MULTIVARIATE LOCAL POLYNOMIAL REGRESSION FOR TIME SERIES: UNIFORM STRONG CONSISTENCY AND RATES

By Elias Masry

Department of Electrical and Computer Engineering, University of California at San Diego

First version received October 1995

Abstract. Local high-order polynomial fitting is employed for the estimation of the multivariate regression function $m(x_1, \ldots, x_d) = E\{\psi(Y_d)|X_1 = x_1, \ldots, X_d = x_d\}$, and of its partial derivatives, for stationary random processes $\{Y_i, X_i\}$. The function ψ may be selected to yield estimates of the conditional mean, conditional moments and conditional distributions. Uniform strong consistency over compact subsets of R^d, along with rates, are established for the regression function and its partial derivatives for strongly mixing processes.

Keywords. Multivariate regression estimation; local polynomial fitting; mixing processes; uniform strong consistency; rates of convergence.

1. INTRODUCTION

Let $\{Y_i, X_i\}_{i=-\infty}^{\infty}$ be jointly stationary processes on the real line and let ψ be an arbitrary measurable function. Assume that $E|\psi(Y_1)| < \infty$ and define the multivariate regression function

$$m(x_1, \ldots, x_d) = E\{\psi(Y_d)|X_1 = x_1, \ldots, X_d = x_d\} \tag{1.1}$$

where the dimension $d \geqslant 1$. The regression function $m(x_1, \ldots, x_d)$ plays an important role in data analysis, filtering ($X_i = Y_i + \varepsilon_i$) and prediction ($Y_i = X_{i+r}$ —r-step prediction) of time series. Our goal is to estimate the regression function $m(x_1, \ldots, x_d)$ and its partial derivatives from the observations $\{Y_i, X_i\}_{i=1}^{n}$. There is fairly extensive literature on the use of the Nadaraya–Watson (1964) estimator in connection with regression estimation. See, for example, Rosenblatt (1969), Robinson (1983, 1986), Collomb and Härdle (1986), Roussas (1990), Roussas and Tran (1992) and Fan and Masry (1992) among others.

The local polynomial fitting approach was introduced originally by Stone (1977) and has been studied in recent years by many researchers (see Fan (1993) and the references therein). Local polynomial fitting has significant advantages over the Nadaraya–Watson regression estimator: it has been shown that local linear fitting in the univariate case reduces the bias (Fan, 1992); when the underlying probability density has a compact support, no boundary

0143-9782/96/06 571–599 JOURNAL OF TIME SERIES ANALYSIS Vol. 17, No. 6

degradation occurs in regions near the end points (Fan and Gijbels, 1992; Ruppert and Wand, 1994); it is superior to the Nadaraya–Watson estimator in the context of estimating the derivatives of the regression function (Fan and Gijbels, 1992; Ruppert and Wand, 1994). All the above-cited works consider an independently and identically distributed (i.i.d.) setting. We mention here that in the i.i.d. case $\{Y_i, X_i\}$ and $m(\boldsymbol{x}) = E\{Y_1 | X_1 = \boldsymbol{x}\}$ where X_i are R^d-valued random variables, Ruppert and Wand (1994) considered local quadratic fit and provided leading bias and variance terms for $m(\boldsymbol{x})$ and its derivatives. In a recent work, Masry (1995) formulated the multivariate regression estimation problem in the general setting given in (1.1), in conjunction with local higher-order polynomial fitting, and established the joint asymptotic normality of the estimator $\hat{m}(x_1, \ldots, x_d)$ and its partial derivatives up to a fixed total order p. Expressions for the bias and variance–covariance matrix (of the joint asymptotic distribution) of these estimators were given.

The purpose of this paper is to establish the strong consistency of the estimator $\hat{m}(x_1, \ldots, x_d)$ and its partial derivatives up to a fixed total order p, and to obtain sharp rates of almost sure convergence which are uniform over compact sets. It should be noted that, to the best of our knowledge, no strong convergence results are available in the literature for local polynomial regression fitting either in the general framework considered in this paper or for the special case of linear fitting with i.i.d. observations. Moreover, as will become clear from the analysis in the following sections, establishing strong rates of convergence for the multivariate local polynomial fitting estimators is technically more involved than for the classical Nadaraya–Watson regression estimator.

An important area of application of the results of this paper is the estimation/identification of the functional structure of nonlinear time series commonly encountered in econometric time series (Tjostheim, 1994). Consider, for example, the popular autoregressive conditional heteroskedastic (ARCH) time series

$$X_j = f_1(X_{j-1}, \ldots, X_{j-d}) + f_2(X_{j-1}, \ldots, X_{j-d})e_j$$

where the functions f_1 and $f_2 \geqslant 0$ are to be determined via estimation. When the $\{e_j\}$ are i.i.d. with zero mean and variance σ^2 then

$$E\{X_j | X_{j-1} = x_1, \ldots, X_{j-d} = x_d\} = f_1(x_1, \ldots, x_d)$$

$$\operatorname{var}(X_j | X_{j-1} = x_1, \ldots, X_{j-d} = x_d) = \sigma^2 f_2^2(x_1, \ldots, x_d)$$

and regression estimation is the natural approach (Masry and Tjostheim, 1995). The general framework (1.1) of local polynomial fitting considered in this paper can be used to provide estimates of f_1 and of $\sigma^2 f_2^2$ and of their partial derivatives which are uniformly strongly convergent over compact sets of R^d.

The local high-order polynomial regression estimation problem is formulated as follows. Let

$$\boldsymbol{X}_j = (X_{j+1}, \ldots, X_{j+d}) \tag{1.2}$$

and

$$m(\boldsymbol{x}) = E\{\psi(Y_d)|\boldsymbol{X}_0 = \boldsymbol{x}\}. \tag{1.3}$$

Assuming that $m(\boldsymbol{x})$ has derivatives of total order $p+1$ at the point $\boldsymbol{x}$ we can approximate $m(\boldsymbol{z})$ locally by a multivariate polynomial of total order p:

$$m(\boldsymbol{z}) \simeq \sum_{0 \leqslant |\boldsymbol{k}| \leqslant p} \frac{1}{\boldsymbol{k}!} D^{\boldsymbol{k}} m(\boldsymbol{y})|_{\boldsymbol{y}=\boldsymbol{x}} (\boldsymbol{z}-\boldsymbol{x})^{\boldsymbol{k}} \tag{1.4}$$

where we use the notation

$$\boldsymbol{k} = (k_1, \ldots, k_d) \qquad \boldsymbol{k}! = k_1! \times \cdots \times k_d! \qquad |\boldsymbol{k}| = \sum_{i=1}^{d} k_i. \tag{1.5}$$

$$\boldsymbol{x}^{\boldsymbol{k}} = x_1^{k_1} \times \cdots \times x_d^{k_d} \tag{1.6}$$

$$\sum_{0 \leqslant |\boldsymbol{k}| \leqslant p} = \sum_{j=0}^{p} \sum_{k_1=0}^{j} \cdots \sum_{\substack{k_d=0 \\ k_1+\cdots+k_d=j}}^{j} \tag{1.7}$$

and

$$(D^{\boldsymbol{k}} m)(\boldsymbol{y}) = \frac{\partial^{\boldsymbol{k}} m(\boldsymbol{y})}{\partial y_1^{k_1} \cdots \partial y_d^{k_d}}. \tag{1.8}$$

