Analysis of chaotic time series with dynamic noise

笛田，薫

https://doi.org/10.11501/3180673

出版情報：九州大学, 2000, 博士（数理学）, 論文博士
バージョン：
権利関係：

Analysis of chaotic time series with dynamic noise

笛田　薫

Analysis of chaotic time series with dynamic noise

Kaoru Fueda
Faculty of Mathematics,
Kyushu University

2001

Abstract

In this thesis we investigate the estimation of the Lyapunov exponent for the nonlinear autoregressive time series model, especially the chaotic time series model with additive dynamic noise. For the deterministic model, which doesn't have a noise, the Lyapunov exponent has been proposed to quantify the sensitive dependence on an initial value. For nonlinear autoregressive time series models with additive noise, some modified Lyapunov-like indexes are proposed. However, they depend not only on the sensitive dependence on initial value, but also on the additive noise. We investigate in this thesis the estimator of the Lyapunov exponent which isn't influenced by the additive noise.

First we introduce delay time to the nonlinear autoregressive model considered in Cheng and Tong (1995). We find it important to take into account the delay time in the embedding dimension from the view point of curse of dimensionality. We develop a method of estimating the embedding dimension and delay time by using Nadaraya-Watson kernel estimator and Cross-Validation, and prove that the proposed estimator is consistent.

Next we consider a skeleton of the nonlinear autoregressive model with dynamic noise by deleting the dynamic noise term. By the Lyapunov exponent of the skeleton, we judge whether a randomness of the observed data is caused only by the dynamic noise or also by the nonlinearity of the autoregressive model. We propose an estimator of the Lyapunov exponent of the skeleton based on the observed data from the nonlinear autoregressive model with dynamic noise, when the embedding dimension is 1 and the skeleton has the Kolmogorov measure. And the consistency of the estimator is proved.

Acknowledgements

I am indebted to my adviser, Professor Takashi Yanagawa for various suggestion and guidance.

I would like to express my sincere gratitude to Professors Sadanori Konishi, Yoshihiko Maesono, Hiroto Hyakutake, Masayuki Uchida, Yuzo Maruyama and Gan Ohama for their encouragement and suggestions. The talk with the students of the Graduate school of Mathematics, Kyushu University gave me a lot of ideas. I also would like to express my deep gratitude to Professor Masanobu Taniguchi of Osaka University for the lecture on asymptotic theory for nonlinear stochastic models.

I wish to express my thanks to financial support by the Grant-in Aid for General Scientific Research from Japan Society for the Promotion of Science.

Finally, thanks to my parents and my wife's parents for raising up us, and special thanks to my wife, HsingYi, for her support and positive-thinking.

Contents

Chapter 1

Introduction

In analysis of data from nonlinear autoregressive time series with dynamic noise, it is a central issue whether randomness of the data is caused only by the dynamic noise or also by the nonlinearity of the autoregressive model. This thesis investigates the estimation of the Lyapunov exponent for the nonlinear autoregressive time series model to quantify the sensitive dependence on an initial value.

1.1 The embedding dimension and delay time

Cheng and Tong (1995) considered a nonlinear autoregressive model with additive dynamic noise

$$X_t = F(X_{t-1}, X_{t-2}, \cdots, X_{t-d}) + \varepsilon_t, \text{ (for } t \geq d). \tag{1.1}$$

Cheng and Tong (1995) also proposed to embed $(X_t, X_{t-1}, \cdots, X_{t-d})$ into $d+1$-dimensional Euclidean space, called d as *the embedding dimension*, and related the intuitive geometric reconstruction of phase space in theoretical physics with statistical theory of the determination of order of a nonlinear autoregressive model.

Although the delay time was not considered in Cheng and Tong (1995), we find it important to take into account the delay time in estimating the embedding dimension. For example, Yonemoto and Yanagawa (1998) show that, if the method of Cheng and Tong (1995) is applied to data generated by $X_t = F(X_{t-2}, X_{t-4}, \cdots, X_{t-2d}) + \varepsilon_t$, $t = 1, 2, \cdots$, the embedding dimension is estimated to be 2d, that is, we should embed $(X_t, X_{t-1}, X_{t-2}, \cdots, X_{t-(2d-1)}, X_{t-2d})$ into $2d+1$-dimensional Euclidean space. But we may represent the dynamics of $\{X_t\}$ by embedding $\{(X_t, X_{t-2}, \cdots, X_{t-2d}), t = 1, 2, \cdots\}$ in $d+1$ dimensional space, thus better to consider 2 as the delay time. This finding indicates that by also selecting the delay time we may embed the dynamics in a lower dimensional space, which is desirable from the view point of *curse of dimensionality*.

1.2 The Lyapunov exponent

Nonlinear dynamical systems which exhibit chaos are characterised by the phenomenon that a small perturbation in the initial condition can lead to a considerable divergence of the states of the system in the short term. In a deterministic dynamical system, which takes the form of a nonlinear autoregressive model without noise,

$$X_t = F(X_{t-1}, X_{t-2}, \cdots, X_{t-d}), \text{ (for } t \geq d), \tag{1.2}$$

this phenomenon has been very well documented and is usually analyzed by the well-known Lyapunov exponents (Eckmann and Ruelle (1985), Chatterjee and Yilmaz (1992), Berliner (1992)). However, for a stochastic, i.e. the dynamic noise is involved, it is well known that the estimates of the Lyapunov exponent by conventional methods is unreliable. Several methods have been developed to overcome the difficulty. Kostelich and York (1990) approximated F by polynomials and separated the signal from noise, and Pikovsky (1986), Landa and Rosenblum (1989), Cawley and Hsu (1992), and Sauer (1992) filtered out the noise by using linear filters. McCaferey et al.(1992) employed nonparametric estimation of F, but they assumed identical noises. Yao and Tong (1994a) explored alternative measures of detecting chaos in

observational data. In this thesis we estimate F, the empirical distribution of $\{X_t\}$ of the model (1.2), and the Lyapunov exponent of the model (1.2) using the observed data from model (1.1).

The plan of the rest of the paper is as follows. In Chapter 2, we give a brief sketch of local polynomial regression, which is used for estimating F and its derivative in Chapter 4.

Chapter 3 provides the estimation of the embedding dimension and delay time from chaotic time series with dynamic noise, on which the Lyapunov exponent depends. In Section 3.1 we introduce the delay time to (1.1), and explore the mathematical properties of the embedding dimension and delay time. In Section 3.2 a method of estimating the embedding dimension and delay time is proposed based on Cross-Validation, a similar technique as Cheng and Tong (1995). Consistency of the proposed estimators is proved in Section 3.3.

Chapter 4 provides the estimation of the Lyapunov exponent from chaotic time series with dynamic noise. In Section 4.1, we review the basics of chaos and the Lyapunov exponent. In Section 4.2, we define the class of the chaotic time series that we investigate. Finally in Section 4.3 we give a method of estimating the Lyapunov exponent and prove consistency of the proposed estimator.

1.3 Basic definitions and condition

In this section, we give basic definitions and condition, used throughout this thesis.

Definition 1.1 (Stationary)
The stochastic process $\{X_t; t \geq 0\}$ is said to be stationary if the random variables

$$X_{t_1}, X_{t_2}, \ldots, X_{t_m}$$

have the same joint probability distribution as the random variables

$$X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_m+h}$$

for any positive integer m, *any* $t_1, \ldots, t_m$ *and* h.

Definition 1.2 (Nonlinear autoregressive time series model)
The stochastic model $\{X_t; t \geq 0\}$ *is said to be a nonlinear autoregressive time series with dynamic noise if* $\{X_t\}$ *is stationary with* $EX_t^2 < \infty$ *and if for every integer* t $(t \geq d)$,

$$X_t = F(X_{t-1}, X_{t-2}, \cdots, X_{t-d}) + \varepsilon_t, \tag{1.3}$$

where d *is positive integer,* $F : R^d \to R$ *is a measurable function and* $\{\varepsilon_t\}$ *is a sequence of random noise and for any* t,

$$E\left[\varepsilon_t | \mathcal{A}_0^{t-1}(X)\right] = 0, \text{ almost surely.}$$

and

$$E\left[\varepsilon_t^2 | \mathcal{A}_0^{t-1}(X)\right] = \sigma^2, (\sigma > 0), \text{ almost surely,}$$

where $\mathcal{A}_s^t(X)$ *denotes the sigma algebra generated by* $(X_s, \ldots, X_t)$, *for* $s \leq t$. *Further, the integer* d *is called the degree of the nonlinear autoregressive time series.*

Definition 1.3 (Skeleton)
The deterministic system $\{X_t(x); t \geq 0\}$ *is said to be a skeleton of the nonlinear autoregressive time series with dynamic noise (1.3) if* $\{X_t(x); t \geq 0\}$ *is generated by*

$$X_t(x) = F(X_{t-1}(x), X_{t-2}(x), \cdots, X_{t-d}(x)), \; (\text{for } t \geq d) \tag{1.4}$$

where $x = {}^t(x_0, \ldots, x_{d-1}) \in R^d$ *is a fixed vector and* $X_t(x) = x_t$ *for* $t = 0, \ldots, d-1$.

Definition 1.4 (Chaos)
The deterministic system $\{X_t(x); t \geq 0\}$ *is said to be chaotic if* $\{X_t(x); t \geq 0\}$ *is bounded and there exists* $\delta > 0$ *such that for all* $x, \varepsilon \in R^d$, *there exists positive integer* n *such that* $|X_n(x) - X_n(x + \varepsilon)| > \delta$.

Definition 1.5 (Chaotic time series model)
The nonlinear autoregressive time series with dynamic noise is said to be chaotic time series if its skeleton is chaotic.

For the function $F : R^d \to R$ in (1.3), we define $\mathbf{F} : R^d \to R^d$ as

$$\mathbf{F}(x_1, x_2, \ldots, x_d) = \begin{pmatrix} F(x_1, x_2, \ldots, x_d) \\ x_1 \\ \vdots \\ x_{d-1} \end{pmatrix},$$

and put

$$\mathbf{X}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-(d-1)} \end{pmatrix}, \epsilon_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Then the model (1.3) implies

$$\mathbf{X}_t = \mathbf{F}(\mathbf{X}_{t-1}) + \epsilon_t, \; (\text{for } t \geq d), \tag{1.5}$$

and the model (1.4) implies

$$\mathbf{X}_t = \mathbf{F}(\mathbf{X}_{t-1}), \; (\text{for } t \geq d). \tag{1.6}$$

The model (1.6) is also said to be a skeleton of the model (1.5).

In this thesis, we assume the following condition.

Condition 1.1
Let the support of $\{\varepsilon_t\}$ *be* S. *We suppose that there exists a set* $M \subset R^d$ *such that* $\mathbf{X}_{d-1} \in M$ *and*

$$\mathbf{F}(x + e) \in M,$$

for all $x \in M$ *and* $e = {}^t(e_1, 0, \ldots, 0)$ *where* $e_1 \in S$.

Chapter 2

Local polynomial regression

In this chapter, refering to Wand and Jones (1995), Simonoff (1996) and Fan and Gijbels (1996), we review the local polynomial regression to estimate F in (1.3) and its derivative.

2.1 Kernel Estimation

First of all, we consider the density estimation problem. Let Y be a random variable that has probability density function $g(y)$ and let $G(y)$ be the distribution function of the random variable Y, and $\{Y_1, \ldots, Y_n\}$ represent a random sample of size n from the density g.

Consider the definition of $g(y)$:

$$g(y) \equiv \frac{d}{dy}G(y) \equiv \lim_{h \to 0} \frac{G(y+h) - G(y-h)}{2h}.$$

Replacing $G(y)$ with the empirical distribution function gives

$$\hat{g}(y) = \frac{\#\{Y_i \in (y-h, y+h]\}}{2nh}.$$

This can be rewritten as

$$\hat{g}(y) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{Y_i - y}{h}), \tag{2.1}$$

where

$$K(u) = \begin{cases} \frac{1}{2}, & \text{if } -1 < u \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

The form (2.1) is that of the kernel density estimator, with kernel function K. Note that this kernel function is a uniform density function on $(-1, 1]$.

