第 27 卷第 3 期　信 息 与 控 制　Vol. 27, No. 3
1998 年 6 月　Information and Control　June, 1998

度量两个序列非线性相关性的一种方法*

樊重俊　王浣尘
(上海交通大学系统工程研究所　上海　200052)

摘　要　基于分数维数, 提出了非线性相关度的概念, 用于度量两列经济数据的非线性相关程度, 以解决非线性经济预测中的变量选择问题. 仿真结果与应用结果说明该方法效果较好.

关键词　非线性动态系统, 分数维数, 经济预测, 非线性相关

1　引言

经济系统是一个极为复杂的非线性系统, 基于线性理论的新古典经济学在解释一些经济现象时往往显得不很有效, 而一些非线性方法以其解释复杂经济系统的能力而越来越引起人们的重视. 然而, 在非线性经济学研究方面所作的工作, 大多是利用低阶的非线性模型, 在现有的经济学框架内来讨论经济模型中产生分叉、浑沌等复杂动态的可能性, 而涉及到对观测数据的统计推断工作并不多.

本文考虑如下的离散时间非线性动态系统$(h, F, x(0))$:

$$\begin{aligned} x(t) &= F(x(t-1)) \quad x(0) \text{ 给定} \\ y(t) &= h(x(t)) \quad t = 1, 2, \cdots, n \end{aligned} \tag{1}$$

其中

$$x(t) = (x_1(t), \cdots, x_r(t))' \in R^r, \quad y(t) = (y_1(t), \cdots, y_p(t))' \in R^p$$

分别为状态变量和观测变量, 通常 $p = 1$. 已有很多文献对 $F(\cdot)$ 函数形式已知时的非线性统计特征进行了讨论. 但通常对 $F(\cdot)$ 的具体形式一无所知, 也不知道该动态系统应该由哪些状态变量组成, 所能得到的只是部分观测变量的数据. 然而观测数据做为多维系统的一个侧面, 却包含着原来的经济系统所有的变量的痕迹, 可以根据它所提供的信息, 通过相空间重构方法[1] 来研究原来的经济系统的动力学特征. 已有很多文献讨论了系统的奇异吸引子维数的估计方法[1~3], 并且, 基于相关积分的概念, Brock, Dechert 和 Scheinkman[4] 给出了一种检验一维时间序列是否具有非线性相关性(dependence)的方法(BDS 检验), 这一方法可用来检验经济观测数据的可预测性, 得到了广泛的应用. 本文研究经济预测中的另一个重要问题: 变量选择问题. 事实上, 从数学上讲, 变量选择问题就是要推断两个序列的非线性相关程度.

当我们确定了需要预测的经济指标后, 首先要分析与该指标有关的经济指标都有那些, 然后再选择有效的定量预测方法进行预测. 对于变量选择问题目前大多采用定性分析的方法, 即通过经济理论, 进行经济因素的相关性分析. 另一种常用的做法是, 先预选若干个与待预测变量可能有关的变量, 然后选择定量预测方法进行预测, 若预测结果较好, 则认为预选的变量有效, 事实上, 这是一种事后推断的方法. 目前也有一些文献涉及变量选择的定量方法, 但均是在线性的意义下进行讨论, 关于非线性意义下的变量选择的定量方法, 尚未见诸文献. 本文基于

1997-03-18 收稿
* 国家自然科学基金与中国博士后科学基金资助项目

相关维数, 提出了非线性相关度的概念, 用于推断两列经济数据的非线性相关程度, 以解决非线性经济预测中的变量选择问题.

随着现代计算机技术与信息处理技术的飞速发展, 国内外的先进企业已相继开发和使用了各具特色的管理信息系统与决策支持系统. 市场分析与预测方法是决策支持系统中所使用的重要方法之一. 本文方法已用于我们开发的"金川有色金属公司经营计划决策支持系统"(简称 JCDSS) 中, 取得了较好的效果.

2 非线性相关度

对于系统$(h, F, x(0))$, 定义

$$C_n(x, \epsilon) = \#\{t, s), 1 \leq t < s \leq n: \| x(t) - x(s) \| < \epsilon\} / n(n-1)$$

若下列极限存在

$$D(x) = \lim_{\epsilon \to +0} \lim_{n \to \infty} \ln C_n(x, \epsilon) / \ln \epsilon$$

则称$D(x)$是系统$(h, F, x(0))$的相关维数. 通常范数取为

$$\| x(t) - x(s) \| = \max_{i=0, \cdots, r-1} x_i(t) - x_i(s) \tag{2}$$

$\#$ 表示集合的势. 此时有

$$C_n(x, \epsilon) = \#\{(t, s), 1 \leq t < s \leq n: \| x_i(t) - x_i(s) \| < \epsilon, i = 1, \cdots, r\} / n(n-1)$$