Let $K(\boldsymbol{u})$ be a nonnegative weight function on R^d and let h be a bandwidth parameter. Given the observations $\{Y_i, X_i\}_{i=0}^{n}$, consider the multivariate weighted least squares

$$\sum_{i=0}^{n-d} \left\{ \psi(Y_{d+i}) - \sum_{0 \leqslant |\boldsymbol{k}| \leqslant p} b_{\boldsymbol{k}}(\boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})^{\boldsymbol{k}} \right\}^2 K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right). \tag{1.9}$$

Minimizing (1.9) with respect to each $b_{\boldsymbol{k}}$ gives an estimate $\hat{b}_{\boldsymbol{k}}(\boldsymbol{x})$ and, by (1.4), $\boldsymbol{k}!\hat{b}_{\boldsymbol{k}}(\boldsymbol{x})$ estimates $(D^{\boldsymbol{k}} m)(\boldsymbol{x})$ so that $(\widehat{D^{\boldsymbol{k}} m})(\boldsymbol{x}) = \boldsymbol{k}!\hat{b}_{\boldsymbol{k}}(\boldsymbol{x})$. The minimization of (1.9) leads to the set of equations

$$t_{n,\boldsymbol{j}}(\boldsymbol{x}) = \sum_{0 \leqslant |\boldsymbol{k}| \leqslant p} h^{|\boldsymbol{k}|} \hat{b}_{\boldsymbol{k}}(\boldsymbol{x}) s_{n,\boldsymbol{j}+\boldsymbol{k}}(\boldsymbol{x}) \qquad 0 \leqslant |\boldsymbol{j}| \leqslant p \tag{1.10}$$

where

$$t_{n,\boldsymbol{j}}(\boldsymbol{x}) = \frac{1}{n-d+1} \sum_{i=0}^{n-d} \psi(Y_{d+i}) \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right)^{\boldsymbol{j}} K_h(\boldsymbol{X}_i - \boldsymbol{x}) \tag{1.11}$$

$$s_{n,\boldsymbol{j}}(\boldsymbol{x}) = \frac{1}{n-d+1} \sum_{i=0}^{n-d} \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right)^{\boldsymbol{j}} K_h(\boldsymbol{X}_i - \boldsymbol{x}) \tag{1.12}$$

and

$$K_h(\boldsymbol{u}) = \frac{1}{h^d} K\left(\frac{\boldsymbol{u}}{h}\right). \tag{1.13}$$

The organization of the paper is as follows. Section 2 recasts the problem of estimating $m(\boldsymbol{x})$ and its derivatives in a convenient vector form, and decomposes the estimation error as a sum of vectors by centering the $t_{n,\boldsymbol{j}}$. Section 3 establishes the uniform almost sure convergence of $s_{n,\boldsymbol{j}}$. Section 4 derives the strong consistency and rates of the centered $t_{n,\boldsymbol{j}}$ and of the regression function $m(\boldsymbol{x})$ and its partial derivatives.

2. DECOMPOSITION OF THE ESTIMATION ERROR

We first note that the set of equations (1.10) can be cast in matrix form by using a lexicographical order in the following manner. Let

$$N_i = \binom{i+d-1}{d-1}$$

be the number of distinct d-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}| = i$. Arrange these N_i d-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, \ldots, 0, i)$ is the first element in the sequence and $(i, 0, \ldots, 0)$ the last element) and let g_i^{-1} denote this one-to-one map. Arrange the $N_{|\boldsymbol{j}|}$ values of $t_{n,\boldsymbol{j}}$ in a column vector $\boldsymbol{\tau}_{n,|\boldsymbol{j}|}$ according to the above order. Then

$$(\boldsymbol{\tau}_{n,|\boldsymbol{j}|})_k = t_{n,\, g_{|\boldsymbol{j}|}(k)}. \tag{2.1}$$

Define

$$\boldsymbol{\tau}_n = \begin{bmatrix} \boldsymbol{\tau}_{n,0} \\ \boldsymbol{\tau}_{n,1} \\ \vdots \\ \boldsymbol{\tau}_{n,p} \end{bmatrix} \tag{2.2}$$

and note that the column vector $\boldsymbol{\tau}_n$ is of dimension $N = \sum_{i=0}^{p} N_i \times 1$. Similarly arrange the distinct values of $h^{|\boldsymbol{k}|}\hat{b}_{\boldsymbol{k}}$, $0 \leq |\boldsymbol{k}| \leq p$, as a column vector of dimension $N \times 1$ in the form

$$\hat{\boldsymbol{\beta}}_n = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{n,0} \\ \hat{\boldsymbol{\beta}}_{n,1} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{n,p} \end{bmatrix} \tag{2.3}$$

Similarly, define the vector $\boldsymbol{\beta}$ by a lexicographical arrangement of $h^{|\boldsymbol{k}|}b_{\boldsymbol{k}}$ for $0 \leq |\boldsymbol{k}| \leq p$. Finally, arrange the possible values of $s_{n,\boldsymbol{j}+\boldsymbol{k}}$ by a matrix $\boldsymbol{S}_{n,|\boldsymbol{j}|,|\boldsymbol{k}|}$ in a lexicographical order with the (l, m)th element of $\boldsymbol{S}_{n, |j|, |k|}$ given by

$$(\boldsymbol{S}_{n, |j|, |k|})_{l, m} = s_{n, g_{|j|}(l)+g_{|k|}(m)}. \tag{2.4}$$

The matrix $\boldsymbol{S}_{n, |j|, |k|}$ has dimension $N_{|j|} \times N_{|k|}$. Now define the $N \times N$ matrix $\boldsymbol{S}_n$ by

$$\boldsymbol{S}_n = \begin{bmatrix} \boldsymbol{S}_{n,0,0} & \boldsymbol{S}_{n,0,1} & \cdots & \boldsymbol{S}_{n,0,p} \\ \boldsymbol{S}_{n,1,0} & \boldsymbol{S}_{n,1,1} & \cdots & \boldsymbol{S}_{n,1,p} \\ \vdots & & & \vdots \\ \boldsymbol{S}_{n,p,0} & \boldsymbol{S}_{n,p,1} & \cdots & \boldsymbol{S}_{n,p,p} \end{bmatrix}. \tag{2.5}$$

Then the set of equations (1.10) can be written in the matrix form

$$\boldsymbol{\tau}_n(\boldsymbol{x}) = \boldsymbol{S}_n(\boldsymbol{x})\hat{\boldsymbol{\beta}}_n(\boldsymbol{x}).$$

Because of the functional form of $s_{n, j+k}$ we have

$$\sum_{0\leqslant|j|\leqslant p}\sum_{0\leqslant|k|\leqslant p} c_j c_k s_{n, j+k} = \frac{1}{n-d+1}\sum_{i=0}^{n-d}\left\{\sum_{0\leqslant|j|\leqslant p} c_j\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h}\right)^j\right\}^2 K_h(\boldsymbol{X}_i-\boldsymbol{x}) > 0$$

for $K > 0$. Thus we assume henceforth that the matrix $\boldsymbol{S}_n$ is positive definite and we write

$$\hat{\boldsymbol{\beta}}_n(\boldsymbol{x}) = \boldsymbol{S}_n^{-1}(\boldsymbol{x})\boldsymbol{\tau}_n(\boldsymbol{x}) \tag{2.6}$$

as the solution of the set of equations (1.10).

