Wavelet Based Estimate for Non－Linear and Non－ Stationary Auto－Regressive Model<br>Fujii，Toru<br>Graduate School of Mathematics，Kyushu University<br>Yanagawa，Takashi<br>Graduate School of Mathematics，Kyushu University

https：／／hdl．handle．net／2324／3342

[^0]
# 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<br>Graduate School of Mathematics, Kyushu University<br>Hakozaki 6-10-1 Higashi-Ku Fukuoka 812-8581 Japan<br>fujii@math.kyushu-u.ac.jp<br>and<br>Takashi YANAGAWA<br>Graduate School of Mathematics, Kyushu University<br>Hakozaki 6-10-1 Higashi-Ku Fukuoka 812-8581 Japan<br>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 $\alpha$-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 $\left\{X_{t}\right\}$ generated by the following auto-regressive model.

$$
\begin{equation*}
X_{t}=\mu\left(X_{t-1}\right)+\varepsilon_{t} \tag{1}
\end{equation*}
$$

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

$$
\mu(x)=E\left[X_{t} \mid X_{t-1}=x\right] .
$$

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 $\left\{X_{t}\right\}$ in this paper, but assuming $\alpha$-mixing condition for $\left\{X_{t}\right\}$, we consider a method to estimate unknown function $\mu$ 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 $\alpha$-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 $\alpha$-mixing condition. In Section 2 we review briefly the wavelet expansion based on Chui $(1992,1997)$ and Daubechies (1992). Wavelet estimator of $\mu$ 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: $\phi$ (father wavelet, left) and $\psi$ (mother wavelet, right)

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

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

and

$$
\begin{equation*}
\phi(x)=\sum_{k \in \mathcal{Z}} p_{k} \phi(2 x-k) \tag{2}
\end{equation*}
$$

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

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

where $\left\{p_{k}\right\}$ is the same sequence as $\phi . \psi$ is called the mother wavelet.

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

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

In addition, if $\sum_{k \in \mathcal{Z}}(-1)^{k} k^{v} p_{k}=0,(1 \leq v \leq r)$ for some integer $r \geq 1$, then the moment condition $\int x^{v} \psi(x) \mathrm{d} x=0,(1 \leq v \leq r)$ is satisfied. Generally, a pair of
wavelets $\phi, \psi$ is not represented as specific functions, but their wave patterns are fixed by the peculiar sequence $\left\{p_{k}\right\},(2)$ 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 $\phi$ and $\psi$, define

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

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

$$
\begin{align*}
\int \phi_{j k}(x) \phi_{j m}(x) \mathrm{d} x & =\delta_{k m} \\
\int \psi_{j k}(x) \psi_{l m}(x) \mathrm{d} x & =\delta_{j l} \delta_{k m} \\
\int \phi_{j k}(x) \psi_{l m}(x) \mathrm{d} x & =0, \quad j \leq l \tag{4}
\end{align*}
$$

### 2.2 Multiresolution Analysis and Wavelet Expansion

Using $\phi_{j k}(x)$ and $\psi_{j k}(x)$ as base functions, one can construct the following subspaces $V_{j}$ and $W_{j}$ in $\mathrm{L}_{2}(\mathcal{R})$.

$$
\begin{aligned}
V_{j} & =\operatorname{clos}_{\mathrm{L}_{2}}\left\langle\phi_{j k}(x) ; k \in \mathcal{Z}\right\rangle \\
W_{j} & =\operatorname{clos}_{\mathrm{L}_{2}}\left\langle\psi_{j k}(x) ; k \in \mathcal{Z}\right\rangle, \quad j \in \mathcal{Z}
\end{aligned}
$$

$V_{j}$ and $W_{j}$ are $\mathrm{L}_{2}(\mathcal{R})$-closures spanned by the linear combinations of $\phi_{j k}(x)$ and $\psi_{j k}(x)$ 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 $\left\{V_{j}\right\}$ have the following hierarchical structure,

$$
\cdots \subset V_{j-1} \subset V_{j} \subset V_{j+1} \subset \cdots \rightarrow \mathrm{~L}_{2}(\mathcal{R})
$$

and that $\left\{W_{j}\right\}$ 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 $\phi$ and $\psi$.
From this MRA, any function $f \in \mathrm{~L}_{2}(\mathcal{R})$ can be expressed as a series

$$
\begin{equation*}
f(x)=\sum_{k} \alpha_{j_{0} k} \phi_{j_{0} k}(x)+\sum_{j \geq j_{0}} \sum_{k} \beta_{j k} \psi_{j k}(x) . \tag{5}
\end{equation*}
$$

This is called the wavelet expansion of $f$. From orthonormality (4), each coefficients in (5) are uniquely expressed by the $\mathrm{L}_{2}$-products of $f$ and $\phi_{j k}$, and $f$ and $\psi_{j k}$, respectively

$$
\alpha_{j k}=\int f(x) \phi_{j k}(x) \mathrm{d} x, \quad \beta_{j k}=\int f(x) \psi_{j k}(x) \mathrm{d} x .
$$

## 3. ESTIMATION OF $\mu(x)$

In this section, we describe a wavelet method for estimating the underlying regression function from $\left\{X_{t}\right\}$ given by (1). We develop a new technique to cope with the nonstationality.

### 3.1 Method for the Non-Stationality

For time series data $\left\{X_{0}, X_{1}, \ldots, X_{n}\right\}$ 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 $\mu$ by using relationship $\mu(x)=g(x) / f(x)$.

The wavelet expansion of $f$ is represented as follows,

$$
f(x)=\sum_{k} \alpha_{j_{0} k}^{f} \phi_{j_{0} k}(x)+\sum_{j \geq j_{0}} \sum_{k} \beta_{j k}^{f} \psi_{j k}(x),
$$

where $\alpha_{j k}^{f}=\int f(x) \phi_{j k}(x) \mathrm{d} x, \beta_{j k}^{f}=\int f(x) \psi_{j k}(x) \mathrm{d} x$. The wavelet expansion of $g$ may be given in the same way, using $\alpha_{j k}^{g}=\int g(x) \phi_{j k}(x) \mathrm{d} x, \beta_{j k}^{g}=\int g(x) \psi_{j k}(x) \mathrm{d} x$.

Put

$$
\begin{equation*}
\hat{\alpha}_{j k}^{f}=\frac{1}{n} \sum_{i=1}^{n} \phi_{j k}\left(X_{i-1}\right), \quad \hat{\beta}_{j k}^{f}=\frac{1}{n} \sum_{i=1}^{n} \psi_{j k}\left(X_{i-1}\right) . \tag{6}
\end{equation*}
$$

Then

$$
\begin{aligned}
E\left[\hat{\alpha}_{j k}^{f}\right] & =\frac{1}{n} \sum_{i=1}^{n} E\left[\phi_{j k}\left(X_{i-1}\right)\right] \\
& =\frac{1}{n} \sum_{i=1}^{n} \int \phi_{j k}(x) f_{i}(x) \mathrm{d} x \\
& =\int\left[\frac{1}{n} \sum_{i=1}^{n} f_{i}(x)\right] \phi_{j k}(x) \mathrm{d} x \\
& =\int f(x) \phi_{j k}(x) \mathrm{d} x \\
& =\alpha_{j k}^{f}, \quad \text { and } \quad E\left[\hat{\beta}_{j k}^{f}\right]=\beta_{j k}^{f}
\end{aligned}
$$

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

$$
\begin{equation*}
\hat{\alpha}_{j k}^{g}=\frac{1}{n} \sum_{i=1}^{n} X_{i} \phi_{j k}\left(X_{i-1}\right), \quad \hat{\beta}_{j k}^{g}=\frac{1}{n} \sum_{i=1}^{n} X_{i} \psi_{j k}\left(X_{i-1}\right) . \tag{7}
\end{equation*}
$$