设已知p 维观测向量y 的观测数据为$\{y(t): t = 1, \cdots, n\}$, 构造

$$y^m(t) = [y(t)', \cdots, y(t+m-1)']'$$

其中m 称为嵌入维数, 定义相关积分:

$$C_{m,n}(y, \epsilon) = \#\{(t, s), 1 \leq t < s \leq N: \| y^m(t) - y^m(s) \| < \epsilon\} / N(N-1)$$

其中$N = n - m + 1$, 由此给出相关维数估计为

$$D_{m,n}(y, \epsilon) = \ln C_{m,n}(y, \epsilon) / \ln \epsilon \tag{3}$$

或对小样本情形取

$$D_{m,n}(y, \epsilon_1, \epsilon_2) = \ln(C_{m,n}(y, \epsilon_1) / C_{m,n}(y, \epsilon_2) / \ln(\epsilon_1 / \epsilon_2) \tag{4}$$

显然有

$$\lim_{\epsilon_1 \to 0, \epsilon_2 \to 0} \lim_{n \to \infty} D_{m,n}(y, \epsilon_1, \epsilon_2) = \lim_{\epsilon \to 0} \lim_{n \to \infty} D_{m,n}(y, \epsilon)$$

为了构造非线性相关度, 需要用到下面的结果(证明见后文),

假设 1 对于系统$(h, F, x(0))$, (1) h 和 F 是 C^2 光滑的; (2) F 具有唯一的紧吸引子L, 且存在轨道在 L 上稠密; (3) F 具有唯一的遍历不变测度 m, 且 m 具有连续密度 $m(dx) = r(x)dx$; (4) 由 $x(0)$ 所确定的轨道在 L 上稠密; (5) 最大 Lyapunov 指数为正.

定理 1 对于系统$(h, F, x(0))$, 在假设 1 的意义下, 设吸引子 L 是一个 d 维紧流性, 且

$$r(x) > 0 \quad \forall x \in L$$

则当 $m \geq 2d + 1$ 时, 通有地有

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} D_{m,n}(y, \epsilon) = d \tag{5}$$

由定理知, 若 y_1 和 y_2 来自于同一动态系统, 则有

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} D_{m,n}(y_1, y_2)', \epsilon) = \lim_{\epsilon \to 0} \lim_{n \to \infty} D_{m,n}(y_1, \epsilon) \tag{6}$$

而我们注意到若 y_1 和 y_2 来自于不同的动态系统, 应有

$$P(\| y^m(t) - y^m(s) \| < \epsilon) = P(\| y_1^m(t) - y_1^m(s) \| < \epsilon, \| y_2^m(t) - y_2^m(s) \| < \epsilon) =$$

$$P(\|y_1^m(t)-y_1^m(s)\|<\epsilon)P(\|y_2^m(t)-y_2^m(s)\|<\epsilon)$$

又

$$C_{m,n}(y,\epsilon)\xrightarrow{P}P(\|y^m(t)-y^m(s)\|<\epsilon)\quad n\to\infty$$

由此得

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}((y_1,y_2)',\epsilon)/\ln(\epsilon)=\lim_{\epsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}(y_1,\epsilon))/\ln(\epsilon)+$$
$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}(y_2,\epsilon))/\ln(\epsilon)$$

即

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}D_{m,n}(y_1,y_2)',\epsilon)=\lim_{\epsilon\to 0}\lim_{n\to\infty}D_{m,n}(y_1,\epsilon)+\lim_{\epsilon\to 0}\lim_{n\to\infty}D_{m,n}(y_2,\epsilon)\tag{7}$$

可利用(6)与(7)式的结果,对观测变量 y_1 和 y_2 的非线性相关性进行统计推断. 若(6)成立,则可认为观测变量 y_1 和 y_2 来自于同一动态系统,即 y_1 和 y_2 具有很强的非线性相关性;若(7)成立,则可认为 y_1 和 y_2 是完全没有关系的两个指标,即 y_1 和 y_2 是非线性不相关的. 然而,对于经济数据,大多数情况则是具于两者之间. 事实上,在统计理论中,采用相关系数来反映两个变量的线性相关程度,为此,我们定义如下的两个序列的非线性相关度的概念.

