在线性理论的基础上,即认为股指时间序列是线性的,但是实际上,股指时间序列具有很强的非线性特征,因此在线性理论基础上得到的相关性结果具有一定的局限性。应用非线性的混沌理论,通过对沪深股指混沌时间序列进行相空间重构,建立了一个多维的股指时间序列系统。运用典型相关分析对构建的系统进行相关性计算,得到沪深股指之间的相关性。

关键词:时间序列;相空间重构;典型相关分析;股指相关性
中图分类号:TP31 **文献标识码**:A **文章编号**:1673-629X(2010)08-0017-04

Correlations Analysis Between Stock Index Time Serials Based on Reconstructed Phase Space

ZHA Chun-sheng, NI Zhi-wei, NI Li-ping, GONG Wei-feng
(School of Management, Hefei University of Technology, Hefei 230009, China;
Ministry of Education Key Laboratory of Process Optimization and Intelligent Decision-making,
Hefei University of Technology, Hefei 230009, China)

Abstract: Correlations of stock index is very important for derivative product pricing, risk evaluation, hedge and optimal portfolio choice. Research of correlations of stock index is almost based on linear theory. They think the stock index time serials have a linear characteristics. But in fact, the stock index time serials have a strong non-linear characteristics. Correlations based on linear theory will hava certain limitations. Adopt non-linear chaology to build a multidimensional time serials system by reconstructing gamma space of Shanghai and Shenzhen stock index time serials. Calculate correlations by canonical correlation analysis and achieve correlations between Shanghai stock index and Shenzhen stock index.

Key words: time serials; reconstructed phase space; canonical correlation analysis; correlations between stock index

0 引 言

股指波动的研究是金融研究的基础。上证综指和深圳成指同处在中国内地,不可避免地会互相产生影响,即上证综指的波动将会影响深圳成指的波动,同时深圳成指的波动也会影响上证综指的波动。研究中国各个股市股指之间的相关性对于中国这样一个不完善的股票市场具有重要的现实意义。但是现有的相关研究并不多,Chui 和 Kwork 研究了中国股票市场 A 股和 B 股之间的因果关系[1]。樊智和张世英利用遗传算法来分析上证综指和深圳成指之间的相关性[2]。郑振龙等运用 ADDC-MGARCH 模型对中国四个主要股票指数市场的收益相关性进行了实证分析[3]。杨兴民等讨论了 Gaussion Copula 与 t-Copula 的密度函数,并进行了相关性建模,采用二步估计法对所建模型进行参数估计并给出了相关性指标[4]。陈守东等运用 Granger 因果检验及 GARCH-M 模型对沪深股市的相关性进行了分析和检验,得出在滞后一阶的情况下,沪深股市收益率互为影响,但随着阶数的增加,上海股市收益率对深圳股市收益率的影响十分的显著,即收益率相关性具有滞后性[5]。

大量实证研究表明,股指时间序列具有混沌分形的特征。混沌时间序列分析是非线性时间序列的最新发展,它是在近 20 年来非线性科学蓬勃发展的基础

收稿日期:2009-12-07;修回日期:2010-03-02
基金项目:国家高技术研究发展计划(863)项目(2007AA04Z116);国家自然科学基金(70871033);安徽省自然科学基金(090416246);合肥工业大学校科学研究发展基金(2009HGXJ0040)
作者简介:查春生(1988-),男,硕士生,研究方向为分形理论及应用;倪志伟,教授,博士生导师,研究方向为人工智能、机器学习。

上,将非线性动力学即混沌理论和分形理论应用于非线性空间序列的研究上所产生的新的分析方法[6]。

混沌时间序列的研究是以 Takens 嵌入定理[7]为基础的,即对于一维时间序列,可以重构一个与原动力系统在拓扑意义下等价的相空间,从而把握混沌时间序列的性质与规律。混沌时间序列的判定、分析与预测都是在这个重构相空间中进行的。

