Journal of Statistical Planning and Inference 142 (2012) 1678–1689

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Polynomial spline confidence bands for time series trend

Q. Shao [a,*], L. Yang [b,c,1]

[a] Department of Mathematics and Statistics, University of Toledo, Toledo OH 43606, USA
[b] Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, China
[c] Department of Statistics and Probability, Michigan State University, East Lansing MI 48824, USA

ARTICLE INFO

Article history:
Received 25 February 2010
Received in revised form
19 December 2011
Accepted 14 February 2012
Available online 22 February 2012

Keywords:
Autoregressive time series
B-splines
Confidence band

ABSTRACT

The paper considers the construction of a confidence band for the trend function of a stationary time series. An explicit formula is derived based on polynomial splines and Sunklodas (1984). The performance of the confidence band is illustrated by simulation studies. The proposed method is applied to the analysis of the annual yields of wheat in the United States.

1. Introduction

Inference on trend functions is one of the classic topics in time series analysis. In this paper, we are interested in constructing a confidence band for a smooth trend function of time series observations $\{y_i, i=1,\ldots,n\}$ as follows:

$$y_i = g(u_i) + x_i, \tag{1.1}$$

where $g(\cdot)$ represents the trend function defined in the interval $[0,1]$ with $u_i = i/n$ and the zero-mean error term x_i is an autoregressive time series of order p (AR(p)) defined by

$$x_t = \sum_{k=1}^{p} \phi_k x_{t-k} + \epsilon_t,$$

where $\{\epsilon_t\}$ is independent and identically distributed (IID) white noise with mean 0 and variance σ^2. To facilitate the discussion, we introduce the vector format of model (1.1) as follows:

$$\mathbf{y} = \mathbf{g} + \mathbf{x},$$

where $\mathbf{y} = (y_1,\ldots,y_n)'$, $\mathbf{g} = (g(u_1),\ldots,g(u_n))'$. Throughout this paper, we will use bold lower-case letters to denote vectors, bold upper-case letters to denote matrices, and lower-case letters to denote both time series and their realizations.

Analysis of the trend function in model (1.1) with an autoregressive error term has received intensive attention due to its wide applications. The classic approach in stationary time series analysis is that the trend is assumed to be a parametric function with known analytical form and unknown parameters. See for example Chapter 9 of Fuller (1996) for details.

* Corresponding author.
E-mail addresses: qin.shao@utoledo.edu (Q. Shao), yanglijian@suda.edu.cn, yang@stt.msu.edu (L. Yang).
[1] His research was supported by NSF awards DMS 0706518 and DMS 1007594, and funding from the "Jiangsu-Specially Appointed Professor Program", Jiangsu Province, China.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2012.02.028

Although this parametric approach is appropriate for many applications in practice, its major drawback is that the assumption about the trend function is usually artificial. It is desirable to extend the analysis to nonparametric settings which do not need to specify the analytical form of the trend functions. In recent years as a result of advances of computing technology, a great deal of work has been devoted to local kernels and polynomial splines, two of the most commonly used nonparametric methods. It is impossible to cover the vast literature on these two methods here. Fan and Gijbels (1996) provided an overview and some details about local kernels, especially for data from independent and identical distributions. Opsomer et al. (2001), on the other hand, summarized the development of research about local kernels for correlated data. For an overview and basic theoretical results of polynomial splines, interested readers can refer to, for example, Stone (1994) and Huang (2003).

While most research concentrated on estimation, a few authors attacked construction of a confidence band for a smooth unknown function. Bickel and Rosenblatt (1973) is a pioneer work on nonparametric confidence bands for a density curve of independent and identically distributed observations. Since then several authors, such as Hall and Titterington (1988), Xia (1998), Claeskens and Van Keilegom (2003), and Wang and Yang (2009), have investigated this issue for independent observations. Wu and Zhao (2007) proposed a confidence band based on local kernels for the trend with a stationary time series error term. In this paper, we will extend the method of Wang and Yang (2009) to stationary time series trend analysis. The major contribution of this paper is to provide practitioners with the theoretical foundation of a confidence bound and a fast as well as easy to implement algorithm. The band is simultaneous and conservative in the sense that it covers the whole trend function at least with the probability of the given confidence level. It is worth mentioning that compared to the local smoothing obtained by using a kernel, spline smoothing is global, i.e., only a single optimization is needed for the unknown function over an entire range, instead of optimization at every point in the range. As a result, polynomial splines used in this paper can be thousands of times faster than kernel smoothing, which was discussed in detail in, for example, Xue and Yang (2006) and Wang and Yang (2007).

The paper will be organized as follows: in Section 2, we will introduce polynomial splines and confidence bands; in Section 3, we will illustrate by simulation studies the performance of the proposed confidence bands for the trend function in model (1.1) with several AR(1) terms, and analyze the annual yields of wheat in the United States; finally in Section 4, we will provide the details of the proofs of the theoretical results based on which confidence bands are constructed. To facilitate reading, the existing theorems that play important roles in deriving the theoretical results of this paper are provided in the Appendix.

2. Construction of confidence band

2.1. Polynomial splines

Suppose that m is a positive integer. Consider a sequence of equally spaced points or knots $(-m+1)h \le \cdots \le 0 \le h \le 2h \cdots \le Nh \le 1$. Notice that there are $N+m$ knots that divide the interval $[(-m+1)h,1]$ into subintervals $J_j=[jh,(j+1)h), j=-m+1,-m+2,\ldots,N-1$ and $J_N=[Nh,1]$, of width h. For any given $u \in [0,1]$, $j(u)$ is the knot corresponding to the interval that includes u. Let $G_N^{(m-2)}=G_N^{(m-2)}[0,1]$ denote the space of functions that are polynomial of degree $m-1$ on each J_j and have continuous $(m-2)$th derivatives. The B-spline basis of $G_N^{(m-2)}$ is $\mathbf{b}_m(u)=(b_{j,m}(u), j=-m+1,\ldots,N)'$. For any function $\varphi(\cdot)$ in $L^2[0,1]$ define the norm as

$$\|\varphi\|_2^2=\int_0^1 \varphi^2(x)\,dx.$$

The B-spline standardized basis $\mathbf{c}_m(u)=(c_{j,m}(u), j=-m+1,\ldots,N)'$ is defined as

$$c_{j,m}(u)=\frac{b_{j,m}(u)}{\|b_{j,m}\|_2}=\frac{b_{j,m}(u)}{\{\int_0^1 b_{j,m}^2(u)\,du\}^{1/2}}. \tag{2.1}$$

For a realization of time series $\mathbf{y}$, define a vector $\mathbf{c}_{j,m}=(c_{j,m}(u_1),\ldots,c_{j,m}(u_n))'$ with

$$c_{j,m}(u_i)=\frac{b_{j,m}(u_i)}{\|b_{j,m}\|_2},$$

and an $n\times(N+m)$ matrix

$$\mathbf{C}_m=(\mathbf{c}_{-m+1},\ldots,\mathbf{c}_N).$$

We will focus on the cases of $m=1,2$: $G_N^{(-1)}$ is the space of functions that are constant on each J_j, and $G_N^{(0)}$ is the space of functions that are linear on each J_j and continuous on [0,1].

