Wavelet Based Estimate for Non-Linear and Non-Stationary Auto-Regressive Model

Fujii, Toru
Graduate School of Mathematics, Kyushu University

Yanagawa, Takashi
Graduate School of Mathematics, Kyushu University

出版情報：MHF Preprint Series. MHF2003-9, 2003-12-31. Faculty of Mathematics, Kyushu University

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Wavelet based estimate for non-linear and non-stationary auto-regressive model

T. Fujii & T. Yanagawa

MHF 2003-9

(Received December 31, 2003)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Wavelet Based Estimate for Non-Linear and Non-Stationary Auto-Regressive Model

Toru FUJII
Graduate School of Mathematics, Kyushu University
Hakozaki 6-10-1 Higashi-Ku Fukuoka 812-8581 Japan
fujii@math.kyushu-u.ac.jp
and
Takashi YANAGAWA
Graduate School of Mathematics, Kyushu University
Hakozaki 6-10-1 Higashi-Ku Fukuoka 812-8581 Japan
yanagawa@math.kyushu-u.ac.jp

Summary:

The wavelet estimator of regression function in a non-linear auto-regressive model for non-stationary time series data is proposed. A convergence theorem of the estimators, and also related theorems, is developed. Cross-Validation criterion is proposed for the optimum selection of parameters. The criterion is justified by the α-mixing condition. Finally, the method is applied to the electroencephalograph (EEG) data.

Key Words:

Alpha-mixing; Cross-validation; Linear wavelet estimator; Smoothing function; Speed of convergence; Time series.

1. INTRODUCTION

We consider the time series $\{X_t\}$ generated by the following auto-regressive model.

$$X_t = \mu(X_{t-1}) + \varepsilon_t, \tag{1}$$

where μ is an unknown non-linear function that is independent of t and $\{\varepsilon_t\}$ are independent random variables which satisfy $E(\varepsilon_t) = 0,\ E(\varepsilon_t^2) < \infty$. Therefore,

$$\mu(x) = E[X_t | X_{t-1} = x].$$

The series has neen considered by many authors under the assumption of stationarity, for example, see Taniguchi and Kakizawa (2000) Section 3.2; and also see Lasoka and Mackey (1989) for a sufficient condition of the stationarity. We do not assume staionarity for $\{X_t\}$ in this paper, but assuming α-mixing condition for $\{X_t\}$, we consider a method to estimate unknown function μ from given data. We deal with regression estimate using wavelets. Wavelets is a mathematical tool to express the function in orthonormal series expansion. Unlike the traditional Fourier series expansion, wavelets method offers a simultaneous localization of the function in space and frequency domain. This enables the multiresolution analysis of function. Wavelet method has been largely applied in sound and image analysis with its usefulness for the detection of edges and singularities. See Chui (1997) for detail.

In statistics, wavelet methods have been introduced in nonparametric estimation by Donoho, Johnston, Kerkyacharian, and Picard (1995, 1996). These papers focused on density estimation or regression estimation for IID Gaussian white noise model, and demonstrated remarkable local adaptivity against discontinuities and spatially varying degree of oscillations in function. This local adaptivity had not been accomplished by kernel estimator or other orthogonal series estimators. Recently, wavelet method is applied in nonparametric regression for time series data with non-linear dependent structure. Truong and Patil (2001) show that the local adaptivity of wavelet estimate can also overcome the sensitivity of the other methods such as kernel-based method, against discontinuities. They focused on wavelet estimate for stationary α-mixing time series generated by model (1).

In this paper we generalize Truong and Patil (2001) to non-stationary multivariate auto-regressive model. It is developed a convergence theorem similar to Truong and Patil (2001). In wavelet estimates, non-linear functions such as "hard thresholding" and "soft thresholding" are generally used for smoothing in addition to linear wavelet estimator. The usefulness of these techniques have been well studied mainly from the view point of local adaptivity over large function classes (See Donoho *et al.* (1995, 1996), Hall and Patil (1996)). We also introduce a simple linear function for smoothing, and show its superiority to the linear wavelet estimator. Furthermore, we propose a criterion similar to Cross-Validation for auto-regressive time series, and show that our criterion is asymptotically unbiased for MSE under α-mixing condition. In Section 2 we review briefly the wavelet expansion based on Chui (1992, 1997) and Daubechies (1992). Wavelet estimator of μ is developed in Section 3. The estimator is generalized in Section 4 for multivariate auto-regressive model. Theorems for the asymptotic convergence of the estimator are given in Section 5. In Section 6 we propose a MSE-based criterion and develop a theorem for its consistency. In Section 7 we apply the method to EEG data by computer simulation.

2. WAVELETS

In this section, we briefly describe the basic concepts of wavelets. Roughly speaking, wavelets is one of the tool to obtain the series expansion of $L_2(\mathcal{R})$ function, which is called the wavelet expansion. It can be said that the main characteristic of wavelets is explained by the "multiresolution analysis" which is not shared by Fourier analysis. For detail, see Chui (1992, 1997), or Daubechies (1992). We start with introducing two orthonormal functions $\phi(x)$ (father wavelet) and $\psi(x)$ (mother wavelet).

2.1 Base Functions

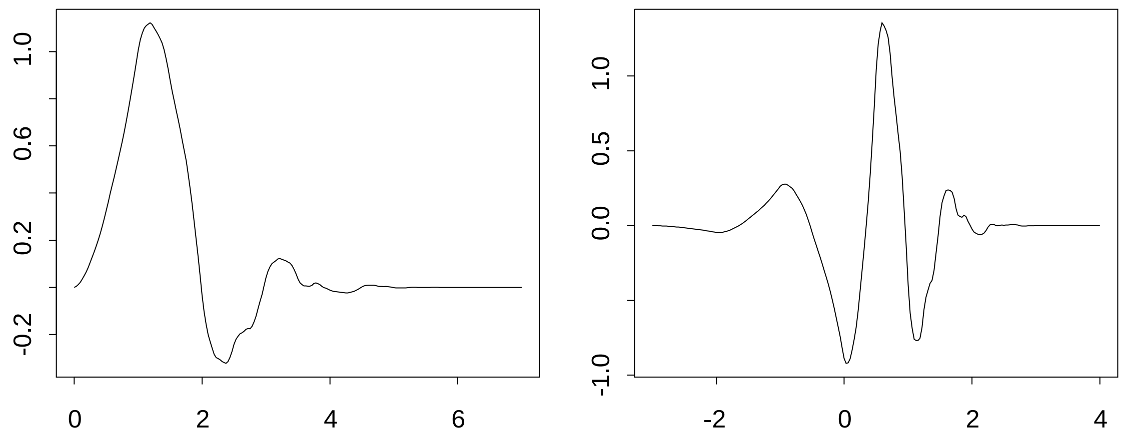

Figure 1: ϕ (father wavelet, left) and ψ (mother wavelet, right)

Definition 1 *Let ϕ be the orthonormal function with compact support on $\mathcal{R}$, which satisfy*

$$\int \phi(x)\,\mathrm{d}x = 1, \qquad \int \phi(x)\phi(x-l)\,\mathrm{d}x = \delta_{0l},$$

and

$$\phi(x) = \sum_{k\in\mathcal{Z}} p_k \phi(2x-k), \tag{2}$$

where $\{p_k\}$ is a finite sequence satisfying $\sum_{k\in\mathcal{Z}} p_k = 2$, $\sum_{k\in\mathcal{Z}} p_k p_{k+2l} = 2\delta_{0l}$ and $\sum_{k\in\mathcal{Z}}(-1)^k p_{1-k} = 0$. ϕ is called the father wavelet.
On the other hand, let define ψ by

$$\psi(x) = \sum_{k\in\mathcal{Z}} (-1)^k p_{1-k} \phi(2x-k), \tag{3}$$

where $\{p_k\}$ is the same sequence as ϕ. ψ is called the mother wavelet.

It follows that ψ has compact support on $\mathcal{R}$ and satisfies the orthonormality

$$\int \psi(x)\,\mathrm{d}x = 0.$$

In addition, if $\sum_{k\in\mathcal{Z}}(-1)^k k^v p_k = 0,\ (1 \le v \le r)$ for some integer $r \ge 1$, then the moment condition $\int x^v \psi(x)\,\mathrm{d}x = 0,\ (1 \le v \le r)$ is satisfied. Generally, a pair of wavelets ϕ, ψ is not represented as specific functions, but their wave patterns are fixed by the peculiar sequence $\{p_k\}$, (2) and (3). For example, Figure 1 shows the wave patterns of $\phi(x)$ and $\psi(x)$ which are called "Daubechies father and mother wavelets" of 3rd order moment condition.

Next, as the translations about scale $j \in \mathcal{Z}$ and shift $k \in \mathcal{Z}$ of ϕ and ψ, define

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

It follows that $\phi_{jk}(x)$ and $\psi_{jk}(x)$ are orthonormal, i.e.

$$\begin{aligned}
\int \phi_{jk}(x)\phi_{jm}(x)\,\mathrm{d}x &= \delta_{km}, \\
\int \psi_{jk}(x)\psi_{lm}(x)\,\mathrm{d}x &= \delta_{jl}\delta_{km}, \\
\int \phi_{jk}(x)\psi_{lm}(x)\,\mathrm{d}x &= 0, \qquad j \leq l.
\end{aligned} \tag{4}$$

2.2 Multiresolution Analysis and Wavelet Expansion

Using $\phi_{jk}(x)$ and $\psi_{jk}(x)$ as base functions, one can construct the following subspaces V_j and W_j in $\mathrm{L}_2(\mathcal{R})$.

$$\begin{aligned}
V_j &= \mathrm{clos}_{\mathrm{L}_2}\langle \phi_{jk}(x)\ ;\ k \in \mathcal{Z}\rangle, \\
W_j &= \mathrm{clos}_{\mathrm{L}_2}\langle \psi_{jk}(x)\ ;\ k \in \mathcal{Z}\rangle, \qquad j \in \mathcal{Z}.
\end{aligned}$$

V_j and W_j are $\mathrm{L}_2(\mathcal{R})$-closures spanned by the linear combinations of $\phi_{jk}(x)$ and $\psi_{jk}(x)$ about the shift parameter $k \in \mathcal{Z}$. The scale parameter $j \in \mathcal{Z}$ represents the resolution level of each subspaces. It is known that $\{V_j\}$ have the following hierarchical structure,

$$\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots \to \mathrm{L}_2(\mathcal{R}),$$

and that $\{W_j\}$ is the orthocomplement, i.e.

$$V_{j+1} = V_j \oplus W_j.$$

From these facts, $\mathrm{L}_2(\mathcal{R})$ can be orthogonally decomposed in countable subspaces with different resolutions based on the level $j_0 \in \mathcal{Z}$ as follows

$$\begin{aligned}\mathrm{L}_2(\mathcal{R}) &= V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \\ &= V_{j_0} \oplus \bigoplus_{j \geq j_0} W_j.\end{aligned}$$

This idea is called MultiResolution Analysis (MRA) of $\mathrm{L}_2(\mathcal{R})$ generated by ϕ and ψ.

