非线性自回归模型的非参数方法及应用

陈耀辉 著

科学出版社

非线性自回归模型的非参数方法及应用

陈耀辉　著

教育部人文社会科学研究一般项目（08JA790063）
江苏高校优势学科建设工程资助项目（PAPD）
南京财经大学学术著作出版基金
联合资助

科学出版社
北京

内 容 简 介

非线性自回归模型是时间序列分析中一类重要的模型，而且和实际应用有密切关系。本书用非参数方法系统深入地研究非线性自回归模型的基本理论、方法及应用，其中包括核估计的中心极限定理、自助估计法在非线性自回归模型中的应用、线性和非线性自回归模型阶的确定、欧式期权的非参数方法以及美式期权的非线性分析等内容。

本书可作为高等院校统计学专业硕士研究生、博士研究生教材，也可供相关专业研究生、教师、科研人员和统计工作者参考。

图书在版编目(CIP)数据

非线性自回归模型的非参数方法及应用/陈耀辉著. —北京：科学出版社，2012

ISBN 978-7-03-035952-0

Ⅰ. ①非… Ⅱ. ①陈… Ⅲ. ①非线性回归-自回归模型-非参数法-研究 Ⅳ. ①O212. 1

中国版本图书馆 CIP 数据核字(2012)第 260977 号

责任编辑：刘燕春 罗 吉 / 责任校对：朱光兰
责任印制：赵德静 / 封面设计：许 瑞

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

新科印刷有限公司 印刷
科学出版社发行 各地新华书店经销

*

2013 年 1 月第 一 版 开本：B5(720×1000)
2013 年 1 月第一次印刷 印张：9
字数：180 000

定价：48.00 元

前　　言

近年来，非线性时间序列研究正处在突飞猛进的发展时刻。过去，时间序列模型主要是参数化的，但是现在，非参数化、半参数化的模型正在发挥着无与伦比的威力，把整个时间序列领域提升到一个前所未有的新境界。作者从线性到非线性，从参数模型到非参数、半参数模型，不断地汲取营养，但又感叹于非线性时间序列模型内容的浩瀚，自知以己之力，绝不可能用非参数方法对非线性时间序列模型的全部内容进行全面系统的研究，只能从中选择一个具体的时间序列模型——非线性自回归模型作为本书的研究对象，采用非参数核估计的方法进行细致的分析，得到几个重要的结论。

本书第 1 章简要介绍非线性自回归模型的非参数方法的研究背景，论述本书的理论和实践意义，给出本书的研究框架。

在第 2 章里，导出一般 k 阶非线性自回归模型在最优窗宽选择下，非参数核估计量一个新的中心极限定理。得到非线性自回归模型的核估计量的渐近正态性及估计量的逐点 MSE 和 MISE 的计算公式；由此构建在回归函数未知的情况下自回归曲线的逐点置信区域；通过一阶和二阶非线性自回归模型的数值研究和模拟计算进一步验证这一理论。

第 3 章将平稳自助法和无序自助法应用到非参数核自回归估计量。得到在非线性自回归的情形下，对于核自回归估计，平稳自助法是有效的。利用平稳自助法和无序自助法，在一阶和二阶非线性自回归过程的情况下，获得非线性自回归曲线的自助置信区域并与利用正态近似得到的逐点置信区域进行比较。

第 4 章讨论在非线性和线性自回归模型中阶的误设效应问题。一般情况下，时间序列模型的真实阶是未知的，所以非常容易产生阶的误设。首先，讨论在非线性自回归模型中，考虑在拟合阶大于和小于真实阶的两种情况对非参数核估计和平稳自助估计的影响；其次，讨论在线性自回归模型中，最小二乘估计和自助估计的拟合阶以一个更慢的速度趋于无穷时的效应。特别地，如果拟合阶趋于无穷的速度明确给定，则正态近似法和自助近似法在非线性和线性两种情况下仍然成立。

第 5 章介绍欧式期权定价的非参数方法，用非参数方法研究股票价格在不服从几何布朗运动下欧式期权价值的评估。首先从理论上论证基于非参数的欧式看涨期权评价方法，然后从上海证券市场收集数据，实证研究用该方法评价欧式看涨期权与经典的 Black-Scholes 定价的结果有所不同，但非参数定价方法更贴近市场。

第 6 章介绍美式期权定价的非线性分析。首先在最优停时理论中，利用动态规划原则，得到关于美式(看涨或看跌)期权定价的一个非线性 Black-Scholes 型偏微分方程，利用黏性解的概念证明该偏微分方程的解的存在性和唯一性，然后给出一种基于最小二乘法的美式期权定价的最优停时的数值算法，由此得到美式期权定价的一个非线性方法。