定义 1　对于序列 $\{y_1(t),t=1,\cdots,n\}$ 与 $\{y_2(t),t=1,\cdots,n\}$,定义:

$$R_{m,n}(y_1,y_2,\epsilon)=\frac{I_{m,n}(y_1,y_2,\epsilon)}{D_{m,n}((y_1,y_2)',\epsilon)}\tag{8}$$

其中

$$I_{m,n}(y_1,y_2,\epsilon)=D_{m,n}(y_1,\epsilon)+D_{m,n}(y_2,\epsilon)-D_{m,n}((y_1,y_2,)'\epsilon)$$

则称(8)为序列 $\{y_1(t),t=1,\cdots,n\}$ 与 $\{y_2(t),t=1,\cdots,n\}$ 的非线性相关度.

不难看出,当(6)成立时非线性相关度为 1,当(7)成立时非线性相关度为 0,其它情况则是具于 0、1 之间. 因此,可利用非线性相关度来反映序列 $y_1(t)$ 和 $y_1(t)$ 的非线性相关程度:非线性相关度越大,非线性相关程度越强.

3　结果证明

引理 1　对于系统 $(h,F,x(0))$,在假设 1 和 $m\geq 2d+1$ 下,存在 K 使得

$$K\|x(t)-x(s)\|<\|y_1^m(t)-y_1^m(s)\|<K^{-1}\|x(t)-x(s)\|$$

证明　见[3]定理 6 的证明过程.

定理 1 的证明:由引理 1 可得,存在 $\{K_i,i=1,\cdots,p\}$,使得

$$K_i\|x(t)-x(s)\|<\|y_1^m(t)-y_1^m(s)\|<K_i^{-1}\|x(t)-x(s)\|,\quad i=1,\cdots,P$$

取

$$K=\min\{K_i,\quad i=1,\cdots,p\}$$

在范数(2)的意义下有

$$\|y^m(t)-y^m(s)\|=\max_{i=1,\cdots,p}\|y_i^m(t)-y_i^m(s)\|$$
$$K\|x(t)-x(s)\|<\|y^m(t)-y^m(s)\|<K^{-1}\|x(t)-x(s)\|$$

因此

$$C_n(x,K\epsilon)<C_{m,n}(y,\epsilon)<C_n(x,K^{-1}\epsilon)$$

而

$$\lim_{\epsilon\to+0}\lim_{n\to\infty}\ln C_n(x,K\epsilon)/\ln\epsilon=\lim_{\epsilon\to+0}\lim_{n\to\infty}\ln C_n(x,K\epsilon)/\ln K\epsilon=\lim_{\epsilon\to+0}\lim_{n\to\infty}\ln C_n(x,\epsilon)/\ln\epsilon$$

定理获证.

4 仿真结果

考虑如下的 Henon 映射[5]:

$$x_1(t+1) = 1 - ax_1(t)^2 + x_2(t)$$
$$x_2(t+1) = bx_1(t)$$
$$y_1(t+1) = 0.2x_2(t+1) + 0.2x_1(t+1)$$
$$y_2(t+1) = 0.3x_2(t+1) - 0.3x_1(t+1)$$

y_1 和 y_2 是两个观测变量, 另设观测数据$\{V(t): t=1,\cdots,n\}$来自于如下的 Logistic 映射:

$$u(t+1) = Au(t)(1-u(t))$$
$$v(t+1) = u(t+1)$$

参数与初始值取

$$a = 1.4 \quad b = 0.3 \quad x_1(-10000) = x_2(-10000) = 0$$
$$A = 4 \quad u(-10000) = 0.1$$

以保证观测数据$\{y_1(t): t=1,\cdots,n\}$、$\{y_2(t): t=1,\cdots,n\}$和$\{v(t): t=1,\cdots,n\}$稳定, $n=2000$. 基于公式(4)的数值结果如表 1.

仿真结果说明, 对于来自于同一动态系统的两个观测变量 y_1 和 y_2, 即 y_1 和 y_2 具有很强的非线性关系, 则 $D_{m,n}(y_1,y_2)$、$D_{m,n}(y_1)$、$D_{m,n}(y_2)$ 三项差别不大, 此时 y_1 和 y_2 的非线性相关度接近于1. 对于来自于不同动态系统的两个观测变量 y_1 和 v, 即 y_1 和 v 之间不存在非线性关系, 则近似的有

$$D_{m,n}(y_1,v) = D_{m,n}(y_1) + D_{m,n}(v)$$

此时 y_1 和 v 的非线性相关度接近于 0.