We next provide a fundamental decomposition for the error $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$. We first center $t_{n,j}$ as follows. Let

$$t^*_{n,j}(\boldsymbol{x}) = \frac{1}{n-d+1}\sum_{i=0}^{n-d}\{\psi(Y_{d+i}) - m(\boldsymbol{X}_i)\}\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h}\right)^j K_h(\boldsymbol{X}_i-\boldsymbol{x}). \tag{2.7}$$

We have

$$t_{n,j}(\boldsymbol{x}) - t^*_{n,j}(\boldsymbol{x}) = \frac{1}{n-d+1}\sum_{i=0}^{n-d} m(\boldsymbol{X}_i)\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h}\right)^j K_h(\boldsymbol{X}_i-\boldsymbol{x}). \tag{2.8}$$

Expanding $m(\boldsymbol{X}_i)$ in a Taylor series around $\boldsymbol{x}$ with an integral remainder and assuming that $m(\boldsymbol{x})$ has continuous derivatives of total order $p+1$, we have

$$\begin{aligned} m(\boldsymbol{X}_i) = {} & \sum_{0\leqslant|k|\leqslant p+1}\frac{1}{\boldsymbol{k}!}(D^k m)(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})^k \\ & + (p+1)\sum_{|k|=p+1}\frac{1}{\boldsymbol{k}!}(\boldsymbol{X}_i-\boldsymbol{x})^k\int_0^1[(D^k m)\{\boldsymbol{x}+w(\boldsymbol{X}_i-\boldsymbol{x})\} - (D^k m)(\boldsymbol{x})] \\ & \times(1-w)^p\,dw. \end{aligned} \tag{2.9}$$

Substituting in (2.8) and using (1.12) we find

$$t_{n,j}(\boldsymbol{x}) - t^*_{n,j}(\boldsymbol{x}) = \sum_{0 \leq |\boldsymbol{k}| \leq p+1} \frac{1}{\boldsymbol{k}!} h^{|\boldsymbol{k}|}(D^{\boldsymbol{k}}m)(\boldsymbol{x}) s_{n,\boldsymbol{j}+\boldsymbol{k}}(\boldsymbol{x}) + e_{n,\boldsymbol{j}}(\boldsymbol{x})$$

where

$$e_{n,\boldsymbol{j}}(\boldsymbol{x}) = (p+1) \sum_{|\boldsymbol{k}|=p+1} \frac{h^{p+1}}{\boldsymbol{k}!} \frac{1}{n-d+1} \sum_{i=0}^{n-d} \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h} \right)^{\boldsymbol{k}+\boldsymbol{j}} K_h(\boldsymbol{X}_i - \boldsymbol{x})$$

$$\times \int_0^1 [(D^{\boldsymbol{k}}m)\{\boldsymbol{x} + w(\boldsymbol{X}_i - \boldsymbol{x})\} - (D^{\boldsymbol{k}}m)(\boldsymbol{x})](1-w)^p \, dw. \tag{2.10}$$

Using (1.10) and $D^{\boldsymbol{k}}m = \boldsymbol{k}! b_{\boldsymbol{k}}$ we obtain

$$t^*_{n,\boldsymbol{j}}(\boldsymbol{x}) = \sum_{0 \leq |\boldsymbol{k}| \leq p} h^{|\boldsymbol{k}|} \{\hat{b}_{\boldsymbol{k}}(\boldsymbol{x}) - b_{\boldsymbol{k}}(\boldsymbol{x})\} s_{n,\boldsymbol{j}+\boldsymbol{k}}(\boldsymbol{x})$$

$$- h^{p+1} \sum_{|\boldsymbol{k}|=p+1} \frac{1}{\boldsymbol{k}!} (D^{\boldsymbol{k}}m)(\boldsymbol{x}) s_{n,\boldsymbol{j}+\boldsymbol{k}}(\boldsymbol{x}) + e_{n,\boldsymbol{j}}(\boldsymbol{x}) \tag{2.11}$$

Now arrange the N_{p+1} elements of the derivatives $(1/\boldsymbol{j}!)(D^{\boldsymbol{j}}m)(\boldsymbol{x})$ for $|\boldsymbol{j}| = p+1$ as a column vector $\boldsymbol{m}_{p+1}(\boldsymbol{x})$ using the lexicographical order introduced earlier; similarly define the $N \times 1$ vector $\boldsymbol{e}_n$ for the elements $\{e_{n,\boldsymbol{j}}\}$ with $0 \leq |\boldsymbol{j}| \leq p$. Also, let the $N \times N_{p+1}$ matrix $\boldsymbol{B}_n$ be defined by

$$\boldsymbol{B}_n(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{S}_{n,0,p+1} \\ \boldsymbol{S}_{n,1,p+1} \\ \vdots \\ \boldsymbol{S}_{n,p,p+1} \end{bmatrix} \tag{2.12}$$

where the matrix $\boldsymbol{S}_{n,i,p+1}$ is defined in (2.4). Then we can write (2.11), using (2.3) and the centered version $\boldsymbol{\tau}^*_n$ of (2.2), in the vector form

$$\boldsymbol{\tau}^*_n(\boldsymbol{x}) = \boldsymbol{S}_n(\boldsymbol{x})(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})(\boldsymbol{x}) - h^{p+1}\boldsymbol{B}_n(\boldsymbol{x})\boldsymbol{m}_{p+1}(\boldsymbol{x}) + \boldsymbol{e}_n(\boldsymbol{x}).$$

Thus

$$(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})(\boldsymbol{x}) = \boldsymbol{S}_n^{-1}(\boldsymbol{x})\boldsymbol{\tau}^*_n(\boldsymbol{x}) + h^{p+1}\boldsymbol{S}_n^{-1}(\boldsymbol{x})\boldsymbol{B}_n(\boldsymbol{x})\boldsymbol{m}_{p+1}(\boldsymbol{x}) + \boldsymbol{S}_n^{-1}(\boldsymbol{x})\boldsymbol{e}_n(\boldsymbol{x}). \tag{2.13}$$

This is the fundamental equation needed to establish the uniform strong consistency and rates for the estimator $\hat{\boldsymbol{\beta}}_n$. We establish in Section 3 the uniform strong consistency of the matrix $\boldsymbol{S}_n$ and of the vector $\boldsymbol{e}_n$. In Section 4 we derive uniform rates of strong convergence of the vector $\boldsymbol{\tau}^*_n$ (note that $\boldsymbol{\tau}^*_n$ is centered since $E\{\boldsymbol{\tau}^*_n\} = \boldsymbol{0}$) and of $m(\boldsymbol{x})$ and its partial derivatives up to a total order p. The first term on the right-hand side of (2.13) corresponds to the 'variance' and the second and third terms correspond to the 'bias'.