The B-spline basis for $G_N^{(-1)}$ is $\mathbf{b}_1(u)=(b_{j,1}(u), j=0,\ldots,N)'$, where $b_{j,1}(u)$ is the indicator function of J_j, i.e.

$$b_{j,1}(u)=\begin{cases}1, & j=j(u),\\ 0 & \text{otherwise}.\end{cases}$$

It is straightforward to obtain that the standardized basis $c_{j,1}(u)$ satisfies

$$c_{j,1}(u)=\begin{cases} h^{-1/2}, & j=j(u),\\ 0 & \text{otherwise.}\end{cases}$$

Each vector $\mathbf{c}_{j,1}$ has at most $\lceil nh\rceil+1$ entries that have value $h^{-1/2}$, the other at least $n-\lceil nh\rceil-1$ entries being zero, where $\lceil u\rceil$ is the smallest integer such that $u\le\lceil u\rceil$. The column vectors of $\mathbf{C}_1$ are orthogonal and there is only one nonzero entry in each row of $\mathbf{C}_1$. In particular,

$$\left(\frac{1}{n}\mathbf{C}_1'\mathbf{C}_1\right)_{j,j'=0}^{N}=\begin{cases}1+r_{j1}, & j=j',\\ 0 & \text{otherwise,}\end{cases}$$

where $\max_{0\le j\le N}|r_{j1}|=O\{(nh)^{-1}\}$.

The B-spline basis for the piecewise linear spline space $G_N^{(0)}$ is $\mathbf{b}_2(u)=(b_{j,2}(u),j=-1,\ldots,N)'$, where $b_{j,2}(u)$ is defined as follows:

$$b_{j,2}(u)=K\left(\frac{u-(j+1)h}{h}\right),$$

where $K(x)=(1-|x|)_+$ with $(x)_+=\max(x,0)$; or equivalently

$$b_{j,2}(u)=\begin{cases} j(u)+1-\dfrac{u}{h}, & j=j(u)-1,\\ \dfrac{u}{h}-j(u), & j=j(u),\\ 0 & \text{otherwise.}\end{cases} \tag{2.2}$$

According to (2.2), after some manipulation, the standardized basis $c_{j,2}(u)$ satisfies

$$c_{j,2}(u)=\begin{cases}\dfrac{u/h-j(u)}{w_j\sqrt{h/3}}, & j=j(u),\\ \dfrac{j(u)+1-u/h}{w_{j-1}\sqrt{h/3}}, & j=j(u)-1,\\ 0 & \text{otherwise,}\end{cases}\quad \text{with } w_j=\begin{cases}\sqrt{2}, & 0\le j\le N-1,\\ 1, & j=-1,N.\end{cases}$$

Each vector $\mathbf{c}_{j,2}$ has at most $2\lceil nh\rceil+2$ nonzero entries. The column vectors of $\mathbf{C}_2$ are not orthogonal. There are at most two nonzero entries in each row of $\mathbf{C}_2$. Following a similar argument to that of Lemma 2 of Shao and Yang (2011), it can be shown that

$$\left(\frac{1}{n}\mathbf{C}_2'\mathbf{C}_2\right)_{j,j'=-1}^{N}=\begin{cases}1+r_{jj'2}, & j=j',\\ \frac{1}{4}+r_{jj'2}, & |j-j'|=1 \text{ and } 0\le j,j'\le N-1,\\ \dfrac{\sqrt{2}}{4}+r_{jj'2}, & |j-j'|=1 \text{ and } j \text{ or } j'=-1,N,\\ 0 & \text{otherwise,}\end{cases}$$

where $\max_{-1\le j,j'\le N}|r_{jj'2}|=o(1)$.

Denote the number of observations in J_j by n_j and define $j_i=\sum_{k=0}^{j-1}n_k+i$. Thus x_{j_i} is the ith observation in J_j. We have $\lceil nh\rceil-1\le n_j\le\lceil nh\rceil+1$ and $\sum_{j=0}^{N}n_j=n$.

2.2. Confidence band

At any $u\in[0,1]$, the polynomial spline estimator of $g(u)$ based on a realization of time series $\mathbf{y}=(y_1,\ldots,y_n)'$ is

$$\hat{g}_m(u)=\sum_{j=1-m}^{N}c_{j,m}(u)\hat{\beta}_{j,m}, \tag{2.3}$$

where $c_{j,m}(u)$ is defined in (2.1) and $\hat{\boldsymbol{\beta}}_m=(\hat{\beta}_{-m+1,m},\cdots,\hat{\beta}_{N,m})'$ is obtained by

$$\hat{\boldsymbol{\beta}}_m=\underset{\boldsymbol{\beta}}{\operatorname{argmin}}(\mathbf{y}-\mathbf{C}_m\boldsymbol{\beta})'(\mathbf{y}-\mathbf{C}_m\boldsymbol{\beta}). \tag{2.4}$$

According to linear model theory, (2.3) and (2.4) are equivalent to

$$\hat{g}_m(u)=\mathbf{c}_m'(u)\left(\frac{1}{n}\mathbf{C}_m'\mathbf{C}_m\right)^{-1}\left(\frac{1}{n}\mathbf{C}_m'\mathbf{y}\right).$$

The error of the polynomial spline estimator is defined as

$$\hat{x}_{nu,m} = \mathbf{c}'_m(u)\left(\frac{1}{n}\mathbf{C}'_m\mathbf{C}_m\right)^{-1}\left(\frac{1}{n}\mathbf{C}'_m\mathbf{x}\right). \tag{2.5}$$

Define

$$\tilde{g}_m(u) = \mathbf{c}'_m(u)\left(\frac{1}{n}\mathbf{C}'_m\mathbf{C}_m\right)^{-1}\left(\frac{1}{n}\mathbf{C}'_m\mathbf{g}\right).$$

Thus

$$\hat{g}_m(u) = \tilde{g}_m(u) + \hat{x}_{nu,m}. \tag{2.6}$$

We need the following assumptions to propose a confidence band for a smooth trend function $g(u)$.

1. The trend function $g(\cdot) \in C^{(m)}[0,1], m=1,2$; that is the trend function has m continuous derivatives.
2. The number of interior knots $(n/\log n)^{1/(2m+1)} \ll N \ll n^{1/3}$.
3. The errors $\{\epsilon_t\}$ are iid with a density function that is positive everywhere. Time series $\{x_t\}$ is causal; that is there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$x_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

4. $E|x_i|^3 < M_0 < \infty$.