本书适用对象为统计专业的高年级硕士研究生、博士研究生、教师以及相关专业的科研工作者。

本书研究的课题得到教育部人文社会科学研究一般项目的鼎力支持，在写作的过程中得到了华中科技大学李楚霖教授和胡适耕教授的大力支持，还得到了东南大学韦博成教授和林金官教授许多有益的建议。此外，本书的出版得到了南京财经大学学术著作出版基金的资助，得到了科学出版社的大力支持和帮助，在此一并表示衷心的感谢。

由于非线性自回归模型的非参数方法及其应用是理论性和实践性都很强的研究领域，尚有许多问题有待深入分析和研究，加上本人水平所限，书中不妥之处在所难免，恳请读者批评指正。

陈耀辉

2012 年 4 月于南京财经大学

目　　录

第 1 章　引　　论

1.1　绪言

现实世界的运动规律往往是非线性的。事实上,在一个线性的世界里,量变永远都不能产生质变。换言之,物理学的相变、生物学的细胞突变、经济学的收益递减等都会消失于线性的世界里。可想而知,线性的世界显得过于简单而单调。

非参数方法的研究是当前计量经济学研究中的一个重要方向,是继协整理论之后,计量经济学的又一个热点研究方向。线性和非线性回归模型都假定经济变量的关系已知,现实中,经济变量之间的关系未必是线性或可线性化的非线性关系,而变量之间的参数非线性关系又很难确定。所以传统线性或非线性计量经济模型在实际应用中往往存在模型的设定误差,不能满足经济和管理应用研究的需要。非参数方法假定经济变量的关系未知,要对整个回归函数进行估计,因而非参数回归模型是较线性和非线性回归更符合现实的模型。本书采用非参数核估计方法对非线性自回归模型进行研究。

线性范围之外,尚有无穷多的非线性形式有待挖掘。一方面,非线性时间序列分析的早期发展重点是各种非线性参数模型(Tong,1990;Tjstheim et al.,1994a,1994b)。成功的例子包括金融数据的波动波幅性的 ARCH 模型(Engle,1982;Bollerslev,1986)和生态及经济数据的门限模型(Tong,1990;Tiao et al.,1994)。另一方面,非参数回归方法的最新进展为建立非线性时间序列模型提供了另一种手段(Tjstheim et al.,1994a,1994b)。这种方法的一个明显优点是对模型结构的先验信息要求很少,而且为进一步的参数拟合提供有用的感性认识。此外,近几年来随着计算能力的增强,存取和试图分析海量的和复杂的时间序列数据已变成平凡之事,随之而来的是对那些能够识别内在结构和按照新的精度标准预报将来的非参数和半参数分析工具的需求日益增大,这无疑为本书增加了几分理论价值,也体现了本书的应用价值。

1.2　国内外研究现状

在 20 世纪八九十年代以来取得过重要进展的统计学诸多领域中,非线性时间

序列和非参数技术几乎是沿着两个不同的轨迹发展起来的，尽管在时间序列分析中运用非参数的技术可追溯到 20 世纪 40 年代，但将非参数核估计方法运用到非线性自回归模型中至今还没有得到一般性的结论（更详细的综述见每章内容的论述）。

本书考虑一个非线性 $k(k\geqslant 1)$ 阶自回归过程 $\{X_n \mid n\geqslant 0\}$：

$$X_{n+1} = H(X_{n+1-k},\cdots,X_n) + \eta_{n+1} \quad (n \geqslant k-1)$$

其中，H 是 $\mathbf{R}^k$ 上的一个实值 Borel 可测函数，$\{\eta_n \mid n\geqslant 1\}$ 是具有零均值和有限方差的独立同分布实值随机变量序列，$\{X_0, X_1,\cdots,X_{k-1}\}$ 是独立于 $\{\eta_n \mid n\geqslant 1\}$ 的任意实值随机变量。

利用 $\mathbf{Y}_n = (X_n, X_{n+1},\cdots,X_{n+k-1})$ 定义一个 Markov 过程 $\{\mathbf{Y}_n \mid n\geqslant 0\}$。设 $P^{(n)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$［或 $P^{(n)}(\boldsymbol{x},A)$，$A\in\mathbf{B}^k$ 是 $\mathbf{R}^k$ 上的 Borel σ-域］表示 $\mathbf{Y}_n(n\geqslant 1)$ 的 n 步转移概率，即

$$P^{(n)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y}) = P(Y_n \in \mathrm{d}\boldsymbol{y} \mid \mathbf{Y}_0 = \boldsymbol{x})$$

对 $P^{(1)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$，如果$\int_{\mathbf{R}^k} P^{(1)}(\boldsymbol{x},A)\pi(\mathrm{d}\boldsymbol{x}) = \pi(A)$，则关于$\{\mathbf{Y}_n \mid n\geqslant 0\}$ 在$(\mathbf{R}^k,\mathbf{B}^k)$ 上的一个概率测度 π 是不变的。

