ESTIMATING THE EMBEDDING DIMENSION AND DELAY TIME FROM CHAOTIC TIME SERIES WITH DYNAMIC NOISE

Kaoru Fueda* and Takashi Yanagawa**

This paper considers the chaotic time series with additive dynamic noise. Defining the embedding dimension and delay time, we discuss their mathematical properties. A method of estimating the embedding dimension and delay time is proposed, and the consistency of the estimators is proved.

Key words and phrases: Chaotic time series, non-linear autoregressive model, embedding dimension, delay time, cross-validation.

1. Introduction

Appropriate selection of embedding dimension and delay time has been a central issue of chaotic dynamical systems in theoretical physics (see, for example, Abarbanel 1996). Cheng and Tong (1995) considered a non-linear autoregressive model with additive dynamic noise

$$(1.1) \quad X_t = F_d(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + \varepsilon_t$$

and related the intuitive geometric reconstruction of phase space in theoretical physics with statistical theory of the determination of order of a non-linear 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 $X_t = F(X_{t-2}, X_{t-4}) + \varepsilon_t$, $t = 1, 2, \dots$, the embedding dimension is estimated to be 4, that is, we should embed $(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4})$ into 4-dimensional Euclidean space. But we may represent the dynamics of $\{X_t\}$ by embedding it in 2 dimensional space $\{(X_{t-2}, X_{t-4}), t = 1, 2, \dots\}$, 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*. In this paper we introduce the delay time in (1.1), and consider the estimation of embedding dimension and delay time. Our basic model is given in Section 2 where the embedding dimension and delay time are defined and their mathematical properties explored. In Section 3 a method of estimating the embedding dimension and delay

Received January, 2000. Revised June, 2000. Accepted July, 2000.

*Graduate School of Mathematics, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581 Japan. E-mail: fueda@math.kyushu-u.ac.jp

**Graduate School of Mathematics, Kyushu University, Hakozaki 6-10-1, Higashiku, Fukuoka 812-8581, Japan. E-mail: yanagawa@math.kyushu-u.ac.jp

This research is partially supported by Grant-in-Aid for scientific Research (B) No. 11440033 from Japan Society for the Promotion of Science.

time is proposed based on Cross-Validation, a similar technique as Cheng and Tong (1995). Finally consistency of the proposed estimators is proved in Section 4.

2. The embedding dimension and the delay time

We consider the stochastic model given by

$$(2.1) \quad X_t = F(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) + \varepsilon_t$$

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 ,

$$(2.2) \quad E[\varepsilon_t \mid \mathcal{A}_1^{t-1}(X)] = 0, \quad \text{almost surely.}$$

and

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

where $\mathcal{A}_s^t(X)$ denotes the sigma algebra generated by $(X_s, \dots, X_t)$, for $s \leq t$. Note that from (2.1) and (2.2), it follows that

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

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

For simplicity we put

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

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

DEFINITION 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*

$$(2.3) \quad F_d(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) \neq F_{d_0}(X_{t-\tau_0}, X_{t-2\tau_0}, \dots, X_{t-d_0\tau_0}) \text{ a.e.}$$

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

$$(2.4) \quad F_d(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) = F_{d_0}(X_{t-\tau_0}, X_{t-2\tau_0}, \dots, X_{t-d_0\tau_0}) \text{ a.e.}$$

for any $(d, \tau) \in \mathbf{B}(d_0, \tau_0)$,

where $\mathbf{B}(d_0, \tau_0) = \{(d, \tau) \mid \{\tau_0, 2\tau_0, \dots, d_0\tau_0\} \subset \{\tau, 2\tau, \dots, d\tau\}\}$.

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

We have the following theorem.

THEOREM 1. *Suppose that for any $\tau > 0$ there exists $d_0(\tau) < \infty$ such that*

$$(2.5) \quad F_d(X_{t-\tau}, \dots, X_{t-d\tau}) \neq F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}) \text{ a.e.}$$

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

$$(2.6) \quad F_d(X_{t-\tau}, \dots, X_{t-d\tau}) = F_{d_0(\tau)}(X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}) \text{ a.e.}$$

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 (2.6)

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

but this contradicts (2.3).

ii). If $d_0 < d_0(\tau_0)$, we have from (2.5)

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

but since $d_0 < d_0(\tau_0)$ and $(d_0(\tau_0), \tau_0) \in \mathbf{B}(d_0, \tau_0)$, this contradicts (2.4).