From this MRA, any function $f \in \mathrm{L}_2(\mathcal{R})$ can be expressed as a series

$$f(x) = \sum_k \alpha_{j_0 k} \phi_{j_0 k}(x) + \sum_{j \geq j_0} \sum_k \beta_{jk} \psi_{jk}(x). \tag{5}$$

This is called the wavelet expansion of f. From orthonormality (4), each coefficients in (5) are uniquely expressed by the L_2-products of f and ϕ_{jk}, and f and ψ_{jk}, respectively

$$\alpha_{jk} = \int f(x)\, \phi_{jk}(x)\, \mathrm{d}x, \qquad \beta_{jk} = \int f(x)\, \psi_{jk}(x)\, \mathrm{d}x.$$

3. ESTIMATION OF $\mu(x)$

In this section, we describe a wavelet method for estimating the underlying regression function from $\{X_t\}$ given by (1). We develop a new technique to cope with the non-stationality.

3.1 Method for the Non-Stationality

For time series data $\{X_0, X_1, \ldots, X_n\}$ given by model (1), define each p.d.f of X_{t-1} by $f_t(x)$ and let

$$\begin{aligned} f(x) &= \frac{1}{n} \sum_{i=1}^n f_i(x), \\ g(x) &= \mu(x) f(x). \end{aligned}$$

We first estimate f and g, then estimate μ by using relationship $\mu(x) = g(x)/f(x)$.

The wavelet expansion of f is represented as follows,

$$f(x) = \sum_k \alpha^f_{j_0 k} \phi_{j_0 k}(x) + \sum_{j \geq j_0} \sum_k \beta^f_{jk} \psi_{jk}(x),$$

where $\alpha^f_{jk} = \int f(x)\, \phi_{jk}(x)\, \mathrm{d}x$, $\beta^f_{jk} = \int f(x)\, \psi_{jk}(x)\, \mathrm{d}x$. The wavelet expansion of g may be given in the same way, using $\alpha^g_{jk} = \int g(x)\, \phi_{jk}(x)\, \mathrm{d}x$, $\beta^g_{jk} = \int g(x)\, \psi_{jk}(x)\, \mathrm{d}x$.

Put

$$\hat{\alpha}^f_{jk} = \frac{1}{n} \sum_{i=1}^n \phi_{jk}(X_{i-1}), \qquad \hat{\beta}^f_{jk} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_{i-1}). \tag{6}$$

Then

$$\begin{aligned} E[\hat{\alpha}^f_{jk}] &= \frac{1}{n} \sum_{i=1}^n E[\phi_{jk}(X_{i-1})] \\ &= \frac{1}{n} \sum_{i=1}^n \int \phi_{jk}(x)\, f_i(x)\, \mathrm{d}x \\ &= \int \left[\frac{1}{n} \sum_{i=1}^n f_i(x) \right] \phi_{jk}(x)\, \mathrm{d}x \\ &= \int f(x)\, \phi_{jk}(x)\, \mathrm{d}x \\ &= \alpha^f_{jk}, \qquad \text{and} \qquad E[\hat{\beta}^f_{jk}] \;=\; \beta^f_{jk}. \end{aligned}$$

Thus $\hat{\alpha}^f_{jk}$ and $\hat{\beta}^f_{jk}$ are unbiased estimators of the coefficients in wavelet expansion of f. In estimating g, let

$$\hat{\alpha}^g_{jk} = \frac{1}{n} \sum_{i=1}^n X_i \phi_{jk}(X_{i-1}), \qquad \hat{\beta}^g_{jk} = \frac{1}{n} \sum_{i=1}^n X_i \psi_{jk}(X_{i-1}). \tag{7}$$

Then

$$\begin{aligned} E[\hat{\alpha}^g_{jk}] &= \frac{1}{n} \sum_{i=1}^n E[X_i \phi_{jk}(X_{i-1})] \\ &= \frac{1}{n} \sum_{i=1}^n E[E[X_i | X_{i-1}] \phi_{jk}(X_{i-1})] \\ &= \frac{1}{n} \sum_{i=1}^n E[\mu(X_{i-1}) \phi_{jk}(X_{i-1})] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\int \mu(x)\,\phi_{jk}(x)\,f_i(x)\,\mathrm{d}x \\
&= \int \mu(x)\left[\frac{1}{n}\sum_{i=1}^{n} f_i(x)\right]\phi_{jk}(x)\,\mathrm{d}x \\
&= \int \mu(x)\,f(x)\,\phi_{jk}(x)\,\mathrm{d}x \\
&= \int g(x)\,\phi_{jk}(x)\,\mathrm{d}x \\
&= \alpha^g_{jk}, \qquad \text{and} \qquad E[\hat{\beta}^g_{jk}] \;=\; \beta^g_{jk}.
\end{aligned}
$$

Thus $\hat{\alpha}^g_{jk}$ and $\hat{\beta}^g_{jk}$ are also unbiased estimators of α^g_{jk} and β^g_{jk}. Note that the estimators $\hat{\alpha}^f_{jk}$, $\hat{\beta}^f_{jk}$, $\hat{\alpha}^g_{jk}$ and $\hat{\beta}^g_{jk}$ are identical to those estimators given in Truong and Patil (2001) who assumed the stationality for $\{X_t\}$.

3.2 Wavelet Estimator and Smoothing Functions

Using those estimators given in (6) and (7), we define the estimator of f and g by

$$
\begin{aligned}
\hat{f}(x) &= \sum_k \hat{\alpha}^f_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{j_1-1}\sum_k \hat{\beta}^f_{jk}\psi_{jk}(x), \\
\hat{g}(x) &= \sum_k \hat{\alpha}^g_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{j_1-1}\sum_k \hat{\beta}^g_{jk}\psi_{jk}(x),
\end{aligned} \tag{8}
$$

and μ by $\hat{\mu} = \hat{g}/\hat{f}$.

Estimators $\hat{f}$ and $\hat{g}$ have two tuning parameters j_0 and j_1; j_0 is the base resolution level, and j_1 is the truncation parameter. Because the second term of right hand side of equation (8) is truncated at level $j_1 - 1$, the smaller j_1 causes the bias but the larger causes the variance. This type of estimators are called the linear wavelet estimators. Note that there are different types of wavelet estimators.

$$
\begin{aligned}
\hat{f}_\lambda(x) &= \sum_k \hat{\alpha}^f_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{j_1-1}\sum_k \eta_\lambda(\hat{\beta}^f_{jk})\psi_{jk}(x), \\
\hat{g}_\lambda(x) &= \sum_k \hat{\alpha}^g_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{j_1-1}\sum_k \eta_\lambda(\hat{\beta}^g_{jk})\psi_{jk}(x),
\end{aligned} \tag{9}
$$

where $\eta_\lambda(\cdot)$ is called the smoothing function. Generally, the following functions are used for smoothing functions.

$$\eta_\lambda^h(t) \;=\; t\,\mathrm{I}\{|\,t\,| > \lambda\},$$

is the hard thresholding and

$$\eta_\lambda^s(t) \;=\; \mathrm{sgn}(t)\,(|\,t\,| - \lambda)\,\mathrm{I}\{|\,t\,| > \lambda\},$$

is the soft thresholding, where $\lambda > 0$ is a smoothing parameter.

The estimator with above non-linear $\eta_\lambda(\cdot)$ is called the threshold wavelet estimator or the nonlinear wavelet estimator. This type of estimator can control the balance of bias and variance of (9) by tuning the parameters j_1 and λ.

As for smoothing function, we introduce

$$\eta_\lambda(t) = \lambda t, \quad \lambda \in \mathcal{R},$$

in this paper. It speed-ups the computation of such algorithm as CV, and also has such nice property as will be shown in Theorem 1.

4. MULTIVARIATE MODEL

In the foregoing sections, we considered univariate model for simplicity. In this section, we extend the model to multivariate auto-regressive models.

4.1 Extention of the Model

We consider the following model.

$$X_t = \mu(X_{t-d}, \ldots, X_{t-1}) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are independent random variables, and

$$\mu(x_1, \ldots, x_d) = E[X_t | X_{t-d} = x_1, \ldots, X_{t-1} = x_d].$$

Let $f_t(x_1,\ldots,x_d)$ be the joint p.d.f of $(X_{t-d},\ldots,X_{t-1})$ and put

$$\begin{aligned} f(x_1,\ldots,x_d) &= \frac{1}{n}\sum_{i=1}^{n} f_i(x_1,\ldots,x_d), \\ g(x_1,\ldots,x_d) &= \mu(x_1,\ldots,x_d)f(x_1,\ldots,x_d). \end{aligned}$$

4.2 Estimation of f, g and μ by Multivariate Wavelets

The wavelet expansions of f and g are represented by the tensor product of univariate case. We use the notation $\boldsymbol{x}$ for $(x_1,\cdots,x_d)$ and $\int_{\mathcal{R}^d}\mathbf{d}\boldsymbol{x}$ for $\int_{\mathcal{R}^d}\mathrm{d}x_1\cdots\mathrm{d}x_d$, for simplicity.