The problem is that the additive form of (2.1) implies that the estimate $\hat{g}$ retains the continuity and differentiability properties of K. Since the uniform density is discontinuous, so is the kernel density estimate based on a uniform kernel function. A smoother kernel function will thus lead to a smoother kernel density estimate.

In this thesis, we assume that the kernel function $K(u)$ is an arbitrary density function satisfying the conditions:

1. $\sup_{-\infty<u<\infty} K(u) < \infty$,
2. $\lim_{|u|\to\infty} |u|K(u) = 0$,
3. $K(u) = K(-u)$ for all $u \in R$,
4. $\int u^2 K(u)du = \sigma_K^2 < \infty$.

The bias and variance of the kernel density estimator are given as follows.

Theorem 2.1 (Parzen (1962))
Assume that $g''(y)$ is absolutely continuous and square integrable. Then we have

$$Bias[\hat{g}(y)] = \frac{h^2\sigma_K^2 g''(y)}{2} + O(h^4)$$

and

$$Var[\hat{g}(y)] = \frac{g(y)R(K)}{nh} + O(n^{-1}),$$

where $R(K) = \int K(u)^2 du$.

The degree to which the data are smoothed has a strong effect on the appearance of $\hat{g}(y)$ through the setting of the bandwidth h. Theorem 2.1 shows the tradeoff of bias versus variance.

Remark 2.1 *Combining variance and squared bias, we have the mean squared error*

$$MSE[\hat{g}(y)] = \frac{g(y)R(K)}{nh} + \frac{h^4\sigma_K^4(g''(y))^2}{4} + O(n^{-1}) + O(h^6).$$

Integrating over the entire line then we have the asymptotic MISE

$$AMISE = \frac{R(K)}{nh} + \frac{h^4\sigma_K^4 R(g'')}{4},$$

where $R(g'') = \int g''(u)^2 du$. The asymptotically optimal bandwidth satisfies

$$h_0 = \left(\frac{R(K)}{n\sigma_K^4 R(g'')}\right)^{1/5},$$

implying minimal AMISE

$$AMISE_0 = \frac{5}{4}(\sigma_K R(K))^{4/5} R(g'')^{1/5} n^{-4/5}.$$

The term $R(g'')$ measures the roughness of the true underlying density. In general, rougher densities are more difficult to estimate and require a smaller bandwidth.

2.2 Kernel Regression

Next we consider the nonparametric regression problem. Let (Y, Z) be a random vector that has joint density function $g(y, z)$, and $\{(Y_1, Z_1), \ldots, (Y_n, Z_n)\}$ represent a random sample of size n from the density g. We consider the nonparametric regression model

$$Z_i = m(Y_i) + \varepsilon_i,$$

where the regression curve $m(y)$ is the conditional expectation $m(y) = E(Z|Y = y)$ with $E(\varepsilon|Y = y) = 0$, and $Var(\varepsilon|Y = y) = \sigma^2(y)$ not necessarily constant.

By definition we have

$$\begin{aligned} m(y) &= E[Z|Y=y] \\ &= \int zg(z|y)dz \\ &= \int z\frac{g(y,z)}{g_Y(y)}dz, \qquad (2.2) \end{aligned}$$

where $g_Y(y)$ and $g(z|y)$ are the marginal density of Y and the conditional density of Z given Y, respectively. A product kernel estimate of $g(y,z)$ is

$$\hat{g}(y,z) = \frac{1}{nh_yh_z}\sum_{i=1}^{n} K_y\left(\frac{Y_i-y}{h_y}\right)K_z\left(\frac{Z_i-z}{h_z}\right),$$

while a kernel estimation of $g_Y(y)$ is

$$\hat{g}_Y(y) = \frac{1}{nh_y}\sum_{i=1}^{n} K_y\left(\frac{Y_i-y}{h_y}\right).$$

Substituting into (2.2), and noting that $\int K_z(u) = 1$ and $\int uK_z(u)du = 0$, yields the Nadayara-Watson kernel estimator,

$$\hat{m}_{NW}(y) = \frac{\sum_{i=1}^{n} Z_iK_y\left(\frac{Y_i-y}{h_y}\right)}{\sum_{i=1}^{n} K_y\left(\frac{Y_i-y}{h_y}\right)}.$$

The Nadaraya-Watson kernel estimator is most natural for data using a random design. If the design is not random, but is rather a fixed set of ordered nonrandom numbers $y_1,\ldots,y_n$, a different form of kernel estimator is considered. Gasser and Müller (1979) proposed the Gasser-Müller kernel estimator,

$$\hat{m}_{GM}(y) = \frac{1}{h}\sum_{i=1}^{n} Z_i\int_{s_{i-1}}^{s_i} K\left(\frac{u-y}{h}\right)du,$$

where $y_{i-1} < s_{i-1} < y_i$. Fan(1992) summarized the asymptotic bias and variance of these estimators as follows:

$$\begin{aligned} Bias[\hat{m}_{NW}(y)] &= \left(\frac{1}{2}m''(y)+\frac{m'(y)g_Y'(y)}{g_Y(y)}\right)h^2\int u^2K(u)du \\ Var[\hat{m}_{NW}(y)] &= \frac{\sigma^2(y)}{g_Y(y)nh}\int K^2(u)du \\ Bias[\hat{m}_{GM}(y)] &= \frac{1}{2}m''(y)h^2\int u^2K(u)du \\ Var[\hat{m}_{GM}(y)] &= \frac{3\sigma^2(y)}{2g_Y(y)nh}\int K^2(u)du \end{aligned}$$

As Fan(1992) showed, $Bias[\hat{m}_{NW}(y)] > Bias[\hat{m}_{GM}(y)]$ and $Var[\hat{m}_{NW}(y)] < Var[\hat{m}_{GM}(y)]$. Fan(1992) also showed that the bias of the local linear regression estimator, which was proposed by Stone (1977), is equal to the bias of the Gasser-Müller estimator and the variance of the local linear regression estimator is equal to the variance of the Nadaraya-Watson estimator. Fan, Hu and Troung(1994) considered a class of kernel estimators based on local linear regression estimator, and showed the asymptotic normality of these estimators. Cleveland (1979) proposed the local polynomial regression estimator, which is the extension of the local linear regression estimator.

2.3 Local polynomial regression

In this section, we review the local polynomial regression estimator. Let (Y,Z) be a random vector that has joint density function $g(y,z)$, and $\{(Y_1,Z_1),\ldots,(Y_n,Z_n)\}$ represent a random sample of size n from the density g. We are interested in to estimate the regression function $m(y_0) = E(Z|Y=y_0)$ and its derivatives $m'(y_0), m''(y_0),\ldots,m^{(p)}(y_0)$, where $m^{(j)}$ represents the j-th derivative of m.

Suppose that the $(p+1)$-th derivative of $m(y)$ at the point y_0 exists. We approximate the unknown regression function $m(y)$ locally by a polynomial of order p. A Taylor expansion gives, for y in a neighborhood of y_0,

$$m(y) \approx m(y_0) + m'(y_0)(y-y_0) + \frac{m''(y_0)}{2!}(y-y_0)^2 + \cdots + \frac{m(p)(y_0)}{p!}(y-y_0)^p. \qquad (2.3)$$

Cleveland (1979) considered the following weighted least square problem:

minimize

$$\sum_{i=1}^{n}\{Z_i - \sum_{j=0}^{p}\beta_j(Y_i - y_0)^j\}^2 K_h(Y_i - y_0), \tag{2.4}$$

with respect to $\beta_0, \ldots, \beta_p$, where h is a bandwidth controlling the size of the local neighborhood, and $K_h(y) = \frac{1}{h}K(\frac{y}{h})$ with K a kernel function assigning weights to each datum point. Denote the minimizer by $\hat{\beta}_0, \ldots, \hat{\beta}_p$. Note that if $p = 0$, then $\hat{\beta}_0$ coincides with the Nadaraya-Watson estimator of $m(y_0)$.

Compare (2.4) with (2.3), an estimator for $m^{(\nu)}(y_0)$ is given by $\hat{m}_\nu(y_0) = \nu!\hat{\beta}_\nu$. To estimate the entire function $m^{(\nu)}(y)$, we denote by $\mathbf{Y}$ the design matrix of problem (2.4):

$$\mathbf{Y} = \begin{pmatrix} 1 & (Y_1 - y_0) & \cdots & (Y_1 - y_0)^p \\ 1 & (Y_2 - y_0) & \cdots & (Y_2 - y_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (Y_n - y_0) & \cdots & (Y_n - y_0)^p \end{pmatrix},$$

and put

$$\mathbf{z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}, \text{ and } \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}.$$

Further, let $\mathbf{W}$ be the $n \times n$ diagonal matrix of weights:

$$\mathbf{W} = \begin{pmatrix} K_h(Y_1 - y_0) & 0 & \cdots & 0 \\ 0 & K_h(Y_2 - y_0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K_h(Y_n - y_0) \end{pmatrix}.$$

Then the weighted least squares problem (2.4) can be written as: minimize

$$^t(\mathbf{z} - \mathbf{Y}\beta)\mathbf{W}(\mathbf{z} - \mathbf{Y}\beta),$$

with respect to β, where $\beta =^t (\beta_0, \beta_1, \ldots, \beta_p)$. The solution vector is provided by weighted least squares theory and is given by

$$\hat{\beta} = (^t\mathbf{YWY})^{-1\,t}\mathbf{YWz}. \tag{2.5}$$

The conditional bias and variance of the estimator $\hat{\beta}$ are derived from its definition (2.5):

$$\begin{aligned} E(\hat{\beta}|\mathrm{Y}) &= (^t\mathbf{YWY})^{-1\,t}\mathbf{YWm} \\ &= \beta + (^t\mathbf{YWY})^{-1\,t}\mathbf{YWr} \\ Var(\hat{\beta}|\mathrm{Y}) &= (^t\mathbf{YWY})^{-1}(^t\mathbf{Y}\textstyle\sum\mathbf{Y})(^t\mathbf{YWY})^{-1}, \end{aligned}$$

where

$$\mathbf{m} = \begin{pmatrix} m(Y_1) \\ m(Y_2) \\ \vdots \\ m(Y_n) \end{pmatrix}, \beta = \begin{pmatrix} m(y_0) \\ \dfrac{m'(y_0)}{1!} \\ \vdots \\ \dfrac{m^{(p)}(y_0)}{p!} \end{pmatrix}$$

and $\mathbf{r} = \mathbf{m} - \mathbf{Y}\beta$, the vector of residuals of the local polynomial approximation, and

$$\sum = \begin{pmatrix} K_h^2(Y_1 - y_0)\sigma^2(Y_1) & 0 & \cdots & 0 \\ 0 & K_h^2(Y_2 - y_0)\sigma^2(Y_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K_h^2(Y_n - y_0)\sigma^2(Y_n) \end{pmatrix}.$$

Since the residual $\mathbf{r}$ and the diagonal matrix $\sum$ is unknown, there is a need for approximating bias and variance. Ruppert and Wand (1994) obtained the result in the following theorem. Denote the moments of K and K^2 respectively by

$$\mu_j = \int u^j K(u)du$$

and

$$\nu_j = \int u^j K^2(u)du.$$

Some matrices and vectors of moment appear in the asymptotic expressions. Let

$$\begin{aligned} S &= (\mu_{j+l})_{0\le j,l\le p}, \quad c_p =^t (\mu_{p+1}, \ldots, \mu_{2p+1}), \\ \tilde{S} &= (\mu_{j+l+1})_{0\le j,l\le p}, \quad \tilde{c}_p =^t (\mu_{p+2}, \ldots, \mu_{2p+2}), \\ S^* &= (\nu_{j+l})_{0\le j,l\le p}. \end{aligned}$$

Further, we consider the unit vector $e_{\nu+1} =^t (0, \ldots, 0, 1, 0, \ldots, 0) \in R^{p+1}$, with 1 on the $(\nu+1)$-th position for $\nu = 0, 1, \ldots, p$.