文献（Bhattacharya et al.，1995a）中定理 1 给出 Markov 过程$\{\mathbf{Y}_n \mid n\geqslant 0\}$是几何哈里斯遍历的，即在下列假设下存在唯一一个不变概率 π 和一个常数 $\rho(0<\rho<1)$，使得当 $n\to+\infty$时，有

$$\rho^{-n} \| P^{(n)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y}) - \pi(\mathrm{d}\boldsymbol{y}) \| \to 0, \quad x \in \mathbf{R}^k$$

（1）H 是有限的；

（2）η_n 的分布函数关于 Lebesgue 测度 λ_1 是绝对连续的，同时几乎处处有一个正的密度；

（3）$E\eta_n=0$ 且 $E|\eta_n|^3<+\infty$；

（4）存在常数 $c\geqslant 0$，$R>0$，$a_i>0(i=1,\cdots,k)$使得 $\sum_{i=1}^{k} a_i < 1$，且对于 $\boldsymbol{y} = (y_1, y_2,\cdots,y_k) \in \mathbf{R}^k$，下列不等式成立

$$| H(\boldsymbol{y}) | \leqslant \sum_{i=1}^{k} a_i \mid y_i \mid + c, \quad | \boldsymbol{y} | \geqslant R$$

其中，$\|\cdot\|$ 表示$(\mathbf{R}^k,\mathbf{B}^k)$上有限广义测度的 Banach 空间上的变范数。

如果 $\mathbf{Y}_n$ 有一个初始分布 π，则$\{\mathbf{Y}_n \mid n\geqslant 0\}$是一个平稳的遍历 Markov 过程，认为 X_n 与 $\mathbf{Y}_n$ 是同等的，即$\{X_n \mid n\geqslant 0\}$与$\{(\mathbf{Y}_n, X_{n+k}) \mid n\geqslant 0\}$一样都是平稳过程。假设在 $\mathbf{R}^1$ 上 η_n 具有密度函数 g，则

（1）$\mathbf{Y}_n$ 的 k 步转移概率 $P^{(k)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$有一个关于 Lebesgue 测度 λ_k 的密度函数

$$p^{(k)}(\boldsymbol{x},\boldsymbol{y}) = g(y_1 - H(\boldsymbol{x}))\prod_{j=2}^{k} g(y_j - H(x_j,\cdots,x_k,y_1,\cdots,y_{j-1}))$$

其中，$\boldsymbol{x}=(x_1,\cdots,x_k)$，$\boldsymbol{y}=(y_1,\cdots,y_k)\in\mathbf{R}^k$。由文献(Bhansali，1981)中引理 1 以及：

(2) 如果 H 和 g 是连续的，则 $p^{(k)}(\boldsymbol{x},\boldsymbol{y})$ 在 $\mathbf{R}^k\times\mathbf{R}^k$ 上是连续的。如果 g 是有界的，则 $p^{(k)}(\boldsymbol{x},\boldsymbol{y})$ 也是有界的，所以 π 关于 λ_k 是绝对连续的且 $\boldsymbol{Y}_0$ 的密度 $f(\boldsymbol{x})$ 存在。

假设有观测值 $\{X_0,X_1,\cdots,X_{n+k}\}$，即

$$\{(\boldsymbol{Y}_1,X_{k+1}),(\boldsymbol{Y}_2,X_{k+2}),\cdots,(\boldsymbol{Y}_n,X_{n+k})\}$$

其中，$\boldsymbol{Y}_i=(X_i,X_{i+1},\cdots,X_{i+k-1})$，$i=1,2,\cdots,n$，希望估计自回归函数

$$H(\boldsymbol{u}) = E(X_{n+k}\mid \boldsymbol{Y}_n=\boldsymbol{u}) = \frac{\int x\gamma(\boldsymbol{u},\boldsymbol{x})\mathrm{d}\boldsymbol{x}}{f(\boldsymbol{u})} = \frac{\int x\gamma(\boldsymbol{u},\boldsymbol{x})\mathrm{d}\boldsymbol{x}}{\int \gamma(\boldsymbol{u},\boldsymbol{x})\mathrm{d}\boldsymbol{x}}$$

其中，$\gamma(\boldsymbol{u},\boldsymbol{x})$ 是 $(\boldsymbol{Y}_n,X_{n+k})$ 的连接函数，本书将用核方法讨论 H 的非参数估计。

对于 $\mathbf{R}^{k+1}$ 上具有独立观测值 $\{(\boldsymbol{Y}_i,X_i)\mid i=1,2,\cdots,n\}$ 的非参数回归模型

$$X_i = m(\boldsymbol{Y}_i)+\varepsilon_i,\quad i=1,2,\cdots,n$$

其中，$m(\cdot)$ 是未知的回归函数，ε_i 是随机误差项，1964 年 Nadaraya 和 Watson 提出的非参数回归估计量为

$$\hat{m}_h(\boldsymbol{u}) = \frac{\sum_{i=1}^{n} X_i K\left(\frac{\boldsymbol{u}-\boldsymbol{Y}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\boldsymbol{u}-\boldsymbol{Y}_i}{h}\right)},\quad \boldsymbol{u}\in\mathbf{R}^k$$