Note that the ith element of $\hat{\boldsymbol{\beta}}_n$ represents an estimate of the derivative of $m(\boldsymbol{x})$ via the relationship

$$(\hat{\boldsymbol{\beta}}_n)_i = \frac{h^{|j|}(\widehat{D^j m})(\boldsymbol{x})}{\boldsymbol{j}!} \qquad i = g_{|j|}^{-1}(\boldsymbol{j}) + \sum_{k=0}^{|j|-1} N_k. \tag{2.14}$$

Finally, we introduce in this section the mixing coefficient. Let $\mathscr{F}_a^b$ be the σ-algebra of events generated by the random variables $\{Y_j, X_j, a \leqslant j \leqslant b\}$. The stationary processes $\{Y_j, X_j\}$ are called strongly mixing (Rosenblatt, 1956) if

$$\sup_{\substack{A \in \mathscr{F}_{-\infty}^0 \\ B \in \mathscr{F}_k^\infty}} |P[AB] - P[A]P[B]| = \alpha(k) \to 0 \text{ as } k \to \infty.$$

$\alpha(k)$ is the strong mixing coefficient.

3. UNIFORM STRONG CONSISTENCY OF THE MATRIX $\boldsymbol{S}_n(\boldsymbol{x})$ AND OF THE VECTOR $\boldsymbol{e}_n(\boldsymbol{x})$

3.1. *Main results*

We first consider the bias of $s_{n,j}(\boldsymbol{x})$. We make the following assumptions on the kernel function K and on the probability density function f of $\boldsymbol{X}_0$ which is assumed to exist.

CONDITION 1.

(a) The kernel $K \in L_1$ satisfies $\|\boldsymbol{u}\|^{2p} K(\boldsymbol{u}) \in L_1$ and f is bounded and uniformly continuous on R^d or
(b) the density f is Lipschitz of order θ,

$$|f(\boldsymbol{x}) - f(\boldsymbol{u})| \leqslant C_1 \|\boldsymbol{x} - \boldsymbol{u}\|^\theta \qquad \text{for some } 0 < \theta \leqslant 1$$

and $\|\boldsymbol{u}\|^{2p+\theta} K(\boldsymbol{u}) \in L_1$.

PROPOSITION 1. Let $h_n \to 0$ as $n \to \infty$. Then, as $n \to \infty$ we have under Condition 1(a)

$$\sup_{\boldsymbol{x} \in R^d} |E\{s_{n,j}(\boldsymbol{x})\} - f(\boldsymbol{x})\mu_j| = o(1) \tag{3.1a}$$

and under Condition 1(b) we have

$$\sup_{\boldsymbol{x} \in R^d} |E\{s_{n,j}(\boldsymbol{x})\} - f(\boldsymbol{x})\mu_j| = O(h_n^\theta) \tag{3.1b}$$

for each $\boldsymbol{j}$ with $0 \leqslant |\boldsymbol{j}| \leqslant 2p$, where

$$\mu_j = \int_{R^d} \boldsymbol{u}^j K(\boldsymbol{u})\, d\boldsymbol{u}. \tag{3.2}$$

Next we obtain an upper bound on the variance of $s_{n,j}(\boldsymbol{x})$ which is uniform over R^d. Put

$$U_{i,j} = \left(\frac{X_i - x}{h}\right)^j K_h(X_i - x) \tag{3.3}$$

Then, by stationarity,

$$\operatorname{var}\{s_{n,j}(x)\} = \frac{1}{n-d+1}\left\{\operatorname{var}(U_{0,j}) + 2\sum_{l=1}^{n-d}\left(1 - \frac{l}{n-d+1}\right)\operatorname{cov}(U_{0,j}, U_{l,j})\right\}$$

$$= J_1(x) + J_2(x). \tag{3.4}$$

The following assumption is made on the kernel K and on the random processes $\{Y_i, X_i\}$.

CONDITION 2.

(a) The kernel $K \in L_1$ is bounded and $\|u\|^{4p} K(u) \in L_1$.

(b)

$$f(u) \leqslant C_2 < \infty \qquad f(u, v;\, l) \leqslant C_3 < \infty$$

for all $l \geqslant 1$ where $f(u)$ and $f(u, v; l)$ denote the probability density of, respectively, X_0 and (X_0, X_l).

(c) The processes $\{Y_i, X_i\}$ are strongly mixing with $\sum_{j=1}^{\infty} j^a \{\alpha(j)\}^{1-2/\nu} < \infty$ for some $\nu > 2$ and $a > 1 - 2/\nu$. Also, we assume further that $\|u\|^{2\nu p} K(u) \in L_1$.

REMARK 1. Note that for $1 \leqslant l \leqslant d-1$ the components of X_0 and X_l overlap. The joint density $f(u, v; l)$ in Condition 2(b) is then the density of the random variables $(X_1, \ldots, X_{d+l})$.

THEOREM 1. *Under Condition 2 and the assumption that $h_n \to 0$, $nh_n^d \to \infty$ as $n \to \infty$, we have, for each j with $0 \leqslant |j| \leqslant 2p$, that*

(i)

$$nh_n^d \sup_{x \in R^d} J_1(x) \leqslant C_2 \gamma_{2j} + O(h_n^d)$$

(ii)

$$nh_n^d \sup_{x \in R^d} J_2(x) = o(1)$$

(iii)

$$nh_n^d \sup_{x \in R^d} \operatorname{var}\{s_{n,j}(x)\} \leqslant C_2 \gamma_{2j}\{1 + o(1)\} \tag{3.5}$$

where

$$\gamma_j = \int_{R^d} u^j K^2(u)\, du. \tag{3.6}$$

Next we establish the strong consistency and rates for $s_{n,j}(\boldsymbol{x})$ which are uniform over compact sets. For each $\boldsymbol{j}$ with $0 \leqslant |\boldsymbol{j}| \leqslant 2p+1$ define the function

$$H_j(\boldsymbol{u}) = \boldsymbol{u}^j K(\boldsymbol{u}).$$

The boundedness of H_j is needed in the sequel. Because of the growth of the multivariate monomial $\boldsymbol{u}^j$, we assume that the kernel K has compact support and is bounded. We make the following assumption on the kernel K.