Theorem 2.2 (Ruppert and Wand (1994))
Assume that $g_Y(y_0) > 0$ and that $g_Y(y), m^{(p+1)}(y)$ and $\sigma^2(y)$ are continuous in a neighborhood of y_0. Further, assume that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Then the asymptotic conditional variance of $\hat{m}_\nu(y_0)$ is given by

$$Var(\hat{m}_\nu(y_0)|\mathbf{Y}) =^t e_{\nu+1} S^{-1} S^* S^{-1} e_{\nu+1} \frac{\nu!^2 \sigma^2(y_0)}{g_Y(y_0) n h^{1+2\nu}} + o_p(\frac{1}{nh^{1+2\nu}}).$$

The asymptotic conditional bias for $p - \nu$ odd is given by

$$Bias(\hat{m}_\nu(y_0)|\mathbf{Y}) =^t e_{\nu+1} S^{-1} c_p \frac{\nu!}{(p+1)!} m^{(p+1)}(y_0) h^{p+1-\nu} + o_p(h^{p+1-\nu}).$$

Further, for $p - \nu$ even the asymptotic conditional bias is given by

$$Bias(\hat{m}_\nu(y_0)|\mathbf{Y})$$
$$=^t e_{\nu+1} S^{-1} \tilde{c}_p \frac{\nu!}{(p+2)!} \left(m^{(p+2)}(y_0) + (p+2) m^{(p+1)}(y_0) \frac{g'_Y(y_0)}{g_Y(y_0)} \right) h^{p+2-\nu} + o_p(h^{p+2-\nu}),$$

provided that $g'_Y(y)$ and $m^{(p+2)}(y)$ are continuous in a neighborhood of y_0 and $nh^3 \to \infty$.

This theorem shows that the degree of the polynomial being fit determines the order of the bias of $\hat{m}_p$, with polynomials of adjacent pairs of degree being conceptually similar. For estimating the $m(y_0)$ (i.e. $\nu = 0$), if $p = 0$, which coincides with the Nadaraya-Watson estimator, or $p = 1$, which coincides with the local linear fit considered in Fan, Hu and Troung (1994), then estimation yields $O_p(h^2)$ bias, and if $p = 2, 3$ then estimation yields $O_p(h^4)$ bias.

2.4 Local polynomial regression for time series

In this section, we study the local polynomial estimator when the sample is not independent. First of all, we define the following mixing conditions.

Let $\{(X_j, Y_j)\}$ be a stationary sequence of random vectors, and $\mathcal{F}_i^k$ be the $\sigma-$algebra of events generated by the random variables $\{(X_j, Y_j), i \leq j \leq k\}$. Denote by $L_2(\mathcal{F}_i^k)$ the collection of all random variables which are $\mathcal{F}_i^k-$measurable and have finite second moment.

Definition 2.1 (Strongly mixing)
The stationary process $\{(X_j, Y_j)\}$ is called strongly mixing if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(A \cap B) - P(A)P(B)| = \alpha(k) \to 0 \text{ as } k \to \infty.$$

Definition 2.2 (Uniformly mixing)
The stationary process $\{(X_j, Y_j)\}$ is called uniformly mixing if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(B|A) - P(B)| = \phi(k) \to 0 \text{ as } k \to \infty.$$

Definition 2.3 (ρ-mixing)
The stationary process $\{(X_j, Y_j)\}$ is called ρ-mixing if

$$\sup_{U \in L_2(\mathcal{F}_{-\infty}^0), V \in L_2(\mathcal{F}_k^\infty)} |Corr(U, V)| = \rho(k) \to 0 \text{ as } k \to \infty,$$

where $Corr(U, V)$ denotes the correlation coefficient between the random variables U and V

The key usage of mixing conditions is contained in the following lemma. The lemma shows that dependent random variables can be approximated by a sequence of independent random variables having the same marginal distribution.

Lemma 2.1 (Volkonskii and Rozanov (1959))
Let $V_1, \ldots, V_n$ be random variables with $|V_j| \leq 1$ for $j = 1, \ldots, n$, and $\mathcal{F}_{i_1}^{j_1}, \ldots, \mathcal{F}_{i_n}^{j_n}$ be the σ-algebra of events generated by the random variables $V_1, \ldots, V_n$ respectively. Suppose that $i_1 < j_1 < \cdots < i_n < j_n$ and there exists $w \geq 1$ such that $i_{k+1} - j_k \geq w$, for $k = 1, \ldots, n-1$. Then

$$\left| E \prod_{j=1}^n V_j - \prod_{j=1}^n E(V_j) \right| \leq 16(n-1)\alpha(w).$$

Now we consider observations $\{X_1, \ldots, X_{n+1}\}$ from the non-linear autoregressive model $X_t = m(X_{t-1}) + \varepsilon_t$, and construct data $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ as $Y_i = X_{i+1}$ for $i = 1, \ldots, n$. We are interested in to estimate

$$m(x) = E(Y_i|X_i = x) \text{ and its derivative } m^{(\nu)}(x).$$

Masry and Fan (1993) approximated $m(x)$ as in (2.3) and fits locally a polynomial as in (2.4). Denote $\hat{\beta}(x)$ the solution to the weighted least squares problem (2.4). Then, an estimator for $m^{(\nu)}(x)$ is $\hat{m}_\nu(x) = \nu!\hat{\beta}_\nu(x)$. Masry and Fan (1993) state that under certain mixing conditions, local polynomial estimators for dependent data have the same asymptotic behavior as for independent data.

Let $f(x)$ be the density of X_1 and $\sigma^2(x) = Var(Y_1|X_1 = x)$. Let S, S^* and c_p denote the same moment matrices and vector as those introduced in previous section, and let

$$\mu_\nu^* = \int {}^t e_{\nu+1} S^{-1}{}^t(1, u, \ldots, u^p) K(u) u^{p+1} du$$

and

$$\xi_\nu^* = \int \left({}^t e_{\nu+1} S^{-1}{}^t(1, u, \ldots, u^p) K(u)\right)^2 du.$$

Masry and Fan (1993) gave the following result.

Condition 2.1

1. *The kernel K is bounded with bounded support.*
2. *For all $l \in N$, $f_{X_0, X_l|Y_0, Y_l}(x_0, x_l|y_0, y_l)$ is bounded, where $f_{X_0, X_l|Y_0, Y_l}(x_0, x_l|y_0, y_l)$ is a conditional density of (X_0, X_l) given (Y_0, Y_l).*
3. *The stationary process $\{(X_j, Y_j)\}$ is strongly mixing.*
4. *For some $\delta > 2$ and $a > 1 - 2/\delta$,*
$$\sum_l l^a (\alpha(l))^{1-2/\delta} < \infty, E|Y_1|^\delta < \infty, f_{X_1|Y_1}(x|y) \text{ is bounded.}$$

5. *There exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o\left(\sqrt{nh}\right)$ such that*
$$\sqrt{\frac{n}{h}}\alpha(s_n) \to 0, \text{ as } n \to \infty.$$

Condition 2.2

1. *The kernel K is bounded with bounded support.*
2. *For all $l \in N$, $f_{X_0, X_l|Y_0, Y_l}(x_0, x_l|y_0, y_l)$ is bounded, where $f_{X_0, X_l|Y_0, Y_l}(x_0, x_l|y_0, y_l)$ is a conditional density of (X_0, X_l) given (Y_0, Y_l).*
3. *The stationary process $\{(X_j, Y_j)\}$ is ρ-mixing.*
4. $$\sum_l \rho(l) < \infty, EY_1^2 < \infty$$
5. *There exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o\left(\sqrt{nh}\right)$ such that*
$$\sqrt{\frac{n}{h}}\rho(s_n) \to 0, \text{ as } n \to \infty.$$

Theorem 2.3
Under Condition 2.1 or Condition 2.2, if $h = O(n^{1/(2p+3)})$, then the estimator $\hat{m}_\nu(x)$ based on the local polynomial fitting is asymptotically normal:as $n \to \infty$,

$$\sqrt{nh^{2\nu+1}}\left(\hat{m}_\nu(x) - m^{(\nu)}(x) - \mu_\nu^* \frac{\nu! m^{(p+1)}(x) h^{p+1-\nu}}{(p+1)!}\right) \xrightarrow{D} N\left(0, \xi_\nu^* \frac{(\nu!)^2 \sigma^2(x)}{f(x)}\right).$$

Chapter 3

The embedding dimension and delay time

3.1 The embedding dimension and the delay time

We consider the stochastic model given by

$$X_t = F(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) + \varepsilon_t \tag{3.1}$$

where d and τ are positive integers and ε_t is the dynamic noise. We assume that $\{X_t\}$ is a discrete-time strictly stationary time series with $EX_t^2 < \infty$ and for any t,

$$E\left[\varepsilon_t | \mathcal{A}_1^{t-1}(X)\right] = 0, \text{ almost surely.} \tag{3.2}$$

and

$$E\left[\varepsilon_t^2 | \mathcal{A}_1^{t-1}(X)\right] = \sigma^2, (\sigma > 0), \text{ almost surely,}$$

where $\mathcal{A}_s^t(X)$ denotes the sigma algebra generated by $(X_s, \dots, X_t)$, for $s \leq t$. Note that from (3.1) and (3.2), it follows that

$$F(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) = E[X_t | X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}]$$

with $E[\varepsilon_t | X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}] = 0$.

For simplicity we put

$$F_d(X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}) = E[X_t | X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}].$$

The embedding dimension and the delay time are defined as follows.

Definition 3.1 *The time series $\{X_t\}$ is said to have the embedding dimension d_0 with the delay time τ_0 if and only if there exist non-negative integers $d_0 < \infty$ and $\tau_0 < \infty$ such that*

$$F_d(X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}) \neq F_{d_0}(X_{t-\tau_0}, X_{t-2\tau_0}, \ldots, X_{t-d_0\tau_0}) \ a.e. \tag{3.3}$$

for any $d < d_0$, and any $\tau > 0$, and

$$F_d(X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}) = F_{d_0}(X_{t-\tau_0}, X_{t-2\tau_0}, \ldots, X_{t-d_0\tau_0}) \ a.e. \tag{3.4}$$

for any $(d, \tau) \in \mathrm{B}(d_0, \tau_0)$,
where $\mathrm{B}(d_0, \tau_0) = \{(d, \tau) | \{\tau_0, 2\tau_0, \ldots, d_0\tau_0\} \subset \{\tau, 2\tau, \ldots, d\tau\}\ \}$.

The definition is identical to that given in Cheng and Tong(1995) when $\tau = 1$.

We have the following theorem.

Theorem 3.1 *Suppose that for any $\tau > 0$ there exists $d_0(\tau) < \infty$ such that*

$$F_d(X_{t-\tau}, \ldots, X_{t-d\tau}) \neq F_{d_0(\tau)}(X_{t-\tau}, \ldots, X_{t-d_0(\tau)\tau}) \ a.e. \tag{3.5}$$

for any $d < d_0(\tau)$, and

$$F_d(X_{t-\tau}, \ldots, X_{t-d\tau}) = F_{d_0(\tau)}(X_{t-\tau}, \ldots, X_{t-d_0(\tau)\tau}) \ a.e. \tag{3.6}$$

for any $d \geq d_0(\tau)$. Then the embedding dimension d_0 and the delay time τ_0 of $\{X_t\}$ satisfy

$$d_0 = \min_\tau d_0(\tau) = d_0(\tau_0).$$

Proof. It is clear that $\min_\tau d_0(\tau) \leq d_0(\tau_0)$, so we show that

$$i) \quad d_0 \leq \min_\tau d_0(\tau)$$

and

$$ii) \quad d_0 \geq d_0(\tau_0).$$

$i)$. If $d_0 > \min_\tau d_0(\tau)$, then there exist τ^* such that $d_0 > d_0(\tau^*)$. Thus we have from (3.6)

$$F_{d_0}(X_{t-\tau^*}, \ldots, X_{t-d_0\tau^*}) = F_{d_0(\tau^*)}(X_{t-\tau^*}, \ldots, X_{t-d_0(\tau^*)\tau^*}) \ a.e.,$$

but this contradicts (3.3).

$ii)$. If $d_0 < d_0(\tau_0)$, we have from (3.5)

$$F_{d_0}(X_{t-\tau_0}, \ldots, X_{t-d_0\tau_0}) \neq F_{d_0(\tau_0)}(X_{t-\tau_0}, \ldots, X_{t-d_0(\tau_0)\tau_0}) \ a.e.,$$

but since $d_0 < d_0(\tau_0)$ and $(d_0(\tau_0), \tau_0) \in \mathrm{B}(d_0, \tau_0)$, this contradicts (3.4).