The wavelet expansion of f is

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{Z}^d}\alpha^f_{j_0\boldsymbol{k}}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j\ge j_0}\sum_{\boldsymbol{k}\in\mathcal{Z}^d}\sum_{\boldsymbol{i}\in\mathcal{I}^d}\beta^f_{j\boldsymbol{k}\boldsymbol{i}}\Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \tag{10}$$

where $\boldsymbol{k}=(k_1,\ldots,k_d)$, $\mathcal{I}^d=\{\boldsymbol{i}=(i_1,\ldots,i_d)\ :\ i_s=0 \text{ or } 1\ (s=1,\ldots,d),\ \text{and}\ \ i_1+\cdots+i_d\ge 1\}$, and

$$\begin{aligned} \Phi_{j\boldsymbol{k}}(\boldsymbol{x}) &= \phi_{jk_1}(x_1)\cdots\phi_{jk_d}(x_d), \\ \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}) &= \chi^{i_1}_{jk_1}(x_1)\cdots\chi^{i_d}_{jk_d}(x_d). \end{aligned} \tag{11}$$

In (11), we used

$$\chi^{i_s}_{jk_s}(\cdot) = \begin{cases} \phi_{jk_s}(\cdot) & \text{if}\quad i_s=0, \\ \psi_{jk_s}(\cdot) & \text{if}\quad i_s=1, \end{cases} \qquad s=1,\ldots,d.$$

It follows from the following orthonormalities

$$\begin{aligned} \int_{\mathcal{R}^d}\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\Phi_{j\boldsymbol{m}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x} &= \prod_{s=1}^{d}\int\phi_{jk_s}(x_s)\phi_{jm_s}(x_s)\,\mathrm{d}x_s \\ &= \delta_{\boldsymbol{k}\boldsymbol{m}}, \\ \int_{\mathcal{R}^d}\Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x})\,\Psi_{l\boldsymbol{m}\boldsymbol{i}'}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x} &= \prod_{s=1}^{d}\int\chi^{i_s}_{jk_s}(x_s)\chi^{i'_s}_{lm_s}(x_s)\,\mathrm{d}x_s \\ &= \delta_{jl}\,\delta_{\boldsymbol{k}\boldsymbol{m}}\,\delta_{\boldsymbol{i}\boldsymbol{i}'}, \\ \int_{\mathcal{R}^d}\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\Psi_{l\boldsymbol{m}\boldsymbol{i}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x} &= \prod_{s=1}^{d}\int\phi_{jk_s}(x_s)\chi^{i_s}_{lm_s}(x_s)\,\mathrm{d}x_s \\ &= 0, \qquad j\le l, \end{aligned} \tag{12}$$

that the coefficients of (10) are represented as

$$\alpha^f_{j\boldsymbol{k}} \;=\; \int_{\mathcal{R}^d} f(\boldsymbol{x})\,\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}, \quad \beta^f_{j\boldsymbol{k}\boldsymbol{i}} \;=\; \int_{\mathcal{R}^d} f(\boldsymbol{x})\,\Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}.$$

The wavelets expansion of g is similarly given.

It is easy to show that

$$\begin{aligned}\hat{\alpha}^f_{j\boldsymbol{k}} &= \frac{1}{n}\sum_{i=1}^{n}\Phi_{j\boldsymbol{k}}(X_{i-d},\ldots,X_{i-1}),\\ \hat{\beta}^f_{j\boldsymbol{k}\boldsymbol{i}} &= \frac{1}{n}\sum_{i=1}^{n}\Psi_{j\boldsymbol{k}\boldsymbol{i}}(X_{i-d},\ldots,X_{i-1}),\end{aligned} \tag{13}$$

are the unbiased estimators of $\alpha^f_{j\boldsymbol{k}}$ and $\beta^f_{j\boldsymbol{k}\boldsymbol{i}}$.

Also

$$\begin{aligned}\hat{\alpha}^g_{j\boldsymbol{k}} &= \frac{1}{n}\sum_{i=1}^{n}X_i\Phi_{j\boldsymbol{k}}(X_{i-d},\ldots,X_{i-1}),\\ \hat{\beta}^g_{j\boldsymbol{k}\boldsymbol{i}} &= \frac{1}{n}\sum_{i=1}^{n}X_i\Psi_{j\boldsymbol{k}\boldsymbol{i}}(X_{i-d},\ldots,X_{i-1}),\end{aligned} \tag{14}$$

are unbiased estimators of $\alpha^g_{j\boldsymbol{k}}$ and $\beta^g_{j\boldsymbol{k}\boldsymbol{i}}$. For example, the unbiasedness of $\hat{\alpha}^g_{j\boldsymbol{k}}$ is shown as follows.

$$\begin{aligned}E[\hat{\alpha}^g_{j\boldsymbol{k}}] &= \frac{1}{n}\sum_{i=1}^{n}E\Big[X_i\Phi_{j\boldsymbol{k}}(X_{i-d},\ldots,X_{i-1})\Big]\\ &= \frac{1}{n}\sum_{i=1}^{n}E\Big[E[X_i|X_{i-d},\ldots,X_{i-1}]\Phi_{j\boldsymbol{k}}(X_{i-d},\ldots,X_{i-1})\Big]\\ &= \frac{1}{n}\sum_{i=1}^{n}E\Big[\mu(X_{i-d},\ldots,X_{i-1})\Phi_{j\boldsymbol{k}}(X_{i-d},\ldots,X_{i-1})\Big]\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{R}^d}\mu(\boldsymbol{x})\,\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\ &= \int_{\mathcal{R}^d}\mu(\boldsymbol{x})\left[\frac{1}{n}\sum_{i=1}^{n}f_i(\boldsymbol{x})\right]\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\ &= \int_{\mathcal{R}^d}\mu(\boldsymbol{x})\,f(\boldsymbol{x})\,\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\ &= \int_{\mathcal{R}^d}g(\boldsymbol{x})\,\Phi_{j\boldsymbol{k}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\ &= \alpha^g_{j\boldsymbol{k}}.\end{aligned}$$

Using (13), (14), the linear wavelet estimators of f are g are given as follows;

$$
\begin{aligned}
\hat{f}(\boldsymbol{x}) &= \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \hat{\alpha}^f_{j_0\boldsymbol{k}} \Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) \\
&\quad + \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \hat{\beta}^f_{j\boldsymbol{k}\boldsymbol{i}} \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \\
\hat{g}(\boldsymbol{x}) &= \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \hat{\alpha}^g_{j_0\boldsymbol{k}} \Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) \\
&\quad + \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \hat{\beta}^g_{j\boldsymbol{k}\boldsymbol{i}} \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \qquad (15)
\end{aligned}
$$

Also the threshold wavelet estimators are given as follows;

$$
\begin{aligned}
\hat{f}_\lambda(\boldsymbol{x}) &= \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \hat{\alpha}^f_{j_0\boldsymbol{k}} \Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) \\
&\quad + \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \eta_\lambda(\hat{\beta}^f_{j\boldsymbol{k}\boldsymbol{i}}) \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \\
\hat{g}_\lambda(\boldsymbol{x}) &= \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \hat{\alpha}^g_{j_0\boldsymbol{k}} \Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) \\
&\quad + \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\mathcal{Z}^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \eta_\lambda(\hat{\beta}^g_{j\boldsymbol{k}\boldsymbol{i}}) \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}). \qquad (16)
\end{aligned}
$$

5. ASYMPTOTIC PROPERTY

In this section, we consider the convergence of multivariate wavelet estimators for non-stationary time series that satisfy α-mixing condition. In addition, we consider the behavior of the smoothing function $\eta_\lambda(t) = \lambda\, t$ introduced in Section 3.. For the independent observations, a lot of studies have been conducted on asymptotic properties of wavelet estimator by Donoho *et al.* (1995, 1996), Hall and Patil (1996) among others. For the stationary time series, Truong and Patil (2001) have derived the convergence rate of wavelet estimator under the α-mixing settings. We derive a similar rate for our estimators under non-stationarity.

To begin with we define α-mixing condition for d variables case.

Definition 2 *Time series* $\{X_t\}$ *is* α-mixing *when*

$$\lim_{u\to\infty} \alpha(u) = 0,$$

where

$$\alpha(u) = \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_j \text{ and } B \in \mathcal{F}^{j+u}\},$$

where $\mathcal{F}_j$ is the σ-algebra generated by $(X_{i-d}, \ldots, X_i),\ -\infty < i \le j$ and $\mathcal{F}^j$ is the σ-algebra generated by $(X_{i-d}, \ldots, X_i),\ j \le i < \infty$.

The next lemma is useful for studying the α-mixing time series.

Lemma 1 *Let* X *be a* $\mathcal{F}_j$-measurable *random variable, and* Y *be* $\mathcal{F}^{j+u}$-measurable. *And* $X,\ Y$ *are bounded satisfying* $|X| \le C_1,\ |Y| \le C_2$, *respectively. Then,*

$$|\mathrm{Cov}(X,Y)| \ \le\ 4\ C_1 C_2\ \alpha(u).$$

The proof of Lemma 1 and details about mixing conditions are given in Hall and Heyde (1980) and Bosq (1998) among others.

We assume the following conditions.

(a) μ and f_t $(t \in \mathcal{Z})$ belong to C^r-class and their derivatives up to rth order are all bounded and continuous on $\mathcal{R}^d$.

(b) There exists bounded and closed interval $Q \subset \mathcal{R}^d$ such that $\mathrm{supp}(f_t) \subset Q$ for all $t \in \mathcal{Z}$, where $\mathrm{supp}(f_t)$ is the support of f_t.

(c) There exists a subset $U \subset Q$ and a positive constant M such that for all $t \in \mathcal{Z}$, $M^{-1} \le f_t(x) \le M$ for $x \in U$.

(d) ϕ and ψ have compact support on $\mathcal{R}$.

(e) It holds that $\int x^v \psi(x)\,\mathrm{d}x = 0$ $(0 \le v \le r-1)$ and $(\,r!\,)^{-1} \int x^r \psi(x)\,\mathrm{d}x < \infty$.

(f) $\alpha(u) = O(\rho^u)$ as $u \to \infty$ for some $0 < \rho < 1$.