文中的沪深股指是非线性的时间序列,运用混沌时间序列中常用的重构相空间方法,分别对沪深股指收益时间序列进行相空间重构,得到一个多维的收益时间序列系统,把一维时间序列之间的相关性计算转化为多维时间序列之间的相关性计算。通过典型相关分析的方法获得相关性值。从算法的过程看,充分考虑到了时间序列之间相关的滞后性,以及时间序列的自相关对时间序列之间相关性的影响。

1 相关理论

1.1 相空间重构

混沌时间序列重构相空间始于 Packard 等[8],可以分解为求时间延迟 τ 和嵌入维数 m,τ 和 m 的选择一般都基于 3 种准则:

(1)序列相关法,如自相关法、高阶相关法等;

(2)相空间扩展法,如填充因子法、摆动量法等;

(3)复自相关法和去偏复自相关法[9]。

自相关法计算时间延迟 τ,对于一个时间序列 x_1,x_2,x_3,…,序列跨度为 $j\tau$ 的自相关函数为[10]:

$$R_{xx}(j\tau) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)x(t+j\tau) \tag{1}$$

可固定 j,当相关函数 R_{xx} 降到初始值$(1-1/e)$倍时,得到最佳时间延迟 τ。

Grassberger 和 Procaccia 提出的 G - P 算法[11],基本步骤如下[10]:

(1) 对于时间序列 $x_1,x_2,x_3,\cdots$,先尝试取一个较小的嵌入维数值 m_0,则在相应的重构相空间中有:

$$X(t_i) = [x(t_i), x(t_i+\tau), x(t_i+2\tau), \cdots, x(t_i+(m-1)\tau)], i = 1,2,\cdots \tag{2}$$

(2) 计算关联积分。

$$C(r) = \lim_{n\to\infty} \frac{2}{N(N-1)}\sum_{i=1,j<1}^{N} \theta(r - | X(t_i) - X(t_j) |) \tag{3}$$

其中:

$| X(t_i) - X(t_j) |$ 表示相点 $X(t_i)$ 和 $X(t_j)$ 的距离,$\theta(z)$ 是 Heaviside 函数:

$$\theta(z) = \begin{cases} 0, z < 0 \\ 1, z > 0 \end{cases} \tag{4}$$

$C(r)$ 是一个关联积分,表示相空间距离小于 r 的概率。

(3) 对于关联球半径 r 的某个适当范围,吸引子的关联积分 $C(r)$ 应该随 r 的变化呈现幂函数标度律,即在双对数坐标图上,$C(r)$ 随 r 的变化曲线呈现为一段直线,此直线的斜率给出吸引子维数 $d(m) = \ln C(r)/\ln r$。对于所有$(r, C(r))$点进行拟合,从而求出对应于 m_0 的关联维数估计值 $d(m_0)$。

(4) 增加嵌入维数,取 $m_1 > m_0$,重复计算步骤(2)和(3),直到相应的维数估计值 $d(m)$ 不再随 m 的增长而增长(在一定误差范围内)。此时得到的 $d(m)$ 即为吸引子的关联维,m 即为嵌入维。如果 d 随 m 的增长而增长,并不收敛于一个稳定的值,则表明所考虑的系统是一个相应于随机时间序列的随机系统,它并不是少自由度的决定论性非线性动力学系统,它在有限维数的相空间中不存在吸引子。

1.2 典型相关分析

典型相关分析(Canonical Correlation Analysis, CCA)是一种常用的研究变量之间相关性的多元统计方法,由 Hotelling 于 1936 年提出。

典型相关分析(CCA)的基本思路[12]为:以最大限度地提取两组变量 X 和 Y 之间的主要特征为准则,从 X 中提取组合变量 U,从 Y 中提取组合变量 V,如式(5)所示:

$$\begin{cases} U_{n\times l} = X_{p\times n}A_{n\times l} \\ V_{n\times l} = Y_{q\times n}B_{n\times l} \end{cases} \tag{5}$$