CONDITION 3.

(a) The kernel K is bounded with compact support ($K(\boldsymbol{u}) = 0$ for $\|\boldsymbol{u}\| > A_0$).
(b)

$$|H_j(\boldsymbol{u}) - H_j(\boldsymbol{v})| \leqslant C_4 \|\boldsymbol{u} - \boldsymbol{v}\| \tag{3.7}$$

for all $\boldsymbol{j}$ with $0 \leqslant |\boldsymbol{j}| \leqslant 2p+1$.

THEOREM 2. *Let D be any compact subset of R^d. Let Conditions 2 and 3 hold. Assume the bandwidth $h_n \to 0$ such that $nh_n^d/\ln n \to \infty$ as $n \to \infty$. Define*

$$r(n) \triangleq \left(\frac{nh_n^d}{\ln n}\right)^{1/2} \qquad L(n) = \left(\frac{n}{h_n^{d+2}\ln n}\right)^{d/2}$$

and

$$\psi(n) \triangleq \frac{nL(n)}{r(n)} \left(\frac{n}{h_n^d \ln n}\right)^{1/4} \alpha\{r(n)\}. \tag{3.8a}$$

If the strong mixing coefficient $\alpha(k)$ satisfies

$$\sum_{n=1}^{\infty} \psi(n) < \infty \tag{3.8b}$$

then for each $\boldsymbol{j}$ with $0 \leqslant |\boldsymbol{j}| \leqslant 2p$ we have

$$\sup_{\boldsymbol{x} \in D} |s_{n,j}(\boldsymbol{x}) - E\{s_{n,j}(\boldsymbol{x})\}| = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \text{ almost surely.}$$

Combining Proposition 1 and Theorem 2 we obtain the following.

COROLLARY 1.

(i) Under the conditions of Theorem 2 and Condition 1(a) we have

$$\sup_{\boldsymbol{x} \in D} |s_{n,j}(\boldsymbol{x}) - f(\boldsymbol{x})\mu_j| = \mathrm{o}(1) \text{ almost surely.}$$

(ii) Under the conditions of Theorem 2 and Condition 1(b) we have

$$\sup_{x\in D} |s_{n,j}(x) - f(x)\mu_j| = \left(\frac{\ln n}{nh_n^d}\right)^{1/2} + \mathrm{O}(h_n^\theta) \text{ almost surely}$$

and if the bandwidth h_n is chosen to satisfy $h_n \sim [(\ln n)/n]^{1/(d+2\theta)}$ then

$$\sup_{x\in D} |s_{n,j}(x) - f(x)\mu_j| = \mathrm{O}\left\{\left(\frac{\ln n}{n}\right)^{\theta/(d+2\theta)}\right\} \text{ almost surely.}$$

We can now state the strong consistency of the matrices $\boldsymbol{S}_n$ and $\boldsymbol{B}_n$ defined in (2.5) and (2.12) respectively. Define the $N \times N$ dimensional matrix $\boldsymbol{M}$ and the $N \times N_{p+1}$ dimensional matrix $\tilde{\boldsymbol{M}}$ by

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{0,0} & \boldsymbol{M}_{0,1} & \dots & \boldsymbol{M}_{0,p} \\ \boldsymbol{M}_{1,0} & \boldsymbol{M}_{1,1} & \dots & \boldsymbol{M}_{1,p} \\ \vdots & & & \vdots \\ \boldsymbol{M}_{p,0} & \boldsymbol{M}_{p,1} & \dots & \boldsymbol{M}_{p,p} \end{bmatrix} \qquad \tilde{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{M}_{0,p+1} \\ \boldsymbol{M}_{1,p+1} \\ \vdots \\ \boldsymbol{M}_{p,p+1} \end{bmatrix} \tag{3.9}$$

where $\boldsymbol{M}_{i,j}$ is an $N_i \times N_j$ dimensional matrix whose (l, m)th element is $\mu_{g_i(l)+g_j(m)}$ and μ_j is defined in (3.2). Note that the elements of the matrices $\boldsymbol{M}$ and $\tilde{\boldsymbol{M}}$ are simply multivariate moments of the kernel K.

COROLLARY 2.

(i) Under the conditions of Theorem 2 and Condition 1(a) we have, uniformly in $\boldsymbol{x} \in D$,

$$\boldsymbol{S}_n(\boldsymbol{x}) \to \boldsymbol{M}f(\boldsymbol{x}) \qquad \boldsymbol{B}_n(\boldsymbol{x}) \to \tilde{\boldsymbol{M}}f(\boldsymbol{x}) \qquad \text{almost surely as } n \to \infty.$$

(ii) Under the conditions of Theorem 2 and Condition 1(b) we have, uniformly in $\boldsymbol{x} \in D$,

$$\boldsymbol{S}_n(\boldsymbol{x}) - \boldsymbol{M}f(\boldsymbol{x}) = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} + \mathrm{O}(h_n^\theta)$$

$$\boldsymbol{B}_n(\boldsymbol{x}) - \tilde{\boldsymbol{M}}f(\boldsymbol{x}) = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} + \mathrm{O}(h_n^\theta)$$

and if the bandwidth h_n is chosen to satisfy $h_n \sim \{(\ln n)/n\}^{1/(d+2\theta)}$ then

$$\boldsymbol{S}_n(\boldsymbol{x}) - \boldsymbol{M}f(\boldsymbol{x}) = \mathrm{O}\left\{\left(\frac{\ln n}{n}\right)^{\theta/(d+2\theta)}\right\} \text{ almost surely}$$

$$\boldsymbol{B}_n(\boldsymbol{x}) - \tilde{\boldsymbol{M}}f(\boldsymbol{x}) = \mathrm{O}\left\{\left(\frac{\ln n}{n}\right)^{\theta/(d+2\theta)}\right\} \text{ almost surely.}$$

Finally we consider in this section the uniform strong consistency of the error term $e_{n,j}(x)$ given in (2.10). We write it in the form

$$e_{n,j}(x) = (p+1)h^{p+1} \sum_{|k|=p+1} \frac{1}{k!} G_{n,j+k}(x) \qquad 0 \leqslant |j| \leqslant p \tag{3.10}$$

where

$$G_{n,j+k}(x) = \frac{1}{n-d+1} \sum_{i=0}^{n-d} \left(\frac{X_i - x}{h} \right)^{k+j} \rho(X_i - x; x) K_h(X_i - x) \tag{3.11}$$

with

$$\rho(X_i - x; x) = \int_0^1 [(D^k m)\{x + w(X_i - x)\} - (D^k m)(x)](1-w)^p \, dw. \tag{3.12}$$

We establish the uniform strong consistency of $G_{n,j+k}$ from which the result for $e_{n,j}$ follows.

CONDITION 4.