Remark. Assumption 3 entails that the zero spectrum of AR(p) time series $\{x_t\}$ is finite; i.e. $\sum_{|k|=0}^{\infty} \gamma_k < \infty$. In addition, Theorem 2.4(i) and Eqs. (2.58) and (2.59) of Fan and Yao (2003) established that $\{x_t\}$ is geometrically β-mixing and thus geometrically α-mixing as well. In other words, $\alpha(n) \le Ke^{-\lambda n}$, for some constants $K>0$ and $\lambda > 0$, where

$$\alpha(n) = \sup_{A \in \mathcal{F}^0_{-\infty}, B \in \mathcal{F}^n_{\infty}} |P(A)P(B) - P(AB)| \to 0 \quad \text{as } n \to \infty,$$

and $\mathcal{F}^i_j$ denotes the σ-algebra generated by $\{x_t, i \le t \le j\}$. In addition, under Assumptions 1 and 2, according to Theorem 5.1 of Huang (2003),

$$\|\tilde{g}_m(u) - g(u)\|_\infty = O_p(h^m). \tag{2.7}$$

A confidence band is constructed based on the following theorem which is the major result of this paper. Its proof will be provided in Section 4.

Theorem 2.1. *Under Assumptions* 1–4, *an asymptotic* $100(1-\alpha)\%$ *conservative confidence band for the trend function* $g(u)$ *is*

$$\hat{g}_m(u) \pm \hat{\eta}_{n,m}(u) m^{1/2} \{2\log(N+1)\}^{1/2} d_{n,\alpha/m}, \quad m=1,2, \tag{2.8}$$

or in other words,

$$\liminf_{n\to\infty} P(g(u) \in \hat{g}_m(u) \pm \hat{\eta}_{n,m}(u) m^{1/2}\{2\log(N+1)\}^{1/2} d_{n,\alpha/m}) \ge 1-\alpha, \quad m=1,2, \tag{2.9}$$

where $\hat{\eta}_{n,1}(u)$ *and* $\hat{\eta}_{n,2}(u)$ *are respectively defined in* (4.1), (4.16), *and*

$$d_{n,\alpha} = 1 - \{2\log(N+1)\}^{-1}[\log(\alpha/2) + \tfrac{1}{2}\{\log\log(N+1) + \log(4\pi)\}]. \tag{2.10}$$

The above asymptotically conservative confidence band has width of order $\{\log n/(nh)\}^{1/2}$ at all points $u \in [0,1]$, which is the same as the asymptotically correct band in Wang and Yang (2009). This partially alleviates the concern that the above band can be too conservative, while an asymptotically correct band needs further investigation for time series trend.

2.3. Implementation of confidence band

One of the critical steps in constructing the confidence band in (2.8) is to estimate the zero spectrum $\sum_{|k|=0}^{\infty} \gamma_k$ of a stationary time series $\{x_t\}$. Two approaches are considered in the next section. The first method is to estimate it indirectly from the observations. Specifically, the trend function $g(u)$ is first estimated with polynomial B-splines and then the residuals are obtained by removing $\hat{g}_m$ from the observations. The autoregressive parameters are estimated with the residuals. Shao and Yang (2011) showed that such estimators of autoregressive parameters are consistent under certain conditions. The second method is to obtain $\hat{\gamma}_k$ directly from observations $\mathbf{y}$ assuming the distribution of x_t has finite fourth moment, which was proposed by Hall and Keilegom (2003). Specifically, first the autocovariances are estimated by

$$\hat{\gamma}_0 = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \frac{1}{2(n-k)} \sum_{i=k+1}^{n} \{(D_k y)_i\}^2,$$

$$\hat{\gamma}_k = \hat{\gamma}_0 - \frac{1}{2(n-k)} \sum_{i=k+1}^{n} \{(D_k y)_i\}^2, \quad k < n,$$

where $(D_k y)_i = y_i - y_{i-k}$ and k_1, k_2 are two integers satisfying $k_1 \le k_2$, $k_1/\log(n) \to \infty$, and $k_2 = O(n^{1/2})$, and then $\hat{\phi}_k$'s are calculated by the Yule–Walker equations. The estimate of the zero spectrum $\sum_{|k|=0}^{\infty} \gamma_k$ is obtained based on $\hat{\phi}_k$. See Hall and Keilegom (2003) for details.

3. Simulation study and applications

3.1. Simulation study

We simulate 500 samples of the time series in (1.1) with AR(1) errors. The autoregressive parameters ϕ_1 are chosen as -0.8, -0.4, -0.2, 0.2, 0.4, 0.8 so that the time series range from the relatively weakly to the relatively highly correlated, and the white noise variance $\sigma^2 = 1$. The sample sizes for each sample path are respectively $n = 100, 200, 300, 400$. The trend function is defined as the same function in Shao and Yang (2011),

$$g(u) = \sin(2\pi u), \quad u \in [0,1].$$

To satisfy Assumption 2, the number of knots here as well as the real data analysis in the next section is chosen among $N = c\lceil n^{1/(2m+1)} \rceil$, where c is an integer ranging from 1 to 5. The best N is determined by BIC or equivalently minimizing

$$\log(\text{MSE}) + (N+m)\log(n)/n, \tag{3.1}$$

where following Hall and Keilegom (2003), $\text{MSE} = (1/n)\sum_{i=1}^{n} \{\hat{g}_m(u_i) - g(u_i)\}^2$ for the simulations in this section and $\text{MSE} = (1/n)\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ for the real data analysis in the next section. For each sample path, the band in (2.8) is calculated. The coverage probability in (2.9) is estimated by the empirical ratio between the number of bands that cover the true curve $g(u)$ and the total number of simulations. While mathematically the true curve is covered by a band if and only if $g(u) \in \hat{g}_m(u) \pm \hat{\eta}_{n,m}(u) m^{1/2} \{2\log(N+1)\}^{1/2} d_{n,\alpha/m}$ for all $u \in [0,1]$, numerically this can only be verified at a finite number of points, which have been chosen as $u_1, \ldots, u_n$. The significance levels are $\alpha = 0.1, 0.05$. The results based on the B-spline basis of the piecewise linear function space $G^{(0)}$ are summarized in Table 1.

When calculating the zero spectrum, the two methods mentioned in the previous section are used. The results under M1 and M2 in Table 1 are obtained by Shao and Yang (2011) and Hall and Keilegom (2003), respectively. Overall both methods provide a similar conclusion: the empirical probabilities increase as the sample sizes rise from 100 to 400, which coincides with Theorem 2.1.

3.2. Application to real data

The dots in Fig. 1 is the annual yields of wheat in the United States from 1908 through 1991. The scatter plot shows a pronounced nonlinear upward trend. The wheat yields remained unchanged for the first several years, constant increasing

Table 1
Simulation.