其中,A,B 为线性变换,又称为空间特征向量。

按照式(5)将 X 与 Y 之间的相关转化为 U 与 V 之间的相关。

2 基于相空间重构后的相关性计算

沪深股指时间序列是比较复杂的时间序列系统,它们之间的相关性也是比较复杂的,复杂性体现在很多方面,那么文中提出相空间重构后计算相关性,旨在还原沪深股指时间序列的复杂度。沪深股指之间相关具有滞后性这样一个特点,也就是说,例如,当上证综指在 t_1 时刻到 t_2 时刻产生波动时,深圳成指在 t_3 时刻到 t_4 时刻才出现一个类似(正相关)或者相反(负相关)的波动,而 $t_1 < t_3$,如果此时利用传统的计算相关性的方法可能就得到一个不理想的甚至是相反的相关性。

复杂性还体现在上证综指和深圳成指各自内部前一段时间的股指波动对后面的波动会产生影响,即具有自相关的特点,各自的自相关也会影响到两者之间的相关性。

沪深股指时间序列相关性的研究本就不多,更没有学者考虑这种复杂性对相关性的影响。为了把沪深股指的复杂性体现在具体的算法中,文中引入相空间重构的概念。沪深股指时间序列,都是一维的,但是对于股指这样一个复杂的时间序列系统,一维不能很好地描述它的所有特性。通过相空间重构以后,就构成是多维的了,这里将每一维都看成新构成的时间序列的属性。首先各个维上的时间起点是不同的,不仅仅是单个时间序列的起点不同,两个序列之间各个对应属性起点也是不同的,这样就能很好地描述时间序列的相关滞后性;再次形成多维后,各自时间序列内部各个属性之间的相关性也可以求,这个相关性其实就是自相关。

那么在这样一个多维的时间序列系统中,要计算两个系统中的相关性,明显已经不是简单相关或者复相关可以计算的了。

文中采用典型相关分析的方法得到沪深股指时间序列重构后的系统的相关性。从典型相关分析(CCA)的过程也可以看出,CCA 充分地考虑到了内部属性之间的相关对两者相关性的影响。

文中提出的基于相空间重构的相关性计算的模型,基本步骤如下:

Step1:分别计算上证综指时间序列 $X=(x_1,x_2,\cdots,x_n)$ 和深圳成指时间序列 $Y=(y_1,y_2,\cdots,y_n)$ 的时间延迟 τ_1,τ_2。文中时间延迟的选择采用自相关法。

Step2:基于 Step1 计算出的时间延迟,分别求出 $X=(x_1,x_2,\cdots,x_n)$ 和 $Y=(y_1,y_2,\cdots,y_n)$ 的嵌入维数 m_1,m_2。文中嵌入维数的选择使用 GP 算法。

Step3:根据 Step1 和 Step2 求出的时间延迟 τ_1,τ_2 和嵌入维数 m_1,m_2,整理 $X=(x_1,x_2,\cdots,x_n)$ 和 $Y=(y_1,y_2,\cdots,y_n)$ 得到新的多维的时间序列:

$$\begin{cases} X_1=(x_1,x_{1+\tau_1},\cdots,x_{1+(m_1-1)\tau_1}) \\ X_2=(x_2,x_{2+\tau_1},\cdots,x_{2+(m_1-1)\tau_1}) \\ \cdots \\ X_i=(x_i,x_{i+\tau_1},\cdots,x_{i+(m_1-1)\tau_1}) \\ \cdots \end{cases} \tag{6}$$

$$\begin{cases} Y_1=(y_1,y_{1+\tau_2},\cdots,y_{1+(m_2-1)\tau_2}) \\ Y_2=(y_2,y_{2+\tau_2},\cdots,y_{2+(m_2-1)\tau_2}) \\ \cdots \\ Y_i=(y_i,y_{i+\tau_2},\cdots,y_{i+(m_2-1)\tau_2}) \\ \cdots \end{cases} \tag{7}$$