如果分母不为零，则 K 是一个核函数。设 K 是一个定义在 $\mathbf{R}^k$ 上的概率密度函数(核函数)，满足下列条件：

(K1) $K(\boldsymbol{x})\leqslant M<+\infty$，$|\boldsymbol{x}|K(\boldsymbol{x})\to 0(|\boldsymbol{x}|\to+\infty)$，$\boldsymbol{x}\in\mathbf{R}^k$；

(K2) $K(-\boldsymbol{x}) = K(\boldsymbol{x})$，$\int |\boldsymbol{x}|^2K(\boldsymbol{x})\mathrm{d}\boldsymbol{x}<+\infty$。

设 $h=h_n>0$ 是依赖于 n 的一个正常数，为以后表述方便，特将下标省略，当 $n\to+\infty$ 时，$h\to 0$ 且 $nh^k\to+\infty$。h 是决定估计曲线光滑性的量，称为窗宽。

1989 年 Härdle 和 Nadaraya 已经导出 $m(\boldsymbol{u})=E(X_i\mid\boldsymbol{Y}_i=\boldsymbol{u})$ 的 N-W 核估计 $\hat{m}_h(\boldsymbol{u})$ 的有用结果。例如，$k=1$，$\boldsymbol{u}\in\mathbf{R}^1$ 时，$\hat{m}_h(\boldsymbol{u})$ 是 $m(\boldsymbol{u})$ 的一致估计，且对于 $\delta\in\left(\frac{1}{2},1\right]$，若 $E|X_i|^{2+\delta}<+\infty$，则

$$\sqrt{nh}(\hat{m}_h(\boldsymbol{u})-m(\boldsymbol{u}))\xrightarrow{L} N(0,\sigma^2(\boldsymbol{u}))$$

其中，$\sigma^2(u)=\frac{1}{f(u)}\mathrm{Var}(X_i \mid Y_i=u)\int K^2(z)\mathrm{d}z$。在假设($K1$)和假设($K2$)的前提下，若 $n^\delta h^{2+\delta}\to+\infty$，则 $h=o(n^{-\frac{1}{3}})$；同时文献(Härdle et al.，1991)中的定理 1 与引理 1 给出 $k\geqslant1,\boldsymbol{u}\in\mathbf{R}^k$ 时，有 $\sqrt{nh^k}(\hat{m}_h(\boldsymbol{u})-m(\boldsymbol{u}))\xrightarrow{L}N(b(\boldsymbol{u}),v(\boldsymbol{u}))$。其中，

$$b(\boldsymbol{u})=\left(\nabla^2 m(\boldsymbol{u})+2\nabla m(\boldsymbol{u})\cdot\frac{\nabla f(\boldsymbol{u})}{f(\boldsymbol{u})}\right)\int \mid \boldsymbol{z}\mid^2 K(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$$

$$v(\boldsymbol{u})=\frac{1}{f(\boldsymbol{u})}\mathrm{Var}(X_i \mid \boldsymbol{Y}_i=\boldsymbol{u})\int K^2(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$$

如果 $h=O(n^{-\frac{1}{4+k}})$及在条件($K2$)和假设 $\sup\limits_{\boldsymbol{u}}E[\varepsilon_i^2 \mid \boldsymbol{Y}_i=\boldsymbol{u}]<+\infty$ 的情况下，$m(\boldsymbol{u}),f(\boldsymbol{u})$和 $\mathrm{Var}(X_i \mid \boldsymbol{Y}_i=\boldsymbol{u})$都是连续可微的。

对于自回归函数 H，如果分母不为 0，则 N-W 非参数核估计可表示为

$$\hat{H}_n(\boldsymbol{u})=\frac{\sum\limits_{i=1}^{n}X_{k+i}K\left(\frac{\boldsymbol{u}-\boldsymbol{Y}_i}{h}\right)}{\sum\limits_{i=1}^{n}K\left(\frac{\boldsymbol{u}-\boldsymbol{Y}_i}{h}\right)},\quad \boldsymbol{u}\in\mathbf{R}^k$$

本书讨论自回归函数 $H(\boldsymbol{u})$的核估计 $\hat{H}_n(\boldsymbol{u})$的渐近性质，文献(Györfi et al.，1989)已经给出在一些混合条件下核估计一致收敛的结果，假设 H 和 f 是 s 阶可微的(对某一 $s\in\mathbf{N}$)，该文献中定理 3.4.1 指出

$$\sum_{n=1}^{+\infty}P(\sup_{u\in D}\mid \hat{H}_n(\boldsymbol{u})-H(\boldsymbol{u})\mid>\varepsilon)<+\infty$$