(a) The kernel $K \in L_1$ satisfies $\|u\|^{2p+1} K(u) \in L_1$ and $f \leqslant C_2 < \infty$.
(b) $(D^k m)(x)$ is bounded and uniformly continuous on R^d for $|k| = p+1$.

PROPOSITION 2. Let $h_n \to 0$ as $n \to \infty$, and let Condition 4 hold. Then for each j with $0 \leqslant |j| \leqslant p$ we have

$$\sup_{x \in R^d} |E\{G_{n,j+k}(x)\}| = o(1) \qquad \sup_{x \in R^d} |E\{e_{n,j}(x)\}| = o(h_n^{p+1}).$$

Next we obtain a uniform bound on the variance of $G_{n,j+k}(x)$ under the following condition.

CONDITION 5.

(a) Same as Condition 4(a).
(b) Same as Condition 2(b).
(c) Same as Condition 2(c).
(d)

$$\sup_{u \in R^d} |(D^k m)(u)| \leqslant C_5 < \infty \text{ for all } |k| = p+1.$$

THEOREM 3. *Under Condition 5 and the assumption that $h_n \to 0$, $nh_n^d \to \infty$ as $n \to \infty$, we have for each j with $0 \leqslant |j| \leqslant p$ and k with $|k| = p+1$ that*

$$nh_n^d \sup_{x \in R^d} \operatorname{var}\{G_{n,j+k}(x)\} \leqslant A_4\{1 + o(1)\} \qquad A_4 \triangleq C_2 \left(\frac{2C_5}{p+1} \right)^2 \gamma_{2(j+k)} \tag{3.13}$$

where γ_j is given in (3.6).

In order to establish uniform rates of almost sure convergence for $G_{n,j+k}(x)$, we need a Lipschitz condition on $D^k m$ as follows.

CONDITION 6. For k with $|k| = p+1$,

$$|(D^k m)(u) - (D^k m)(v)| \leqslant C_6 \|u - v\|. \tag{3.14}$$

THEOREM 4. *Let D be any compact subset of R^d. Let Conditions 3, 5, and 6 hold. Assume the bandwidth $h_n \to 0$ such that $nh_n^d/\ln n \to \infty$ as $n \to \infty$. Let the strong mixing coefficient $\alpha(k)$ satisfy the summability condition (3.8) of Theorem 2. Then for each j with $0 \leqslant |j| \leqslant p$ and k with $|k| = p+1$ we have*

$$\sup_{x \in D} |G_{n,j+k}(x) - E\{G_{n,j+k}(x)\}| = \mathrm{O}\left\{ \left(\frac{\ln n}{nh_n^d} \right)^{1/2} \right\} \text{ almost surely}$$

$$\sup_{x \in D} |e_{n,j}(x) - E\{e_{n,j}(x)\}| = \mathrm{O}\left\{ \left(\frac{\ln n}{nh_n^d} \right)^{1/2} \right\} h_n^{p+1} = \mathrm{o}(h_n^{p+1}) \text{ almost surely.}$$

Combining Theorem 4 and Proposition 2 we obtain the following.

COROLLARY 3. Under the conditions of Theorem 4 and Condition 4 we have

$$\sup_{x \in D} |e_{n,j}(x)| = \mathrm{o}(h_n^{p+1}) \text{ almost surely}$$

and similarly for the vector $e_n(x)$.

3.2. *Proofs*

PROOF OF PROPOSITION 1. By (1.12) we have

$$E(s_{n,j}) - f(x)\mu_j = \int_{R^d} \left(\frac{u - x}{h} \right)^j K_h(u - x)\{f(u) - f(x)\}\, du. \tag{3.15}$$

Thus, under Condition 1(a), the right-hand side is bounded by

$$\int_{R^d} |H_j(u)| \omega_f(h\|u\|)\, du \tag{3.16}$$

where ω_f is the modulus of continuity of f and $H_j(u) = u^j K(u)$. Since f is bounded, so is ω_f and $\omega_f(h_n\|u\|) \to 0$ as $n \to \infty$. The expression (3.16) tends to zero by dominated convergence. Under Condition 1(b), the right-hand side of (3.15) is bounded by

$$C_1 h_n^\theta \int_{R^d} \|u\|^\theta |H_j(u)|\, du = \mathrm{O}(h_n^\theta).$$

□

PROOF OF THEOREM 1. For part (i) we have under Condition 2(b)

$$\sup_{x \in R^d} |E\{s_{n,j}(x)\}| \leqslant C_2 \int_{R^d} \|u\|^{|j|} K(u)\, du = O(1).$$

Then

$$(n-d+1)J_1 = E(U_{0,j}^2) + O(1)$$

where the O(1) term is uniform in x over R^d so that

$$(n-d+1)h_n^d J_1 = \int_{R^d} \left(\frac{u-x}{h_n}\right)^{2j} \left\{\frac{1}{h^d} K^2\left(\frac{u-x}{h_n}\right)\right\} f(u)\, du + O(h_n^d)$$

and by Condition 2(b)

$$nh_n^d \sup_{x \in R^d} J_1(x) \leqslant C_2 \int_{R^d} u^{2j} K^2(u)\, du + O(h_n^d). \tag{3.17}$$

For part (ii) we decompose the sum in J_2 of (3.4) in three sums

$$\begin{aligned} \frac{1}{n-d+1} \sum_{l=1}^{n-d} |\operatorname{cov}\{U_{0,j}, U_{l,j}\}| &= \frac{1}{n-d+1} \sum_{l=1}^{d-1} + \frac{1}{n-d+1} \sum_{l=d}^{\pi_n} \\ &\quad + \frac{1}{n-d+1} \sum_{l=\pi_n+1}^{n-d} \\ &= J_{21} + J_{22} + J_{23} \end{aligned} \tag{3.18}$$

where $\pi_n \to \infty$ such that $\pi_n h_n^d \to 0$ as $n \to \infty$. For J_{21} there is an overlap between the components of X_0 and X_l. Let $f(u', u'', u''')$ be the joint density of the $d+l$ distinct random variables in (X_0, X_l), where u', u'' and u''' are of dimensions l, $d-l$ and l respectively. Then

$$\begin{aligned} \operatorname{cov}(U_{0,j}, U_{l,j}) = \frac{1}{h^{d-l}} \int_{R^{d+l}} & (u')^{j'} (u'')^{j''} (u''')^{j'''} K(u', u'') K(u'', u''') \\ & \times \{f(x' - hu', x'' - hu'', x''' - hu''') - f(x' - hu', x'' - hu'') \\ & \times f(x'' - hu'', x''' - hu''')\}\, du'\, du''\, du'''. \end{aligned}$$

By Condition 2(b) we have

$$h^{d-l} |\operatorname{cov}(U_{0,j}, U_{l,j})| \leqslant$$

$$(C_3 + C_2^2) \int_{R^{d+l}} \prod_{i=1}^{l} |u_i|^{j_i} \prod_{i=l+1}^{d} |u_i|^{2j_i} \prod_{i=d+1}^{d+l} |u_i|^{j_i} K(u', u'') K(u'', u''')\, du'\, du''\, du'''.$$