Then

$$
\begin{aligned}
E\left[\hat{\alpha}_{j k}^{g}\right] & =\frac{1}{n} \sum_{i=1}^{n} E\left[X_{i} \phi_{j k}\left(X_{i-1}\right)\right] \\
& =\frac{1}{n} \sum_{i=1}^{n} E\left[E\left[X_{i} \mid X_{i-1}\right] \phi_{j k}\left(X_{i-1}\right)\right] \\
& =\frac{1}{n} \sum_{i=1}^{n} E\left[\mu\left(X_{i-1}\right) \phi_{j k}\left(X_{i-1}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
& =\frac{1}{n} \sum_{i=1}^{n} \int \mu(x) \phi_{j k}(x) f_{i}(x) \mathrm{d} x \\
& =\int \mu(x)\left[\frac{1}{n} \sum_{i=1}^{n} f_{i}(x)\right] \phi_{j k}(x) \mathrm{d} x \\
& =\int \mu(x) f(x) \phi_{j k}(x) \mathrm{d} x \\
& =\int g(x) \phi_{j k}(x) \mathrm{d} x \\
& =\alpha_{j k}^{g}, \quad \text { and } \quad E\left[\hat{\beta}_{j k}^{g}\right]=\beta_{j k}^{g}
\end{aligned}
$$

Thus $\hat{\alpha}_{j k}^{g}$ and $\hat{\beta}_{j k}^{g}$ are also unbiased estimators of $\alpha_{j k}^{g}$ and $\beta_{j k}^{g}$. Note that the estimators $\hat{\alpha}_{j k}^{f}, \hat{\beta}_{j k}^{f}, \hat{\alpha}_{j k}^{g}$ and $\hat{\beta}_{j k}^{g}$ are identical to those estimators given in Truong and Patil (2001) who assumed the stationality for $\left\{X_{t}\right\}$.

### 3.2 Wavelet Estimator and Smoothing Functions

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

$$
\begin{align*}
& \hat{f}(x)=\sum_{k} \hat{\alpha}_{j_{0} k}^{f} \phi_{j_{0} k}(x)+\sum_{j=j_{0}}^{j_{1}-1} \sum_{k} \hat{\beta}_{j k}^{f} \psi_{j k}(x), \\
& \hat{g}(x)=\sum_{k} \hat{\alpha}_{j_{0} k}^{g} \phi_{j_{0} k}(x)+\sum_{j=j_{0}}^{j_{1}-1} \sum_{k} \hat{\beta}_{j k}^{g} \psi_{j k}(x), \tag{8}
\end{align*}
$$

and $\mu$ 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.

$$
\begin{align*}
& \hat{f}_{\lambda}(x)=\sum_{k} \hat{\alpha}_{j_{0} k}^{f} \phi_{j_{0} k}(x)+\sum_{j=j_{0}}^{j_{1}-1} \sum_{k} \eta_{\lambda}\left(\hat{\beta}_{j k}^{f}\right) \psi_{j k}(x), \\
& \hat{g}_{\lambda}(x)=\sum_{k} \hat{\alpha}_{j_{0} k}^{g} \phi_{j_{0} k}(x)+\sum_{j=j_{0}}^{j_{1}-1} \sum_{k} \eta_{\lambda}\left(\hat{\beta}_{j k}^{g}\right) \psi_{j k}(x), \tag{9}
\end{align*}
$$

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

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

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 $\lambda$.

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.

$$
X_{t}=\mu\left(X_{t-d}, \ldots, X_{t-1}\right)+\varepsilon_{t}
$$

where $\left\{\varepsilon_{t}\right\}$ are independent random variables, and

$$
\mu\left(x_{1}, \ldots, x_{d}\right)=E\left[X_{t} \mid X_{t-d}=x_{1}, \ldots, X_{t-1}=x_{d}\right] .
$$

Let $f_{t}\left(x_{1}, \ldots, x_{d}\right)$ be the joint p.d.f of $\left(X_{t-d}, \ldots, X_{t-1}\right)$ and put

$$
\begin{aligned}
f\left(x_{1}, \ldots, x_{d}\right) & =\frac{1}{n} \sum_{i=1}^{n} f_{i}\left(x_{1}, \ldots, x_{d}\right) \\
g\left(x_{1}, \ldots, x_{d}\right) & =\mu\left(x_{1}, \ldots, x_{d}\right) f\left(x_{1}, \ldots, x_{d}\right)
\end{aligned}
$$

### 4.2 Estimation of $f, g$ and $\mu$ 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 $\left(x_{1}, \cdots, x_{d}\right)$ 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

$$
\begin{equation*}
f(\boldsymbol{x})=\sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \alpha_{j_{0}}^{f} \boldsymbol{k}^{\Phi_{j_{0}}} \boldsymbol{k}(\boldsymbol{x})+\sum_{j \geq j_{0}} \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \beta_{j \boldsymbol{k} i}^{f} \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}), \tag{10}
\end{equation*}
$$

where $\boldsymbol{k}=\left(k_{1}, \ldots, k_{d}\right), \mathcal{I}^{d}=\left\{\boldsymbol{i}=\left(i_{1}, \ldots, i_{d}\right): i_{s}=0\right.$ or $1(s=1, \ldots, d)$, and $i_{1}+$ $\left.\cdots+i_{d} \geq 1\right\}$, and

$$
\begin{align*}
\Phi_{j \boldsymbol{k}}(\boldsymbol{x}) & =\phi_{j k_{1}}\left(x_{1}\right) \cdots \phi_{j k_{d}}\left(x_{d}\right), \\
\Psi_{j \boldsymbol{k i}}(\boldsymbol{x}) & =\chi_{j k_{1}}^{i_{1}}\left(x_{1}\right) \cdots \chi_{j k_{d}}^{i_{d}}\left(x_{d}\right) . \tag{11}
\end{align*}
$$

In (11), we used

$$
\chi_{j k_{s}}^{i_{s}}(\cdot)=\left\{\begin{array}{ll}
\phi_{j k_{s}}(\cdot) & \text { if } \quad i_{s}=0, \\
\psi_{j k_{s}}(\cdot) & \text { if } \quad i_{s}=1,
\end{array} \quad s=1, \ldots, d\right.
$$

It follows from the following orthonormalities

$$
\begin{align*}
\int_{\mathcal{R}^{d}} \Phi_{j \boldsymbol{k}}(\boldsymbol{x}) \Phi_{j} \boldsymbol{m}(\boldsymbol{x}) \mathrm{d} \boldsymbol{x} & =\prod_{s=1}^{d} \int \phi_{j k_{s}}\left(x_{s}\right) \phi_{j m_{s}}\left(x_{s}\right) \mathrm{d} x_{s} \\
& =\delta_{\boldsymbol{k} \boldsymbol{m}}, \\
\int_{\mathcal{R}^{d}} \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}) \Psi_{l \boldsymbol{m i}} \boldsymbol{i}^{\prime}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x} & =\prod_{s=1}^{d} \int \chi_{j k_{s}}^{i_{s}}\left(x_{s}\right) \chi_{l m_{s}}^{i_{s}^{\prime}}\left(x_{s}\right) \mathrm{d} x_{s} \\
& =\delta_{j l} \delta_{\boldsymbol{k} \boldsymbol{m}} \delta_{\boldsymbol{i} \boldsymbol{i}^{\prime}}, \\
\int_{\mathcal{R}^{d}} \Phi_{j \boldsymbol{k}}(\boldsymbol{x}) \Psi_{l \boldsymbol{m} \boldsymbol{i}}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x} & =\prod_{s=1}^{d} \int \phi_{j k_{s}}\left(x_{s}\right) \chi_{l m_{s}}^{i_{s}}\left(x_{s}\right) \mathrm{d} x_{s} \\
& =0, \quad j \leq l, \tag{12}
\end{align*}
$$

that the coefficients of (10) are represented as

$$
\alpha_{j}^{f} \boldsymbol{k}=\int_{\mathcal{R}^{d}} f(\boldsymbol{x}) \Phi_{j \boldsymbol{k}}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}, \quad \beta_{j \boldsymbol{k} \boldsymbol{i}}^{f}=\int_{\mathcal{R}^{d}} f(\boldsymbol{x}) \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}
$$