Step4:(6) 和(7) 式分别是相空间重构后的 m_1 和 m_2 维的时间序列,每一维看成是一个属性,对此多维时间序列系统进行典型相关分析。

3　实验及对比

3.1　数据选取

文中分别选取沪深股指 2000 年 1 月 4 日至 2008 年 12 月 31 日的每日收盘价各 2171 个数据。股价波动如图 1 所示。

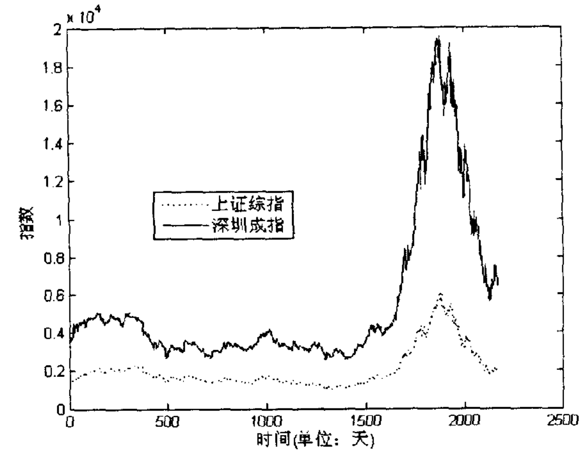

图 1　上证综指和深圳成指的股价波动图

3.2　时间延迟 τ 和嵌入维数 m 的计算

采用自相关法求得上证指数收益时间序列的最佳时间延迟 $\tau_1=297$,深圳成指收益时间序列的最佳时间延迟 $\tau_2=240$。

采用 GP 算法求嵌入维数,对于上证综指取 $\tau_1=297$,$r_1=2000,1800,1600,1400,1200,1000,800,600,400,200,100,50$,嵌入维数从 $m_1=2$ 开始计算,分别得到:

$m_1=2$ 时,$d_1=1.1189$;

$m_1=3$ 时,$d_1=1.3096$;

$m_1=4$ 时,$d_1=1.4219$;

$m_1=5$ 时,$d_1=1.4289$;

$m_1=6$ 时,$d_1=1.4689$。

$\ln C_1$ 与 $\ln r_1$ 的关系如图 2 所示。

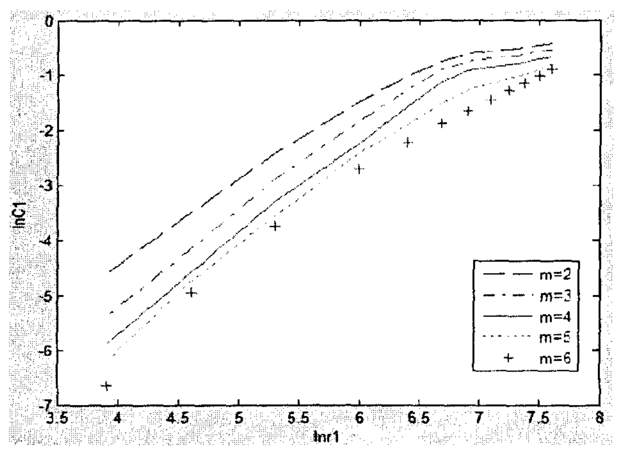

图 2　上证综指 $\ln C_1$ 与 $\ln r_1$ 的关系

对于深圳成指取 $\tau_2 = 240, r_2 = 6000,5000,4000,3500,3000,2500,2000,1000,800,600,400,200$，嵌入维数从 $m_2 = 2$ 开始计算，分别得到：