Hence

$$nh_n^d \sup_{x\in R^d} |J_{21}(x)| \le \text{constant} \times \sum_{l=1}^{d-1} h_n^l = \mathrm{O}(h_n) \to 0. \tag{3.19a}$$

For J_{22}, there is no overlap between the components of X_0 and X_l so that by Condition 2(b) we have

$$|\mathrm{cov}(U_{0,j}, U_{l,j})| \le (C_3 + C_2^2)\left\{\int_{R^d} \prod_{i=1}^d |u_i|^{j_i} K(u)\,du\right\}^2 < \infty.$$

Hence

$$nh_n^d \sup_{x\in R^d} |J_{22}(x)| \le \text{constant} \times \sum_{l=d}^{\pi_n} h_n^d = \mathrm{O}(\pi_n h_n^d) \to 0 \tag{3.19b}$$

by the choice of π_n. For J_{23} we have by Davydov's lemma (Hall and Heyde, 1980, Corollary A2)

$$|\mathrm{cov}(U_{0,j}, U_{l,j})| \le 8\{\alpha(l-d+1)\}^{1-2/\nu}(E|U_{0,j}|^\nu)^{2/\nu}.$$

It is easily seen that

$$E|U_{0,j}|^\nu \le \int_{R^d} \left\|\frac{u-x}{h}\right\|^{\nu|j|} \frac{1}{h^{\nu d}} K^\nu\left(\frac{u-x}{h}\right) f(u)\,du$$

so that, by Condition 2(b),

$$h_n^{(\nu-1)d} E|U_{0,j}|^\nu \le C_2 \int_{R^d} \|u\|^{\nu|j|} K^\nu(u)\,du.$$

Thus

$$\begin{aligned} nh_n^d \sup_{x\in R^d} |J_{23}(x)| &\le \frac{\text{constant}}{h_n^{d(1-2/\nu)}} \sum_{l=\pi_n}^\infty \{\alpha(l-d+1)\}^{1-2/\nu} \\ &\le \frac{\text{constant}}{h_n^{d(1-2/\nu)} \pi_n^a} \sum_{l=\pi_n}^\infty l^a \{\alpha(l)\}^{1-2/\nu}. \end{aligned}$$

Choose $\pi_n = h_n^{d(1-2/\nu)/a}$ and note that since $a > 1 - 2/\nu$ we indeed have $\pi_n h_n^d \to 0$ as required. Then

$$nh_n^d \sup_{x\in R^d} |J_{23}(x)| \le \text{constant} \times \sum_{l=\pi_n}^\infty l^a \{\alpha(l)\}^{1-2/\nu} \to 0 \tag{3.19c}$$

as $n \to \infty$ by Condition 2(c). Part (ii) of the theorem now follows from (3.18) and (3.19). Part (iii) of the theorem follows from parts (i) and (ii). □

PROOF OF THEOREM 2. Since D is compact, it can be covered by a finite number $L = L(n)$ of cubes $I_k = I_{n,k}$ with centers $x_k = x_{n,k}$ having sides of length l_n for $k = 1, \ldots, L(n)$. Clearly $l_n = \text{constant}/L^{1/d}(n)$ since D is compact. Write

$$\begin{aligned}\sup_{x\in D} |s_{n,j}(x) - E\{s_{n,j}(x)\}| &= \max_{1\leqslant k\leqslant L(n)} \sup_{x\in D\cap I_k} |s_{n,j}(x) - E\{s_{n,j}(x)\}| \\ &\leqslant \max_{1\leqslant k\leqslant L(n)} \sup_{x\in D\cap I_k} |s_{n,j}(x) - s_{n,j}(x_k)| \\ &\quad + \max_{1\leqslant k\leqslant L(n)} |s_{n,j}(x_k) - E\{s_{n,j}(x_k)\}| \\ &\quad + \max_{1\leqslant k\leqslant L(n)} \sup_{x\in D\cap I_k} |E\{s_{n,j}(x_k)\} - E\{s_{n,j}(x)\}| \\ &\equiv Q_1 + Q_2 + Q_3. \end{aligned} \tag{3.20}$$

Now by (1.12) and Condition 3,

$$|s_{n,j}(x) - s_{n,j}(x_k)| \leqslant (C_4/h_n^{d+1})\|x - x_k\| \tag{3.21}$$

so that

$$Q_1 \leqslant \frac{C_4 l_n}{h_n^{d+1}} = \frac{\text{constant}}{L^{1/d}(n)h_n^{d+1}} = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \text{ almost surely.} \tag{3.22}$$

From (3.21) we find immediately that

$$Q_3 = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\}. \tag{3.23}$$

The main task is to show that

$$Q_2 = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \text{ almost surely.}$$

Write

$$W_n \triangleq s_{n,j}(x) - E\{s_{n,j}(x)\} = \frac{1}{n-d+1}\sum_{i=0}^{n-d} Z_{n,i} \tag{3.24}$$

where

$$Z_{n,i} = \frac{1}{h_n^d}\left[H_j\left(\frac{X_i - x}{h_n}\right) - E\left\{H_j\left(\frac{X_i - x}{h_n}\right)\right\}\right]. \tag{3.25}$$

Partition the set $\{0, 1, \ldots, n-d\}$ into $2q = 2q(n)$ consecutive blocks of size $r(n)$; $n - d + 1 = 2q(n)r(n) + v(n)$ with $0 \leqslant v(n) < r(n)$. Henceforth we replace the sample size $n - d$ by n; asymptotically this has no effect on the result. Write

$$V_n(j) = \frac{1}{n}\sum_{i=(j-1)r+1}^{jr} Z_{n,i} \qquad j = 1, \ldots, 2q \tag{3.26}$$

and

$$W_n' = \sum_{j=1}^{q} V_n(2j-1) \qquad W_n'' = \sum_{j=1}^{q} V_n(2j) \qquad W_n''' = \sum_{i=2qr+1}^{n} Z_{n,i} \tag{3.27}$$

$$W_n = W_n' + W_n'' + W_n''' \tag{3.28}$$

so that W_n' and W_n'' are the sums of the odd-numbered and even-numbered blocks, respectively. The contribution of the remainder term W_n''' is negligible (and is subsequently ignored) since it consists of at most $r(n)$ terms whereas W_n' and W_n'' each consist of $q(n)r(n)$ terms and $q(n) \to \infty$ as $n \to \infty$. Then, for each $\eta > 0$,

$$\begin{aligned} P[Q_2 > \eta] &\leqslant P\left[\max_{1\leqslant k\leqslant L(n)} |W_n'(x_k)| > \frac{\eta}{2}\right] + P\left[\max_{1\leqslant k\leqslant L(n)} |W_n''(x_k)| > \frac{\eta}{2}\right] \\ &\leqslant 2L(n) \sup_{x\in R^d} P\left[|W_n'(x)| > \frac{\eta}{2}\right]. \end{aligned} \tag{3.29}$$