Denoting the residuals and their variances by

$$\varepsilon_t^{(d,\tau)} = \begin{cases} X_t & (d = 0) \\ X_t - F_d(X_{t-\tau}, \ldots, X_{t-d\tau}) & (d > 0), \end{cases}$$
$$\sigma^2(d, \tau) = E\left[\varepsilon_t^{(d,\tau)}\right]^2.$$

We may show the following lemma.

Lemma 3.1
i) For any positive integers d_1, d_2, τ_1, τ_2 such that $(d_1, \tau_1) \in \mathrm{B}(d_2, \tau_2)$,

$$\sigma^2(d_2, \tau_2) - \sigma^2(d_1, \tau_1) = E\left[F_{d_2}(X_{t-\tau_2}, \ldots, X_{t-d_2\tau_2}) - F_{d_1}(X_{t-\tau_1}, \ldots, X_{t-d_1\tau_1})\right]^2$$

ii) For any $d > 0$ and $\tau > 0$ such that $(d, \tau) \in \mathrm{B}(d_0, \tau_0)$,

$$\sigma^2(d_0, \tau_0) - \sigma^2(d, \tau) = 0.$$

Proof.

i). For simplicity, let $Z_t^{(d,\tau)} = (X_{t-\tau}, \dots, X_{t-d\tau})$.

$$
\begin{aligned}
&E\left[F_{d_2}(X_{t-\tau_2}, \dots, X_{t-d_2\tau_2}) - F_{d_1}(X_{t-\tau_1}, \dots, X_{t-d_1\tau_1})\right]^2 \\
&= E\left[\left\{X_t - F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right)\right\} - \left\{X_t - F_{d_2}\left(Z_t^{(d_2,\tau_2)}\right)\right\}\right]^2 \\
&= \sigma^2(d_1,\tau_1) + \sigma^2(d_2,\tau_2) \\
&-2E\left[\left\{X_t - F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right)\right\}\left\{X_t - F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right) + F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right) - F_{d_2}\left(Z_t^{(d_2,\tau_2)}\right)\right\}\right] \\
&= \sigma^2(d_1,\tau_1) + \sigma^2(d_2,\tau_2) \\
&\quad -2\left(\sigma^2(d_1,\tau_1) + E\left[\left\{X_t - F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right)\right\}\left\{F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right) - F_{d_2}\left(Z_t^{(d_2,\tau_2)}\right)\right\}\right]\right) \\
&= \sigma^2(d_2,\tau_2) - \sigma^2(d_1,\tau_1) \\
&\quad -2E\left[\left\{F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right) - F_{d_2}\left(Z_t^{(d_2,\tau_2)}\right)\right\} E\left[X_t - F_{d_1}\left(Z_t^{(d_1,\tau_1)}\right) \middle| Z_t^{(d_1,\tau_1)}\right]\right] \\
&= \sigma^2(d_2,\tau_2) - \sigma^2(d_1,\tau_1)
\end{aligned}
$$

ii). From the definition of d_0 and τ_0, $(d,\tau) \in \mathbf{B}(d_0,\tau_0)$ implies

$$F_d(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) = F_{d_0}(X_{t-\tau_0}, X_{t-2\tau_0}, \dots, X_{t-d_0\tau_0}) \ a.e.$$

and from Lemma 1 *i*) we have

$$
\begin{aligned}
&\sigma^2(d_0,\tau_0) - \sigma^2(d,\tau) \\
&= E\left[F_d(X_{t-\tau}, \dots, X_{t-d\tau}) - F_{d_0}(X_{t-\tau_0}, \dots, X_{t-d_0\tau_0})\right]^2 \\
&= 0
\end{aligned}
$$

From Lemma 1 we have the following theorem.

Theorem 3.2 *For any $\tau > 0$ and $d_0(\tau)$ defined in Theorem 3.1,*

i) $\sigma^2(d,\tau) > \sigma^2(d_0(\tau),\tau)$ *for any* $d < d_0(\tau)$,

ii) $\sigma^2(d,\tau) = \sigma^2(d_0(\tau),\tau)$ *for any* $d \geq d_0(\tau)$,

iii) $\sigma^2(d_0,\tau_0) \leq \sigma^2(d,\tau)$ *for any* $d > 0$ *and* $\tau > 0$.

Proof. *i*). From the definition of $d_0(\tau)$, for $d < d_0(\tau)$, we have

$$F_d(X_{t-\tau}, \dots, X_{t-d\tau}) \neq F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}) \ a.e.,$$

and $d < d_0(\tau)$ implies $(d_0(\tau),\tau) \in \mathbf{B}(d,\tau)$. Thus from Lemma 1 *i*),

$$
\begin{aligned}
&\sigma^2(d,\tau) - \sigma^2(d_0(\tau),\tau) \\
&= E\left[F_d(X_{t-\tau}, \dots, X_{t-d\tau}) - F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau})\right]^2 \\
&> 0.
\end{aligned}
$$

ii). From the definition of $d_0(\tau)$, for $d \geq d_0(\tau)$, we have

$$F_d(X_{t-\tau}, \dots, X_{t-d\tau}) = F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}) \ a.e.,$$

and $d \geq d_0(\tau)$ implies $(d,\tau) \in \mathbf{B}(d_0(\tau),\tau)$. Thus from Lemma 1 *i*),

$$
\begin{aligned}
&\sigma^2(d,\tau) - \sigma^2(d_0(\tau),\tau) \\
&= -E\left[F_d(X_{t-\tau}, \dots, X_{t-d\tau}) - F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau})\right]^2 \\
&= 0.
\end{aligned}
$$

iii). For any $\tau > 0$, we may rewrite $\sigma^2(d_0(\tau),\tau) - \sigma^2(d_0,\tau_0)$ as

$$\left(\sigma^2(d_0(\tau),\tau) - \sigma^2(d_0(\tau)\tau,1)\right) + \left(\sigma^2(d_0(\tau)\tau,1) - \sigma^2(d_0\tau_0,1)\right) + \left(\sigma^2(d_0\tau_0,1) - \sigma^2(d_0,\tau_0)\right).$$

- Since $(d_0(\tau)\tau, 1) \in \mathbf{B}(d_0(\tau),\tau)$, from Lemma 1 *i*), we have

$$
\begin{aligned}
&\sigma^2(d_0(\tau),\tau) - \sigma^2(d_0(\tau)\tau,1) \\
&= E\left[F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}) - F_{d_0(\tau)\tau}(X_{t-1}, \dots, X_{t-d_0(\tau)\tau})\right]^2 \\
&\geq 0.
\end{aligned}
$$

- When $d_0\tau_0 > d_0(\tau)\tau$, we have $(d_0\tau_0, 1) \in \mathbf{B}(d_0(\tau)\tau, 1)$. Thus from Lemma 1 *i*), we have

$$
\begin{aligned}
&\sigma^2(d_0(\tau)\tau,1) - \sigma^2(d_0\tau_0,1) \\
&= E\left[F_{d_0(\tau)\tau}(X_{t-1}, \dots, X_{t-d_0(\tau)\tau}) - F_{d_0\tau_0}(X_{t-1}, \dots, X_{t-d_0\tau_0})\right]^2 \\
&\geq 0.
\end{aligned}
$$

When $d_0\tau_0 \leq d_0(\tau)\tau$, we have $(d_0(\tau)\tau, 1) \in \mathbf{B}(d_0\tau_0, 1) \subset \mathbf{B}(d_0, \tau_0)$. Thus from Lemma 1 *ii*), we have

$$\begin{aligned}&\sigma^2(d_0(\tau)\tau, 1) - \sigma^2(d_0\tau_0, 1)\\ &= -E\left[F_{d_0(\tau)\tau}(X_{t-1}, \ldots, X_{t-d_0(\tau)\tau}) - F_{d_0\tau_0}(X_{t-1}, \ldots, X_{t-d_0\tau_0})\right]^2\\ &= 0.\end{aligned}$$

- Since $(d_0\tau_0, 1) \in \mathbf{B}(d_0, \tau_0)$, from Lemma 1 *ii*), we have

$$\begin{aligned}&\sigma^2(d_0\tau_0, 1) - \sigma^2(d_0, \tau_0)\\ &= -E\left[F_{d_0\tau_0}(X_{t-1}, \ldots, X_{t-d_0\tau_0}) - F_{d_0}(X_{t-\tau_0}, \ldots, X_{t-d_0\tau_0})\right]^2\\ &= 0\end{aligned}$$

So $\sigma^2(d_0, \tau_0) \leq \sigma^2(d_0(\tau), \tau)$.

Thus from Theorem 3.2 i), ii), we have $\sigma^2(d_0, \tau_0) \leq \sigma^2(d_0(\tau), \tau) \leq \sigma^2(d, \tau)$.

3.2 Estimation of the embedding dimension and delay time

In this section we propose the procedure for determining the embedding dimension and the delay time suggested by Theorem 3.2. This procedure is based on the cross-validation approach developed by Cheng and Tong (1995) for determining the embedding dimension.

Let $\{X_1, \ldots, X_N\}$ be the observed data, D, T be sufficiently large for $d_0 \leq D$ and $\tau_0 \leq T$ and $L = DT$.

Put

$$CV(d,\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}\left(X_t - \hat{F}_{\backslash t(d,\tau)}(X_{t-\tau}, \ldots, X_{t-d\tau})\right)^2,$$

where $\hat{F}_{\backslash t(d,\tau)}$ denotes the estimated regression function with the t-th point deleted. That is,

$$\hat{F}_{\backslash t(d,\tau)}(z) = \frac{1}{N-L}\sum_{s=L, s\neq t}^{N} X_s K_{d,h}\left(z - (X_{s-\tau}, \ldots, X_{s-d\tau})\right)\left(\hat{f}_{\backslash t(d,\tau)}\right)^{-1},$$

where the summation over s omit t in each case, and

$$\hat{f}_{\backslash t(d,\tau)}(z) = \frac{1}{N-L}\sum_{s=L, s\neq t}^{N} K_{d,h}\left(z - (X_{s-\tau}, \ldots, X_{s-d\tau})\right),$$

and $K_{d,h}$ is a kernel with constant bandwidth h that decreases toward 0 as N tends to infinity, i.e.,

$$K_{d,h}(z) = \frac{1}{h^d}K_d\left(\frac{z}{h}\right).$$

K_d is usually taken to be a probability density function on $\mathbf{R}^d$.

Now we describe our procedure for determining the embedding dimension and the delay time. First, minimize $CV(d,\tau)$ with respect to d over $1 \leq d \leq D$ for each $\tau \leq T$. Denoting the minimizer by $\hat{d}_0(\tau)$, then the estimators of embedding dimension and the delay time are given by $\hat{d}_0 = \min_{1\leq\tau\leq T}\hat{d}_0(\tau)$ and $\hat{\tau}_0 = \mathrm{argmin}_{1\leq\tau\leq T}\hat{d}_0(\tau)$.

Theorem 3.3 *Under conditions (c),(d) and (f)–(r) which are listed in Section 3.3.1,*

i) For any $\tau = 1, \ldots, T$, $\lim_{N\to\infty} P\left\{\hat{d}_0(\tau) = d_0(\tau)\right\} = 1$,

ii) $\lim_{N\to\infty} P\{\hat{\tau}_0 = \tau_0\} = 1$.

The proof of Theorem 3.3 is given in the next section.

3.3 Proof of Theorem 3.3

3.3.1 Basic conditions and theorems

We use the following conditions for Theorem 3.3.

(a) $E\left[\varepsilon_t | \mathcal{A}_{-\infty}^{t-1}(X)\right] = 0$, almost surely.

(b) $E\left[\varepsilon_t^2 | \mathcal{A}_{-\infty}^{t-1}(X)\right] = \sigma^2, (\sigma > 0)$, almost surely.

(c) $K_d(u) = \Pi_{i=1}^d k(u_i)$ for $u = (u_1, \ldots, u_d) \in \mathbf{R}^d$.

(d) F is Hölder continuous, i.e. There exists $c_1 > 0$ and $0 < \mu \leq 1$ such that for all $x, y \in \mathbf{R}^d$, $|F(x) - F(y)| \leq c_1 ||x - y||^\mu$, where $||\cdot||$ denotes the Euclidean norm in $\mathbf{R}^d$.