ϕ_1	α	$n=100$		$n=200$		$n=300$		$n=400$	
		M1	M2	M1	M2	M1	M2	M1	M2
−0.8	0.10	0.978	0.958	0.992	0.974	0.970	0.948	0.980	0.980
	0.05	0.994	0.976	0.996	0.982	0.992	0.972	0.992	0.994
−0.4	0.10	0.986	0.994	0.978	0.984	0.978	0.984	0.984	0.992
	0.05	0.998	1.000	0.996	0.996	0.994	0.998	0.998	0.998
−0.2	0.10	0.982	0.994	0.984	0.986	0.976	0.990	0.982	0.988
	0.05	1.000	1.000	0.992	0.994	0.996	0.998	0.994	0.996
0.2	0.10	0.980	0.996	0.974	0.982	0.992	0.996	0.990	0.990
	0.05	1.000	1.000	0.996	0.998	0.998	0.998	0.998	0.998
0.4	0.10	0.978	0.990	0.974	0.980	0.994	0.998	0.992	1.000
	0.05	0.996	0.998	0.996	0.994	0.998	0.998	1.000	1.000
0.8	0.10	0.930	0.980	0.970	0.982	0.976	0.990	0.988	0.996
	0.05	0.956	0.992	0.986	0.994	0.988	0.994	0.992	0.998

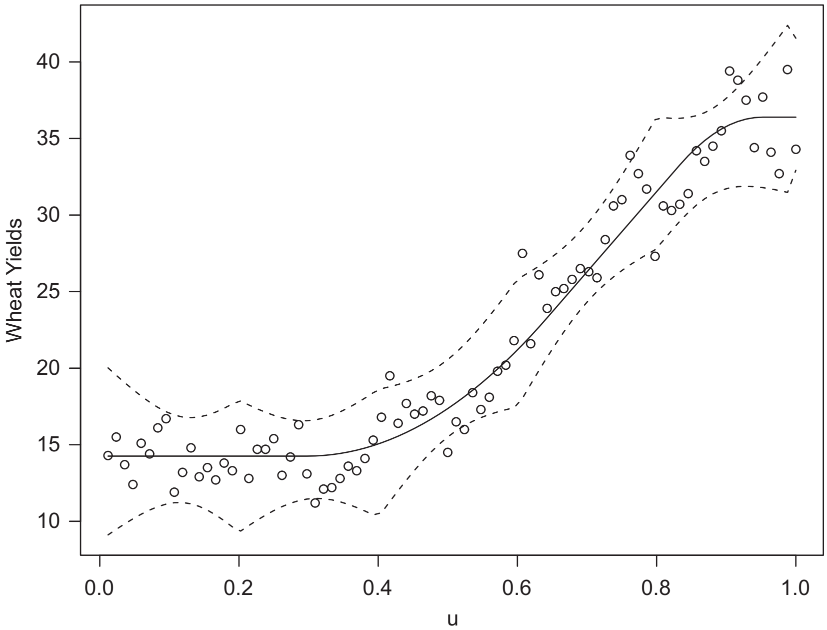

Fig. 1. United States wheat yields from 1908 through 1991.

rate for several years, polynomial increasing rate in the other years and stayed constant for the last several years. According to Fuller (1996), the null hypothesis is that the trend function is a grafted polynomial:

$$H_0 : g(u) = a_1 + a_2\varphi(u),$$

where a_1 and a_2 are two constants and

$$\varphi(u) = \begin{cases} 0, & 0 \le u \le 25/84, \\ (84u-25)^2, & 25/84 \le u \le 54/84, \\ 841+58(84u-54), & 54/84 \le u \le 70/84, \\ 841+58(84u-54)-2.9(84u-70)^2, & 70/84 \le u \le 80/84, \\ 2059, & 80/84 \le u \le 1. \end{cases}$$

Under H_0, Fuller (1996) estimated the trend function $g(u)$ in model (1.1) with an AR(1) error term. The solid line in Fig. 1 is Fuller's $\hat{g}(u)$. To test whether the curve in the null hypothesis is appropriate, we construct a 95% confidence band in (2.8) using the B-spline basis of the piecewise linear function space $G^{(0)}$ and plot them using the two dashed lines in Fig. 1. The zero spectrum is estimated by Shao and Yang (2011). Fig. 1 indicates that the solid line is included in the band and H_0 cannot be rejected. As the referee points out, this conclusion should be taken with caution as the inclusion of Fuller's $\hat{g}(u)$ in the band may be due to the conservativeness of the band.

4. Proofs

In this section, we will show Theorem 2.1 for $m=1$ and 2, respectively. Throughout this section, $U(\cdot)$, $u(\cdot)$, $U_p(\cdot)$ and $u_p(\cdot)$ denote the orders of quantities that are $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ uniformly over $[0,1]$, respectively.

4.1. Proof for $m=1$

We replace $n^{-1}\mathbf{C}_1'\mathbf{C}_1$ by the identity matrix $\mathbf{I}$ in (2.5) and define $\tilde{x}_{nu,1}$ by

$$\tilde{x}_{nu,1} = \mathbf{c}_1'(u)\left(\frac{1}{n}\mathbf{C}_1'\mathbf{x}\right) = \frac{1}{nh^{1/2}}\sum_{i=1}^{n_{j(u)}} c_{j(u),1}(u_{j(u)_i})x_{j(u)_i}.$$

Lemma 4.1. *The pointwise variance of* $\tilde{x}_{nu,1}$ $(u \in [0,1])$ *is*

$$E\{\tilde{x}_{nu,1}\}^2 = \frac{1}{nh}\left(\sum_{|k|=0}^{\lceil nh \rceil} \gamma_k\right)[1+u\{(1)\}].$$

Proof.

$$E\{\tilde{x}_{nu,1}\}^2 = \frac{1}{n^2h}E\left\{\sum_{i=1}^{n_{j(u)}} c_{j(u),1}(u_{j(u)_i})x_{j(u)_i}\right\}^2 = \frac{1}{n^2h^2}\sum_{i=1}^{n_{j(u)}}\sum_{k=1}^{n_{j(u)}}\gamma_{i-k} = \frac{1}{nh}\left(\sum_{|k|=0}^{\lceil nh \rceil}\gamma_k\right)[1+u(1)]. \qquad \square$$

Define

$$\hat{\eta}^2_{n,1}(u) = \frac{1}{nh}\left(\sum_{|k|=0}^{\lceil nh \rceil}\hat{\gamma}_k\right), \tag{4.1}$$

where $\hat{\gamma}_k$ is a consistent estimator γ_k. According to Lemma 4.1, it can be concluded that

$$E\{\tilde{x}_{nu,1}\}^2 - \hat{\eta}^2_{n,1}(u) = u_p\{(nh)^{-1}\}. \tag{4.2}$$

Lemma 4.2. *Under Assumptions* 2 *and* 3,

$$\hat{\eta}^{-1}_{n,1}(u)\sup_{u\in[0,1]}\left|\hat{x}_{nu,1} - \tilde{x}_{nu,1}\right| = U_P\{(nh)^{-1}\} = u_p(1).$$

Proof.

$$\begin{aligned}
\left|\hat{x}_{nu,1} - \tilde{x}_{nu,1}\right| &= \left|\mathbf{c}_1'(u)\left(\frac{1}{n}\mathbf{C}_1'\mathbf{C}_1\right)^{-1}\left(\mathbf{I} - \frac{1}{n}\mathbf{C}_1'\mathbf{C}_1\right)\left(\frac{1}{n}\mathbf{C}_1'\mathbf{x}\right)\right| \\
&= \frac{1}{h^{1/2}}\left|\sum_{i=1}^{n_{j(u)}}\left(\frac{1}{n}\mathbf{c}'_{j(u),1}\mathbf{c}_{j(u),1}\right)^{-1}\left(1 - \frac{1}{n}\mathbf{c}'_{j(u),1}\mathbf{c}_{j(u),1}\right)\left(\frac{1}{n}c_{j(u),1}(u_{j(u)_i})x_{j(u)_i}\right)\right| \\
&= \frac{1}{h^{1/2}}\mathrm{O}\{(nh)^{-1}\}\left|\left(\frac{1}{nh^{1/2}}\sum_{i=1}^{n_{j(u)}}x_{j(u)_i}\right)\right|.
\end{aligned}$$