The wavelets expansion of $g$ is similarly given.
It is easy to show that

$$
\begin{align*}
\hat{\alpha}_{j \boldsymbol{k}}^{f} & =\frac{1}{n} \sum_{i=1}^{n} \Phi_{j \boldsymbol{k}}\left(X_{i-d}, \ldots, X_{i-1}\right) \\
\hat{\beta}_{j \boldsymbol{k} \boldsymbol{i}}^{f} & =\frac{1}{n} \sum_{i=1}^{n} \Psi_{j \boldsymbol{k}}\left(X_{i-d}, \ldots, X_{i-1}\right) \tag{13}
\end{align*}
$$

are the unbiased estimators of $\alpha_{j}^{f} \boldsymbol{k}$ and $\beta_{j}^{f} \boldsymbol{k i}$.
Also

$$
\begin{align*}
\hat{\alpha}_{j \boldsymbol{k}}^{g} & =\frac{1}{n} \sum_{i=1}^{n} X_{i} \Phi_{j \boldsymbol{k}}\left(X_{i-d}, \ldots, X_{i-1}\right) \\
\hat{\beta}_{j \boldsymbol{k} \boldsymbol{i}}^{g} & =\frac{1}{n} \sum_{i=1}^{n} X_{i} \Psi_{j \boldsymbol{k}}\left(X_{i-d}, \ldots, X_{i-1}\right) \tag{14}
\end{align*}
$$

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

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

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

$$
\begin{align*}
\hat{f}(\boldsymbol{x})= & \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \hat{\alpha}_{j_{0}}^{f} \boldsymbol{k}^{\Phi_{j_{0}}} \boldsymbol{k}^{(\boldsymbol{x})} \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j}^{f} \boldsymbol{k} \boldsymbol{i}_{j} \Psi_{j \boldsymbol{i}}(\boldsymbol{x}), \\
\hat{g}(\boldsymbol{x})= & \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \hat{\alpha}_{j_{0}}^{g} \boldsymbol{k}^{\Phi_{j_{0}} \boldsymbol{k}^{(\boldsymbol{x})}} \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j}^{g} \boldsymbol{k} \boldsymbol{i} \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}), \tag{15}
\end{align*}
$$

Also the threshold wavelet estimators are given as follows;

$$
\begin{align*}
\hat{f}_{\lambda}(\boldsymbol{x})= & \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \hat{\alpha}_{j_{0}}^{f} \boldsymbol{k}^{\Phi_{j_{0}} \boldsymbol{k}}(\boldsymbol{x}) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \eta_{\lambda}\left(\hat{\beta}_{j}^{f} \boldsymbol{k} \boldsymbol{i}\right) \Psi_{j \boldsymbol{k}}(\boldsymbol{x}), \\
\hat{g}_{\lambda}(\boldsymbol{x})= & \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \hat{\alpha}_{j_{0}}^{g} \boldsymbol{k}^{\Phi_{j_{0}} \boldsymbol{k}}(\boldsymbol{x}) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \mathcal{Z}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \eta_{\lambda}\left(\hat{\beta}_{j \boldsymbol{k} \boldsymbol{i}}^{g}\right) \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}) . \tag{16}
\end{align*}
$$

## 5. ASYMPTOTIC PROPERTY

In this section, we consider the convergence of multivariate wavelet estimators for non-stationary time series that satisfy $\alpha$-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 $\alpha$-mixing settings. We derive a similar rate for our estimators under non-stationarity.

To begin with we define $\alpha$-mixing condition for $d$ variables case.

Definition 2 Time series $\left\{X_{t}\right\}$ is $\alpha$-mixing when

$$
\lim _{u \rightarrow \infty} \alpha(u)=0
$$

where

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

where $\mathcal{F}_{j}$ is the $\sigma$-algebra generated by $\left(X_{i-d}, \ldots, X_{i}\right),-\infty<i \leq j$ and $\mathcal{F}^{j}$ is the $\sigma$-algebra generated by $\left(X_{i-d}, \ldots, X_{i}\right), j \leq i<\infty$.

The next lemma is useful for studying the $\alpha$-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 $|X| \leq C_{1},|Y| \leq C_{2}$, respectively. Then,

$$
|\operatorname{Cov}(X, Y)| \leq 4 C_{1} C_{2} \alpha(u)
$$

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) $\mu$ and $f_{t}(t \in \mathcal{Z})$ belong to $C^{r}$-class and their derivatives up to $r$ th order are all bounded and continuous on $\mathcal{R}^{d}$.
(b) There exists bounded and closed interval $Q \subset \mathcal{R}^{d}$ such that $\operatorname{supp}\left(f_{t}\right) \subset Q$ for all $t \in \mathcal{Z}$, where $\operatorname{supp}\left(f_{t}\right)$ is the support of $f_{t}$.
(c) There exists a subset $U \subset Q$ and a positive constant $M$ such that for all $t \in \mathcal{Z}$, $M^{-1} \leq f_{t}(x) \leq M$ for $x \in U$.
(d) $\phi$ and $\psi$ have compact support on $\mathcal{R}$.
(e) It holds that $\int x^{v} \psi(x) \mathrm{d} x=0(0 \leq v \leq r-1)$ and $(r!)^{-1} \int x^{r} \psi(x) \mathrm{d} x<\infty$.
(f) $\alpha(u)=O\left(\rho^{u}\right)$ as $u \rightarrow \infty$ for some $0<\rho<1$.

Theorem 1 Assume the $\alpha$-mixing condition and conditions (a), (b), (c), (d), (e) and $(f)$. Furthermore, suppose $j_{0}, j_{1} \rightarrow \infty$ as $n \rightarrow \infty$ in such a manner that $2^{d j_{0}}, 2^{d j_{1}}=o(n)$. Then
(i) Linear wavelet estimator $\hat{g}$ of $g$ given in (15) satisfies,

$$
\begin{aligned}
E\left[\int_{\mathcal{R}^{d}}(\hat{g}-g)^{2} \mathbf{d} \boldsymbol{x}\right] & =O\left(\frac{2^{d j_{0}}}{n}+\frac{2^{d j_{1}}}{n}+2^{-2 j_{1} r}\right) \\
& =o(1) \text { as } n \rightarrow \infty
\end{aligned}
$$

(ii) If $2^{d j_{0}}, 2^{d j_{1}}=o\left(n^{1 / 3}\right)$, then estimator $\hat{\mu}=\hat{g} / \hat{f}$ of $\mu$ 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

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

for arbiterary $j_{0}$ and $j_{1}$.
proof). Proof is given in Appendix.
Theorem 2 Assume the $\alpha$-mixing condition and conditions (a), (b), (c), (d), (e) and $(f)$. Furthermore, suppose that $d<r$ and $j_{0}, j_{1} \rightarrow \infty$ as $n \rightarrow \infty$ in such a manner that $2^{(d-r) j_{0}}, 2^{(d-r) j_{1}}=o(1)$. Then as $n \rightarrow \infty$ we have for $\boldsymbol{x} \in U$

$$
\begin{aligned}
& E(\hat{f}(\boldsymbol{x}))-f(\boldsymbol{x})=O\left(2^{(d-r) j_{0}}\right)=o(1) \\
& E(\hat{g}(\boldsymbol{x}))-g(\boldsymbol{x})=O\left(2^{(d-r) j_{1}}\right)=o(1) \\
& E(\mu(\boldsymbol{x}))-\mu(\boldsymbol{x})=o(1)
\end{aligned}
$$

proof). Proof is given in Appendix.