对某些 $\varepsilon>0$ 和一个 $\mathbf{R}^k$ 上的紧子集 D 是成立的。根据 Borel-Cantelli 引理，意味着 $\hat{H}_n(\boldsymbol{u})$依概率收敛到 $H(\boldsymbol{u})$以及对于 $\boldsymbol{u}\in D$，几乎必然是一致的。

作为近似抽样分布的强有力的非参数工具，自助过程已在文献(Efron，1979)中用过，用来进行参数回归的曲线估计，Bose(1988)，Freedman(1981，1984)，Hall(1992a)和 Härdle(1991)分别研究了非参数回归、密度估计和线性自回归。自助法在某种意义上比正态近似要好，例如，在观测值来自一个连续分布的前提下，自助法的覆盖误差要比使用中心极限定理得到的经典正态近似的误差小，同样，与经典过程相比较它还有其他的一些优点。

对于具有独立观测值的非参数回归模型 $X_i=m(\boldsymbol{Y}_i)+\varepsilon_i,i=1,\cdots,n$，Härdle 等(1991)综述了 N-W 核估计量 $\hat{m}_h(\boldsymbol{u})$，Härdle 等(1993)已经建议使用无序自助法的思想。在无序自助法中，每一个自助残差产生于两点分布，该分布具有均值为零，方差等于残差的平方，三阶矩等于残差的立方的特性。对于残差 $\hat{\varepsilon}_i=X_i-\hat{m}_h(\boldsymbol{Y}_i)$，它们定义一个具有两点分布 $\hat{G}_i$ 的新的随机变量 $\hat{\varepsilon}_i^*$，其中 $\hat{G}_i=\gamma\delta_a+(1-\gamma)\delta_b$，且算

出 $a=\frac{\hat{\varepsilon}_i(1-\sqrt{5})}{2}$，$b=\frac{\hat{\varepsilon}_i(1+\sqrt{5})}{2}$，$\gamma=\frac{5+\sqrt{5}}{10}$，使得$E\varepsilon_i^*=0$，$E\varepsilon_i^{*2}=\hat{\varepsilon}_i^2$，$E\varepsilon_i^{*3}=\hat{\varepsilon}_i^3$。然后抽样，定义一个新的样本观测值

$$X_i^* = \hat{m}_g(\mathbf{Y}_i)+\varepsilon_i^*$$

其中，$\hat{m}_g(\boldsymbol{u})$是一个具有窗宽为 g（g 比 h 趋于 0 的速度要慢）的 N-W 核估计量。令

$$\hat{m}_g(\boldsymbol{u}) = \frac{\sum_{i=1}^{n} X_i^* K\left(\frac{\boldsymbol{u}-\mathbf{Y}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\boldsymbol{u}-\mathbf{Y}_i}{h}\right)}$$

其主要结果是对于 $\boldsymbol{u}\in\mathbf{R}^k$，$x\in\mathbf{R}^1$ 以及 $h=O(h^{-\frac{1}{4+k}})$和 g 一致地有

$$\sup_x \mid P^*(\sqrt{nh^k}[\hat{m}_h^*(\boldsymbol{u})-\hat{m}_g(\boldsymbol{u})]<x)-P(\sqrt{nh^k}[\hat{m}_h(\boldsymbol{u})-m(\boldsymbol{u})]<x)\mid \to 0$$

其中，g 趋于 0 的速度比 h 慢，如取 g 满足 $n^{-\frac{1}{4+k}+\delta}\leqslant g\leqslant n^{-\delta}$，$\delta>0$。

近年来我国学者对于时间序列的研究取得了极其丰硕的成果，主要体现在基础理论研究的不断加强，应用领域的不断拓展，在应用中求创新求发展，在部分应用领域中已经跟上了国际步伐。

理论上的进展主要表现在两个方面：一是单位根理论；二是非线性模型理论。非线性模型理论的进展集中在几何遍历性问题和非线性过程的平稳性这两方面。我国学者在非线性时间序列分析方面取得了一系列高水平的成果。

汤家豪教授将有关非线性时间序列分析的研究与动力系统科学的模型相连接而备受赞赏。现在他着眼于非参数时间序列模型的发展，并与生态学家进行大量的合作研究。

姚琦伟教授基于信息量，首次提出了描述一般随机系统对初始条件敏感性的度量及估计方法。在高维模型领域，姚琦伟教授提出用复系数线性模型近似高维非线性回归函数的新方法，以此克服高维非参数回归中样本量短缺的困难问题。此方法在生物、经济、金融等应用中获得了成功。在时间序列模型的最大似然估计方法的研究中，他完整地建立了在金融风险管理中有直接应用的 ARCH 模型和 GARCH 模型为最大似然估计的极限理论。对于重尾部(heavy-tailed)分布模型，提出了基于 bootstrap 的新的估计方法以及稳健统计方法。他还首次建立了在空间域上空间 ARMA 过程的最大似然估计理论，这一工作同时也对 Hannan 在 1973 年给出的关于时间序列的最大似然估计理论首次给出了一个完整的时域上的证明。