According to Assumption 3, $(1/nh)\sum_{i=1}^{n_{j(u)}} x_{j(u)_i} = \mathrm{O}_P\{(nh)^{-1/2}\}$. Therefore,

$$\left|\tilde{x}_{nu,1} - \hat{x}_{nu,1}\right| = U_P\{(nh)^{-3/2}\}. \tag{4.3}$$

The proof is complete from (4.1), (4.3) and Assumption 2. $\square$

Lemma 4.3. *Under Assumptions* 1–4,

$$\liminf_{n\to\infty} P\left(\sup_{u\in[0,1]}\{(E\tilde{x}^2_{nu,1})^{-1/2}\left|\tilde{x}_{nu,1}\right|\} \le \{2\log(N+1)\}^{1/2}d_{n,\alpha}\right) \ge 1-\alpha, \tag{4.4}$$

where $d_{n,\alpha}$ *is defined in* (2.10).

Proof. Define

$$\xi_{i,j} = h^{1/2}c_{j,1}(u_{j_i})x_{j_i},$$

and $s^2_{\xi,j} = E(\sum_{i=1}^{n_j}\xi_{i,j})^2$. Then

$$s^2_{\xi,j} = nh\left(\sum_{|k|=0}^{\infty}\gamma_k\right)[1+\mathrm{o}(1)]. \tag{4.5}$$

It is straightforward to show that

$$\sup_{u\in[0,1]}\{(E\tilde{x}^2_{nu,1})^{-1/2}\left|\tilde{x}_{nu,1}\right|\} = \max_{0\le j\le N} s^{-1}_{\xi,j}\left|\sum_{i=1}^{n_j}\xi_{i,j}\right|.$$

We will apply Sunklodas's Theorem in the Appendix to $\{\xi_{i,j}\}$ to establish (4.4).

For any fixed j, $\{\xi_{i,j}; 1 \le i \le n_j\}$ is an α-mixing sequence. In addition,

$$\max_{1 \le i \le n_j} E|\xi_{i,j}|^3 = h^{3/2} \max_{1 \le i \le n_j} E|c_{j,1}(u_{j_i})x_{j_i}|^3 = E|x_1|^3 < M_0,$$

under Assumption 4. From (4.5), there exists a constant a_1 $(0 < a_1 < \infty)$ such that $s^2_{\xi,j} \ge a_1 nh$.

According to Sunklodas's Theorem with $s=3$, there exist constants a_2, a_3 and a_4 such that for all λ $(\lambda_1 \le \lambda \le \lambda_2)$ and all $n_j > 1$,

$$\sup_z \left| P\left(s^{-1}_{\xi,j} \sum_{i=1}^{n_j} \xi_{i,j} < z \right) - \Phi(z) \right| \le a_2 \frac{M_0}{a_1 s_{\xi,j}} \left\{ \frac{1}{\lambda} \log(s_{\xi,j} a_1^{-1/2}) \right\}^2, \tag{4.6}$$

where

$$\lambda_1 = \frac{a_3}{n_j} \{\log(s_{\xi,j} a_1^{-1/2})\}^{a_4}, \quad a_4 > 4,$$

$$\lambda_2 = 12 \log(s_{\xi,j} a_1^{-1/2}).$$

In particular, if $\lambda = \lambda_2$, it is obvious from (4.6) and Assumption 2

$$\sup_z \left| P\left(s^{-1}_{\xi,j} \sum_{i=1}^{n_j} \xi_{i,j} < z \right) - \Phi(z) \right| \le a_0 (nh)^{-1/2} \to 0, \tag{4.7}$$

where a_0 is a constant.

From Theorem 1.5.3 of Leadbetter et al. (1983) in the Appendix,

$$P(|Z| \le \tau / a_{N+1} + b_{N+1}) \ge 1 - \frac{2e^{-\tau}}{N+1} + o(N^{-1}), \tag{4.8}$$

where Z is the standard normal random variable and

$$a_{N+1} = \{2 \log (N+1)\}^{1/2},$$

$$b_{N+1} = \{2\log(N+1)\}^{1/2} - \tfrac{1}{2}\{2\log(N+1)\}^{-1/2}\{\log\log(N+1) + \log(4\pi)\}.$$

In particular, if $\tau = -\log(\alpha/2)$ and $\tau/a_{N+1} + b_{N+1} = \{2\log(N+1)\}^{1/2} d_{n,\alpha}$, (4.8) gives

$$P(|Z| \le \{2 \log(N+1)\}^{1/2} d_{n,\alpha}) \ge 1 - \frac{\alpha}{N+1} + o(N^{-1}). \tag{4.9}$$

Notice

$$P\left(\max_{0 \le j \le N} s^{-1}_{\xi,j} \left| \sum_{i=1}^{n_j} \xi_{i,j} \right| \le \{2\log(N+1)\}^{1/2} d_{n,\alpha} \right) \ge 1 - \sum_{j=0}^{N} P\left(\left| s^{-1}_{\xi,j} \sum_{i=1}^{n_j} \xi_{i,j} \right| > \{2\log(N+1)\}^{1/2} d_{n,\alpha} \right) \ge 1 - \sum_{j=0}^{N} (1 + R_{1j}) + (N+1)R_2, \tag{4.10}$$

where

$$R_{1j} = \left| P\left(s^{-1}_{\xi,j} \left| \sum_{i=1}^{n_j} \xi_{i,j} \right| \le \{2 \log(N+1)\}^{1/2} d_{n,\alpha} \right) - P(|Z| \le \{2 \log(N+1)\}^{1/2} d_{n,\alpha}) \right|,$$

$$R_2 = P(|Z| \le \{2 \log (N+1)\}^{1/2} d_{n,\alpha}).$$

Therefore, from (4.7), (4.9) and (4.10), we obtain

$$P\left(\sup_{u \in [0,1]} \{(E\tilde{x}^2_{nu,1})^{-1/2} |\tilde{x}_{nu,1}|\} \le \{2 \log(N+1)\}^{1/2} d_{n,\alpha} \right) \ge 1 - (N+1)[1 + O\{(nh)^{-1/2}\}] + (N+1)\left\{ 1 - \frac{\alpha}{N+1} + o(N^{-1}) \right\}$$
$$= 1 - \alpha + O\{(nh^3)^{-1/2}\} + o(1).$$

Under Assumption 2, the last two terms are negligible as n approaches infinity. Hence (4.4) holds and the proof is complete. □