Denoting the residuals and their variances by

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

We may show the following lemma.

LEMMA 1.

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

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

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

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

PROOF.

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

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

$$\begin{aligned}
&= \sigma^2(d_1, \tau_1) + \sigma^2(d_2, \tau_2) \\
&\quad - 2(\sigma^2(d_1, \tau_1) + E[\{X_t - F_{d_1}(Z_t^{(d_1, \tau_1)})\} \\
&\quad \quad \{F_{d_1}(Z_t^{(d_1, \tau_1)}) - F_{d_2}(Z_t^{(d_2, \tau_2)})\}]) \\
&= \sigma^2(d_2, \tau_2) - \sigma^2(d_1, \tau_1) \\
&\quad - 2E[\{F_{d_1}(Z_t^{(d_1, \tau_1)}) - F_{d_2}(Z_t^{(d_2, \tau_2)})\} \\
&\quad \quad E[X_t - F_{d_1}(Z_t^{(d_1, \tau_1)}) \mid Z_t^{(d_1, \tau_1)}]] \\
&= \sigma^2(d_2, \tau_2) - \sigma^2(d_1, \tau_1).
\end{aligned}$$

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

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

and from Lemma 1 i) we have

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

From Lemma 1 we have the following theorem.

THEOREM 2. For any $\tau > 0$ and $d_0(\tau)$ defined in Theorem 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

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

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

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

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

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

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

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

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

$$\begin{aligned}
&(\sigma^2(d_0(\tau), \tau) - \sigma^2(d_0(\tau)\tau, 1)) + (\sigma^2(d_0(\tau)\tau, 1) - \sigma^2(d_0\tau_0, 1)) \\
&\quad + (\sigma^2(d_0\tau_0, 1) - \sigma^2(d_0, \tau_0)).
\end{aligned}$$

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

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

- 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

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

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

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

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

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

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

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

3. 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 2. This procedure is based on the cross-validation approach developed by Cheng and Tong (1995) for determining the embedding dimension.

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

Put

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

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

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

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

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

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 $\hat{\tau}_0 = \operatorname{argmin}_{1 \leq \tau \leq T} \hat{d}_0(\tau)$.

THEOREM 3. Under conditions (c), (d) and (f)–(q) which are listed in Section 4.1,

- i) For any $\tau = 1, \dots, T$, $\lim_{N \rightarrow \infty} P\{\hat{d}_0(\tau) = d_0(\tau)\} = 1$,
- ii) $\lim_{N \rightarrow \infty} P\{\hat{\tau}_0 = \tau\} = 1$.

The proof of Theorem 3 is given in the next section.

4. Proof of Theorem 3

4.1. Basic conditions and theorems

We use the following conditions for Theorem 3.

- (a) $E[\varepsilon_t | \mathcal{A}_{-\infty}^{t-1}(X)] = 0$, almost surely.
- (b) $E[\varepsilon_t^2 | \mathcal{A}_{-\infty}^{t-1}(X)] = \sigma^2$, ($\sigma > 0$), almost surely.
- (c) $K_d(u) = \prod_{i=1}^d k(u_i)$ for $u = (u_1, \dots, u_d) \in \mathbf{R}^d$.
- (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

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

- (f) For all $d < D$ and $\tau < T$, let $f_{(d,\tau)}$ denote the probability density function of $(X_{t-\tau}, \dots, X_{t-d\tau})$, which is strictly positive on S , and there exists $c_2 > 0$ such that for all $x, y \in \mathbf{R}^d$, $|f_{(d,\tau)}(x) - f_{(d,\tau)}(y)| \leq c_2 \|x - y\|$.
- (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 $(Z_t^{(d,\tau)}, Z_s^{(d,\tau)}, Z_u^{(d,\tau)}, Z_{t'}^{(d,\tau)}, Z_{s'}^{(d,\tau)}, Z_{u'}^{(d,\tau)})$ 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 $E|F(X_{t-\tau}, \dots, X_{t-d\tau})|^{2p(1+\delta)} < \infty$.
- (j) For δ in condition (i) and some $\varepsilon > 0$, $\beta_j^{\delta/(1+\delta)} = O(j^{-2+\varepsilon})$, where