Theorem 1 *Assume the α-mixing condition and conditions (a), (b), (c), (d), (e) and (f). Furthermore, suppose $j_0,\ j_1 \to \infty$ as $n \to \infty$ in such a manner that $2^{dj_0},\ 2^{dj_1} = o(n)$. Then*

(i) *Linear wavelet estimator $\hat{g}$ of g given in (15) satisfies,*

$$\begin{aligned} E\Big[\int_{\mathcal{R}^d} (\hat{g}-g)^2\,\mathbf{d}\boldsymbol{x}\Big] &= O\Big(\frac{2^{dj_0}}{n}+\frac{2^{dj_1}}{n}+2^{-2j_1 r}\Big) \\ &= o(1) \text{ as } n\to\infty. \end{aligned}$$

(ii) *If $2^{dj_0},\ 2^{dj_1} = o(n^{1/3})$, then estimator $\hat{\mu}=\hat{g}/\hat{f}$ of μ satisfies,*

$$\int_U (\hat{\mu}-\mu)^2\,\mathbf{d}\boldsymbol{x} = o_p(1).$$

(iii) *Moreover, for smoothing function $\eta_\lambda(t)=\lambda t$, there exists some $\lambda\in\mathcal{R}$ such that*

$$\int_{\mathcal{R}^d} (\hat{g}-g)^2\,\mathbf{d}\boldsymbol{x} > \int_{\mathcal{R}^d} (\hat{g}_\lambda-g)^2\,\mathbf{d}\boldsymbol{x},$$

for arbiterary j_0 and j_1.

proof). Proof is given in Appendix.

Theorem 2 *Assume the α-mixing condition and conditions (a), (b), (c), (d), (e) and (f). Furthermore, suppose that $d<r$ and $j_0,\ j_1\to\infty$ as $n\to\infty$ in such a manner that $2^{(d-r)j_0},\ 2^{(d-r)j_1} = o(1)$. Then as $n\to\infty$ we have for $\boldsymbol{x}\in U$*

$$\begin{aligned} E\Big(\hat{f}(\boldsymbol{x})\Big)-f(\boldsymbol{x}) &= O\Big(2^{(d-r)j_0}\Big)=o(1), \\ E\Big(\hat{g}(\boldsymbol{x})\Big)-g(\boldsymbol{x}) &= O\Big(2^{(d-r)j_1}\Big)=o(1), \\ E\Big(\mu(\hat{\boldsymbol{x}})\Big)-\mu(\boldsymbol{x}) &= o(1). \end{aligned}$$

proof). Proof is given in Appendix.

6. SELECTION OF PARAMETERS

In this section, we discuss a criterion for choosing the optimum $\theta=(j_0,j_1,\lambda)$ involved in the estimator by assuming α-mixing condition.

6.1 Basic Criterion

At first, we assume auto-regressive model $X_t = \mu(\boldsymbol{X}_{t-1}) + \varepsilon_t$ where $\boldsymbol{X}_{t-1}$ denote a vector of random variables $(X_{t-d}, \ldots, X_{t-1})$, and consider mean squared error

$$\mathrm{R}(\theta) \;=\; \frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})]^2,$$

where $\hat{\mu}_\theta(\,\cdot\,)$ is the estimator of μ with $\theta = (j_0, j_1, \lambda) \in \Theta$ based on given data $\{(X_1, \boldsymbol{X}_0), \ldots, (X_n, \boldsymbol{X}_{n-1})\}$. Define the residual sum of squares by

$$\mathrm{RSS}(\theta) \;=\; \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})]^2,$$

then we have the following lemma.

Lemma 2

$$E[\mathrm{RSS}(\theta)] \;=\; \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \mathrm{R}(\theta) - \frac{2}{n}\sum_{i=1}^{n}\mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\}, \tag{17}$$

where σ_i^2 denote the variance of ε_i.

proof).

$$\begin{aligned}
E[\mathrm{RSS}(\theta)] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) + \varepsilon_i - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E[\varepsilon_i^2 + \{\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\}^2 \\
&\quad + 2\varepsilon_i\{\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\}] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \mathrm{R}(\theta) + \frac{2}{n}\sum_{i=1}^{n} E\{\varepsilon_i\mu(\boldsymbol{X}_{i-1})\} \\
&\quad - \frac{2}{n}\sum_{i=1}^{n} E\{\varepsilon_i\hat{\mu}_\theta(\boldsymbol{X}_{i-1})\}.
\end{aligned}$$

Since ε_i and $\boldsymbol{X}_{i-1}$ are independent, and $\hat{\mu}_\theta(\boldsymbol{X}_{i-1})$ is the random variable of $(X_{1-d}, \ldots,$ $X_n)$, it hold that $E\{\varepsilon_i\mu(\boldsymbol{X}_{i-1})\} = 0$ and $E\{\varepsilon_i\hat{\mu}_\theta(\boldsymbol{X}_{i-1})\} = \mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\}$. Note that, the first term in the right hand side of (17) is independent of parameter θ, however the third term is not, so $\arg\min_\theta\{\mathrm{RSS}(\theta)\}$ may not coincide with $\arg\min_\theta\{\mathrm{R}(\theta)\}$. This is because the time series data has dependent structure.

6.2 A Criterion for α-mixing Time Series

We develop a criterion whose miminum argument agree with an $\arg\min_\theta\{\mathrm{R}(\theta)\}$ under α-mixing condition.

Let $\hat{\mu}_\theta^{\setminus(i,M_n)}(\,\cdot\,)$ be the estimator from the data which delete M_n data points $\{(X_i, \boldsymbol{X}_{i-1}), \ldots, (X_{i+M_n-1}, \boldsymbol{X}_{i+M_n-2})\}$ from an original data $\{(X_1, \boldsymbol{X}_0), \ldots, (X_n, \boldsymbol{X}_{n-1})\}$ for each i, and let

$$\mathrm{CV}_{M_n}(\theta) \;=\; \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{\mu}_\theta^{\setminus(i,M_n)}(\boldsymbol{X}_{i-1})]^2.$$

We have the following theorem.

Theorem 3 *Assume* $\{X_t\}$ *be the* α-mixing *time series generated by model* $X_t = \mu(\boldsymbol{X}_{t-1}) + \varepsilon_t$*. Set* $M_n = o(n)$ as $n \to \infty$*, then*

$$E[\mathrm{CV}_{M_n}(\theta)] \;=\; \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \mathrm{R}(\theta) + o(1).$$

proof). Proof is given in Appendix.
It follows from Theorem 3 that

$$\arg\min_\theta\{\mathrm{CV}_{M_n}(\theta)\} \;\longrightarrow\; \arg\min_\theta\{\mathrm{R}(\theta)\} \text{ as } n \to \infty,$$

thus we may use $\mathrm{CV}_{M_n}(\theta)$ for selecting the optimum parameters.

7. APPLICATION TO EEG DATA

In this section, we apply the wavelet estimation method to a real time series data. We treat here a time series of EEG(human brain waves) which is recorded from an 11 years old female patient suffering from epilepsy. Figure 2 and Figure 3 show the 631 data points $(X_1, \ldots, X_{631})$ sampled by 200 points per second frequency.

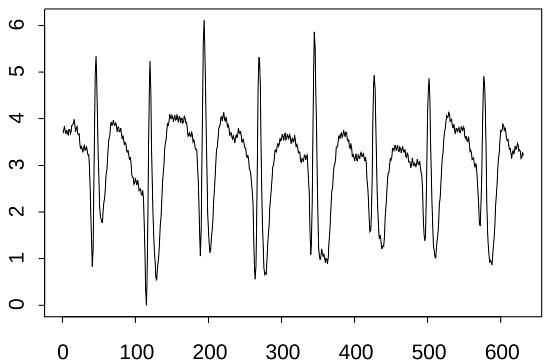

Figure 2: Original data

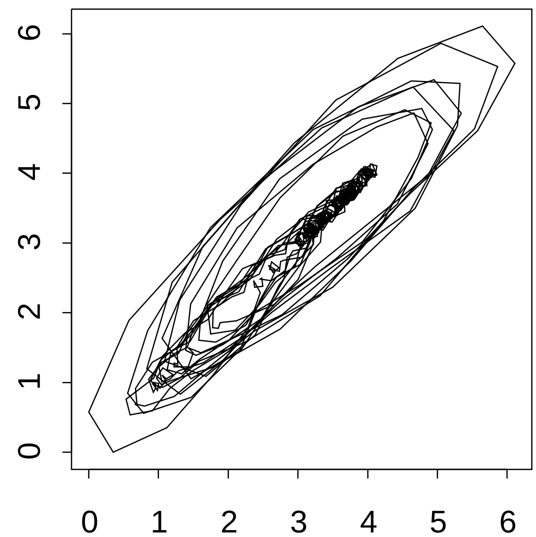

Figure 3: Lorenz plot of original data

7.1 EEG Data and Non-Parametric Models

It is well-known that EEG data have non-linear probability structure. In particular, EEG recorded from an epilepsy patient is characterized by the Spike and Wave (SW) pattern as be seen in Figure 2 and suggested to have strong non-linearlity (see Theiler (1995)). In order to search for the deterministic structure of EEG, Miwakeichi, Ramirez-Padron, Valdes-Sosa, and Ozaki (2001) applied nonparametric models based on Nadaraya-Watson, Local-Linear regression and other kernel based regression methods.

We start with assuming the model (10) for the EEG data. Yanagawa and Yonemoto estimated $d = 2$ from the data and we assume in the sequel that $d = 2$.