According to the definition (2.6), it is obvious that

$$\hat{\eta}^{-1}_{n,m}(u)\{\hat{g}_m(u) - g(u)\} = \hat{\eta}^{-1}_{n,m}(u)\{\tilde{g}_m(u) + \hat{x}_{nu,m} - g(u) + \tilde{x}_{nu,m} - \tilde{x}_{nu,m}\}$$
$$= \hat{\eta}^{-1}_{n,m}(u)\{\tilde{g}_m(u) - g(u)\} + \hat{\eta}^{-1}_{n,m}(u)(\hat{x}_{nu,m} - \tilde{x}_{nu,m}) + \{\hat{\eta}^{-1}_{n,m}(u) - (E\tilde{x}^2_{nu,m})^{-1/2}\}\tilde{x}_{nu,m} + (E\tilde{x}^2_{nu,m})^{-1/2}\tilde{x}_{nu,m}. \tag{4.11}$$

Proof of Theorem 2.1 when $m=1$. Based on Assumptions 1 and 2 and (2.7), we obtain

$$\hat{\eta}_{n,1}^{-1}(u)\sup_{u\in[0,1]}|\tilde{g}_1(u)-g(u)|=U_p\{(nh)^{1/2}\}U(h)=U_p\{(nh^3)^{1/2}\}=u_p\{(\log N)^{1/2}\}.$$

On the other hand, according to Lemmas 4.1 and 4.2, (4.2), $\sup_{u\in[0,1]}\{(E\tilde{x}_{nu,1}^2)^{-1/2}|\tilde{x}_{nu,1}|\}$ is the only term that is not negligible in (4.11). Hence, it has the same distribution as $\hat{\eta}_{n,1}^{-1}(u)\sup_{u\in[0,1]}\{\hat{g}_1(u)-g(u)\}$. Therefore, (2.9) is obtained from Lemma 4.3.

4.2. Proof for $m=2$

Define an $(N+2)\times(N+2)$ matrix $\mathbf{V}$ by

$$(\mathbf{V})_{j,j'=-1}^{N}=\begin{cases}1, & j=j',\\ \frac{1}{4}, & |j-j'|=1 \text{ and } 0\le j,j'\le N-1,\\ \frac{\sqrt{2}}{4}, & |j-j'|=1 \text{ and } j \text{ or } j'=-1,N,\\ 0 & \text{otherwise.}\end{cases}$$

The inverse of $\mathbf{V}$ is denoted by $\mathbf{S}$. Matrix $\mathbf{V}$ is symmetric and tri-diagonal and $n^{-1}\mathbf{C}_2'\mathbf{C}_2=\mathbf{V}+o(1)$. We replace $n^{-1}\mathbf{C}_2'\mathbf{C}_2$ by $\mathbf{V}$ in (2.5) and define $\tilde{x}_{nu,2}$ by

$$\tilde{x}_{nu,2}=\mathbf{c}_2'(u)\mathbf{S}\left(\frac{1}{n}\mathbf{C}_2'\mathbf{x}\right). \tag{4.12}$$

In the following, we use $|\boldsymbol{\Omega}|$ to denote the maximal absolute value of any vector or matrix $\boldsymbol{\Omega}$. We have the following property regarding a matrix such as $n^{-1}\mathbf{C}_2'\mathbf{C}_2$.

Lemma 4.4. *Given any matrix $\boldsymbol{\Omega}=\mathbf{V}+\boldsymbol{\Gamma}$, in which $\boldsymbol{\Gamma}=(\gamma_{j,j'})_{j,j'=-1}^{N}$ satisfies $\gamma_{j,j'}\equiv 0$ if $|j-j'|>1$ and $|\boldsymbol{\Gamma}|\to 0$, there exist constants $a_1>0$ and $a_2>0$ independent of N and $\boldsymbol{\Gamma}$, such that*

$$a_1|\zeta|\le|\boldsymbol{\Omega}\zeta|\le a_2|\zeta|,\quad a_2^{-1}|\zeta|\le|\boldsymbol{\Omega}^{-1}\zeta|\le a_1^{-1}|\zeta|,\ \forall\zeta\in R^{N+2}.$$

The proof is similar to Lemma B.2 of Wang and Yang (2009) and thus is omitted.

Lemma 4.5. *The pointwise variance of $\tilde{x}_{nu,2}$ $(u\in[0,1])$ is*

$$E\{\tilde{x}_{nu,2}\}^2=\frac{3}{nh}\left(\sum_{|k|=0}^{\lceil nh\rceil}\gamma_k\right)\{\boldsymbol{\delta}'(u)\mathbf{S}_{j(u)}\boldsymbol{\delta}(u)+u(1)\}, \tag{4.13}$$

where

$$\boldsymbol{\delta}(u)=\left(\frac{j(u)+1-u/h}{w_{j(u)-1}},\frac{u/h-j(u)}{w_{j(u)}}\right)',$$

and $\mathbf{S}_{j(u)}$ is a2×2 submatrix of $\mathbf{S}$,

$$\mathbf{S}_{j(u)}=\begin{pmatrix}s_{j(u)-1,j(u)-1} & s_{j(u)-1,j(u)}\\ s_{j(u),j(u)-1} & s_{j(u),j(u)}\end{pmatrix}.$$

Proof. Following Lemma 2 of Shao and Yang (2011), after some manipulation, we obtain

$$\mathbf{C}_2'\boldsymbol{\Gamma}\mathbf{C}_2=n\left(\sum_{|k|=0}^{\lceil nh\rceil}\gamma_k\right)\{\mathbf{V}+o(1)\}, \tag{4.14}$$

where $\boldsymbol{\Gamma}=Cov(\mathbf{x})$. Thus, according to (4.12) and (4.14),

$$E\{\tilde{x}_{nu,2}\}^2=n^{-2}\mathbf{c}_2'(u)\mathbf{S}\mathbf{C}_2'\boldsymbol{\Gamma}\mathbf{C}_2\mathbf{S}\mathbf{c}_2(u)=\frac{1}{n}\left(\sum_{|k|=0}^{\lceil nh\rceil}\gamma_k\right)\{\mathbf{c}_2'(u)\mathbf{S}\mathbf{c}_2(u)+u(1)\}. \tag{4.15}$$

After some manipulation, (4.13) can be obtained from (4.15). The proof is complete. □

Define

$$\hat{\eta}_{n,2}^2(u)=\frac{3}{nh}\left(\sum_{|k|=0}^{\lceil nk\rceil}\hat{\gamma}_k\right)\boldsymbol{\delta}'(u)\mathbf{S}_{j(u)}\boldsymbol{\delta}(u). \tag{4.16}$$

According to Lemma 4.5,

$$E\{\tilde{x}_{nu,2}\}^2 - \hat{\eta}^2_{n,2}(u) = u_p\{(nh)^{-1}\}. \tag{4.17}$$

Lemma 4.6. *Under Assumption*3,

$$\hat{\eta}^{-1}_{n,2}(u)\sup_{u\in[0,1]}|\hat{x}_{nu,2}-\tilde{x}_{nu,2}| = u_p(1). \tag{4.18}$$