$m_2 = 2$ 时，$d_2 = 0.7971$；

$m_2 = 3$ 时，$d_2 = 1.1020$；

$m_2 = 4$ 时，$d_2 = 1.2835$；

$m_2 = 5$ 时，$d_2 = 1.3460$；

$m_2 = 6$ 时，$d_2 = 1.3508$；

$m_2 = 7$ 时，$d_2 = 1.3253$。

$\ln C_2$ 与 $\ln r_2$ 的关系如图3所示。

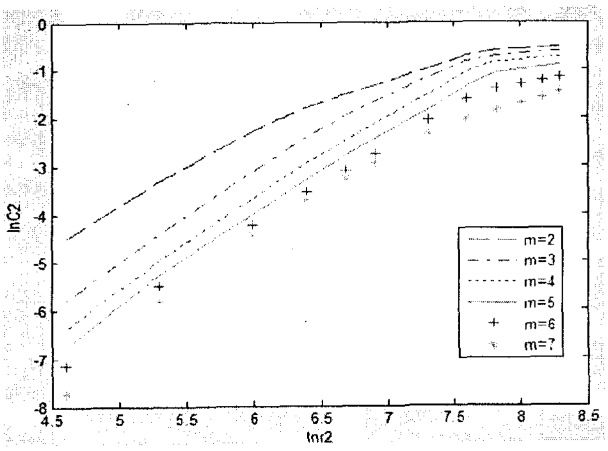

图3 深圳成指 $\ln C_2$ 与 $\ln r_2$ 的关系

通过对比，文中选取上证综指的嵌入维数 $m_1 = 4$，此时关联维 $d_1 = 1.4219$；选取深圳成指嵌入维数 $m_2 = 5$，此时关联维 $d_2 = 1.3460$。

3.3 相关性计算

沪深股指时间序列经相空间重构后，再运用典型相关分析得到的结果如表1所示。

表1 典型相关分析结果

典型相关性	相关系数
1	0.988
2	0.959
3	0.900
4	0.711

为了对比，文中对未相空间重构的一维时间序列进行相关分析，采用的方法也是典型相关分析(CCA)，得到的相关系数为0.982。

计算得到沪深股指收益具有很强的相关性，第一典型相关系数、第二典型相关系数和第三典型相关系数都在0.9以上。另外，在未重构相空间的基础上，采用典型相关分析计算出的相关系数为0.982，和第一典型相关系数很接近，说明此模型是适用的。

第二典型相关系数和第三典型相关系数均比未重构时的小，并且第一典型相关系数和第二典型相关系数都要略大于文献[6]的计算结果0.9023，第一典型相关系数要略大于二元正态的估计值0.9333，说明了沪深股指时间序列的自身的相关性和相关的滞后性都对两股相关产生了影响。

4 结束语

通过文中建立的相关系数计算模型，在一定程度上还原了沪深股指时间序列的复杂性，将一维时间序列扩展到多维空间，增加了对比的角度，并且充分考虑了时间序列相关的滞后性和自身的相关性对两者之间的相关性的影响。

文中还存在着不少不足之处，相关系数的大小不是一成不变的，会随着时间的增长发生变化，即相关系数是动态变化的，这也是文中需要改进的地方。

参考文献：

[1] Chui A C W, Kwork C C Y. Cross－Autocorrelation Between A Shares and B Shares in the Chinese Stock Market[J]. Journal of Financial Research, 1998(3): 333－353.

[2] 樊 智，张世英. 多元GARCH建模及其在中国股市分析中的应用[J]. 管理科学学报，2003(2): 68－73.

[3] 郑振龙，张 雷. 中国主要股指收益相关性研究[J]. 厦门大学学报，2007(3): 35－39.

[4] 杨兴民，刘保东，李 娟. 基于Gaussian Copula与t－Copula的沪深股指相关性分析[J]. 山东大学学报，2007，42(12): 63－68.

[5] 陈守东，陈 雷，刘艳武. 中国沪深股市收益率及波动性相关分析[J]. 金融研究，2003(7): 80－85.

[6] 陈 铿，韩伯棠. 混沌时间序列分析中的相空间重构技术综述[J]. 计算机科学，2005，32(4): 67－70.

[7] Takens F. Detecting strange attractor in turbulence[J]. Lecture Notes in Mathematics, 1981, 898(2): 366－381.

[8] Packard. Geometry from a Time Series[J]. Physical Review Letters, 1980, 45(9): 712－716.

[9] 马红光，李夕海，王国华，等. 相空间重构中嵌入维和时间延迟的选择[J]. 西安交通大学学报，2004，38(4): 335－338.

[10] 张中华，丁华福. 基于混沌神经网络的股票分析及其预测[J]. 计算机技术与发展，2009，19(3): 185－188.

[11] Grassberger P, Procacia I. Dimension and Entropy of Strange Attractors from a Fluctuating[J]. Physica, 1984, 13: 34－54.

[12] 杨雪梅，董逸生，徐宏炳，等. 高维数据流的在线相关性分析[J]. 计算机研究与发展，2006，43(10): 1744－1750.