(e) W_d is a weight function which has a compact support $S \subset \mathbf{R}^d$ and

$$0 < \int_{\mathbf{R}^d} W_d(x)dx < \infty, 0 \leq W_d(x) \leq 1.$$

(f) For all $d < D$ and $\tau < T$, let $f_{(d,\tau)}$ denote the probability density function of $(X_{t-\tau}, \ldots, X_{t-d\tau})$, which is strictly positive on S, and there exists $c_2 > 0$ such that for all $x, y \in \mathbf{R}^d$, $\left|f_{(d,\tau)}(x) - f_{(d,\tau)}(y)\right| \leq c_2 ||x - y||$.

(g) k has compact support, and there exists $c_3 > 0$ such that for all $x, y \in \mathbf{R}$, $|k(x) - k(y)| \leq c_3 |x - y|$.

(h) For all $d < D$ and $\tau < T$, and for every $t, s, u, t', s', u' \in \mathbf{N}$, the joint probability density function of $\left(Z_t^{(d,\tau)}, Z_s^{(d,\tau)}, Z_u^{(d,\tau)}, Z_{t'}^{(d,\tau)}, Z_{s'}^{(d,\tau)}, Z_{u'}^{(d,\tau)}\right)$ is bounded, where $Z_t^{(d,\tau)}$ is defined in the proof of Lemma 1.

(i) Let $1/p + 1/q = 1$. For some $p > 2$ and $\delta > 0$ such that $\delta < 2/q - 1$, $E|\varepsilon_t|^{2p(1+\delta)} < \infty$ and $E|F(X_{t-\tau}, \ldots, X_{t-d\tau})|^{2p(1+\delta)} < \infty$.

(j) For δ in condition (i) and some $\varepsilon > 0$, $\beta_j^{\delta/(1+\delta)} = O(j^{-2+\varepsilon})$, where

$$\beta_j = \sup_{i \in \mathbf{N}} \left(E \left[\sup_{A \in \mathcal{A}_{i+j}^{\infty}(X)} \left| P(A|\mathcal{A}_1^i(X)) - P(A) \right| \right] \right).$$

(k) Let $j = j(N)$ be a positive integer and $i = i(N)$ be the largest positive integer such that $2ij \leq N$,

$$\limsup_{N \to \infty} \left(1 + 6e^{1/2} \beta_j^{1/(1+i)}\right)^i < \infty.$$

(l) For $i = i(N)$ in condition (k) and the bandwidth $h(N, d)$,

$$\limsup_{N \to \infty} \left(i(N) h(N, d)^d\right) < \infty.$$

(m) $Nh(N, d)^{2d} \to \infty$ as $N \to \infty$.

(n) For μ in assumption (d), $Nh(N, d)^{2d+2\mu} \to 0$ as $N \to \infty$.

(o) For q, δ and ε in condition (i) and (j), $N^\varepsilon h(N, d)^{-2d+\theta} \to 0$ as $N \to \infty$, where $\theta = 4d/(q + q\delta)$.

(p) The set M and S defined in Condition 1.1 are bounded.

(q) $\{X_t\}$ is ergodic.

(r) For $d > d'$,

$$\frac{h(N, d)^d}{h(N, d')^{d'}} \to 0 \text{ as } N \to \infty.$$

Conditions (a)–(o) are needed for Theorem 3.4 and Theorem 3.5 described below. Note that (a) and (b) are assumed in equation (3.2), and that (e) is derived from (p) in the proof of Theorem 3.3. We need the following two theorems which is immediately obtained from Theorem 1 and Theorem 3 in Cheng and Tong (1992) by replacing $(X_{t-1}, \ldots, X_{t-d})$ with $(X_{t-\tau}, \ldots, X_{t-d\tau})$.

Theorem 3.4 *Under conditions (a)–(o),*

$$CV(d, \tau) = RSS(d, \tau) \left(1 + \frac{2\alpha(d)\gamma(d)}{h(N, d)^d N} + o_p \left(\frac{1}{h(N, d)^d N}\right)\right),$$

where

$$RSS(d, \tau) = \frac{1}{N - L + 1} \sum_{t=L}^{N} \left(X_t - \hat{F}_{(d,\tau)}(X_{t-\tau}, \ldots, X_{t-d\tau})\right)^2 W_d(X_{t-\tau}, \ldots, X_{t-d\tau}),$$

where W_d is a non-negative weight function which satisfies the condition (e) and

$$\hat{F}_{(d,\tau)}(z) = \frac{1}{N - L + 1} \sum_{s=L}^{N} X_s K_{d,h}(z - (X_{s-\tau}, \ldots, X_{s-d\tau}))(\hat{f}_{(d,\tau)})^{-1},$$

where

$$\hat{f}_{(d,\tau)}(z) = \frac{1}{N - L + 1} \sum_{s=L}^{N} K_{d,h}(z - (X_{s-\tau}, \ldots, X_{s-d\tau})),$$

and $\alpha(d) = K_d^{1/d}(0), \gamma(d) = \dfrac{\int W_d(x)dx}{\int W_d(x)f(x)dx}$.

Theorem 3.5 *Under conditions (a)–(o),*

$$RSS(d,\tau) = \sigma_N^2(d,\tau)\left(1 - \frac{(2\alpha(d) - \beta(d))\gamma(d)}{h(N,d)^d N} + o_p\left(\frac{1}{h(N,d)^d N}\right)\right),$$

where

$$\sigma_N^2(d,\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(t,\tau)}\right)^2 W_d(X_{t-\tau},\dots,X_{t-d\tau})$$

and $\beta(d) = \int K_d(u)^{2/d}du$.

3.3.2 The proof of Theorem 3.3

To prove part i) of Theorem 3.3, we fix $0 < \tau < T$, and let

$$W_d(x) = \begin{cases} 1 & x \in S_{X_{d,\tau}} \\ 0 & \text{otherwise} \end{cases},$$

where

$$S_{X_{d,\tau}} = \left\{(x_1,\dots,x_d) \in \mathbf{R}^d | P\{||(X_{t-\tau},\dots,X_{t-d\tau}) - (x_1,\dots,x_d)|| < \varepsilon\} > 0 \text{ for all } \varepsilon >\right.$$

Then from boundedness of $\{X_t\}$, $W_d(x)$ satisfies the condition (e) and $W_d(X_{t-\tau},\dots,X_{t-d\tau}) = 1$ with probability 1.
From condition (m) we have

$$\frac{1}{h(N,d)^d N} \to 0 \text{ as } N \to \infty,$$

thus from Theorem 3.4 and Theorem 3.5,

$$CV(d,\tau) = \sigma_N^2(d,\tau) + o_p(1) \text{ for any } d. \tag{3.7}$$

From ergodicity of $\{X_t\}$, we have

$$\sigma_N^2(d,\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(d,\tau)}\right)^2 W_d(X_{t-\tau},\dots,X_{t-d\tau})$$

$$\begin{aligned} &\to E\left[\left(\varepsilon_t^{(d,\tau)}\right)^2 W_d(X_{t-\tau},\dots,X_{t-d\tau})\right] \text{ almost surely as } N \to \infty \\ &- E\left[\varepsilon_t^{(d,\tau)}\right]^2 \\ &= \sigma^2(d,\tau). \end{aligned} \tag{3.8}$$

Thus from (3.7) and (3.8), we have

$$\lim_{N\to\infty} CV(d,\tau) = \sigma^2(d,\tau).$$

For $d < d_0(\tau)$, we have $\sigma^2(d,\tau) - \sigma^2(d_0(\tau),\tau) > 0$ from Theorem 3.2. Thus

$$\begin{aligned} &P\left\{\hat{d}_0(\tau) = d\right\} \\ &= P\left\{CV(d,\tau) = \min_{d'} CV(d',\tau)\right\} \\ &\le P\{CV(d,\tau) \le CV(d_0(\tau),\tau)\} \\ &= P\left\{\sigma^2(d,\tau) + \left(CV(d,\tau) - \sigma^2(d,\tau)\right) \le \sigma^2(d_0(\tau),\tau) + \left(CV(d_0(\tau),\tau) - \sigma^2(d_0(\tau),\tau)\right)\right\} \\ &= P\left\{\sigma^2(d,\tau) - \sigma^2(d_0(\tau),\tau) \le \left(CV(d_0(\tau),\tau) - \sigma^2(d_0(\tau),\tau)\right) - \left(CV(d,\tau) - \sigma^2(d,\tau)\right)\right\} \\ &\le P\left\{\sigma^2(d,\tau) - \sigma^2(d_0(\tau),\tau) \le \left|CV(d_0(\tau),\tau) - \sigma^2(d_0(\tau),\tau)\right| + \left|CV(d,\tau) - \sigma^2(d,\tau)\right|\right\} \\ &\to 0 \text{ as } N \to \infty. \end{aligned}$$

For $d_0(\tau) < d \le D$, we have

$$\varepsilon_t^{(d,\tau)} = X_t - E[X_t|X_{t-\tau},\dots,X_{t-d\tau}] = X_t - E\left[X_t|X_{t-\tau},\dots,X_{t-d_0(\tau)\tau}\right] = \varepsilon_t^{(d_0(\tau),\tau)},$$

and

$$\begin{aligned} &P\left\{\hat{d}_0(\tau) = d\right\} \\ &\le P\{CV(d,\tau) \le CV(d_0(\tau),\tau)\} \\ &= P\left\{CV(d,\tau) \le CV(d_0(\tau),\tau) \text{ and } (X_{t-\tau},\dots,X_{t-d\tau}) \in S_{X_{(d,\tau)}} \text{ for any } t = L, L+1,\dots,N\right\} \\ &+ P\left\{CV(d,\tau) \le CV(d_0(\tau),\tau) \text{ and } (X_{t-\tau},\dots,X_{t-d\tau}) \notin S_{X_{(d,\tau)}} \text{ for some } t = L, L+1,\dots,N\right\} \\ &= P\left\{CV(d,\tau) \le CV(d_0(\tau),\tau) \text{ and } (X_{t-\tau},\dots,X_{t-d\tau}) \in S_{X_{(d,\tau)}} \text{ for any } t = L, L+1,\dots,N\right\} \end{aligned}$$

When $(X_{t-\tau},\dots,X_{t-d\tau}) \in S_{Y_{(d,\tau)}}$ for any $t = L, L+1,\dots,N$, we have

$$\sigma_N^2(d,\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(d,\tau)}\right)^2 W_d(X_{t-d},\dots,X_{t-d\tau})$$

$$= \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(d,\tau)}\right)^2$$
$$= \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(d_0(\tau),\tau)}\right)^2.$$

Note that for any $(x_1,\ldots,x_d) \in \mathrm{R}^d$

$$||(X_{t-\tau},\ldots,X_{t-d\tau})-(x_1,\ldots,x_d)|| \geq ||(X_{t-\tau},\ldots,X_{t-d_0(\tau)\tau})-(x_1,\ldots,x_{d_0(\tau)})||,$$

implies

$$(X_{t-\tau},\ldots,X_{t-d\tau}) \in S_{X_{(d,\tau)}} \Rightarrow (X_{t-\tau},\ldots,X_{t-d_0(\tau)\tau}) \in S_{X_{(d_0(\tau),\tau)}},$$

we have

$$\sigma_N^2(d_0(\tau),\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}\left(\varepsilon_t^{(d_0(\tau),\tau)}\right)^2.$$

So for any $(X_{t-\tau},\ldots,X_{t-d\tau}) \in S_{X_{(d,\tau)}}$ (for any $t = L,\ldots,N$), we have

$$\sigma_N^2(d,\tau) = \sigma_N^2(d_0(\tau),\tau).$$

From Theorem 3.4 and Theorem 3.5, we have

$$CV(d,\tau) = \sigma_N^2(d,\tau)\left(1+\beta(d)\gamma(d)\frac{1}{h(N,d)^dN} + o_p\left(\frac{1}{h(N,d)^dN}\right)\right)$$
$$CV(d_0(\tau),\tau)$$
$$= \sigma_N^2(d_0(\tau),\tau)\left(1+\beta(d_0(\tau))\gamma(d_0(\tau))\frac{1}{h(N,d_0(\tau))^{d_0(\tau)}N} + o_p\left(\frac{1}{h(N,d_0(\tau))^{d_0(\tau)}N}\right)\right),$$

and

$$\frac{1}{\sigma_N^2(d_0(\tau),\tau)/h(N,d)^dN}(CV(d_0(\tau),\tau) - CV(d,\tau))$$
$$= \beta(d_0(\tau))\gamma(d_0(\tau))\frac{1/h(N,d_0(\tau))^{d_0(\tau)}N}{1/h(N,d)^dN} - \beta(d)\gamma(d) + \frac{o_p(1/h(N,d_0(\tau))^{d_0(\tau)}N)}{1/h(N,d)^dN}$$
$$+\frac{o_p(1/h(N,d)^dN)}{1/h(N,d)^dN}.$$