7.2 Estimation by 2 Variables Wavelets

We assume

$$X_t = \mu(X_{t-2}, X_{t-1}) + \varepsilon_t,$$

and estimate $\mu(x,y)$ from $\{X_3,(X_1,X_2)\}$,. . .,$\{X_{631},(X_{629},X_{630})\}$. $\hat{f}$ and $\hat{g}$ in (16) are represented as follows,

$$\begin{aligned}
\hat{f}_\theta(x,y) &= \sum_{k_1k_2} \hat{\alpha}^f_{j_0(k_1,k_2)}\phi_{j_0k_1}(x)\phi_{j_0k_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^f_{j(k_1,k_2)(0,1)}\Big)\phi_{jk_1}(x)\psi_{jk_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^f_{j(k_1,k_2)(1,0)}\Big)\psi_{jk_1}(x)\phi_{jk_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^f_{j(k_1,k_2)(1,1)}\Big)\psi_{jk_1}(x)\psi_{jk_2}(y),
\end{aligned}$$

$$\begin{aligned}
\hat{g}_\theta(x,y) &= \sum_{k_1k_2} \hat{\alpha}^g_{j_0(k_1,k_2)}\phi_{j_0k_1}(x)\phi_{j_0k_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^g_{j(k_1,k_2)(0,1)}\Big)\phi_{jk_1}(x)\psi_{jk_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^g_{j(k_1,k_2)(1,0)}\Big)\psi_{jk_1}(x)\phi_{jk_2}(y) \\
&\quad + \sum_{j=j_0}^{j_1-1}\sum_{k_1k_2}\eta_\lambda\Big(\hat{\beta}^g_{j(k_1,k_2)(1,1)}\Big)\psi_{jk_1}(x)\psi_{jk_2}(y),
\end{aligned}$$

where $\theta = (j_0, j_1, \lambda)$.

The values of $\hat{f}$ may be close to zero, and we suggest to introduce sufficiently small $\delta > 0$ and compute the estimator as follows,

$$\hat{\mu}_\theta(x,y) = \begin{cases} \hat{g}_\theta(x,y)/\hat{f}_\theta(x,y) & \text{if } \quad |\hat{f}_\theta(x,y)| > \delta, \\ 0 & \text{if } \quad |\hat{f}_\theta(x,y)| \leq \delta, \end{cases}$$

7.3 Cross-Validation

For the estimation of the parameter $\theta = (j_0, j_1, \lambda)$ when $\eta_\lambda(t) = \lambda\, t$, we compute the value of

$$\mathrm{CV}(\eta,\theta) \;=\; \frac{1}{629}\sum_{i=3}^{631}[X_i - \hat{\mu}^{\setminus i}_{\eta\theta}(X_{i-2},X_{i-1})]^2.$$

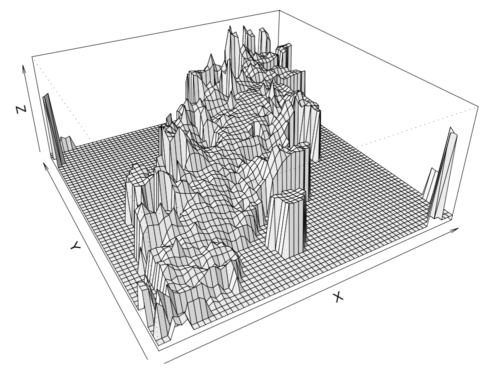

Figure 4: Perspective plots of $\hat{\mu}(x, y)$

for combinations of j_0, j_1 and λ and select the θ which attains the minimum of CV. To save the computational time we set $j_1 = 6$ in this simulation. The results are given in Table 1. The table also contain the value of CV for hard and soft thresholdings when $j_1 = 6$. The Haar wavelets are employed as the base. From the table j_0 and λ of our estimator are estimated to be $j_0 = 2$ and $\lambda = 0.8$ with the value of CV 0.138. On the other hand j_0 and λ for hard and soft thresholdings are estimated to be $j_0 = 2$, $\lambda = 0.3$ (or $\lambda = 0.4$), and $j_0 = 2$, $\lambda = 0.3$ with the values of CV 0.197 and 0.202, respectively. The comparison shows that the CV values of our estimator is smallest, indicating the superiority of our estimator to the others considered in the table.

Figure 4 is the graph of $\{\hat{\mu}_\theta(x, y), (x, y)\}$ obtained when $\delta = 0.001$.

Table 1: Cross-Validation ($j_1 = 6$)

	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	hard	2.367	1.425	0.318	1.440	0.480	16.631	16.181	0.383	0.383
$j_0 = 1$	soft	0.540	68.820	0.429	0.405	0.427	0.395	0.386	0.386	0.549
	new	0.277	0.260	0.253	0.250	0.249	0.248	0.248	0.248	0.248
	hard	2.714	1.314	0.197	0.197	0.234	3.320	1.528	0.210	0.210
$j_0 = 2$	soft	0.993	0.215	0.202	0.205	1.923	0.329	0.219	0.213	0.213
	new	0.166	0.152	0.146	0.143	0.140	0.139	0.139	0.138	0.138
	hard	6.640	1.431	0.288	0.288	0.286	1.114	0.282	0.282	0.282
$j_0 = 3$	soft	0.355	0.322	0.283	0.281	0.281	0.284	0.282	0.282	0.282
	new	0.266	0.259	0.255	0.252	0.250	0.249	0.248	0.248	0.248
	hard	3.905	1.404	5.168	2.894	0.789	0.756	1.007	0.877	0.756
$j_0 = 4$	soft	3.778	0.815	0.785	0.759	0.758	0.759	0.760	0.756	0.756
	new	0.753	0.752	0.751	0.750	0.750	0.750	0.750	0.750	0.751
	hard	1.904	3.175	1.704	1.884	2.218	2.444	3.795	1.517	1.517
$j_0 = 5$	soft	1.635	1.533	1.528	1.528	1.532	1.524	1.518	1.517	1.517
	new	1.517	1.517	1.517	1.517	1.517	1.517	1.517	1.518	1.518

7.4 Reproduction of the New Time Series

To assure our estimator we generated a new time series $\{Y_t\}$ using $\hat{\mu}(x, y)$ as follows,

$$\begin{aligned} Y_3 &= \hat{\mu}(Y_1, Y_2) \\ Y_4 &= \hat{\mu}(Y_2, Y_3) \\ &\vdots \end{aligned}$$

by giving an appropriate initial value for (Y_1, Y_2). Figure 5 gives the plot of $\{Y_t\}$ and Figure 6 is the Lorentz plot. Figures show that Figure 2 and 3 are well reproduced by our method. Furthermore, Figure 5 shows that the present data may be characterized as a repetition of the same dynamics.

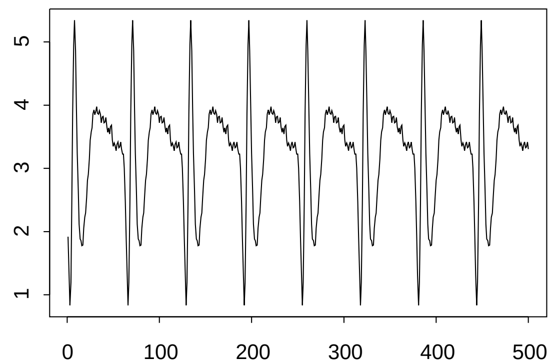

Figure 5: $\{Y_t\}$ generated by $\hat{\mu}(x, y)$

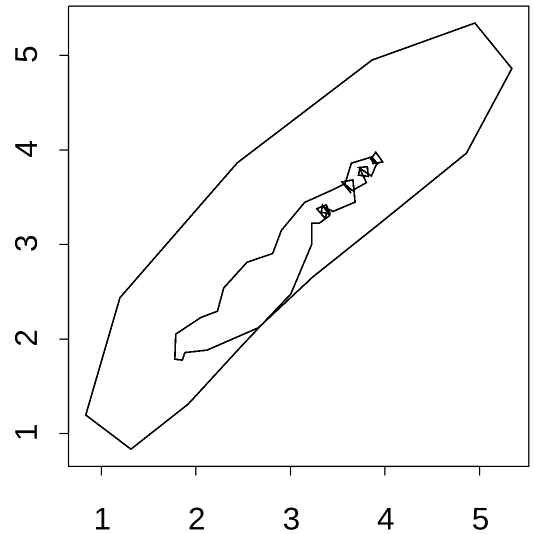

Figure 6: Lorenz plot of $\{Y_t\}$

8. CONCLUSION

In this paper we generalized the wavelet estimator of regression function considered by Truong and Patil (2001) to non-stationary time series in a multivariate autoregressive model. Assuming α-mixing condition for the time series, we developed the convergence theorem. In addition, we introduced a new smoothing function $\eta_\lambda(t) = \lambda\, t$ in the context of thresholding techniques for the wavelet estimate and showed its superiority to the linear wavelet estimator mathematically. For the optimum selection of parameters, we proposed to use a CV-criterion, and gave its relationship to the MSE-based criterion. The method was applied to EEG data. It is indicated that the method captured successfully the non-linear structure of EEG data. Although the new smoothing function was introduced to improve the integrated squared error of estimating g, the inspection of the table indicates that it also improve the value of Cross-Validation of hard and soft thresholding function in estimating μ.

APPENDIX: PROOFS

Proof of Theorem 1 (i)

Without loss of generality, we assume $\mathrm{supp}(\phi),\ \mathrm{supp}(\psi) = [0,1]$ and $Q = [0,1]^d$, for simplicity. Under these settings, it is sufficient to give the proof for the order of $E[\int_Q (\hat{g} - g)^2 \, \mathbf{d}\boldsymbol{x}]$. In the following, we use the notation $\boldsymbol{X}_{i-1}$ for $(X_{i-d}, \ldots, X_{i-1})$, $\boldsymbol{x}$ for $(x_1, \cdots, x_d)$ and $\int_{\mathcal{R}^d} \mathbf{d}\boldsymbol{x}$ for $\int_{\mathcal{R}^d} \mathrm{d}x_1 \cdots \mathrm{d}x_d$.