安鸿志、朱力行、陈敏关于非线性自回归模型的平稳性、遍历性和高阶矩的成

果,获得了有这些性质的最弱条件。关于回归或自回归的非线性检验问题,具有重要的实际意义。他们首次给出了完全对立的假设检验方法,无论从原理和应用都表明此方法有明显优点。他们研究了条件方差为非常数的回归和自回归模型的平稳性、遍历性和检验方法。

时间序列分析研究的一个重要原动力源自于金融市场、信息网络以及电子商务等领域超容量数据的获得,在全球化竞争日益激烈的环境中,这些数据的可利用价值越来越大。对这些数据进行综合分析的迫切性促进了我国时间序列分析应用研究的发展。

尽管我国在时间序列研究领域取得了长足的进步,但是基础领域的研究状况仍不乐观,主要体现在整体研究水平不高,国际领先成果往往集中于个别院校甚至个别人,与国际研究趋势不符。特别是用非参数核估计方法对非线性自回归模型的理论研究鲜见于文献,本书试图在非线性自回归模型的研究中用非参数核估计方法对非线性自回归模型进行系统的研究,得出几个主要结论。

1.3 主要研究内容

本书主要讨论非线性自回归模型的非参数估计,主要研究内容包括以下四点。

(1) 讨论核估计量的中心极限定理,建立非线性自回归模型的核估计量的渐近正态性,给出估计量的逐点 MSE 和 MISE 的计算公式,根据这些内容可以得到核估计量的最优条件以及构造非线性自回归曲线的非参数置信区间。最后对一阶和二阶非线性自回归模型进行数值研究和模拟仿真计算,验证先前得到的基本结论。

(2) 利用文献(Härdle et al.,1989;Politis et al.,1994)中介绍的平稳自助法和无序自助法的基本思想,讨论利用非参数自助过程如何构造非线性自回归曲线的置信区域的问题。证明利用平稳自助法构造的非线性自回归模型的核估计量是有效的,同时将无序自助法应用到非线性自回归核估计量,也得出相应的结果。最后给出由平稳自助法和无序自助法导出的自助置信区间的算法设计,并在一阶和二阶非线性自回归模型的情况下进行数值仿真计算,将计算结果与利用正态近似求出的逐点置信区间进行比较。得出在相同的置信水平下,自助置信区间的精度要高于正态近似置信区间的精度。

(3)研究在非线性和线性自回归模型中阶的误设效应。讨论在非线性自回归模型中,非参数核估计和平稳自助估计在拟合阶大于和小于真实阶的情况下的效应以及拟合阶以比 $n\to+\infty$ 更慢的速度趋于无穷时的效应,也讨论在线性自回归模型中,最小二乘估计和自助估计在拟合阶比 $n\to+\infty$ 更慢的速度趋于无穷时的效应。特别得到,如果拟合阶趋于无穷的速度明确给定,则正态近似法和自助近似

法在非线性和线性两种情况下仍然成立。

(4) 研究欧式期权定价和美式期权定价的非参数和非线性方法。首先在非参数方法框架下研究股票价格不服从几何布朗运动下欧式期权价值的评估方法,并以上海证券市场数据进行实证分析,得出比经典 Black-Scholes 定价更接近真实情况的结果;其次利用最优停时理论和动态规划原则,得到美式期权价值的一个非线性偏微分方程,并证明该偏微分方程解的存在性和唯一性;最后给出一种基于最小二乘法的美式期权定价的最优停时的数值算法,并讨论该算法的收敛速度。

第 2 章　核估计的中心极限定理

对于具有独立观测值的非参数回归模型，Nadaraya(1989)和 Härdle 等(1991)导出了非线性回归曲线的核估计的渐近正态性。本章考虑在非独立的情况下，讨论非线性自回归函数 $H(\boldsymbol{u})$的核估计 $\hat{H}_n(\boldsymbol{u})$的渐近性质，证明模型(2.1.1)的核自回归估计量的渐近正态性，构造估计量的逐点 MSE 和 MISE，根据这些结果，得到核估计量和构造自回归曲线的非参数置信区域的最优条件，且将结果推广到 ARCH 模型，最后利用 SAS 软件和 Matlab 程序对非线性自回归函数在 $k=1$ 和 $k=2$ 的情况下的非参数核估计量进行数值模拟，导出利用正态近似法的逐点置信区间。

2.1　问题的提出及基本假设

设$\{X_n \mid n\geqslant 0\}$是由式(2.1.1)定义的一个非线性 k 阶($k\geqslant 1$)自回归过程：

$$X_{n+1} = H(X_{n+1-k},\cdots,X_n) + \eta_{n+1} \quad (n \geqslant k-1) \tag{2.1.1}$$