Thus

$$P\left\{\hat{d}_0(\tau) = d\right\}$$
$$\leq P\left\{CV(d_0(\tau),\tau) - CV(d,\tau) \geq 0 \text{ and } (X_{t-\tau},\ldots,X_{t-d\tau}) \in S_{X_{(d,\tau)}} \text{ (for any } t = L,\ldots,N)\right\}$$
$$= P\left\{\frac{1}{\sigma_N^2(d_0(\tau),\tau)/h(N,d)^d}(CV(d_0(\tau),\tau) - CV(d,\tau)) \geq 0\right.$$
$$\left.\text{and } (X_{t-\tau},\ldots,X_{t-d\tau}) \in S_{X_{(d,\tau)}} \text{ (for any } t = L,\ldots,N)\right\}$$
$$\leq P\left\{\beta(d_0(\tau))\gamma(d_0(\tau))\frac{1/h(N,d_0(\tau))^{d_0(\tau)}}{1/h(N,d)^d} + \frac{o_p(1/h(N,d_0(\tau))^{d_0(\tau)}N)}{1/h(N,d)^dN}\right.$$
$$\left.+\frac{o_p(1/h(N,d)^dN)}{1/h(N,d)^dN} \geq \beta(d)\gamma(d)\right\}.$$

From assumption (r), we have

$$\beta(d_0(\tau))\gamma(d_0(\tau))\frac{1/h(N,d_0(\tau))^{d_0(\tau)}}{1/h(N,d)^d} + \frac{o_p(1/h(N,d_0(\tau))^{d_0(\tau)}N)}{1/h(N,d)^dN} + \frac{o_p(1/h(N,d)^dN)}{1/h(N,d)^dN} \to 0$$

as $N \to \infty$ in probability and from the definition of $\beta(d),\gamma(d)$, we have $\beta(d)\gamma(d) > 0$. Thus $P\{\hat{d}_0(\tau) = d\} \to 0$ as $N \to \infty$ in probability.

Next we prove part ii) of Theorem 3.3. For $\tau > 0$ such that $d_0(\tau) \neq d_0$, we have $d_0(\tau) > d_0(\tau_0)$ from Theorem 3.1. Thus

$$\begin{aligned}
P\{\hat{\tau} = \tau\} &= P\left\{\hat{d}_0(\tau) = \min_{\tau} \hat{d}_0(\tau)\right\} \\
&\leq P\left\{\hat{d}_0(\tau) \leq \hat{d}_0(\tau_0)\right\} \\
&\leq P\left\{\hat{d}_0(\tau) < d_0(\tau) \text{ or } \hat{d}_0(\tau_0) > d_0(\tau_0)\right\} \\
&\leq P\left\{\hat{d}_0(\tau) < d_0(\tau)\right\} + P\left\{\hat{d}_0(\tau_0) > d_0(\tau_0)\right\} \\
&\to 0 \text{ as } N \to \infty.
\end{aligned}$$

This completes the proof of Theorem 3.3.

Chapter 4

The Lyapunov exponent

In this chapter, we propose the consistent estimator of the skeleton using the data from the non-linear autoregressive time series with dynamic noise. First of all, refering to Taniguchi and Kakizawa (2000), we review the basics of chaos and the Lyapunov exponent.

4.1 Chaos and the Lyapunov exponent

We consider the mapping $\mathbf{F} : M \to M$, where $M \subset R^d$. We denote by $\mathbf{F}^p$ the $p-$fold composition of F, i.e., $\mathbf{F}^p \equiv \mathbf{F} \circ \mathbf{F}^{p-1}$ and $\mathbf{F}^1 \equiv \mathbf{F}$. For each $t \in N$, let x_t denote a $d-$dimensional state vector in M satisfying

$$x_t = \mathbf{F}(x_{t-1}), x_0 \in M, \tag{4.1}$$

and the sequence $\{x_t; t \geq 0\}$ is called the trajectory.

Definition 4.1 (Periodic point)
Let q be a finite positive integer. A d-dimensional vector $x^* \in M$ is called a periodic point with period q of (4.1) if $x^* = \mathbf{F}^q(x^*)$ and $x^* \neq \mathbf{F}^j(x^*)$ for $1 \leq j < q$. The ordered set $\{x^*, \mathbf{F}(x^*), \ldots, \mathbf{F}^{q-1}(x^*)\}$ is called a q-cycle.

Definition 4.2 (Attractor)
A d-dimensional set $A \subset M$ is called an attractor for $\mathbf{F} : M \to M$ if A is a

minimal compact set such that

$$B = \{x; \lim_{n\to\infty} \inf_{y\in A} |\mathbf{F}^n(x) - y| = 0\}$$

has positive Lebesgue measure. The set B is called the basin of attraction for A. If the attractor is a set of q points $\{x_1^*, \ldots, x_q^*\}$ such that

$$x_t^* = \mathbf{F}(x_{t-1}^*),\ t > 1,$$

and

$$\mathbf{F}(x_q^*) = x_1^*,$$

then it is said to be a limit cycle. If the attractor is not a limit cycle, it is said to be a strange attractors.

If the attractor is a limit cycle, this case is regarded as degenerate.

A standard way to quantify the sensitive dependence of $\mathbf{F} : M \to M$, on an initial conditions is to evaluate the so-called Lyapunov exponent. Let x_0 and $x_0' \in M$ denote two initial vectors and put $\delta = x_0' - x_0$. Then, after n iteration

$$\begin{aligned} x_n' - x_n &= \mathbf{F}^n(x_0') - \mathbf{F}^n(x_0) \\ &\approx D\mathbf{F}^n(x_0)(x_0' - x_0), \end{aligned}$$

where $D\mathbf{F}^n$ is the $n \times n$ derivative matrix of $\mathbf{F}^n$. Set $J_t = D\mathbf{F}(x_t)$ and $T_n(x_0) = J_0 \cdot J_1 \cdots J_{n-1}$. By application of the chain rule we obtain

$$|x_n' - x_n| = |T_n(x_0)\delta| + o(|\delta|). \tag{4.2}$$

Let $\mu_n(x_0)$ denote the largest eigenvalue of a positive definite matrix ${}^tT_n(x_0) \cdot T_n(x_0)$. Thus we get the following definition

Definition 4.3 (Lyapunov exponent)
The deterministic system (4.1) is said to have a Lyapunov exponent $\lambda(x_0)$ if

$$\lambda(x_0) = \lim_{n\to\infty} \left(\frac{1}{2n} \log|\mu_n(x_0)|\right). \tag{4.3}$$

exists.

From (4.3) and (4.2) we can see that main order term of $|x_n' - x_n|$ is $\exp(n\lambda(x_0))|\delta|$. Hence positive $\lambda(x_0)$ confirms sensitive dependence of $\mathbf{F}$ on x_0.

Eckmann and Ruelle (1992) propose the method for estimating the Lyapunov exponent from the trajectory $\{x_t; t = 0, \ldots, n\}$ of the deterministic system as follows; For sufficiently small $\delta > 0$, put

$$\mathcal{A}_i = \{x_s; |x_s - x_i| < \delta, s \neq i, n\},\ i = 0, \ldots, n-1$$

and find $D(i) = \hat{D}(i)$ that minimizes

$$\sum_{x_s \in \mathcal{A}_i} \left| x_{s+1} - x_{i+1} - \hat{D}(i)(x_s - x_i) \right|$$

for each $i = 0, 1, \ldots, n-1$. Denote by μ the maximum eigenvalue of ${}^t(\hat{D}(0) \cdot \hat{D}(1) \cdots \hat{D}(n-1)) \cdot (\hat{D}(0) \cdot \hat{D}(1) \cdots \hat{D}(n-1))$. Then the Lyapunov exponent is estimated by

$$\hat{\lambda} = \frac{1}{2n} \log \mu.$$

The concept of a Lyapunov exponent has been developed to characterize the sensitive dependence on the initial value of a deterministic system, for example, a skeleton of the non-linear autoregressive time series with dynamic noise. However, in the case of the non-linear autoregressive time series with dynamic noise, the sequence $\{X_t; t \geq 0\}$ depend not only on the initial value but also on the dynamic noise. For this case, to quantify the sensitive dependence on initial value, the Lyapunov-exponent-type quantities have been proposed.

Definition 4.4 (Local Lyapunov exponent, Wolff(1992))
For the non-linear autoregressive time series model $\{X_t; t \geq 0\}$,

$$\lambda_{i,m} = \frac{1}{m}\frac{1}{n_i} \sum_{j \in S_i} \log \left| \frac{X_{i+m} - Y_{j+m}}{Y_i - Y_j} \right|,$$

where $S_i = \{j; 0 < |Y_i - Y_j| \leq \delta\}, n_i = \#(S_i), m \in N$ and $\delta > 0$, is called the local Lyapunov exponent at Y_i for lag m.

This idea is to estimate the Lyapunov exponent locally at Y_i for lag m and pre-specified δ representing the perturbation. Wolff(1992) studied the statistical properties of $\lambda_{i,m}$ for a variety of data which are from specified models.

Yao and Tong(1994a) proposed $m-$step Lyapunov-like index for the one-dimensional case.

Definition 4.5 (Lyapunov-like index, Yao and Tong(1994a))
For the non-linear autoregressive time series model $\{X_t; t \geq 0\}$,

$$\lambda_m(x) = E\left(\prod_{k=1}^{m} \frac{d}{dx} F(X_{k-1}) | X_0 = x\right)$$

is called the $m-$*step Lyapunov-like index.*

These Lyapunov-exponent-type indexes quantify the sensitive dependence on initial value, but they also depend on the variance of ε_t in the non-linear autoregressive time series model. Thus they aren't consistent estimator of the Lyapunov exponent of the skeleton if $Var(\varepsilon_t) > 0$.

4.2 The ergodic theory of chaos

In this section we study stationarity and ergodicity of nonlinear autoregressive time series model and the ergodic theory of deterministic chaos to give the Lyapunov exponent by space average. First of all, we discuss the fundamental properties of Markov chain.

Definition 4.6 (Markov chain)
An d-vector stochastic process $\{X_t; t \geq 0\}$ *is called a Markov chain with transition probability* $P(\cdot,\cdot)$, *provided that*

$$\mathcal{L}\{X_{t+1}|\mathcal{F}_t\} = \mathcal{L}\{X_{t+1}|X_t\} = P(X_t, \cdot) \quad \textit{for all } t \geq 0,$$

where $\mathcal{L}\{X_{t+1}|*\}$ *is the conditional distribution of* X_{t+1} *given* $*$, *and* $\mathcal{F}_t = \mathcal{F}(X_t, \ldots, X_0)$ *is the* σ*-algebra generated by* $X_t, X_{t-1}, \ldots, X_0$. *The distribution* $\mathcal{F}(X_0)$ *of* X_0 *is called the initial distribution of* $\{X_t\}$.

For a Markov chain $\{X_t\}$, the t-th step transition probability is defined by

$$P^t(x, A) = P(X_t \in A | X_0 = x), x \in R^d, A \in \mathcal{B}^d.$$

Definition 4.7 (Irreducibility)
Let ϕ *be a* σ*-finite measure on* $(R^d, \mathcal{B}^d)$ *with* $\phi(R^d) > 0$. *A Markov chain* $\{X_t\}$ *is said to be* ϕ*-irreducible if*

$$\sum_{t=1}^{\infty} P^t(x, A) > 0,$$

for all $x \in R^d$ *and all* $A \in \mathcal{B}^d$ *with* $\phi(A) > 0$.