## 6. SELECTION OF PARAMETERS

In this section, we discuss a criterion for choosing the optimum $\theta=\left(j_{0}, j_{1}, \lambda\right)$ involved in the estimator by assuming $\alpha$-mixing condition.

### 6.1 Basic Criterion

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

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

where $\hat{\mu}_{\theta}(\cdot)$ is the estimator of $\mu$ with $\theta=\left(j_{0}, j_{1}, \lambda\right) \in \Theta$ based on given data $\left\{\left(X_{1}, \boldsymbol{X}_{0}\right), \ldots,\left(X_{n}, \boldsymbol{X}_{n-1}\right)\right\}$. Define the residual sum of squares by

$$
\operatorname{RSS}(\theta)=\frac{1}{n} \sum_{i=1}^{n}\left[X_{i}-\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right]^{2}
$$

then we have the following lemma.

## Lemma 2

$$
\begin{equation*}
E[\operatorname{RSS}(\theta)]=\frac{1}{n} \sum_{i=1}^{n} \sigma_{i}^{2}+\mathrm{R}(\theta)-\frac{2}{n} \sum_{i=1}^{n} \operatorname{Cov}\left\{\varepsilon_{i}, \hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\}, \tag{17}
\end{equation*}
$$

where $\sigma_{i}^{2}$ denote the variance of $\varepsilon_{i}$.
proof).

$$
\begin{aligned}
E[\operatorname{RSS}(\theta)]= & \frac{1}{n} \sum_{i=1}^{n} E\left[X_{i}-\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right]^{2} \\
= & \frac{1}{n} \sum_{i=1}^{n} E\left[\mu\left(\boldsymbol{X}_{i-1}\right)+\varepsilon_{i}-\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right]^{2} \\
= & \frac{1}{n} \sum_{i=1}^{n} E\left[\varepsilon_{i}^{2}+\left\{\mu\left(\boldsymbol{X}_{i-1}\right)-\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\}^{2}\right. \\
& \left.+2 \varepsilon_{i}\left\{\mu\left(\boldsymbol{X}_{i-1}\right)-\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\}\right] \\
= & \frac{1}{n} \sum_{i=1}^{n} \sigma_{i}^{2}+\mathrm{R}(\theta)+\frac{2}{n} \sum_{i=1}^{n} E\left\{\varepsilon_{i} \mu\left(\boldsymbol{X}_{i-1}\right)\right\} \\
& -\frac{2}{n} \sum_{i=1}^{n} E\left\{\varepsilon_{i} \hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\} .
\end{aligned}
$$

Since $\varepsilon_{i}$ and $\boldsymbol{X}_{i-1}$ are independent, and $\hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)$ is the random variable of $\left(X_{1-d}, \ldots\right.$, $\left.X_{n}\right)$, it hold that $E\left\{\varepsilon_{i} \mu\left(\boldsymbol{X}_{i-1}\right)\right\}=0$ and $E\left\{\varepsilon_{i} \hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\}=\operatorname{Cov}\left\{\varepsilon_{i}, \hat{\mu}_{\theta}\left(\boldsymbol{X}_{i-1}\right)\right\}$. Note that, the first term in the right hand side of (17) is independent of parameter $\theta$, however the third term is not, so $\arg \min _{\theta}\{\operatorname{RSS}(\theta)\}$ 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 $\alpha$-mixing Time Series

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

Let $\hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}(\cdot)$ be the estimator from the data which delete $M_{n}$ data points $\left\{\left(X_{i}, \boldsymbol{X}_{i-1}\right), \ldots,\left(X_{i+M_{n}-1}, \boldsymbol{X}_{i+M_{n}-2}\right)\right\}$ from an original data $\left\{\left(X_{1}, \boldsymbol{X}_{0}\right), \ldots,\left(X_{n}, \boldsymbol{X}_{n-1}\right)\right\}$ for each $i$, and let

$$
\mathrm{CV}_{M_{n}}(\theta)=\frac{1}{n} \sum_{i=1}^{n}\left[X_{i}-\hat{\mu}_{\theta}^{\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right]^{2} .
$$

We have the following theorem.
Theorem 3 Assume $\left\{X_{t}\right\}$ be the $\alpha$-mixing time series generated by model $X_{t}=$ $\mu\left(\boldsymbol{X}_{t-1}\right)+\varepsilon_{t}$. Set $M_{n}=o(n)$ as $n \rightarrow \infty$, then

$$
E\left[\mathrm{CV}_{M_{n}}(\theta)\right]=\frac{1}{n} \sum_{i=1}^{n} \sigma_{i}^{2}+\mathrm{R}(\theta)+o(1)
$$

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

$$
\arg \min _{\theta}\left\{\mathrm{CV}_{M_{n}}(\theta)\right\} \longrightarrow \arg \min _{\theta}\{\mathrm{R}(\theta)\} \text { as } n \rightarrow \infty
$$

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 $\left(X_{1}, \ldots, X_{631}\right)$ sampled by 200 points per second frequency.


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

$$
X_{t}=\mu\left(X_{t-2}, X_{t-1}\right)+\varepsilon_{t}
$$

and estimate $\mu(x, y)$ from $\left\{X_{3},\left(X_{1}, X_{2}\right)\right\}, \ldots,\left\{X_{631},\left(X_{629}, X_{630}\right)\right\} . \hat{f}$ and $\hat{g}$ in (16) are represented as follows,

$$
\begin{aligned}
\hat{f}_{\theta}(x, y)= & \sum_{k_{1} k_{2}} \hat{\alpha}_{j_{0}\left(k_{1}, k_{2}\right)}^{f} \phi_{j_{0} k_{1}}(x) \phi_{j_{0} k_{2}}(y) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(0,1)}^{f}\right) \phi_{j k_{1}}(x) \psi_{j k_{2}}(y) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(1,0)}^{f}\right) \psi_{j k_{1}}(x) \phi_{j k_{2}}(y) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(1,1)}^{f}\right) \psi_{j k_{1}}(x) \psi_{j k_{2}}(y), \\
\hat{g}_{\theta}(x, y)= & \sum_{k_{1} k_{2}}^{\hat{\alpha}_{j_{0}\left(k_{1}, k_{2}\right)}^{g} \phi_{j_{0} k_{1}}(x) \phi_{j_{0} k_{2}}(y)} \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(0,1)}^{g}\right) \phi_{j k_{1}}(x) \psi_{j k_{2}}(y) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(1,0)}^{g}\right) \psi_{j k_{1}}(x) \phi_{j k_{2}}(y) \\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{k_{1} k_{2}} \eta_{\lambda}\left(\hat{\beta}_{j\left(k_{1}, k_{2}\right)(1,1)}^{g}\right) \psi_{j k_{1}}(x) \psi_{j k_{2}}(y),
\end{aligned}
$$

where $\theta=\left(j_{0}, j_{1}, \lambda\right)$.
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,

$$
\hat{\mu}_{\theta}(x, y)=\left\{\begin{array}{cll}
\hat{g}_{\theta}(x, y) / \hat{f}_{\theta}(x, y) & \text { if } & \left|\hat{f}_{\theta}(x, y)\right|>\delta \\
0 & \text { if } & \left|\hat{f}_{\theta}(x, y)\right| \leq \delta
\end{array}\right.
$$

### 7.3 Cross-Validation

For the estimation of the parameter $\theta=\left(j_{0}, j_{1}, \lambda\right)$ when $\eta_{\lambda}(t)=\lambda t$, we compute the value of

$$
\operatorname{CV}(\eta, \theta)=\frac{1}{629} \sum_{i=3}^{631}\left[X_{i}-\hat{\mu}_{\eta \theta}^{\backslash i}\left(X_{i-2}, X_{i-1}\right)\right]^{2}
$$



Figure 4: Perspective plots of $\hat{\mu}(x, y)$
for combinations of $j_{0}, j_{1}$ and $\lambda$ and select the $\theta$ 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 $\lambda$ 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 $\lambda$ 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 $\left\{\hat{\mu}_{\theta}(x, y),(x, y)\right\}$ obtained when $\delta=0.001$.