其中，H 是 $\mathbf{R}^k$ 上的一个实值 Borel 可测函数，$\{\eta_n \mid n\geqslant 1\}$是一列具有零均值和有限方差的独立同分布的实值随机变量，$\{X_0,X_1,\cdots,X_{k-1}\}$是任意一组独立于$\{\eta_n \mid n\geqslant 1\}$的实值随机变量。本章主要讨论非线性自回归函数 $H(\boldsymbol{u})=E\{X_{n+1} \mid (X_{n+1-k},\cdots,X_n)=\boldsymbol{u}\}$，$\boldsymbol{u}\in\mathbf{R}^k$ 的非参数估计及其性质。对于 $\mathbf{Y}_i=\{X_i,X_{i+1},\cdots,X_{i+k-1}\}$，如果式(2.1.2)分母不为零，则自回归函数 $H(\boldsymbol{u})$的核估计为

$$\hat{H}_h(\boldsymbol{u}) = \frac{\sum_{i=1}^{n} X_{k+i} K\left(\frac{\boldsymbol{u}-\mathbf{Y}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\boldsymbol{u}-\mathbf{Y}_i}{h}\right)}, \quad \boldsymbol{u} \in \mathbf{R}^k \tag{2.1.2}$$

其中，$h=h_n$ 是一个依赖于 n 的正数，满足当 $h\to 0$，$n\to+\infty$时，$nh^k\to+\infty$成立，同时 K 是一个核函数，h 是决定估计曲线光滑性的窗宽。在独立观测值的情形下，非线性回归模型的核估计的原始定义可参阅文献(Nadaraya，1964；Watson，1964)，对于具有独立观测值的非参数回归模型，文献(Härdle et al.，1991；Nadaraya，1989)导出了非线性回归曲线的核估计的渐近正态性。正如密度函数的核估计一样，核方法已经被用来估计非参数回归曲线，核回归可参阅文献(Härdle， 1990；Härdle et al.，1991；Roussas，1990)，核密度估计可参阅文献(Hall et al.，1995；Prakasa Rao，1983；Rosenblatt，1970；Roussas，1990；Silerman，1986)。

本章考虑在非独立的情况下，自回归函数 $H(\boldsymbol{u})$ 的核估计 $\hat{H}_n(\boldsymbol{u})$ 的渐近性质。在适当的假设下，Györfi 等(1989)[41] 中定理 3.4.1 已经证明对 $\varepsilon>0$ 及 $\mathbf{R}^k$ 上的一个紧子集 D，有

$$\sum_{n=1}^{\infty} P(\sup_{\boldsymbol{u}\in D} \mid \hat{H}_n(\boldsymbol{u}) - H(\boldsymbol{u}) \mid > \varepsilon) < +\infty \tag{2.1.3}$$

根据博雷尔-坎特利引理(Borel-Cantelli Lemma)，它表示任意 $\boldsymbol{u}\in D$，核估计 $\hat{H}_n(\boldsymbol{u})$ 几乎必然一致地收敛到 $H(\boldsymbol{u})$；在 2.2 节，证明模型(2.1.1)的核自回归估计量的渐近正态性，构造估计量的逐点 MSE 和 MISE，根据这些结果，得到核估计量和构造自回归曲线的非参数置信区域的最优条件；2.3 节将主要定理扩展到 ARCH 模型；2.4 节利用 SAS 和 Matlab 程序对自回归函数在 $k=1$ 和 $k=2$ 的情况下的非参数核估计量进行数值模拟，导出利用正态近似法的逐点置信区间。

为讨论方便起见，作以下假设：

(H) 函数 H 是具有一阶和二阶导数、有界的、二阶连续可微的，且存在常数 $c\geqslant 0, R>0, a_i>0(i=1,\cdots,k)$，使得当 $|\boldsymbol{y}|\geqslant R$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_k)\in\mathbf{R}^k$ 时，有 $\sum_{i=1}^{k} a_i < 1$ 及 $\mid H(\boldsymbol{y})\mid \leqslant \sum_{i=1}^{k} a_i \mid y_i \mid + c$。

(g) η_n 均值为零，$E|\eta_n|^3<+\infty$ 且 η_n 在 $\mathbf{R}^1$ 上具有一个正的二阶连续可微的密度函数 g，且 g, g', g'' 均有界。

这些条件保证了 Markov 过程 $\{\boldsymbol{Y}_n | n\geqslant 0, \boldsymbol{Y}_n=(X_n, X_{n+1},\cdots,X_{n+k+1})\}$ 是几何哈里斯遍历(geometrically Harris ergodic)(Bhattachya et al.，1995a)，即存在一个唯一不变概率 π 和一个常数 $\rho(0<\rho<1)$，使得当 $n\to+\infty$ 时，有

$$\rho^{-n} \parallel P^{(n)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y}) - \pi(\mathrm{d}\boldsymbol{y}) \parallel \to 0, \quad \boldsymbol{x}\in\mathbf{R}^k \tag{2.1.4}$$