If $\{X_t\}$ is ϕ-irreducible for some ϕ, it is simply called irreducible. In this case the measure ϕ is called an irreducibility measure for $\{X_t\}$. A nonnull set $C \in \mathcal{B}^d$ is called to be small if there exists a positive integer k, a constant $b > 0$ and a probability measure ν such that

$$P^k(x, A) > b\nu(A), \text{ for all } x \in C, A \in \mathcal{B}^d. \tag{4.4}$$

Proposition 4.1 (Nummelin(1984))
For a small set C, *define*

$$I(C) = \{k \in N :^{\exists} b, \nu \textit{ such that } P^k(x, A) > b\nu(A), \textit{ for all } x \in C, A \in \mathcal{B}^d\},$$

and let $D(C)$ *be the greatest common divisor of* $I(C)$. *Then* $D(C)$ *is the same for all small sets.*

Thus we may write D instead of $D(C)$. If $D = 1$, the Markov chain is called aperiodic otherwise it is called periodic with period D.

Now we discuss the ergodicity of nonlinear autoregressive models.

Definition 4.8 (Ergodic process)
A Markov chain $\{X_t\}$ *is said to be geometrically ergodic if there exists a probability measure* π *on* $(R^d, \mathcal{B}^d)$ *and a positive constant* $\rho < 1$ *such that*

$$\lim_{t\to\infty} \rho^{-t} \left\| P^t(x, \cdot) - \pi(\cdot) \right\| = 0, \; x \in R^d, \tag{4.5}$$

where $||\cdot||$ *is the total variation norm. If*

$$\lim_{t\to\infty}\left\|P^t(x,\cdot)-\pi(\cdot)\right\|=0,\ x\in R^d$$

holds, then $\{X_t\}$ *is said to be ergodic.*

Definition 4.9 (Invariant measure)
For a Markov chain $\{X_t\}$*, a probability measure* π *on* $(R^d,\mathcal{B}^d)$ *is said to be invariant if*

$$\pi(A)=\int P^1(x,A)\pi(dx),\ \textit{for all}\ A\in\mathcal{B}^d,$$

holds.

Proposition 4.2 (Nummelin(1984))
If a Markov chain $\{X_t\}$ *is geometrically ergodic, and if the initial distribution* $\mathcal{L}\{X_0\}$ *is* π *in (4.5), then* $\{X_t\}$ *is strictly stationary.*

To argue the geometrical ergodicity of nonlinear autoregressive process, we state three preliminary lemmas.

Lemma 4.1 (Chan and Tong (1985))
For a nonlinear autoregressive time series model (1.3), suppose that F *is bounded over bounded sets. Then* $\{X_t\}$ *defined by (1.5) is aperiodic and* μ*-irreducible, where* μ *is the Lebesgue measure. Furthermore,* μ*-nonnull compact sets are small sets.*

Lemma 4.2 (Tweedie's criterion, Tweedie (1975))
Let $\{X_t\}$ *be an aperiodic and ireducible Markov chain. Support that there exist a small set* C*, a nonnegative measurable function* g*, positive constants* c_1, c_2 *and* $\rho<1$ *such that*

$$E\left(g(X_{t+1})|X_t=x\right)<\rho g(x)-c_1,\ \textit{for any}\ x\notin C,$$

and

$$E\left(g(X_{t+1})|X_t=x\right)<c_2,\ \textit{for any}\ x\in C.$$

Then $\{X_t\}$ *is geometrically ergodic.*

Lemma 4.3 (h-step criterion, Tjøstheim(1990))
If there exists a positive integer h *such that a Markov chain* $\{X_{th}\}$ *is geometrically ergodic, then* $\{X_t\}$ *is geometrically ergodic.*

To check the geometrical ergodicity of nonlinear autoregressive model (1.3), An and Huang (1996) gave the following theorems to check the geometrical ergodicty of the nonlinear autoregressive model given in (1.3).

Theorem 4.1
In model (1.3), suppose that there exists a positive number $\lambda<1$ *and a constant* c *such that*

$$|F(x_1,\ldots,x_p)|\le\lambda\max\left(|x_1|,\ldots,|x_p|\right)+c.$$

Then $\{X_t\}$ *defined by (1.3) is geometrically ergodic.*

Theorem 4.2
In the model (1.3), suppose that

$$\sup_{|x|\le K}|F(x)|<\infty\ \textit{for each}\ K>0,$$

and

$$\lim_{|X|\to\infty}\frac{|F(x)-{}^t ax|}{|x|}=0,$$

where $x={}^t(x_1,\ldots,x_p)$*, and* $a={}^t(a_1,\ldots,a_p)$ *satisfies*

$$1-a_1z-\cdots-a_pz^p\ne0\ \textit{for all}\ |z|\le1.$$

Then $\{X_t\}$ *generated by (1.3) is geometrically ergodic.*

In the rest of this chapter, we consider the one-dimensional case ($d=1$). Note that for the one-dimensional case, the largest eigenvalue of ${}^tT_n(x_0)T_n(x_0)$

is $T_n(x_0)^2 = (F'(x_0) \cdot F'(x_1) \cdots F'(x_{n-1}))^2$, and the Lyapunov exponent is given by

$$\begin{aligned} \lambda(x_0) &= \lim_{n\to\infty} \frac{1}{n} \log |T_n(x_0)| \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left| \prod_{t=0}^{n-1} F'(x_t) \right| \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \log |F'(x_t)| \\ &= \lim_{n\to\infty} \int \log |F'(x)|\, dG_n(x), \end{aligned}$$

where

$$G_n(x) = \frac{1}{n} \#\{0 \le t < n; x_t \le x\}.$$

Invariance and ergodicity are also defined for deterministic chaos series.

Definition 4.10 (Invariant measure)
For the deterministic system $x_t = \mathbf{F}(x_{t-1}), x_0 \in M$, a probability measure ρ is said to be invariant if $\rho(A) = \rho(F^{-1}(A))$ for any Borel set $A \subset M$.

Definition 4.11 (Ergodicity)
An invariant probability measure ρ is ergodic if it does not have a nontrivial convex decomposition:

$$\rho = \alpha\rho_1 + (1-\alpha)\rho_2 \text{ with } \alpha \neq 0, 1,$$

where ρ_1 and ρ_2 are again invariant probability measure and $\rho_1 \neq \rho_2$.

Theorem 4.3 (Ergodic theorem, Birkhoff (1931b))
For the deterministic system $x_t = \mathbf{F}(x_{t-1}), x_0 \in M$, let ϕ be an integrable function on M and ρ an invariant probability measure, then for ρ-almost all $x_0 \in M$, the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(x_i)$$

exists. If ρ is ergodic then for ρ-almost all $x_0 \in M$,

$$\int \phi(x)\rho(dx) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(x_i).$$

Ergodic theorem shows that if the invariant measure ρ of the deterministic system (4.1) is ergodic, then the Lyapunov exponent is given by

$$\lambda(x_0) = \int \log |F'(x)| dG(x) \text{ for } \rho\text{-almost all } x_0 \in M, \tag{4.6}$$

where G is the distribution function of the invariant measure ρ.

Ulam and von Neumann (1947) proved that an invariant measure for the logistic-4 map is that of the beta$(\frac{1}{2}, \frac{1}{2})$ distribution. Hall and Wolff (1995) provide the density function of the invariant measure of logistic-θ map whenever the invariant distribution is absolutely continuous, and showed that the Lyapunov exponent of logistic-θ map is not continuous function of θ. Lasota and Yorke (1973) proved the existence of absolutely continuous invariant distributions for maps which are quite different from the logistic.

To estimate the invariant distribution, we assume the existence of the following Kolmogorov measure.

Definition 4.12 (Kolmogorov measure)
The deterministic system $x_t = \mathbf{F}(x_{t-1}), x_0 \in M$ is said to have the Kolomogorov measure ρ if the stationary stochastic system $X_t = \mathbf{F}(X_{t-1}) + \epsilon_t, X_0 \in M$ has only one stationary measure ρ_ϵ and ρ_ϵ converges in law to ρ as $E(\epsilon_t^2) \to 0$.

The assumption of the existence of the Kolmogorov measure seems too strict. However, in a computer study, roundoff errors should play the role of the random noise. Due to sensitive dependence on initial conditions, even a very small level of roundoff errors has importance effects. The existence of the Kolmogorov measure shows that calculated value in computer study converges to the true value when roundoff errors tend to 0. Thus the assumption of the existence of the Kolmogorov measure is reasonable.

4.3 Estimation of the Lyapunov exponent

In this section we propose the procedure for estimating the Lyapunov exponent of the skeleton $x_t = F(x_{t-1}), x_0 \in M$ given by equation (4.6).

Let $\{X_1, \ldots, X_n\}$ be the observed data from the non-linear time series model

$$X_t = F(X_{t-1}) + \varepsilon_t, X_0 \in M. \tag{4.7}$$

Denote by $\hat{F}_n(x)$ and $\hat{f}_n(x)$, the local polynomial estimator of $F(x)$ and its derivative respectively. Next we regenerate the trajectory of the skeleton as follows:

$$Y_t = \hat{F}_n(Y_{t-1}) \text{ for } t > 0, Y_0 = X_0,$$

and denote by $\hat{G}_m$ the empirical distribution of the trajectory $\{Y_0, \ldots, Y_{m-1}\}$. Then we propose an estimator of the Lyapunov exponent of the skeleton as

$$\hat{\lambda}(F) = \int \log|\hat{f}_n(y)|\hat{G}_m(y).$$

We prove that the proposed estimator is consistent in the following theorem.

Theorem 4.4
Let $f(x)$ be the density function of X_0 and λ be the Lyapunov exponent of the skeleton of the non-linear time series (4.7). Suppose that the skeleton of (4.7) has the ergodic Kolmogorov measure with distribution function G, λ satisfies $-\infty < \lambda < \infty$, $f(x)$ has a compact support M_1, $f(x)$ and $F^{(p+1)}(x)$ are continuous on M_1 for some $p \geq 1$, and $h = O(n^{1/(2p+3)})$. Then $\hat{\lambda}(F)$ is a consistent estimator of the Lyapunov exponent of the skeleton of (4.7).

(Proof.) Since M_1 is compact and $f(x)$ and $F^{(p+1)}(x)$ are continuous on M_1, there exist $N_1, N_2 > 0$ such that

$$\left|F^{(p+1)}(x)\right| \leq N_1 \text{ and } \left|\frac{1}{f(x)}\right| \leq N_2 \text{ for } x \in M_1.$$

We may rewrite

$$\begin{aligned} Y_t &= \hat{F}_n(Y_{t-1}) \\ &= F(Y_{t-1}) + \left(\hat{F}_n(Y_{t-1}) - F(Y_{t-1})\right), \end{aligned}$$

and let $\epsilon_t = \hat{F}_n(Y_{t-1}) - F(Y_{t-1})$, then by Theorem 2.3 we have

$$E(\epsilon_t) \leq \mu_0^* \frac{N_1}{(p+1)!} h^{p+1},$$

and

$$Var(\epsilon_t) \leq \xi_0^* \frac{\sigma^2 N_2}{nh}.$$

Thus $E(\epsilon_t^2)$ converges to 0 as n tends to infinity. Then by the assumption of the existence of the Kolmogorov measure, the empirical distribution of $\{Y_t\}$ converges to the Kolmogorov measure.

Now we expand $\left|\hat{\lambda}(F) - \lambda\right|$ as

$$\begin{aligned} &\left|\hat{\lambda}(F) - \lambda\right| \\ &= \left|\int_M \log|\hat{f}_n(y)|d\hat{G}_m(y) - \int_M \log|F'(y)|dG(y)\right| \\ &= \left|\int_M \log|\hat{f}_n(y)|d\hat{G}_m(y) - \int_M \left(\log|\hat{f}_n(y)| + (\log|F'(y)| - \log|\hat{f}_n(y)|)\right) dG(y)\right| \\ &= \left|\int_M \log|\hat{f}_n(y)|d\hat{G}_m(y) - \int_M \log|\hat{f}_n(y)|dG(y) + \int_M \left(\log|\hat{f}_n(y)| - \log|F'(y)|\right) dG(y)\right| \\ &\leq \int_M \left|\log|\hat{f}_n(y)|\right| dV_{\hat{G}_m - G}(y) + \int_M \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y), \end{aligned}$$

where $V_{\hat{G}_m - G}(y)$ is the total variation of $\hat{G}_m - G$. For any $0 < \varepsilon \leq 1/2$, we put $M_\varepsilon = \{y \in M; |F'(y)| < \varepsilon\}$.