表 1 基于公式(4)的观测变量数值结果

m	4	4	5	5	6	6
ϵ_1	0.010	0.008	0.010	0.008	0.010	0.008
ϵ_2	0.005	0.004	0.005	0.004	0.005	0.004
$D_{m,n}(y_1,\epsilon_1,\epsilon_2)$	1.1031	1.1262	1.1050	1.1513	1.1210	1.1846
$D_{m,n}(y_2,\epsilon_1,\epsilon_2)$	1.2125	1.1814	1.2126	1.1448	1.1939	1.1313
$D_{m,n}(y_1,y_2,\epsilon_1,\epsilon_2)$	1.1893	1.1642	1.1823	1.1505	1.1621	1.1586
$R_{m,n}(y_1,y_2,\epsilon_1,\epsilon_2)$	0.9470	0.9822	0.9602	0.9957	0.9919	0.9988
$D_{m,n}(v,\epsilon_1,\epsilon_2)$	1.0161	0.9875	1.0996	1.0332	1.1596	1.0783
$D_{m,n}(y_1,v,\epsilon_1,\epsilon_2)$	2.0325	2.0130	2.0440	2.1632	2.1369	2.1535
$R_{m,n}(y_1,v,\epsilon_1,\epsilon_2)$	0.0420	0.0500	0.0785	0.0098	0.0672	0.0508

5 实用实例

下面以 JCDSS 系统中伦敦有色金属市场镍价格趋势分析与预测问题为例, 说明本文方法的应用. 通过人机对话, 我们可从数据库中抽取从94 年10 月到 96 年1 月的有关数据. 设 x, y 分别表示镍成交价、3 个月期货价, 并剔除趋势项, 做规格化处理. 计算结果如表 2. 从数据结果

可以看出, 非线性相关度 $R_{m,n}(x, y, \epsilon_1, \epsilon_2)$ 在 0.85 左右, 说明镍成交价与 3 个月期货价具有很强的相关性, 预测时应做为一个总体来考虑, 以上结果可以用来帮助确定非线性预测模型的结构[7].

表 **2** 伦敦有色金属市场镍价格数值结果

m	4	4	5	5	6	6
ϵ_1	0.05	0.04	0.06	0.05	0.07	0.06
ϵ_2	0.025	0.02	0.03	0.025	0.035	0.03
$D_{m,n}(x, \epsilon_1, \epsilon_2)$	1.531471	1.904066	1.575236	1.877418	1.599219	1.894866
$D_{m,n}(y, \epsilon_1, \epsilon_2)$	1.477532	1.842138	1.525993	1.815922	1.546460	1.826865
$D_{m,n}(x, y, \epsilon_1, \epsilon_2)$	1.630392	2.057583	1.669857	1.993473	1.688972	2.010571
$R_{m,n}(x, y, \epsilon_1, \epsilon_2)$	0.845611	0.820704	0.856886	0.852764	0.862514	0.855812

6 结束语

本文方法可用于推断两列观测数据是否来自于同一非线性动态系统, 以解决非线性经济预测中的变量选择问题, 仿真结果与应用实例说明, 本文给出的推断方法效果很好.

参 考 文 献

1 Takens F. Detecting Strange Attractors in Turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics 898, Springer- Verlag, Berlin

2 Grassberger P, Procaccia I. Measuring the Strangeness of Strange Attractors. Phys. 9D, 1983

3 Brock W A. Distinguishing Random and Deterministic Systems: Abridged Version. Journal of Economic Theory 40, 1986

4 Brock W, Dechert W D, Scheinkman J. A Tests for Independence Based on the Correlation Dimension. Working Paper, Department of Economics, University of Wisconsin, Madison, 1987

5 Henon M. A Two-dimensional Mapping with a Strange Attractor. Communacations in Mathematical Physics 1976, **50**: 69～77

6 张朋柱, 曹文华, 樊重俊, 韩崇昭. 金川有色金属公司经济活动分析经营计划决策支持系统的快速生成——UCP- IDSS 生成环境应用之二. 决策与决策支持系统, 1995, **5**(3)

7 Fan Chongjun, Han Chongzhao, Hu Baosheng. Forecasting the Behavior of Multivariate Time Series with Neural Networks in a DSS for Business Planning. Proceedings of ISIM'96 (Nanjing, 1996)

MEASURING NONLINEAR DEPENDENCE BETWEEN TIME SERIES

FAN Chongjun WANG Huanchen

(*Systems Engineering Institute, Shanghai Jiaotong University, Shanghai* 200052)

Abstract In this paper, a statistic-Nonlinear Dependence Coefficient based on fractal dimension is proposed. This statistic is designed for measuring nonlinear dependence between time series, and can be used in variable selecting for economical forecasting. Numerical results show the statistic is reasonable.

Key words nonlinear dynamics, fractal dimension, forecasting, statistic

作者简介

樊重俊, 男, 34 岁, 副教授, 博士后. 研究领域为决策理论与决策支持系统, 企业 Internet 策略, 可持续发展, 经济分析与预测等.

王浣尘, 男, 63 岁, 教授. 研究领域为控制理论, 系统工程, 管理科学等.