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

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

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

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

$$\limsup_{N \rightarrow \infty} (i(N) B(N, d)^d) < \infty.$$

- (m) $NB(N, d)^{2d} \rightarrow \infty$ as $N \rightarrow \infty$.
- (n) For μ in assumption (d), $NB(N, d)^{2d+2\mu} \rightarrow 0$ as $N \rightarrow \infty$.
- (o) For q, δ and ε in condition (i) and (j), $N^\varepsilon B(N, d)^{-2d+\theta} \rightarrow 0$ as $N \rightarrow \infty$, where $\theta = 4d/(q + q\delta)$.
- (p) $\{X_t\}$ is stationary, ergodic and bounded.
- (q) For $d > d'$,

$$\frac{B(N, d)^d}{B(N, d')^{d'}} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Conditions (a)–(o) are needed for Theorem 4 and Theorem 5 described below. Note that (a) and (b) are assumed in equation (2.2), and that (e) is derived from (p) in the proof of Theorem 3. We need the following two theorems which can be immediately obtained from Theorem 1 and Theorem 3 in Cheng and Tong (1992) by replacing $(X_{t-1}, \dots, X_{t-d})$ with $(X_{t-\tau}, \dots, X_{t-d\tau})$.

THEOREM 4. *Under conditions (a)–(o),*

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

where

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

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

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

where

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

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

THEOREM 5. Under conditions (a)–(o),

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

where

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

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

4.2. The proof of Theorem 3

To prove part i) of Theorem 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

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

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

$$\frac{1}{B(N, d)^d N} \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

thus from Theorem 4 and Theorem 5,

$$(4.1) \quad CV(d, \tau) = \sigma_N^2(d, \tau) + o_p(1) \quad \text{for any } d.$$

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

$$\begin{aligned}
 (4.2) \quad \sigma_N^2(d, \tau) &= \frac{1}{N - L + 1} \sum_{t=L}^N (\varepsilon_t^{(d, \tau)})^2 W_d(X_{t-\tau}, \dots, X_{t-d\tau}) \\
 &\rightarrow E[(\varepsilon_t^{(d, \tau)})^2 W_d(X_{t-\tau}, \dots, X_{t-d\tau})] \\
 &\quad \text{almost surely as } N \rightarrow \infty \\
 &= E[\varepsilon_t^{(d, \tau)}]^2 \\
 &= \sigma^2(d, \tau).
 \end{aligned}$$

Thus from (4.1) and (4.2), we have

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

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

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

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

$$\begin{aligned}
 \varepsilon_t^{(d, \tau)} &= X_t - E[X_t \mid X_{t-\tau}, \dots, X_{t-d\tau}] = X_t - E[X_t \mid X_{t-\tau}, \dots, X_{t-d_0(\tau)\tau}] \\
 &= \varepsilon_t^{(d_0(\tau), \tau)},
 \end{aligned}$$

and

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

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

$$\begin{aligned}\sigma_N^2(d, \tau) &= \frac{1}{N-L+1} \sum_{t=L}^N (\varepsilon_t^{(d,\tau)})^2 W_d(X_{t-d}, \dots, X_{t-d\tau}) \\ &= \frac{1}{N-L+1} \sum_{t=L}^N (\varepsilon_t^{(d,\tau)})^2 \\ &= \frac{1}{N-L+1} \sum_{t=L}^N (\varepsilon_t^{(d_0(\tau), \tau)})^2.\end{aligned}$$