First we consider $\int_{M_\varepsilon} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y)$. From Theorem 2.3 and continuity of the density of $\{X_t\}$, $Pr\left\{|\hat{f}_n(y)| < \varepsilon\right\}$ converges to $\Phi\left(\frac{\varepsilon - \mu_n}{\sigma_n}\right) - \Phi\left(\frac{-\varepsilon - \mu_n}{\sigma_n}\right)$ uniformly as $n \to \infty$, where

$$\mu_n = F'(y) + \mu_1^* \frac{m^{(p+1)(y)}}{(p+1)!} h^p$$

and

$$\sigma_n = \sqrt{\xi_1^* \frac{\sigma^2}{f(y)}}.$$

Thus for any $\delta > 0$, there exist $\varepsilon_0 > 0$ and $n_1 \in N$ such that

$$Pr\left\{|\hat{f}_n(y)| < \varepsilon\right\} < \delta \text{ for } \varepsilon < \varepsilon_0 \text{ and } n > n_1.$$

And for $y \in M_\varepsilon$, $|F'(y)| \le \varepsilon \le 1/2$ and Theorem 2.3 imply that there exists $n_2 \in N$ such that

$$Pr\left\{|\hat{f}_n(y)| > 1\right\} < \delta \text{ for } n > n_2.$$

Thus for $n > \max(n_1, n_2)$, we have with probability larger than $1 - 2\delta$,

$$\begin{aligned} 0 &\ge \int_{M_\varepsilon} \log|\hat{f}_n(y)| dG(y) \\ &\ge \int_{M_\varepsilon} \log \varepsilon dG(y) \\ &\ge \int_{M_\varepsilon} \log|F'(y)| dG(y). \end{aligned}$$

Then for $n > \max(n_1, n_2)$, we have with probability larger than $1 - 2\delta$,

$$2\int_{M_\varepsilon} \log|F'(y)| dG(y) \le \int_{M_\varepsilon} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y) \le 0.$$

Since $\lambda > -\infty$, we have

$$\int_{M_\varepsilon} \log|F'(y)| dG(y) \to 0 \text{ as } \varepsilon \to 0.$$

Next we consider $\int_{M-M_\varepsilon} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y)$. For $y \in M - M_\varepsilon$, by Taylor Series expansions we have

$$\begin{aligned} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| &\le \frac{\left||\hat{f}_n(y)| - |F'(y)|\right|}{|F'(y)|} + o\left(\left||\hat{f}_n(y)| - |F'(y)|\right|\right) \\ &\le \frac{\left|\hat{f}_n(y) - F'(y)\right|}{\varepsilon} + o\left(\left|\hat{f}_n(y) - F'(y)\right|\right). \end{aligned}$$

By Theorem 2.3 we have

$$E\left(\frac{|\hat{f}_n(y) - F'(y)|}{\varepsilon}\right) \le \mu_1^* \frac{N_1 h^p}{\varepsilon(p+1)!} + o(h^p)$$

and

$$Var\left(\frac{|\hat{f}_n(y) - F'(y)|}{\varepsilon}\right) \le \xi_1^* \frac{\sigma N_2}{\varepsilon^2 n h^3} + o\left(\frac{1}{nh^3}\right).$$

Then we have

$$\begin{aligned} & E\left(\int_{M-M_\varepsilon} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y)\right)^2 \\ \le\ & E\int_{M-M_\varepsilon} \left(\log|\hat{f}_n(y)| - \log|F'(y)|\right)^2 dG(y) \\ \le\ & \left(\mu_1^* \frac{N_1 h^p}{\varepsilon(p+1)!}\right)^2 + \xi_1^* \frac{\sigma N_2}{\varepsilon^2 n h^3} + o(h^{2p}) + o\left(\frac{1}{nh^3}\right) \end{aligned}$$

Then by Chebychev's inequality we have

$$\int_{M-M_\varepsilon} \left|\log|\hat{f}_n(y)| - \log|F'(y)|\right| dG(y) = o_p(1).$$

And by existence of the Kolmogorov measure, we have $V_{\hat{G}_m - G} \to 0$. Therefore $\hat{\lambda}(F) - \lambda = o_p(1)$. This completes the proof of Theorem 4.4.

Bibliography

[1] Abarbanel, H. D. I. (1996). *Analysis of Observed Chaotic Data.* Springer, Berlin.

[2] An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of non-linear autoregressive models. *Statistica Sinica* **6,** 943-956.

[3] An, H. Z., Chen, M. and Huang, F. (1997). The geometric ergodicity and existence of moment for a class of non-linear time series model. *Statist. Probab. Lett.* **31,** 213-224.

[4] Berliner, L. M. (1992). Statistics, probability and chaos. *Statis. Sci.,* **7,** 69-122.

[5] Birkhoff, G. D. (1991a) Proof of a recurrence theorem for strongly transitive systems. *Proc. Natl. Acad. Sci. USA,* **17,** 650-655.

[6] Birkhoff, G. D. (1991b) Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA,* **17,** 656-660.

[7] Brockwel, P. J. and Davis, R. A. (1986) *Time Series: Theory and Methods.* Springer series in statistics, Springer, New York.

[8] Cawley, R. and Hsu, G. (1992). Local-geometric-projection method for noise reduction in chaotic maps and flows. *Physical Review A,* **46,** 3057-3082.

[9] Chatterjee, S. and Yilmaz, M. R. (1992). Chaos, fractals and statistics. *Statis. Sci.*, **7**, 49-68.

[10] Chan, K. S. and Tong, H. (1985). ●n the use of the deterministic. Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666-678.

[11] Cheng, B. and Tong, H. (1992). On consistent nonparametric order determination and chaos. *J. Roy. Statist. Soc.*, **B, 54**, 427-450.

[12] Cheng, B. and Tong, H. (1993). On residual sums of squares in nonparametric autoregression *Stochastic Process. Appl* **48** 157-174.

[13] Cheng, B. and Tong, H. (1995). Orthogonal projection, embedding dimension and sample size in chaotic time series from a statistical perspective. In Tong, H., editor, *Chaos and Forcasting,* World Scientific, Singapore, London.

[14] Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

[15] Eckmann, J. P. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57**, No. 3, Part I, 617–656.

[16] Eckmann, J. P. and Ruelle, D. (1992). Fundamental limitations for estimating dimensions and Lyapunov exponent in dynamical system. *Physica* D **56**, 185

[17] Fan, J. (1992). Design-adaptive Nonparametric Regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.

[18] Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Monographs on statistics and applied probability 66, Chapman and Hall, London.

[19] Fan, J., Hu, T. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.*, **21**, 433-446.

[20] Fan, J., Ya●, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measure in nonlinear dynamical systems. *Biometrika*, **81, 1**, 189-206.

[21] Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. In *Smooting techniques for curve estimation,* Lecture Note in Mathematics 757, 23-68, Springer-Verlag, New York.

[22] Hafner, C. M. (1998). *Nonlinear Time Series Analysis with Applications to Foreigh Exchange Rate Volatility.* Heidelberg: Physica-Verlag.

[23] Hall, P. and Truong, Y. K. (1995). On bandwidth choice for density estimation with dependent data. *Ann. Statist.*, **23, 6**, 2241-2263.

[24] Hall, P. and Wolfe, R. C. L. (1995). Properties of Invariant Distributions and Lyapunov Exponents for Chaotic Logistic Maps. *J. R. Statist. Soc. B* **57**, No. 2, 439–452.

[25] Hall, P. and Wolff, R. C. L. (1995). Estimators of integrals of powers of density derivatives. *Statist. Probab. Lett.* **6**, 109-115.

[26] Hall, P. and Wolff, R. C. L. (1998). Properties of distributions and correlation integrals for generalised versions of the logistic map. *Stochastic Process. Appl* **77**, 123-137.

[27] Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94**, 154-163.

[28] Hamilton, S. A. and Truong, Y. K. (1997). Local linear estimation in partly linear models. *J. Multivariate Anal.* **60**, 1-19.

[29] Kostelich, K. and Yorke, J. A. (1990). Noise reduction-finding the simplest dynamical system consistent with the data. *Physica D,* **41** 183-196.

[30] Landa,P. S. and Rozenblum, M. G. (1989). A comparison of methods for constructing a phase space and determining the dimension of an attractor from experimental data. *Sov. Phys.-Tech. Phys.*, **34** 1229-1232.

[31] Lasota, A. and Mackey, M. C. (1989). Stochastic perturbation of dynamical systems: the weak convergence of measures. *J. Math. Anal. Appl.* **138** 232-248.

[32] Lasota, A. and Mackey, M. C. (1994) *Chaos, fractals, and noise.* Springer-Verlag, New York.

[33] Lasota, A. and Yorke, J. A. (1973). On the existence of invariant measures for piecewise monotonic transformations. *Trans. Am. Math. Soc.*, **186**, 481-485.

[34] Masry, E. and Fan, J. (1993). Local polynomial estimation of regression functions for mixing process. *Institute of Statistics Mimeo Series #2311,* University of North Carolina at Chapel Hill.

[35] McCaffrey, D. F., Ellner, S., Gallant, D., and Nychka, D. (1992). Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *J. of Americ. Statist. Assoc.*, **87** 682-695.

[36] Nadayara, E. A. (1964). On estimating regression. *Theory of probability and its applications,* **9**, 141-142.

[37] Nadayara, E. A. (1988). *Nonparametric estimation of probability densitics and regression curves.* Mathematics and its applications, Kluwer academic publishers, Dordrecht.

[38] Nummelin, E. (1984). *General irreducible Markov chains and nonnegative operators.* Cambridge: Cambridge University Press.

[39] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals Math. Statist.*, **27, 3,** 832-837.

[40] Pikovsky, A. S. (1970). Discrete-time dynamic noise filtering. *Radio. Eng. Electron. Phys.*, **9**, 81.

[41] Rosenblatt, M. (1991). *Stochastic Curve Estimation.* NSF-CBMS Regional Conference Series in Probability and Statistics, **3.** Institute of Mathematical Statistics, California.

[42] Ruelle, D. (1989). *Chaotic Evolution and Strange Attractors.* London: Cambridge University Press.

[43] Ruppert, D. and Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.*, **22**, 1346-1370.

[44] Sauer, T. (1992). A noise reduction method for signals from nonlinear systems. *Physica, D,* **58** 193-201.

[45] Simonoff, J. S. (1996). *Smoothing methods in statistics.* Springer series in statistics, Springer, New York.

[46] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-620.

[47] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

[48] Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Adv. Appl. Probab.* **22**, 587-611.

[49] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford Univ. Press.

[50] Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective (with discussion). *Scand. J. Statis.* **22**, 399-445.

[51] Tweedie, R. L. (1975). Criteria for classifying general Markov chain. *Adv. Appl. Probab.* **8**, 737-771.

[52] Ulam, S. M. and von Neumann, J. (1945). On the combination of stochastic and deterministic processes. *Bull. Am. Math. Soc.,* **53,** 1120.

[53] Volkonskii, V. A. and Rozanov, Yu. A. (1959). Some limit theorems for random functions. *Theory Prob. Appl.* **4,** 178-197.

[54] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Monographs on statistics and applied probability 60, Chapman and Hall, London.

[55] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā,* Ser. A, **26,** 359-372.

[56] Wolff, R. C. L. (1992). Local Lyapunov exponents: looking closely at chaos. *J. R. Statist. Soc.* **B54** 353-371.

[57] Yao, Q. and Tong, H. (1994a). Quantifying the influence of initial value on nonlinear prediction. *J. R. Statist. Soc.* **B 56,** 701-725

[58] Yao, Q. and Tong, H. (1994b). On subset selection in non-parametric stochastic regression. *Statistica Sinica* **4,** 51-70.

[59] Yonemoto, K. and Yanagawa, T. (1998). Graphical representation of observed chaotic time series and its application. In proceedings of chaos and non-linear time series symposium in Hakata.