At first, define

$$\Omega_j^d = \{\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathcal{Z}^d \ : \ 0 \le k_1, \ldots, k_d \le 2^j - 1\}.$$

Then, it follows from $\mathrm{supp}(\phi),\ \mathrm{supp}(\psi) = [0,1]$ that

$$\begin{aligned} \hat{g}(\boldsymbol{x}) - g(\boldsymbol{x}) \quad &= \sum_{\boldsymbol{k} \in \Omega_{j_0}^d} (\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}}) \Phi_{j_0\boldsymbol{k}}(\boldsymbol{x}) \\ &\quad + \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} (\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}}) \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}) \\ &\quad - \sum_{j \ge j_1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \beta_{j\boldsymbol{k}\boldsymbol{i}} \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \qquad \boldsymbol{x} \in Q. \end{aligned} \tag{A.1}$$

According to the orthonormality of wavelets (12),

$$\begin{aligned} \int_Q (\hat{g} - g)^2 \, \mathbf{d}\boldsymbol{x} \quad &= \quad S_1 + S_2 + S_3, \\ S_1 \quad &= \quad \sum_{\boldsymbol{k} \in \Omega_{j_0}^d} (\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}})^2, \\ S_2 \quad &= \quad \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} (\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}})^2, \\ S_3 \quad &= \quad \sum_{j \ge j_1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \beta_{j\boldsymbol{k}\boldsymbol{i}}^2. \end{aligned} \tag{A.2}$$

We evaluate $E(S_1),\ E(S_2),\ E(S_3)$ in (A.2) respectively. It follows from

$$E\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\} \;=\; \int_{\mathcal{R}^d}\mu(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}$$

that

$$\begin{aligned}
E(\hat\alpha_{j_0\boldsymbol{k}}-\alpha_{j_0\boldsymbol{k}})^2 &= E\bigg[\frac{1}{n}\sum_{i=1}^n X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1}) - \frac{1}{n}\sum_{i=1}^n\int_{\mathcal{R}^d}\mu(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\bigg]^2\\
&= \frac{1}{n^2}E\bigg[\sum_{i=1}^n\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1}) - \int_{\mathcal{R}^d}\mu(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\Big\}\bigg]^2\\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}\\
&\quad + \frac{2}{n^2}\sum_{i<j}\mathrm{Cov}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1}),\ X_j\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}. \qquad \text{(A.3)}
\end{aligned}$$

First, we evaluate the Variance term of (A.3);

$$\mathrm{Var}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\} = E\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}^2 - \Big[E\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}\Big]^2, \qquad \text{(A.4)}$$

where

$$\begin{aligned}
E\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\} &= \int_{\mathcal{R}^d}\mu(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\
&= 2^{-dj_0/2}\int_{\mathcal{R}^d}\mu\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_0}}\Big)\phi(x_1)\cdots\phi(x_d)f_i\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_0}}\Big)\,\mathbf{d}\boldsymbol{x}\\
&= O(2^{-dj_0/2}) \qquad \text{(A.5)}
\end{aligned}$$

and

$$\begin{aligned}
E\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}^2 &\le \text{Const.}\,E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}^2\\
&= \text{Const.}\int_{\mathcal{R}^d}\Phi^2_{j_0\boldsymbol{k}}(\boldsymbol{x})f_i(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\\
&= \text{Const.}\int_{\mathcal{R}^d}\Big\{\phi(x_1)\cdots\phi(x_d)\Big\}^2 f_i\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_0}}\Big)\,\mathbf{d}\boldsymbol{x}\\
&= O(1). \qquad \text{(A.6)}
\end{aligned}$$

Note that we used the boundness and the support compactness of $\mu,\ f_i,\ X_i$ in the above evaluations.

Combining (A.4), (A.5) and (A.6), we have

$$\begin{aligned}
\mathrm{Var}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\} &= O(1),\\
\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\} &= O(n^{-1}). \qquad (A.7)
\end{aligned}$$

Next, we evaluate the Covariance term of (A.3). Let f_{ij} denote the joint p.d.f of $(\boldsymbol{X}_{i-1},\boldsymbol{X}_{j-1})$, and let

$$h_{ij}(\boldsymbol{x},\boldsymbol{y}) = f_{ij}(\boldsymbol{x},\boldsymbol{y}) - f_i(\boldsymbol{x})\, f_j(\boldsymbol{y}),$$

then

$$\begin{aligned}
&\mathrm{Cov}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\ ,\ X_j\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\\
&= \mathrm{Const.}\Big[\,E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\\
&\qquad - E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\Big]\\
&= \mathrm{Const.}\int_{\mathcal{R}^d}\int_{\mathcal{R}^d}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{y})\, h_{ij}(\boldsymbol{x},\boldsymbol{y})\,\mathbf{d}\boldsymbol{x}\,\mathbf{d}\boldsymbol{y}\\
&= \mathrm{Const.}\int_{\mathcal{R}^d}\int_{\mathcal{R}^d} 2^{-dj_0}\,\phi(x_1)\cdots\phi(x_d)\phi(y_1)\cdots\phi(y_d)\\
&\qquad \times h_{ij}\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_0}},\frac{\boldsymbol{y}+\boldsymbol{k}}{2^{j_0}}\Big)\,\mathbf{d}\boldsymbol{x}\,\mathbf{d}\boldsymbol{y}\\
&= O(2^{-dj_0}). \qquad (A.8)
\end{aligned}$$

On the other hand, from α-mixing condition, we can also evaluate the covariance term of (A.3) as follows. Since $X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})$ is $\mathcal{F}_i$-measurable, and $X_j\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})$ is $\mathcal{F}^j$-measurable, it follows from lemma 1 that

$$\begin{aligned}
&\mathrm{Cov}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\ ,\ X_j\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\\
&= \mathrm{Const.}\Big[\,E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\\
&\qquad - E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1})\Big\}E\Big\{\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\Big]\\
&= \mathrm{Const.}\int_{\mathcal{R}^d}\int_{\mathcal{R}^d}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{x})\Phi_{j_0\boldsymbol{k}}(\boldsymbol{y})\, h_{ij}(\boldsymbol{x},\boldsymbol{y})\,\mathbf{d}\boldsymbol{x}\,\mathbf{d}\boldsymbol{y}\\
&= \mathrm{Const.}\int_{\mathcal{R}^d}\int_{\mathcal{R}^d} 2^{-dj_0}\,\phi(x_1)\cdots\phi(x_d)\phi(y_1)\cdots\phi(y_d)\\
&\qquad \times h_{ij}\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_0}},\frac{\boldsymbol{y}+\boldsymbol{k}}{2^{j_0}}\Big)\,\mathbf{d}\boldsymbol{x}\,\mathbf{d}\boldsymbol{y}\\
&= O(2^{-dj_0}). \qquad (A.9)
\end{aligned}$$

From (A.8) and (A.9), we have

$$\begin{aligned}
&\frac{2}{n^2}\sum_{i<j}\mathrm{Cov}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1}),\ X_j\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{j-1})\Big\}\\
&\quad= \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\mathrm{Cov}\Big\{X_i\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i-1}),\ X_{i+j}\Phi_{j_0\boldsymbol{k}}(\boldsymbol{X}_{i+j-1})\Big\}\\
&\quad= \frac{2}{n^2}\sum_{i=1}^{n-1}\Big\{\sum_{j\le M_n}O(2^{-dj_0})\ +\ \sum_{j>M_n}O(2^{dj_0}\alpha(j))\Big\}\\
&\quad\le \frac{1}{n}M_nO(2^{-dj_0})\ +\ O(2^{dj_0}\alpha(M_n))\\
&\quad= o(n^{-1}),
\end{aligned}\tag{A.10}$$

where we set $M_n = o(2^{dj_0}) = o(n)$ as $n\to\infty$, and used $\alpha(u) = O(\rho^u),\ (0<\rho<1)$. Combining (A.3), (A.7) and (A.10), we can deduce that

$$\begin{aligned}
E(S_1) &= \sum_{\boldsymbol{k}\in\Omega_{j_0}^d}[O(n^{-1})\ +\ o(n^{-1})]\\
&= O(2^{dj_0}n^{-1}).
\end{aligned}\tag{A.11}$$

Similar evaluation to $E(S_1)$ yields

$$E(S_2)\ =\ O(2^{dj_1}n^{-1}).\tag{A.12}$$

It follows that

$$\begin{aligned}
\beta_{j\boldsymbol{k}\boldsymbol{i}} &= \int_{\mathcal{R}^d} g(\boldsymbol{x})\Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x} \qquad\qquad (A.13)\\
&= 2^{-dj/2}\int_{\mathcal{R}^d} g\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^j}\Big)\chi^{i_1}(x_1)\cdots\chi^{i_d}(x_d)\,\mathbf{d}\boldsymbol{x}.
\end{aligned}$$

In (A.13), using Taylor expansion for g around $\boldsymbol{k}/2^j\in\mathcal{R}^d$, then

$$\begin{aligned}
\beta_{j\boldsymbol{k}\boldsymbol{i}} &= 2^{-dj/2}\sum_{v=0}^{r-1}\frac{1}{v!}\sum_{(u_1\ldots u_d)\in U_v^d}\binom{v}{u_1,\ldots,u_d}\\
&\quad\times\frac{\partial^r g}{\partial^{u_1}x_1\cdots\partial^{u_d}x_d}\Big(\frac{\boldsymbol{k}}{2^j}\Big)\\
&\quad\times\int_{\mathcal{R}^d}2^{-vj}x_1^{u_1}\cdots x_d^{u_d}\chi^{i_1}(x_1)\cdots\chi^{i_d}(x_d)\,\mathbf{d}\boldsymbol{x}\\
&\quad+\ 2^{-dj/2}\int_{\mathcal{R}^d}R_r\Big(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^j}\Big)\\
&\quad\times\chi^{i_1}(x_1)\cdots\chi^{i_d}(x_d)\,\mathbf{d}\boldsymbol{x},
\end{aligned}\tag{A.14}$$

where $R_r(\cdot)$ is residual term of rth order, and
$U_v^d = \{(u_1, \ldots, u_d) \in \mathcal{Z}^d \;:\; 0 \le u_1, \ldots, u_d \le v \quad \text{and} \quad u_1 + \cdots + u_d = v\}$.
Only the rth term of (A.14) remains because $\int x^v \psi(x)\,\mathrm{d}x = 0,\ (0 \le v \le r-1)$. Hence

$$
\begin{aligned}
\beta_{j\boldsymbol{k}\boldsymbol{i}} &= 2^{-j(r+d/2)} \frac{1}{r!} \sum_{(u_1 \ldots u_d) \in U_r^d} \binom{r}{u_1, \ldots, u_d} \\
&\quad \times \int_{\mathcal{R}^d} \left\{ \frac{\partial^r g}{\partial^{u_1} x_1 \cdots \partial^{u_d} x_d} \Big(\frac{\boldsymbol{k} + \theta \boldsymbol{x}}{2^j} \Big) \right\} \\
&\quad \times x_1^{u_1} \cdots x_d^{u_d} \chi^{i_1}(x_1) \cdots \chi^{i_d}(x_d)\, \mathbf{d}\boldsymbol{x}.
\end{aligned}
$$