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

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

implies

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

we have

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

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

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

From Theorem 4 and Theorem 5, we have

$$\begin{aligned}CV(d, \tau) &= \sigma_N^2(d, \tau) \left(1 + \beta(d)\gamma(d) \frac{1}{B(N, d)^d N} + o_p \left(\frac{1}{B(N, d)^d N} \right) \right) \\ CV(d_0(\tau), \tau) &= \sigma_N^2(d_0(\tau), \tau) \left(1 + \beta(d_0(\tau))\gamma(d_0(\tau)) \frac{1}{B(N, d_0(\tau))^{d_0(\tau)} N} \right. \\ &\quad \left. + o_p \left(\frac{1}{B(N, d_0(\tau))^{d_0(\tau)} N} \right) \right),\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{\sigma_N^2(d_0(\tau), \tau)/B(N, d)^d N} (CV(d_0(\tau), \tau) - CV(d, \tau)) \\ &= \beta(d_0(\tau))\gamma(d_0(\tau)) \frac{1/B(N, d_0(\tau))^{d_0(\tau)} N}{1/B(N, d)^d N} - \beta(d)\gamma(d) \\ &\quad + \frac{o_p(1/B(N, d_0(\tau))^{d_0(\tau)} N)}{1/B(N, d)^d N} + \frac{o_p(1/B(N, d)^d N)}{1/B(N, d)^d N}\end{aligned}$$

Thus

$$\begin{aligned}&P\{\hat{d}_0(\tau) = d\} \\ &\leq P\{CV(d_0(\tau), \tau) - CV(d, \tau) \geq 0\} \quad \text{and}\end{aligned}$$

$$\begin{aligned}
& (X_{t-\tau}, \dots, X_{t-d\tau}) \in S_{X_{(d,\tau)}} \quad (\text{for any } t = L, \dots, N) \} \\
& = P \left\{ \frac{1}{\sigma_N^2(d_0(\tau), \tau) / B(N, d)^d} (CV(d_0(\tau), \tau) - CV(d, \tau)) \geq 0 \right. \\
& \quad \left. \text{and } (X_{t-\tau}, \dots, X_{t-d\tau}) \in S_{X_{(d,\tau)}} \text{ (for any } t = L, \dots, N) \right\} \\
& \leq P \left\{ \beta(d_0(\tau)) \gamma(d_0(\tau)) \frac{1/B(N, d_0(\tau))^{d_0(\tau)}}{1/B(N, d)^d} + \frac{o_p(1/B(N, d_0(\tau))^{d_0(\tau)} N)}{1/B(N, d)^d N} \right. \\
& \quad \left. + \frac{o_p(1/B(N, d)^d N)}{1/B(N, d)^d N} \geq \beta(d) \gamma(d) \right\}.
\end{aligned}$$

From assumption (q), we have

$$\begin{aligned}
& \beta(d_0(\tau)) \gamma(d_0(\tau)) \frac{1/B(N, d_0(\tau))^{d_0(\tau)}}{1/B(N, d)^d} + \frac{o_p(1/B(N, d_0(\tau))^{d_0(\tau)} N)}{1/B(N, d)^d N} \\
& + \frac{o_p(1/B(N, d)^d N)}{1/B(N, d)^d N} \rightarrow 0
\end{aligned}$$

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

Next we prove part ii) of Theorem 3. For $\tau > 0$ such that $d_0(\tau) \neq d_0$, we have $d_0(\tau) > d_0(\tau_0)$ from Theorem 1. Thus

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

This completes the proof of Theorem 3.

5. Discussion

One referee commented on the assumption of the boundedness of $\{X_t\}$ which is employed in the proof of Theorem 3. This assumption is reasonable for many chaotic dynamical models of which we are interested. The method in this paper, however, is also applicable to data from other models, including Gaussian AR models. If this is the case, the assumption is too strict and we would like to weaken the assumption in a follow-up paper. Numerical evaluation of the proposed method will also be given in a follow-up paper.

Acknowledgements

The authors would like to thank the referees for their helpful comments.

REFERENCES

- Abarbanel, H. D. I. (1996). *Analysis of Observed Chaotic Data*, Springer, Berlin.
- Cheng, B. and Tong, H. (1992) On consistent nonparametric order determination and chaos. *J. Roy. Statist. Soc., B*, **54**, 427-450.
- Cheng, B. and Tong, H. (1995) Orthogonal projection, embedding demension and sample size in chaotic time series from a statistical perspective, In Tong, H., editor, *Chaos and Forcasting*, World Scientific, Singapore, London.
- 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.