Table 1: Cross-Validation $\left(j_{1}=6\right)$

|  | $\lambda$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
| $j_{0}=1$ | hard | 2.367 | 1.425 | 0.318 | 1.440 | 0.480 | 16.631 | 16.181 | 0.383 | 0.383 |
|  | 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 |
|  |  |  |  |  |  |  |  |  |  |  |
| $j_{0}=2$ | hard | 2.714 | 1.314 | 0.197 | 0.197 | 0.234 | 3.320 | 1.528 | 0.210 | 0.210 |
|  | 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 |
|  |  |  |  |  |  |  |  |  |  |  |
| $j_{0}=3$ | hard | 6.640 | 1.431 | 0.288 | 0.288 | 0.286 | 1.114 | 0.282 | 0.282 | 0.282 |
|  | 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 $\left\{Y_{t}\right\}$ using $\hat{\mu}(x, y)$ as follows,

$$
\begin{aligned}
& Y_{3}=\hat{\mu}\left(Y_{1}, Y_{2}\right) \\
& Y_{4}=\hat{\mu}\left(Y_{2}, Y_{3}\right)
\end{aligned}
$$

by giving an appropriate initial value for $\left(Y_{1}, Y_{2}\right)$. Figure 5 gives the plot of $\left\{Y_{t}\right\}$ 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.


## 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 $\alpha$-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 $\mu$.

## APPENDIX: PROOFS

## Proof of Theorem 1 (i)

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

$$
\Omega_{j}^{d}=\left\{\boldsymbol{k}=\left(k_{1}, \ldots, k_{d}\right) \in \mathcal{Z}^{d}: 0 \leq k_{1}, \ldots, k_{d} \leq 2^{j}-1\right\} .
$$

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

$$
\begin{align*}
\hat{g}(\boldsymbol{x})-g(\boldsymbol{x})= & \sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left(\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x})  \tag{A.1}\\
& +\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}}\left(\hat{\beta}_{j \boldsymbol{k}}-\beta_{j \boldsymbol{k}}\right) \Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x}) \\
& -\sum_{j \geq j_{1}} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \beta_{j} \boldsymbol{k} \boldsymbol{i} \Psi_{j} \boldsymbol{k} \boldsymbol{i}(\boldsymbol{x}), \quad \boldsymbol{x} \in Q .
\end{align*}
$$

According to the orthonormality of wavelets (12),

$$
\begin{align*}
\int_{Q}(\hat{g}-g)^{2} \mathbf{d} \boldsymbol{x} & =S_{1}+S_{2}+S_{3}, \\
S_{1} & =\sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left(\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right)^{2}, \\
S_{2} & =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i}_{\in \mathcal{I}}}\left(\hat{\mathcal{I}}_{j} \boldsymbol{k} \boldsymbol{i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right)^{2}, \\
S_{3} & =\sum_{j \geq j_{1}} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{i \in \mathcal{I}^{d}} \beta_{j}^{2} \boldsymbol{k} \boldsymbol{i} \tag{A.2}
\end{align*}
$$

We evaluate $E\left(S_{1}\right), E\left(S_{2}\right), E\left(S_{3}\right)$ in (A.2) respectively. It follows from

$$
E\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\}=\int_{\mathcal{R}^{d}} \mu(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) f_{i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}
$$

that

$$
\begin{align*}
E\left(\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right)^{2}= & E\left[\frac{1}{n} \sum_{i=1}^{n} X_{i} \Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{i-1}\right)-\frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{R}^{d}} \mu(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) f_{i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}\right]^{2}} \begin{array}{rl}
= & \frac{1}{n^{2}} E\left[\sum_{i=1}^{n}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)-\int_{\mathcal{R}^{d}} \mu(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) f_{i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}\right\}\right]^{2} \\
= & \frac{1}{n^{2}} \sum_{i=1}^{n} \operatorname{Var}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} \\
& +\frac{2}{n^{2}} \sum_{i<j} \operatorname{Cov}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right), X_{j} \Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{j-1}\right)\right\} .}\right.
\end{array},\right. \text { A.3 }
\end{align*}
$$

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

$$
\begin{equation*}
\operatorname{Var}\left\{X_{i} \Phi_{j_{0} \boldsymbol{k}}\left(\boldsymbol{X}_{i-1}\right)\right\}=E\left\{X_{i} \Phi_{j_{0} \boldsymbol{k}}\left(\boldsymbol{X}_{i-1}\right)\right\}^{2}-\left[E\left\{X_{i} \Phi_{j_{0} \boldsymbol{k}}\left(\boldsymbol{X}_{i-1}\right)\right\}\right]^{2}, \tag{A.4}
\end{equation*}
$$

where

$$
\begin{align*}
E\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} & =\int_{\mathcal{R}^{d}} \mu(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) f_{i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x} \\
& =2^{-d j_{0} / 2} \int_{\mathcal{R}^{d}} \mu\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_{0}}}\right) \phi\left(x_{1}\right) \cdots \phi\left(x_{d}\right) f_{i}\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_{0}}}\right) \mathbf{d} \boldsymbol{x} \\
& =O\left(2^{-d j_{0} / 2}\right) \tag{A.5}
\end{align*}
$$

and

$$
\begin{align*}
E\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\}^{2} & \leq \text { Const. } E\left\{\Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\}^{2} \\
& =\text { Const. } \int_{\mathcal{R}^{d}} \Phi_{j_{0}}^{2} \boldsymbol{k}(\boldsymbol{x}) f_{i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x} \\
& =\text { Const. } \int_{\mathcal{R}^{d}}\left\{\phi\left(x_{1}\right) \cdots \phi\left(x_{d}\right)\right\}^{2} f_{i}\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_{0}}}\right) \mathbf{d} \boldsymbol{x} \\
& =O(1) . \tag{A.6}
\end{align*}
$$

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

$$
\begin{align*}
\operatorname{Var}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} & =O(1) \\
\frac{1}{n^{2}} \sum_{i=1}^{n} \operatorname{Var}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} & =O\left(n^{-1}\right) \tag{A.7}
\end{align*}
$$

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

$$
h_{i j}(\boldsymbol{x}, \boldsymbol{y})=f_{i j}(\boldsymbol{x}, \boldsymbol{y})-f_{i}(\boldsymbol{x}) f_{j}(\boldsymbol{y}),
$$

then

$$
\begin{align*}
\operatorname{Cov}\{ & X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right), X_{j} \Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{j-1}\right)\right\}} \\
= & \text { Const. }\left[E \left\{\Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{i-1}\right) \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)\right\}}\right.\right. \\
& \left.-E\left\{\Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} E\left\{\Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)\right\}\right] \\
= & \text { Const. } \int_{\mathcal{R}^{d}} \int_{\mathcal{R}^{d}} \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{y}) h_{i j}(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d} \boldsymbol{x} \mathrm{~d} \boldsymbol{y} \\
= & \text { Const. } \int_{\mathcal{R}^{d}} \int_{\mathcal{R}^{d}} 2^{-d j_{0}} \phi\left(x_{1}\right) \cdots \phi\left(x_{d}\right) \phi\left(y_{1}\right) \cdots \phi\left(y_{d}\right) \\
& \times h_{i j}\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_{0}}}, \frac{\boldsymbol{y}+\boldsymbol{k}}{2^{j_{0}}}\right) \mathbf{d} \boldsymbol{x} \mathrm{d} \boldsymbol{y} \\
= & O\left(2^{-d j 0}\right) . \tag{A.8}
\end{align*}
$$