其中，$P^{(n)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$ 是 $\boldsymbol{Y}_n$ 的 n 步转移概率，$\parallel\cdot\parallel$ 表示 $(\mathbf{R}^k,\mathbf{B}^k)$ 上有限带号测度的 Banach 空间上的全变差。如果 $\boldsymbol{Y}_n$ 有初始分布 π，则 $\{\boldsymbol{Y}_n | n\geqslant 0\}$ 是一个平稳遍历的 Markov 过程。把 X_n 看成是 $\boldsymbol{Y}_n$ 的第一个坐标，则 $\{X_n | n\geqslant 0\}$ 与 $\{(\boldsymbol{Y}_n, X_{n+k}) | n\geqslant 0\}$ 一样也是平稳过程。$\boldsymbol{Y}_n$ 的 k 步转移概率 $P^{(k)}(\boldsymbol{x},\mathrm{d}\boldsymbol{y})$ 具有关于 Lebesgue 测度 λ_k 的密度

$$p^{(k)}(\boldsymbol{x},\boldsymbol{y}) = g(y_1 - H(\boldsymbol{x}))\prod_{j=2}^{k} g(y_j - H(x_j,\cdots,x_k,y_1,\cdots,y_{j-1}))$$

其中，$\boldsymbol{x}=(x_1,\cdots,x_k)$，$\boldsymbol{y}=(y_1,\cdots,y_k)\in\mathbf{R}^k$[(Bhattacharya et al.，1995a)中的引理 1]，且 $p^{(k)}(\boldsymbol{x},\boldsymbol{y})$ 是 $\mathbf{R}^k\times\mathbf{R}^k$ 上连续有界的。所以关于 λ_k，π 是绝对连续的，且 $\boldsymbol{Y}_0$ 的

密度函数 $f(\boldsymbol{x})$存在，如果 $\widetilde{f}$ 是 π 的密度的逆，则 $f(\boldsymbol{x})=\int\widetilde{f}(\boldsymbol{x})p^{(k)}(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}=\widetilde{f}(\boldsymbol{x})a.e.$，在假设($H$)和假设($g$)的前提下，$f$ 是 $\mathbf{R}^k$ 上正的二阶连续可微的。

对于核密度函数 K，作如下假设

(K)：($K1$) $K(\boldsymbol{x})\leqslant M<+\infty$；

($K2$) $K(-\boldsymbol{x})=K(\boldsymbol{x})$；

($K3$) K 有一个紧支撑。

2.2 渐近正态性

本节讨论样本观测值在非独立的情况下，非线性自回归函数的核估计量 $\hat{H}_n(\boldsymbol{u})$的渐近正态性及其误差的计算公式。

2.2.1 本章重要结论

定理 2.2.1 若上述假设(H)、假设(g)和假设(K)都成立，则

(1) 对某常数 c，如果 $c>0,h=cn^{-\frac{1}{4+k}}$，则当 $n\to+\infty$时，有

$$\sqrt{nh^k}[\hat{H}_n(\boldsymbol{u})-H(\boldsymbol{u})]\xrightarrow{L}N(b(\boldsymbol{u}),\sigma^2(\boldsymbol{u}))$$

其中，

$$b(\boldsymbol{u})=\frac{c_1}{f(\boldsymbol{u})}\sum_{i,j=1}^{k}\left[D_iH(\boldsymbol{u})D_jf(\boldsymbol{u})+\frac{1}{2}f(\boldsymbol{u})D_{ij}H(\boldsymbol{u})\right]\int z_iz_jK(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \tag{2.2.1}$$

$$\sigma^2(\boldsymbol{u})=\frac{1}{f(\boldsymbol{u})}E\eta_1^2\int K^2(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \tag{2.2.2}$$

常数 $c_1=c^{\frac{4+k}{2}}>0$，其中，$D_i=\dfrac{\partial}{\partial u_i}$和 $D_{ij}=\dfrac{\partial^2}{\partial u_i\partial u_j}$，$i,j\in\{1,2,\cdots,k\}$

(2) 核估计量 $\hat{H}_n$ 的逐点 MSE(mean squared error)由下式给出

$$\mathrm{MSE}(\hat{H}_n(\boldsymbol{u}))=\frac{1}{nh^k}(\sigma^2(\boldsymbol{u})+b^2(\boldsymbol{u}))+o\left(\frac{1}{nh^k}\right)$$

(3) 核估计量 $\hat{H}_n$ 的 MISE(mean integrated squared error)为

$$\mathrm{MISE}(\hat{H}_n(\boldsymbol{u}))=\frac{1}{nh^k}(s_1+b_1)+o\left(\frac{1}{nh^k}\right)$$

其中，$s_1=\int\sigma^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$，$b_1=\int b^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$。