Proof. According to (2.5) and (4.12),

$$|\hat{x}_{nu,2}-\tilde{x}_{nu,2}| = |\mathbf{c}_2'(u)\mathbf{S}(\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2)(n^{-1}\mathbf{C}_2'\mathbf{C}_2)^{-1}(n^{-1}\mathbf{C}_2'\mathbf{x})| \le \|\mathbf{c}_2'(u)\mathbf{S}(\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2)(n^{-1}\mathbf{C}_2'\mathbf{C}_2)^{-1}\|\|n^{-1}\mathbf{C}_2'\mathbf{x}\|, \tag{4.19}$$

where $\|\cdot\|$ is an Euclidean norm. In the following, we consider the two Euclidean norms in (4.19). According to Lemma 4.4, it is obvious that there exists a constant a such that

$$|\mathbf{c}_2'(u)\mathbf{S}(\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2)(n^{-1}\mathbf{C}_2'\mathbf{C}_2)^{-1}| \le a|\mathbf{c}_2'(u)\mathbf{S}(\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2)|$$
$$\le a\max_{-1\le j\le N}\left|\sum_{i=j-1}^{j+1}\{c_{j(u)-1}(u)S_{j(u)-1,i}+c_{j(u)}(u)S_{j(u),i}\}\Delta_{i,j}\right| = O(h^{-1/2})o(1),$$

where $S_{i,j}$ is the (i,j)th entry of matrix $\mathbf{S}$ and $\Delta_{i,j}$ is the (i,j)th entry of matrix $\mathbf{\Delta}=\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2$ and $\Delta_{i,j}=0$ for $i=-2$ and $i=N+1$. Then

$$\|\mathbf{c}_2'(u)\mathbf{S}(\mathbf{V}-n^{-1}\mathbf{C}_2'\mathbf{C}_2)(n^{-1}\mathbf{C}_2'\mathbf{C}_2)^{-1}\| = U(h^{-1})o(1). \tag{4.20}$$

For $0\le j\le N-1$,

$$E(\mathbf{c}_{j,2}'\mathbf{x})^2 = \frac{3}{2h}E\left\{\sum_{i=1}^{n_j}\left(\frac{j_i}{nh}-j\right)x_{j_i}+\sum_{i=1}^{n_{j+1}}\left(j+2-\frac{j_i}{nh}\right)x_{(j+1)_i}\right\}^2 = O(h^{-1})O(nh) = O(n).$$

Therefore, $\mathbf{c}_{j,2}'\mathbf{x}=O_p(n^{1/2})$. Similarly, it can be shown that $\mathbf{c}_{j,2}'\mathbf{x}=O_p(n^{1/2})$ for $j=-1$ and N. Hence,

$$\frac{1}{n}\|\mathbf{C}_2'\mathbf{x}\| = O_p\{(nh)^{-1/2}\}. \tag{4.21}$$

We conclude (4.18) from (4.20), (4.21) and (4.16). The proof is complete. □

Lemma 4.7. *Under Assumptions* 1–4

$$\liminf_{n\to\infty}P\left(\sup_{u\in[0,1]}\{(E\tilde{x}^2_{nu,2})^{-1/2}|\tilde{x}_{nu,2}|\}\le 2\{\log(N+1)\}^{1/2}d_{n,\alpha/2}\right)\ge 1-\alpha,$$

where $d_{n,\alpha}$ is defined in (2.10).

Proof. Define a vector $\boldsymbol{\lambda}_j$ as follows:

$$\boldsymbol{\lambda}_j = \frac{1}{\sqrt{n}}\mathbf{S}_{j-1,j}(\mathbf{C}_2'\mathbf{x}),$$

where $\mathbf{S}_{j-1,j}$ is a $2\times(N+2)$ submatrix of $\mathbf{S}$,

$$\mathbf{S}_{j-1,j} = \begin{pmatrix} s_{j-1,-1} & \cdots & s_{j-1,N} \\ s_{j,-1} & \cdots & s_{j,N}\end{pmatrix}.$$

Define $\{Cov(\boldsymbol{\lambda}_j)\}^{-1/2}=(\lambda^*_{i,i'})^{j}_{i,i'=j-1}$. The maximal absolute value of $Cov(\boldsymbol{\lambda}_j)$ satisfies

$$|Cov(\boldsymbol{\lambda}_j)| = \frac{1}{n}|\mathbf{S}\mathbf{C}_2'\mathbf{\Gamma}\mathbf{C}_2\mathbf{S}| = O(1),$$

and thus

$$\max_{0\le j\le N}|Cov(\boldsymbol{\lambda}_j)| = O(1). \tag{4.22}$$

Notice that

$$(E\tilde{x}^2_{nu,2})^{-1}\tilde{x}^2_{nu,2} = \frac{n^{-2}\mathbf{c}_2'(u)\mathbf{S}\mathbf{C}_2'\mathbf{x}\mathbf{x}'\mathbf{C}_2\mathbf{S}\mathbf{c}_2(u)}{n^{-2}\mathbf{c}_2'(u)\mathbf{S}\mathbf{C}_2'\mathbf{\Gamma}\mathbf{C}_2\mathbf{S}\mathbf{c}_2(u)} \le \boldsymbol{\lambda}_{j(u)}'\{Cov(\boldsymbol{\lambda}_{j(u)})\}^{-1}\boldsymbol{\lambda}_{j(u)}.$$

Define

$$\xi_{i,j(u),l}=\sum_{j=-1}^{N}(\lambda^{*}_{j(u)+l-2,j(u)-1}s_{j(u)-1,j}+\lambda^{*}_{j(u)+l-2,j(u)}s_{j(u),j})c_{j,2}(u_i)x_i,\quad l=1,2, \tag{4.23}$$

and hence

$$\{Cov(\boldsymbol{\lambda}_{j(u)})\}^{-1/2}\boldsymbol{\lambda}_{j(u)}=\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j(u),1}\\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j(u),2}\end{pmatrix}. \tag{4.24}$$

Therefore,

$$(E\tilde{x}^2_{nu,2})^{-1}\tilde{x}^2_{nu,2}\le\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j(u),1}\right)^2+\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j(u),2}\right)^2.$$

If we can show that

$$\limsup_{n\to\infty}P\left(\max_{0\le j\le N}\left|n^{-1/2}\sum_{i=1}^{n}\xi_{i,j,l}\right|>\{2\log(N+1)\}^{1/2}d_{n,\alpha/2}\right)=\frac{\alpha}{2}, \tag{4.25}$$

then

$$\liminf_{n\to\infty}P\left(\sup_{u\in[0,1]}\{(E\tilde{x}^2_{nu,2})^{-1/2}|\tilde{x}_{nu,2}|\}\le 2\{\log(N+1)\}^{1/2}d_{n,\alpha/2}\right)\ge\liminf_{n\to\infty}P\left(\max_{0\le j\le N}\boldsymbol{\lambda}'_j\{\mathrm{Cov}(\boldsymbol{\lambda}_j)\}^{-1}\boldsymbol{\lambda}_j\le 4\log(N+1)d^2_{n,\alpha/2}\right)$$

$$=1-\limsup_{n\to\infty}P\left(\max_{0\le j\le N}\sum_{l=1,2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j,l}\right)^2>4\log(N+1)d^2_{n,\alpha/2}\right)$$

$$\ge 1-\sum_{l=1,2}\limsup_{n\to\infty}P\left(\max_{0\le j\le N}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j,l}\right)^2>2\log(N+1)d^2_{n,\alpha/2}\right)=1-\alpha.$$