On the other hand, from $\alpha$-mixing condition, we can also evaluate the covariance term of (A.3) as follows. Since $X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)$ is $\mathcal{F}_{i}$-measurable, and $X_{j} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)$ is $\mathcal{F}^{j}$-measurable, it follows from lemma 1 that

$$
\begin{align*}
\operatorname{Cov}\{ & X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right), X_{j} \Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{j-1}\right)\right\}} \\
= & \text { Const. }\left[E \left\{\Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{i-1}\right) \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)\right\}}\right.\right. \\
& \left.-E\left\{\Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right)\right\} E\left\{\Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)\right\}\right] \\
= & \text { Const. } \int_{\mathcal{R}^{d}} \int_{\mathcal{R}^{d}} \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{x}) \Phi_{j_{0}} \boldsymbol{k}(\boldsymbol{y}) h_{i j}(\boldsymbol{x}, \boldsymbol{y}) \mathbf{d} \boldsymbol{x} \mathbf{d} \boldsymbol{y} \\
= & \text { Const. } \int_{\mathcal{R}^{d}} \int_{\mathcal{R}^{d}} 2^{-d j_{0}} \phi\left(x_{1}\right) \cdots \phi\left(x_{d}\right) \phi\left(y_{1}\right) \cdots \phi\left(y_{d}\right) \\
& \times h_{i j}\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j_{0}}}, \frac{\boldsymbol{y}+\boldsymbol{k}}{2^{j_{0}}}\right) \mathbf{d} \boldsymbol{x} \mathbf{d} \boldsymbol{y} \\
= & O\left(2^{-d j_{0}}\right) . \tag{A.9}
\end{align*}
$$

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

$$
\begin{align*}
& \frac{2}{n^{2}} \sum_{i<j} \operatorname{Cov}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i-1}\right), X_{j} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{j-1}\right)\right\} \\
& \quad=\frac{2}{n^{2}} \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \operatorname{Cov}\left\{X_{i} \Phi_{j_{0}} \boldsymbol{k}^{\left.\left(\boldsymbol{X}_{i-1}\right), X_{i+j} \Phi_{j_{0}} \boldsymbol{k}\left(\boldsymbol{X}_{i+j-1}\right)\right\}}\right. \\
& \quad=\frac{2}{n^{2}} \sum_{i=1}^{n-1}\left\{\sum_{j \leq M_{n}} O\left(2^{-d j_{0}}\right)+\sum_{j>M_{n}} O\left(2^{d j_{0}} \alpha(j)\right)\right\} \\
& \quad \leq \frac{1}{n} M_{n} O\left(2^{-d j_{0}}\right)+O\left(2^{d j_{0}} \alpha\left(M_{n}\right)\right) \\
&  \tag{A.10}\\
& \quad=o\left(n^{-1}\right),
\end{align*}
$$

where we set $M_{n}=o\left(2^{d j_{0}}\right)=o(n)$ as $n \rightarrow \infty$, and used $\alpha(u)=O\left(\rho^{u}\right),(0<\rho<1)$. Combining (A.3), (A.7) and (A.10), we can deduce that

$$
\begin{align*}
E\left(S_{1}\right) & =\sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left[O\left(n^{-1}\right)+o\left(n^{-1}\right)\right] \\
& =O\left(2^{d j_{0}} n^{-1}\right) . \tag{A.11}
\end{align*}
$$

Similar evaluation to $E\left(S_{1}\right)$ yields

$$
\begin{equation*}
E\left(S_{2}\right)=O\left(2^{d j_{1}} n^{-1}\right) \tag{A.12}
\end{equation*}
$$

It follows that

$$
\begin{align*}
\beta_{j} \boldsymbol{k i} & =\int_{\mathcal{R}^{d}} g(\boldsymbol{x}) \Psi_{j} \boldsymbol{k i}(\boldsymbol{x}) \mathbf{d} \boldsymbol{x}  \tag{A.13}\\
& =2^{-d j / 2} \int_{\mathcal{R}^{d}} g\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j}}\right) \chi^{i_{1}}\left(x_{1}\right) \cdots \chi^{i_{d}}\left(x_{d}\right) \mathbf{d} \boldsymbol{x} .
\end{align*}
$$

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

$$
\begin{align*}
\beta_{j} \boldsymbol{k i}= & 2^{-d j / 2} \sum_{v=0}^{r-1} \frac{1}{v!} \sum_{\left(u_{1} \ldots u_{d}\right) \in U_{v}^{d}}\binom{v}{u_{1}, \ldots, u_{d}} \\
& \times \frac{\partial^{r} g}{\partial^{u_{1}} x_{1} \cdots \partial^{u_{d}} x_{d}}\left(\frac{\boldsymbol{k}}{2^{j}}\right) \\
& \times \int_{\mathcal{R}^{d}} 2^{-v j} x_{1}^{u_{1}} \cdots x_{d}^{u_{d}} \chi^{i_{1}}\left(x_{1}\right) \cdots \chi^{i_{d}}\left(x_{d}\right) \mathbf{d} \boldsymbol{x} \\
& +2^{-d j / 2} \int_{\mathcal{R}^{d}} R_{r}\left(\frac{\boldsymbol{x}+\boldsymbol{k}}{2^{j}}\right) \\
& \times \chi^{i_{1}}\left(x_{1}\right) \cdots \chi^{i_{d}}\left(x_{d}\right) \mathbf{d} \boldsymbol{x} \tag{A.14}
\end{align*}
$$

where $R_{r}(\cdot)$ is residual term of $r$ th order, and
$U_{v}^{d}=\left\{\left(u_{1}, \ldots, u_{d}\right) \in \mathcal{Z}^{d}: 0 \leq u_{1}, \ldots, u_{d} \leq v \quad\right.$ and $\left.\quad u_{1}+\cdots+u_{d}=v\right\}$.
Only the $r$ th term of (A.14) remains because $\int x^{v} \psi(x) \mathrm{d} x=0, \quad(0 \leq v \leq r-1)$.
Hence

$$
\begin{aligned}
\beta_{j \boldsymbol{k i}}= & 2^{-j(r+d / 2)} \frac{1}{r!} \sum_{\left(u_{1} \ldots u_{d}\right) \in U_{r}^{d}}\binom{r}{u_{1}, \ldots, u_{d}} \\
& \times \int_{\mathcal{R}^{d}}\left\{\frac{\partial^{r} g}{\partial^{u_{1}} x_{1} \cdots \partial^{u_{d}} x_{d}}\left(\frac{\boldsymbol{k}+\theta \boldsymbol{x}}{2^{j}}\right)\right\} \\
& \times x_{1}^{u_{1}} \cdots x_{d}^{u_{d}} \chi^{i_{1}}\left(x_{1}\right) \cdots \chi^{i_{d}}\left(x_{d}\right) \mathbf{d} \boldsymbol{x} .
\end{aligned}
$$