Now, let

$$
\begin{aligned}
\xi_{j\boldsymbol{k}}^{(u_1 \ldots u_d)} &= \frac{\partial^r g}{\partial^{u_1} x_1 \cdots \partial^{u_d} x_d} \Big(\frac{\boldsymbol{k} + \theta \boldsymbol{x}}{2^j} \Big) \\
&\quad - \frac{\partial^r g}{\partial^{u_1} x_1 \cdots \partial^{u_d} x_d} \Big(\frac{\boldsymbol{k}}{2^j} \Big) \quad \text{and} \\
g_{j\boldsymbol{k}}^{(u_1 \ldots u_d)} &= \frac{\partial^r g}{\partial^{u_1} x_1 \cdots \partial^{u_d} x_d} \Big(\frac{\boldsymbol{k}}{2^j} \Big),
\end{aligned}
$$

then, from $\xi_{j\boldsymbol{k}}^{(u_1 \ldots u_d)} = o(1)$ as $j \to \infty$, and $(\,r!\,)^{-1} \int x^r \psi(x)\,\mathrm{d}x < \infty$, we have

$$
\begin{aligned}
\beta_{j\boldsymbol{k}\boldsymbol{i}} &= 2^{-j(r+d/2)} \frac{1}{r!} \sum_{(u_1 \ldots u_d) \in U_r^d} \binom{r}{u_1, \ldots, u_d} \\
&\quad \times \Big\{ g_{j\boldsymbol{k}}^{(u_1 \ldots u_d)} + o(1) \Big\} \prod_{s=1}^{d} \int x_s^{u_s} \chi^{i_s}(x_s)\,\mathrm{d}x_s \\
&= \begin{cases} O(2^{-j(r+d/2)}) & \text{if} \quad i_1 + i_2 + \cdots + i_d = 1, \\ 0 & \text{if} \quad i_1 + i_2 + \cdots + i_d > 1. \end{cases}
\end{aligned} \tag{A.15}
$$

Hence,

$$
S_3 = \sum_{j=j_1}^{\infty} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \beta_{j\boldsymbol{k}\boldsymbol{i}}^2 = \sum_{j=j_1}^{\infty} 2^{dj} d\, O(2^{-j(2r+d)}) = O(2^{-2j_1 r}). \tag{A.16}
$$

From (A.2), (A.11), (A.12) and (A.16), it follows that

$$
\begin{aligned}
E\Big[\int_{\mathcal{R}^d} (\hat{g} - g)^2\, \mathbf{d}\boldsymbol{x} \Big] &= E(S_1 + S_2 + S_3) \\
&= O\Big(\frac{2^{dj_0}}{n} + \frac{2^{dj_1}}{n} + 2^{-2j_1 r} \Big) \text{ as } n \to \infty.
\end{aligned}
$$

Proof of Theorem 1 (ii)

At first, we prove the following result.

$$\sup_{\boldsymbol{x}\in U} |\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})| \;=\; o(1) \qquad \text{a.e.} \tag{A.17}$$

Since $\text{supp}(\phi),\ \text{supp}(\psi) = [0,1]$ and $U \subset Q = [0,1]^d$,

$$\sup_U |\hat{f} - f| \;\leq\; s_1 + s_2 + s_3,$$

where

$$\begin{aligned}
s_1 &= \sum_{\boldsymbol{k}\in\Omega^d_{j_0}} |\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}}|\, \|\Phi_{j_0\boldsymbol{k}}\|_\infty, \\
s_2 &= \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\Omega^d_j} \sum_{\boldsymbol{i}\in\mathcal{I}^d} |\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}}|\, \|\Psi_{j\boldsymbol{k}\boldsymbol{i}}\|_\infty, \\
s_3 &= \sum_{j\geq j_1} \sum_{\boldsymbol{k}\in\Omega^d_j} \sum_{\boldsymbol{i}\in\mathcal{I}^d} |\beta_{j\boldsymbol{k}\boldsymbol{i}}|\, \|\Psi_{j\boldsymbol{k}\boldsymbol{i}}\|_\infty.
\end{aligned}$$

Note that $\|\Phi_{j_0\boldsymbol{k}}\|_\infty = O(2^{dj_0/2})$ and $\|\Psi_{j\boldsymbol{k}\boldsymbol{i}}\|_\infty = O(2^{dj/2})$. It follows from

$$\left(\frac{s_1}{2^{dj_0}}\right)^2 \;\leq\; \frac{1}{2^{dj_0}} \sum_{\boldsymbol{k}\in\Omega^d_{j_0}} \Big(|\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}}|\, \|\Phi_{j_0\boldsymbol{k}}\|_\infty\Big)^2,$$

that

$$s_1 \;=\; O(2^{dj_0})\Big\{\sum_{\boldsymbol{k}\in\Omega^d_{j_0}} (\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}})^2\Big\}^{1/2}.$$

By (A.11), for all sufficiently large n, there exists a positive constant B_1 such that

$$n2^{-dj_0} \sum_{\boldsymbol{k}\in\Omega^d_{j_0}} (\hat{\alpha}_{j_0\boldsymbol{k}} - \alpha_{j_0\boldsymbol{k}})^2 \;<\; B_1 \qquad \text{a.e.}$$

Hence,

$$s_1 \;<\; n^{-1/2}2^{dj_0/2}B_1^{1/2}O(2^{dj_0}) \qquad \text{a.e,}$$

which together with $2^{dj_0} = o(n^{1/3})$ yields that

$$s_1 \;=\; o(1) \qquad \text{a.e.} \tag{A.18}$$

Similally, it follows from (A.12) and $2^{dj_1} = o(n^{1/3})$ that

$$s_2 \;=\; o(1) \qquad \text{a.e.} \tag{A.19}$$

According to the argument leading to (A.16),

$$s_3 \;=\; o(1). \tag{A.20}$$

Thus, equation (A.17) follows from (A.18), (A.19) and (A.20).

Next, we show $\int(\hat{g} - g)^2 = o_p(1)$. By $E\{\int(\hat{g} - g)^2\} = o(1)$, for all sufficiently large n, there exists a positive constants B_2 such that

$$\int(\hat{g} - g)^2 \;<\; B_2 \qquad \text{a.e,}$$

and then, it follows that

$$\begin{aligned}
&\mathrm{Var}\Big\{\int(\hat{g} - g)^2\Big\} \\
&\quad= E\Big[\Big\{\int(\hat{g} - g)^2\Big\}^2\Big] - \Big[E\Big\{\int(\hat{g} - g)^2\Big\}\Big]^2 \\
&\quad< B_2\, E\Big\{\int(\hat{g} - g)^2\Big\} + o(1) \\
&\quad= o(1).
\end{aligned}$$

Therefore, by Chebyshev's inequality,

$$\int(\hat{g} - g)^2 \;=\; o_p(1). \tag{A.21}$$

Now, it follows from (A.17) and Condition (c) that

$$\hat{\mu} - \mu \;=\; \hat{f}^{-1}(\hat{g} - \mu\hat{f}) \;=\; \{1 + o_p(1)\}f^{-1}(\hat{g} - \mu\hat{f}) \qquad \text{on } U.$$

Furthermore, by Condition (a) and (A.21) respectively, g and $(\hat{g} - g)$ are bounded. Therefore, it follows that

$$\hat{g} - \mu\hat{f} \;=\; \hat{g} - \{1 + o_p(1)\}g \;=\; \{1 + o_p(1)\}(\hat{g} - g),$$

and that

$$\hat{\mu} - \mu \;=\; \{1 + o_p(1)\} f^{-1}(\hat{g} - g) \qquad \text{on } U. \tag{A.22}$$

Combining (A.21) and (A.22), we have

$$\begin{aligned}
\int_U (\hat{\mu} - \mu)^2 \;&=\; \{1 + o_p(1)\} \int_U f^{-2}(\hat{g} - g)^2 \\
&\le\; \{1 + o_p(1)\} M^2 \int_U (\hat{g} - g)^2 \\
&=\; o_p(1).
\end{aligned}$$

Proof of Theorem 1 (iii)

We have the following representation.

$$\int_{\mathcal{R}^d} (\hat{g}_\lambda - g)^2 \, \mathbf{d}\boldsymbol{x} \;=\; S_{\lambda 1} + S_{\lambda 2} + S_{\lambda 3},$$

where $S_{\lambda 1} = S_1,\ S_{\lambda 3} = S_3$ in (A.2), and

$$\begin{aligned}
S_{\lambda 2} \;&=\; \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \{\eta_\lambda(\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}}) - \beta_{j\boldsymbol{k}\boldsymbol{i}}\}^2 \\
&=\; \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} (\lambda \hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}})^2 \\
&=\; \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \{(\lambda - 1)\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} + (\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}})\}^2.
\end{aligned}$$

It follows that

$$\begin{aligned}
S_2 - S_{\lambda 2} \;&=\; -(\lambda - 1)^2 \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \hat{\beta}^2_{j\boldsymbol{k}\boldsymbol{i}} \\
&\quad - 2(\lambda - 1) \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k} \in \Omega_j^d} \sum_{\boldsymbol{i} \in \mathcal{I}^d} \hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} (\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}}).
\end{aligned}$$

Now, let

$$\begin{aligned} a &= \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\Omega_j^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \hat{\beta}^2_{j\boldsymbol{k}\boldsymbol{i}}, \\ b &= \sum_{j=j_0}^{j_1-1} \sum_{\boldsymbol{k}\in\Omega_j^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}}(\hat{\beta}_{j\boldsymbol{k}\boldsymbol{i}} - \beta_{j\boldsymbol{k}\boldsymbol{i}}), \end{aligned}$$

then we have

$$\begin{aligned} S_2 - S_{\lambda 2} &= -a(\lambda-1)^2 - 2b(\lambda-1) \\ &= -a\Big(\lambda + \frac{b}{a} - 1\Big)^2 + \frac{b^2}{a}. \end{aligned}$$