The following is needed to show (4.25),

$$\sup_z\left|P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{i,j,l}<z\right)-\Phi(z)\right|=O\{(nh)^{-1/2}\}. \tag{4.26}$$

Since the proofs of (4.26) for $\xi_{i,j,1}$ and $\xi_{i,j,2}$ are very similar, we only consider $\xi_{i,j,1}$. For any fixed j, $\{\xi_{i,j,1};1\le i\le n\}$ is an α-mixing sequence. In addition, from (4.23)

$$\max_{1\le i\le n_j}E|\xi_{i,j,1}|^3=ah^{-3/2},$$

where a is a constant obtained by the definition of matrix **S**, (4.22) and Assumption 4. From (4.24), $E(\sum_{i=1}^{n}\xi_{i,j,1})^2=n$. Thus (4.26) is concluded from Sunklodas' Theorem.

Following a similar procedure to that of Lemma 4.3, it is straightforward to obtain the following:

$$P\left(\max_{0\le j\le N}\left|n^{-1/2}\sum_{i=1}^{n}\xi_{i,j,l}\right|\le\{2\log(N+1)\}^{1/2}d_{n,\alpha/2}\right)\ge 1-(N+1)[1+O\{(nh)^{1/2}\}]+(N+1)\left\{1-\frac{\alpha}{2(N+1)}+o(N^{-1})\right\}$$

$$=1-\frac{\alpha}{2}+O\{(nh^3)^{-1/2}\}+o(1).$$

Under Assumption 2, the last two terms are negligible. Therefore, (4.25) holds. The proof is complete. □

Proof of Theorem 2.1 for $m=2$ is very similar to that of the case $m=1$. First based on Assumptions 1 and 2, and (2.7), it can be shown that

$$\hat{\eta}^{-1}_{n,2}(u)\sup_{u\in[0,1]}|\tilde{g}_2(u)-g(u)|=U_p\{(nh)^{1/2}\}U(h^2)=U_p\{(nh^5)^{1/2}\}=u_p\{(\log N)\}.$$

On the other hand, according to Lemmas 4.5, 4.6, (4.17) and (4.11), $\hat{\eta}^{-1}_{n,2}(u)\sup_{u\in[0,1]}\{\hat{g}_2(u)-g(u)\}$ asymptotically has the same distribution as $\sup_{u\in[0,1]}\{(E\tilde{x}^2_{nu,2})^{-1/2}|\tilde{x}_{nu,2}|\}$. Therefore, (2.9) is obtained from Lemma 4.7.

Acknowledgments

The authors wish to thank the referee for the critical comments and constructive suggestions which improved the paper considerably.

Appendix A

Theorem A.1 (Sunklodas, 1984). *Let $\{\xi_i\}_{i=1}^n$ be an α-mixing sequence with $E\xi_i = 0$, $d = \max_{1 \le i \le n} E|\xi_i|^s < \infty$ $(2 < s \le 3)$, and the α-mixing coefficient satisfying: $\alpha(n) \le Ke^{-\lambda n}$, where $K > 0$ is a finite constant and $\lambda > 0$. Let $S_n = \sum_{i=1}^n \xi_i$ and $B_n^2 = ES_n^2 \ge c_0 n$ for some $0 < c_0 < \infty$. Then there exist constants $c_1(K,s)$ and $c_2(K,s)$ such that for all λ $(\lambda_1 \le \lambda \le \lambda_2)$ and all $n > 1$,*

$$\Delta_n = \sup_z |P(B_n^{-1} S_n < z) - \Phi(z)| \le c_1(K,s) \frac{d}{c_0 B_n^{s-2}} \left\{ \frac{1}{\lambda} \log(B_n c_0^{-1/2}) \right\}^{s-1},$$

where

$$\lambda_1 = c_2(K,s) n^{-1} \{\log(B_n c_0^{-1/2})\}^{c_3}, \quad c_3 > \frac{2(s-1)}{s-2},$$

$$\lambda_2 = \frac{4s}{s-2} \log(B_n c_0^{-1/2}).$$

Theorem A.2 (Leadbetter et al., 1983). *If $\xi_1, \ldots, \xi_N$ are IID standard normal random variables, then for $\tau \in R$, as $N \to \infty$,*

$$P\left(\max_{1 \le j \le N} \xi_j \le \tau / a_N + b_N \right) \to \exp(-e^{-\tau}),$$

and

$$P\left(\max_{1 \le j \le N} |\xi_j| \le \tau / a_N + b_N \right) \to \exp(-2e^{-\tau}),$$

where

$$a_N = (2 \log N)^{1/2}, \quad b_N = (2 \log N)^{1/2} - \tfrac{1}{2}(2 \log N)^{-1/2} \{\log \log N + \log(4\pi)\}.$$

References

Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Annals of Statistics 1, 1071–1095.

Claeskens, G., Van Keilegom, I., 2003. Bootstrap confidence bands for regression curves and their derivatives. Annals of Statistics 31, 1852–1884.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman & Hall.

Fan, J., Yao, Q.W., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York.

Fuller, W.A., 1996. Introduction to Statistical Time Series. Wiley.

Hall, P., Titterington, D.M., 1988. On confidence bands in nonparametric density estimation and regression. Journal of Multivariate Analysis 27, 228–254.

Hall, P., Keilegom, I.V., 2003. Using difference-based methods for inference in nonparametric regression with time series errors. Journal of the Royal Statistical Society. Series B. Statistical Methodology 65, 443–456.

Huang, J.Z., 2003. Local asymptotics for polynomial spline regression. Annals of Statistics 31, 1600–1635.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York.

Opsomer, J., Wang, Y.D., Yang, Y.H., 2001. Nonparametric regression with correlated errors. Statistical Science 16, 134–153.

Shao, Q., Yang, L., 2011. Autoregressive coefficient estimation in nonparametric analysis. Journal of Time Series Analysis 32, 587–597.

Stone, C.J., 1994. The use of polynomial splines and their tensor products in multivariate function estimation. Annals of Statistics 22, 118–184.

Sunklodas, J., 1984. On the rate of convergence in the central limit theorem for strongly mixing random variables. Lithuanian Mathematical Journal 24, 182–190.

Wang, L., Yang, L., 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. Annals of Statistics 35, 2474–2503.

Wang, J., Yang, L., 2009. Polynomial spline confidence bands for regression curves. Statistica Sinica 19, 325–342.

Wu, W.B., Zhao, Z.B., 2007. Inference of trends in time series. Journal of the Royal Statistical Society. Series B. Statistical Methodology 69, 391–410.

Xia, Y., 1998. Bias-corrected confidence bands in nonparametric regression. Journal of the Royal Statistical Society. Series B. Statistical Methodology 60, 797–811.

Xue, L., Yang, L., 2006. Additive coefficient modelling via polynomial spline. Statistica Sinica 16, 1423–1446.