Now, let

$$
\begin{aligned}
\xi_{j \boldsymbol{k}}^{\left(u_{1} \ldots u_{d}\right)}= & \frac{\partial^{r} g}{\partial^{u_{1}} x_{1} \cdots \partial^{u_{d}} x_{d}}\left(\frac{\boldsymbol{k}+\theta \boldsymbol{x}}{2^{j}}\right) \\
& -\frac{\partial^{r} g}{\partial^{u_{1}} x_{1} \cdots \partial^{u_{d}} x_{d}}\left(\frac{\boldsymbol{k}}{2^{j}}\right) \text { and } \\
g_{j \boldsymbol{k}}^{\left(u_{1} \ldots u_{d}\right)}= & \frac{\partial^{r} g}{\partial^{u_{1}} x_{1} \cdots \partial^{u_{d}} x_{d}}\left(\frac{\boldsymbol{k}}{2^{j}}\right),
\end{aligned}
$$

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

$$
\begin{align*}
\beta_{j \boldsymbol{k} \boldsymbol{i}}= & 2^{-j(r+d / 2)} \frac{1}{r!} \sum_{\left(u_{1} \ldots u_{d}\right) \in U_{r}^{d}}\binom{r}{u_{1}, \ldots, u_{d}}  \tag{A.15}\\
& \times\left\{g_{j \boldsymbol{k}}^{\left(u_{1} \ldots u_{d}\right)}+o(1)\right\} \prod_{s=1}^{d} \int x_{s}^{u_{s}} \chi^{i_{s}}\left(x_{s}\right) \mathrm{d} x_{s} \\
= & \left\{\begin{array}{ccc}
O\left(2^{-j(r+d / 2)}\right) & \text { if } & i_{1}+i_{2}+\cdots+i_{d}=1, \\
0 & \text { if } & i_{1}+i_{2}+\cdots+i_{d}>1 .
\end{array}\right.
\end{align*}
$$

Hence,

$$
\begin{equation*}
S_{3}=\sum_{j=j_{1}}^{\infty} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \beta_{j}^{2} \boldsymbol{k} \boldsymbol{i}=\sum_{j=j_{1}}^{\infty} 2^{d j} d O\left(2^{-j(2 r+d)}\right)=O\left(2^{-2 j_{1} r}\right) . \tag{A.16}
\end{equation*}
$$

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

$$
\begin{aligned}
E\left[\int_{\mathcal{R}^{d}}(\hat{g}-g)^{2} \mathbf{d} \boldsymbol{x}\right] & =E\left(S_{1}+S_{2}+S_{3}\right) \\
& =O\left(\frac{2^{d j_{0}}}{n}+\frac{2^{d j_{1}}}{n}+2^{-2 j_{1} r}\right) \text { as } n \rightarrow \infty .
\end{aligned}
$$

## Proof of Theorem 1 (ii)

At first, we prove the following result.

$$
\begin{equation*}
\sup _{\boldsymbol{x} \in U}|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|=o(1) \quad \text { a.e. } \tag{A.17}
\end{equation*}
$$

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

$$
\sup _{U}|\hat{f}-f| \leq s_{1}+s_{2}+s_{3},
$$

where

$$
\begin{aligned}
& s_{1}=\sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left|\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right|\left\|\Phi_{j_{0} \boldsymbol{k}}\right\|_{\infty}, \\
& s_{2}=\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}}\left|\hat{\beta}_{j} \boldsymbol{k i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right|\left\|\Psi_{j} \boldsymbol{k} \boldsymbol{i}\right\|_{\infty}, \\
& s_{3}=\sum_{j \geq j_{1}} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}}\left|\beta_{j} \boldsymbol{k} \boldsymbol{i}\right|\left\|\Psi_{j} \boldsymbol{k} \boldsymbol{i}\right\|_{\infty} .
\end{aligned}
$$

Note that $\left\|\Phi_{j_{0}} \boldsymbol{k}\right\|_{\infty}=O\left(2^{d j / 2}\right)$ and $\left\|\Psi_{j} \boldsymbol{k i}\right\|_{\infty}=O\left(2^{d j / 2}\right)$. It follows from

$$
\left(\frac{s_{1}}{2^{d j_{0}}}\right)^{2} \leq \frac{1}{2^{d j_{0}}} \sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left(\left|\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right|\left\|\Phi_{j_{0}} \boldsymbol{k}\right\|_{\infty}\right)^{2}
$$

that

$$
s_{1}=O\left(2^{d j_{0}}\right)\left\{\sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left(\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right)^{2}\right\}^{1 / 2}
$$

By (A.11), for all sufficiently large $n$, there exists a positive constant $B_{1}$ such that

$$
n 2^{-d j_{0}} \sum_{\boldsymbol{k} \in \Omega_{j_{0}}^{d}}\left(\hat{\alpha}_{j_{0}} \boldsymbol{k}-\alpha_{j_{0}} \boldsymbol{k}\right)^{2}<B_{1} \quad \text { a.e. }
$$

Hence,

$$
s_{1}<n^{-1 / 2} 2^{d j_{0} / 2} B_{1}^{1 / 2} O\left(2^{d j_{0}}\right) \quad \text { a.e }
$$

which together with $2^{d j_{0}}=o\left(n^{1 / 3}\right)$ yields that

$$
\begin{equation*}
s_{1}=o(1) \quad \text { a.e. } \tag{A.18}
\end{equation*}
$$

Similally, it follows from (A.12) and $2^{d j_{1}}=o\left(n^{1 / 3}\right)$ that

$$
\begin{equation*}
s_{2}=o(1) \quad \text { a.e. } \tag{A.19}
\end{equation*}
$$

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

$$
\begin{equation*}
s_{3}=o(1) \tag{A.20}
\end{equation*}
$$

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 $E\left\{\int(\hat{g}-g)^{2}\right\}=o(1)$, for all sufficiently large $n$, there exists a positive constants $B_{2}$ such that

$$
\int(\hat{g}-g)^{2}<B_{2} \quad \text { a.e }
$$

and then, it follows that

$$
\begin{aligned}
& \operatorname{Var}\left\{\int(\hat{g}-g)^{2}\right\} \\
& \quad=E\left[\left\{\int(\hat{g}-g)^{2}\right\}^{2}\right]-\left[E\left\{\int(\hat{g}-g)^{2}\right\}\right]^{2} \\
& \quad<B_{2} E\left\{\int(\hat{g}-g)^{2}\right\}+o(1) \\
& \quad=o(1)
\end{aligned}
$$

Therefore, by Chebyshev's inequality,

$$
\begin{equation*}
\int(\hat{g}-g)^{2}=o_{p}(1) \tag{A.21}
\end{equation*}
$$

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

$$
\hat{\mu}-\mu=\hat{f}^{-1}(\hat{g}-\mu \hat{f})=\left\{1+o_{p}(1)\right\} f^{-1}(\hat{g}-\mu \hat{f}) \quad \text { on } U .
$$

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

$$
\hat{g}-\mu \hat{f}=\hat{g}-\left\{1+o_{p}(1)\right\} g=\left\{1+o_{p}(1)\right\}(\hat{g}-g),
$$

and that

$$
\begin{equation*}
\hat{\mu}-\mu=\left\{1+o_{p}(1)\right\} f^{-1}(\hat{g}-g) \quad \text { on } U . \tag{A.22}
\end{equation*}
$$

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

$$
\begin{aligned}
\int_{U}(\hat{\mu}-\mu)^{2} & =\left\{1+o_{p}(1)\right\} \int_{U} f^{-2}(\hat{g}-g)^{2} \\
& \leq\left\{1+o_{p}(1)\right\} M^{2} \int_{U}(\hat{g}-g)^{2} \\
& =o_{p}(1) .
\end{aligned}
$$

## Proof of Theorem 1 (iii)

We have the following representation.