When $\lambda = 1 - b/a$, it holds that

$$S_2 - S_{\lambda 2} = \frac{b^2}{a} \ > \ 0 \qquad \text{almost surely.}$$

Hence, we have

$$\int_{\mathcal{R}^d} (\hat{g} - g)^2 \, \mathbf{d}\boldsymbol{x} \ > \ \int_{\mathcal{R}^d} (\hat{g}_\lambda - g)^2 \, \mathbf{d}\boldsymbol{x}.$$

Proof of Theorem 2

From (A.2) we have

$$E\Big(\hat{g}(\boldsymbol{x})\Big) - g(\boldsymbol{x}) = -\sum_{j\ge j_1} \sum_{\boldsymbol{k}\in\Omega_j^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \beta_{j\boldsymbol{k}\boldsymbol{i}} \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}), \qquad \boldsymbol{x} \in Q.$$

Now it follows from the definition that

$$\Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}) = O(2^{dj/2}).$$

Thus together with (A.16) we have

$$\begin{aligned} \sum_{j\ge j_1} \sum_{\boldsymbol{k}\in\Omega_j^d} \sum_{\boldsymbol{i}\in\mathcal{I}^d} \beta_{j\boldsymbol{k}\boldsymbol{i}} \Psi_{j\boldsymbol{k}\boldsymbol{i}}(\boldsymbol{x}) &= \sum_{j\ge j_1} O(2^{(d-r)j}) \\ &= O(2^{(d-r)j_1}), \end{aligned}$$

since $r > d$ is assumed. The bias of $\hat{f}$ is evaluated similarly. For $\hat{\mu}$, it follows from (A.17) that $f(\boldsymbol{x})/\hat{f}(\boldsymbol{x}) = 1 + o(1)$ a.e. for $\boldsymbol{x} \in U$. Thus

$$\frac{\hat{g}}{\hat{f}} = \frac{g}{f} + o(1) \qquad \text{a.e,}$$

and the rest of the proof is trivial.

Proof of Theorem 3

Similarly as the proof of Lemma 2,

$$\begin{aligned} & E[\mathrm{CV}_{M_n}(\theta)] \\ & \quad = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})]^2 \\ & \qquad - \frac{2}{n}\sum_{i=1}^{n}\mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})\}. \end{aligned} \tag{A.23}$$

In the following, we evaluate each term of (A.23). Since $\hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})$ is random variable of $(X_{1-d}, \ldots, X_{i-1}, X_{i+M_n-d}, \ldots, X_n)$, we consider the joint distribution of this random variable. Let

$$f^{\backslash(i,M_n)}(x_{1-d}, \ldots, x_{i-1}, x_i, x_{i+M_n-d}, \ldots, x_n)$$

be the joint p.d.f of $(X_{1-d}, \ldots, X_{i-1}, \varepsilon_i, X_{i+M_n-d}, \ldots, X_n)$, and let

$$f_1^{\backslash(i,M_n)}(x_{1-d}, \ldots, x_{i-1}, x_i) \quad \text{and} \quad f_2^{\backslash(i,M_n)}(x_{i+M_n-d}, \ldots, x_n)$$

be the joint p.d.f of $(X_{1-d}, \ldots, X_{i-1}, \varepsilon_i)$ and $(X_{i+M_n-d}, \ldots, X_n)$ respectively. Since $(X_{1-d}, \ldots, X_{i-1}, \varepsilon_i)$ is $\mathcal{F}_i$-measurable, and $(X_{i+M_n-d}, \ldots, X_n)$ is $\mathcal{F}^{i+M_n}$-measurable, it follows from the definition of $\alpha(\cdot)$ that

$$\begin{aligned} & \Big| f^{\backslash(i,M_n)}(x_{1-d}, \ldots, x_{i-1}, x_i, x_{i+M_n-d}, \ldots, x_n) \\ & \qquad - f_1^{\backslash(i,M_n)}(x_{1-d}, \ldots, x_{i-1}, x_i) f_2^{\backslash(i,M_n)}(x_{i+M_n-d}, \ldots, x_n) \Big| \\ & \leq \ \alpha(M_n). \end{aligned} \tag{A.24}$$

Furthermore, let $f_{11}^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1})$ and $f_{12}^{\backslash(i,M_n)}(x_i)$ be the p.d.f of $(X_{1-d},\ldots,X_{i-1})$ and ε_i respectively, then

$$\begin{aligned}
& f_{1}^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1},x_i) \\
& \qquad = \quad f_{11}^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1}) f_{12}^{\backslash(i,M_n)}(x_i). \qquad (A.25)
\end{aligned}$$

Now,

$$\begin{aligned}
& \mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})\} \\
& \quad = \quad E\{\varepsilon_i \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})\} \\
& \quad = \quad \int x_i \hat{\mu}_\theta^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1},x_{i+M_n-d},\ldots,x_n) \\
& \qquad\quad \times f^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1},x_i,x_{i+M_n-d},\ldots,x_n) \\
& \qquad\quad \times \mathrm{d}x_{1-d}\cdots \mathrm{d}x_{i-1}\mathrm{d}x_i \mathrm{d}x_{i+M_n-d}\cdots \mathrm{d}x_n,
\end{aligned}$$

then, using (A.24) and (A.25) for $f^{\backslash(i,M_n)}$, we have

$$\begin{aligned}
& \mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})\} \\
& \quad = \quad \int_{\mathrm{supp}(f^{\backslash(i,M_n)})} x_i \hat{\mu}_\theta^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1},x_{i+M_n-d},\ldots,x_n) \\
& \qquad\quad \times \{f_{11}^{\backslash(i,M_n)}(x_{1-d},\ldots,x_{i-1}) f_{12}^{\backslash(i,M_n)}(x_i) \\
& \qquad\qquad f_{2}^{\backslash(i,M_n)}(x_{i+M_n-d},\ldots,x_n) + O(\alpha(M_n))\} \\
& \quad = \quad \int x_i f_{12}^{\backslash(i,M_n)}(x_i) \int \hat{\mu}_\theta^{\backslash(i,M_n)} f_{11}^{\backslash(i,M_n)} f_{2}^{\backslash(i,M_n)} \\
& \qquad\quad + O(\alpha(M_n)) \int_{\mathrm{supp}(f^{\backslash(i,M_n)})} x_i \hat{\mu}_\theta^{\backslash(i,M_n)}.
\end{aligned}$$

It follows from $\int x_i f_{12}^{\backslash(i,M_n)}(x_i) = E(\varepsilon_i) = 0$ and $\int_{\mathrm{supp}(f^{\backslash(i,M_n)})} x_i \hat{\mu}_\theta^{\backslash(i,M_n)} < \infty$ that

$$\mathrm{Cov}\{\varepsilon_i,\ \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})\} \ = \ O(\alpha(M_n)). \qquad (A.26)$$

Now

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})]^2 \\
& \qquad = \quad \frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})
\end{aligned}$$

$$
\begin{aligned}
& + \hat{\mu}_\theta(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})]^2 \\
= \quad & \mathrm{R}(\theta) + \frac{1}{n}\sum_{i=1}^{n} E[\{\hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\} \\
& \times \{\hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1}) + \hat{\mu}_\theta(\boldsymbol{X}_{i-1}) - 2\mu(\boldsymbol{X}_{i-1})\}].
\end{aligned}
$$

Since $M_n = o(n)$, $\{\hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta(\boldsymbol{X}_{i-1})\} = o(1)$ as $n \to \infty$, and

$$
\frac{1}{n}\sum_{i=1}^{n} E[\mu(\boldsymbol{X}_{i-1}) - \hat{\mu}_\theta^{\backslash(i,M_n)}(\boldsymbol{X}_{i-1})]^2 \;=\; \mathrm{R}(\theta) + o(1), \tag{A.27}
$$

Using (A.23), (A.26) and (A.27), we have

$$
\begin{aligned}
E[\mathrm{CV}_{M_n}(\theta)] &= \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \mathrm{R}(\theta) + o(1) + O(\alpha(M_n)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \mathrm{R}(\theta) + o(1).
\end{aligned}
$$

REFERENCES

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics 110, New York: Springer-Verlag.

Chui, C. (1997), *Wavelets: A Mathematical Tool for Signal Analysis*, Philadelphia: SIAM.

—— (1992), *An Introduction to Wavelets*, Boston: Academic Press.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: SIAM.

Donoho, D., Johnston, I., Kerkyacharian, G. and Picard, D. (1995), "Wavelet shrinkage: Asymptopia?," *Journal of the Royal Statistical Society*, Ser. B, 57, 301-369.

—— (1996), "Density estimation by wavelet thresholding," *The Annals of Statistics*, 24, 508-539.

Fueda, K. and Yanagawa, T. (2001), "Estimating the embedding dimension and delay time from chaotic time series with dynamic noise," *Journal of the Japan Statistical Society*, 31, 27-38.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and its Applications*, New York: Academic Press.

Hall, P. and Patil, P. (1996), "On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods," *Journal of the Royal Statistical Society*, Ser. B, 58, 361-377.

Lasota, A. and Mackey, M.C. (1989), "Stochastic perturbation of dynamical systems: the weak convergence of measures," *Journal of mathematical Analysis and Applications*, 138, 232-248.

Miwakeichi, F., Ramirez-Padron, R., Valdes-Sosa, P. A. and Ozaki, T. (2001), "A comparison of non-linear non-parametric models for epilepsy data," *Computers in Biology and Medicine*, 31, 41-57.

Taniguchi, M. and Kakizawa, Y. (2001), *Asymptotic Theory of Statistical Inference for Time Series*, New York, Springer-Verlag.

Theiler, J. (1995), "On the evidence for low-dimensional chaos in an epileptic electroencephalogram," *Physics Letter*, A, 196, 335-341.

Truong, Y. and Patil, P. (2001), "Asymptotics for wavelet based estimates of pieswise smooth regression for stationary time series," *The Annals of Institute of Statistical Mathematics*, 53, 159-178.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF

2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model