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

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

$$
\begin{aligned}
S_{\lambda 2} & =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{i \in \mathcal{I}^{d}}\left\{\eta_{\lambda}\left(\hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}\right)-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right\}^{2} \\
& =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}}\left(\lambda \hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right)^{2} \\
& =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}}\left\{(\lambda-1) \hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}+\left(\hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right)\right\}^{2} .
\end{aligned}
$$

It follows that

$$
\begin{aligned}
S_{2}-S_{\lambda 2}= & -(\lambda-1)^{2} \sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j}^{2} \boldsymbol{k} \boldsymbol{i} \\
& -2(\lambda-1) \sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}\left(\hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right) .
\end{aligned}
$$

Now, let

$$
\begin{aligned}
a & =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j}^{2} \boldsymbol{k} \boldsymbol{i} \\
b & =\sum_{j=j_{0}}^{j_{1}-1} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}\left(\hat{\beta}_{j} \boldsymbol{k} \boldsymbol{i}-\beta_{j} \boldsymbol{k} \boldsymbol{i}\right),
\end{aligned}
$$

then we have

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

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

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

Hence, we have

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

## Proof of Theorem 2

From (A.2) we have

Now it follows from the definition that

$$
\Psi_{j \boldsymbol{k} \boldsymbol{i}}(\boldsymbol{x})=O\left(2^{d j / 2}\right)
$$

Thus together with (A.16) we have

$$
\begin{aligned}
\sum_{j \geq j_{1}} \sum_{\boldsymbol{k} \in \Omega_{j}^{d}} \sum_{\boldsymbol{i} \in \mathcal{I}^{d}} \beta_{j} \boldsymbol{k} \boldsymbol{i} \Psi_{j \boldsymbol{k i}}(\boldsymbol{x}) & =\sum_{j \geq j_{1}} O\left(2^{(d-r) j}\right) \\
& =O\left(2^{(d-r) j_{1}}\right)
\end{aligned}
$$

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

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

and the rest of the proof is trivial.

## Proof of Theorem 3

Similarly as the proof of Lemma 2,

$$
\begin{align*}
& E\left[\mathrm{CV}_{M_{n}}(\theta)\right] \\
& =\frac{1}{n} \sum_{i=1}^{n} \sigma_{i}^{2}+\frac{1}{n} \sum_{i=1}^{n} E\left[\mu\left(\boldsymbol{X}_{i-1}\right)-\hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right]^{2} \\
& \quad-\frac{2}{n} \sum_{i=1}^{n} \operatorname{Cov}\left\{\varepsilon_{i}, \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right\} . \tag{A.23}
\end{align*}
$$

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

$$
f^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}, x_{i+M_{n}-d}, \ldots, x_{n}\right)
$$

be the joint p.d.f of $\left(X_{1-d}, \ldots, X_{i-1}, \varepsilon_{i}, X_{i+M_{n}-d}, \ldots, X_{n}\right)$, and let

$$
f_{1}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}\right) \quad \text { and } \quad f_{2}^{\backslash\left(i, M_{n}\right)}\left(x_{i+M_{n}-d}, \ldots, x_{n}\right)
$$

be the joint p.d.f of $\left(X_{1-d}, \ldots, X_{i-1}, \varepsilon_{i}\right)$ and $\left(X_{i+M_{n}-d}, \ldots, X_{n}\right)$ respectively. Since $\left(X_{1-d}, \ldots, X_{i-1}, \varepsilon_{i}\right)$ is $\mathcal{F}_{i}$-measurable, and $\left(X_{i+M_{n}-d}, \ldots, X_{n}\right)$ is $\mathcal{F}^{i+M_{n}}$-measurable, it follows from the definition of $\alpha(\cdot)$ that

$$
\begin{align*}
& \mid f^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}, x_{i+M_{n}-d}, \ldots, x_{n}\right) \\
& \quad-\quad f_{1}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}\right) f_{2}^{\backslash\left(i, M_{n}\right)}\left(x_{i+M_{n}-d}, \ldots, x_{n}\right) \mid \\
& \leq  \tag{A.24}\\
& \quad \alpha\left(M_{n}\right) .
\end{align*}
$$

Furthermore, let $f_{11}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}\right)$ and $f_{12}^{\backslash\left(i, M_{n}\right)}\left(x_{i}\right)$ be the p.d.f of $\left(X_{1-d}, \ldots, X_{i-1}\right)$ and $\varepsilon_{i}$ respectively, then

$$
\begin{align*}
& f_{1}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}\right) \\
& \quad=f_{11}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}\right) f_{12}^{\backslash\left(i, M_{n}\right)}\left(x_{i}\right) . \tag{A.25}
\end{align*}
$$

Now,

$$
\begin{aligned}
\operatorname{Cov}\left\{\varepsilon_{i},\right. & \left.\hat{\mu}_{\theta}^{\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right\} \\
= & E\left\{\varepsilon_{i} \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right\} \\
= & \int x_{i} \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i+M_{n}-d}, \ldots, x_{n}\right) \\
& \times f^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i}, x_{i+M_{n}-d}, \ldots, x_{n}\right) \\
& \times \mathrm{d} x_{1-d} \cdots \mathrm{~d} x_{i-1} \mathrm{~d} x_{i} \mathrm{~d} x_{i+M_{n}-d} \cdots \mathrm{~d} x_{n},
\end{aligned}
$$

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

$$
\begin{aligned}
\operatorname{Cov}\left\{\varepsilon_{i},\right. & \left.\hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right\} \\
= & \int_{\operatorname{supp}\left(f \backslash\left(i, M_{n}\right)\right)} x_{i} \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}, x_{i+M_{n}-d}, \ldots, x_{n}\right) \\
& \times\left\{f_{11}^{\backslash\left(i, M_{n}\right)}\left(x_{1-d}, \ldots, x_{i-1}\right) f_{12}^{\backslash\left(i, M_{n}\right)}\left(x_{i}\right)\right. \\
& \left.\quad f_{2}^{\backslash\left(i, M_{n}\right)}\left(x_{i+M_{n}-d}, \ldots, x_{n}\right)+O\left(\alpha\left(M_{n}\right)\right)\right\} \\
= & \int x_{i} f_{12}^{\backslash\left(i, M_{n}\right)}\left(x_{i}\right) \int \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)} f_{11}^{\backslash\left(i, M_{n}\right)} f_{2}^{\backslash\left(i, M_{n}\right)} \\
& +O\left(\alpha\left(M_{n}\right)\right) \int_{\operatorname{supp}\left(f \backslash\left(i, M_{n}\right)\right)} x_{i} \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}
\end{aligned}
$$

It follows from $\int x_{i} f_{12}^{\backslash\left(i, M_{n}\right)}\left(x_{i}\right)=E\left(\varepsilon_{i}\right)=0$ and $\int_{\text {supp }\left(f \backslash\left(i, M_{n}\right)\right)} x_{i} \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}<\infty$ that

$$
\begin{equation*}
\operatorname{Cov}\left\{\varepsilon_{i}, \hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right\}=O\left(\alpha\left(M_{n}\right)\right) . \tag{A.26}
\end{equation*}
$$

Now

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

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

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

$$
\begin{equation*}
\frac{1}{n} \sum_{i=1}^{n} E\left[\mu\left(\boldsymbol{X}_{i-1}\right)-\hat{\mu}_{\theta}^{\backslash\left(i, M_{n}\right)}\left(\boldsymbol{X}_{i-1}\right)\right]^{2}=\mathrm{R}(\theta)+o(1) \tag{A.27}
\end{equation*}
$$

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

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

## 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 <br> 21st Century COE Program <br> 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


[^0]:    出版情報：MHF Preprint Series．MHF2003－9，2003－12－31．九州大学大学院数理学研究院 バージョン：
    権利関係：