We proceed to bound the expression $P[|W_n'(x)| > \eta/2]$. We use the strong approximation theorem of Bradley (1983) to approximate the random variables $V_n(1), V_n(3), \ldots, V_n(2q-1)$ by independent random variables. By enlarging the probability space if necessary, introduce a sequence $(U_1, U_2, \ldots)$ of independent uniform $[0, 1]$ random variables that is independent of $\{V_n(2j-1)\}_{j=1}^{q}$. Define $V_n^*(0) = 0$, $V_n^*(1) = V_n(1)$. Then for each $j > 2$, there exists a random variable $V_n^*(2j-1)$ which is a measurable function of $V_n(1), V_n(3), \ldots, V_n(2j-1), U_j$ such that $V_n^*(2j-1)$ is independent of $V_n(1), V_n(3), \ldots, V_n(2j-3)$, has the same distribution as $V_n(2j-1)$ and satisfies

$$P[|V_n^*(2j-1) - V_n(2j-1)| > \mu] \leqslant$$
$$18(\|V_n(2j-1)\|_\infty/\mu)^{1/2}(\sup |P[AB] - P[A]P[B]|) \tag{3.30}$$

where the supremum is over all sets A, B with A, B in the σ-algebras of events generated by $\{V_n(1), V_n(3), \ldots, V_n(2j-3)\}$ and $V_n(2j-1)$ respectively. Here μ is any positive number such that $0 < \mu \leqslant \|V_n(2j-1)\|_\infty < \infty$. Now

$$\begin{aligned} P\left[|W_n'(x)| > \frac{\eta}{2}\right] &\leqslant P\left[\left|\sum_{j=1}^{q} V_n^*(2j-1)\right| > \frac{\eta}{4}\right] \\ &\quad + P\left[\left|\sum_{j=1}^{q} V_n(2j-1) - V_n^*(2j-1)\right| > \frac{\eta}{4}\right] \\ &\equiv I_1(x) + I_2(x). \end{aligned} \tag{3.31}$$

We bound I_1 as follows. With $A_1 = \sup_{u\in R^d} H_j(u)$ (cf. Condition 3(a)) we have by (3.25)

$$|Z_{n,i}| \leqslant \frac{2A_1}{h_n^d} \qquad i = 1, \ldots, n. \tag{3.32}$$

Moreover by (3.26)

$$|V_n(j)| \leqslant \frac{2A_1 r(n)}{nh_n^d}. \tag{3.33}$$

Define

$$\lambda_n = \frac{1}{4A_1}(nh_n^d \ln n)^{1/2}. \tag{3.34}$$

Then

$$\lambda_n |V_n(j)| \leqslant \tfrac{1}{2} \qquad j = 1, \ldots, 2q. \tag{3.35}$$

Because $\exp(x) \leqslant 1 + x + x^2$ for $|x| \leqslant 1/2$ and $V_n^*(2j-1)$ has the same distribution as $V_n(2j-1)$, it follows by (3.35) that $\lambda_n|V_n^*(2j-1)| \leqslant 1/2$ so that $\exp\{\pm\lambda_n V_n^*(2j-1)\} \leqslant 1 \pm \lambda_n V_n^*(2j-1) + \lambda_n^2\{V_n^*(2j-1)\}^2$. Hence

$$E[\exp\{\pm\lambda_n V_n^*(2j-1\}] \leqslant 1 + \lambda_n^2 E\{V_n^*(2j-1)\}^2 \leqslant \exp[\lambda_n^2 E\{V_n^*(2j-1)\}^2]. \tag{3.36}$$

Now by (3.31), the Markov inequality and the independence of the $\{V_n^*(2j-1)\}_{j=1}^q$,

$$\begin{aligned} I_1 &\leqslant \frac{E[\exp\{\lambda_n \sum_{j=1}^q V_n^*(2j-1)\}] + E[\exp\{-\lambda_n \sum_{j=1}^q V_n^*(2j-1)\}]}{\exp(\lambda_n \eta/4)} \\ &\leqslant 2\exp\left(-\lambda_n \frac{\eta}{4}\right)(\exp[\lambda_n^2 \textstyle\sum_{j=1}^q E\{V_n^*(2j-1)\}^2]) \end{aligned} \tag{3.37}$$

by (3.36). We obtain an upper bound on $E\{V_n^*(2j-1)\}^2$:

$$\begin{aligned} \sum_{j=1}^q E\{V_n^*(2j-1)\}^2 &= \sum_{j=1}^q E\{V_n(2j-1)\}^2 \\ &= \frac{1}{n^2}\sum_{j=1}^q E\left\{\sum_{i=(2j-1)r+1}^{2rj} Z_{n,i}\right\}^2 \\ &\leqslant \frac{1}{n^2}\left\{\sum_{i=1}^n \operatorname{var}(Z_{n,i}) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i\neq j}}^n |\operatorname{cov}(Z_{n,i}, Z_{n,j})|\right\} \\ &\leqslant \frac{A_2}{nh_n^d}\{1 + o(1)\} \qquad A_2 \triangleq C_2\gamma_{2j} \end{aligned} \tag{3.38}$$

by Theorem 1. Thus by (3.37) and (3.38),

$$\sup_{x \in R^d} I_1(x) \leqslant 2\exp\left(-\lambda_n \frac{\eta}{4} + \frac{\lambda_n^2 A_2}{nh_n^d}\right). \tag{3.39}$$

We now bound the term I_2 in the right-hand side of (3.31).

$$I_2 \leqslant \sum_{j=1}^{q} P\left[|V_n(2j-1) - V_n^*(2j-1)| > \frac{\eta}{4q}\right]. \tag{3.40}$$

We make use of (3.30).

(i) If $\|V_n(2j-1)\|_\infty > \eta/(4q)$, then by (3.30)

$$I_2 \leqslant 18q\left\{\|V_n(2j-1)\|_\infty \Big/ \frac{\eta}{4q}\right\}^{1/2} (\sup |P[AB] - P[A]P[B]|) \tag{3.41}$$

where $A \in \sigma\{V_n(1), V_n(3), \ldots, V_n(2j-3)\}$, $B \in \sigma\{V_n(2j-1)\}$. By (1.2), (3.25) and (3.26) we have

$$|P[AB] - P[A]P[B]| \leqslant 4\alpha\{r(n) - d + 2\}.$$

By (3.33) we then have

$$\sup_{x \in R^d} I_2(x) \leqslant \text{constant} \times q(n)\left\{\frac{q(n)}{\eta}\right\}^{1/2}\left\{\frac{r(n)}{nh_n^d}\right\}^{1/2} \alpha\{r(n) - d + 2\}. \tag{3.42i}$$

(ii) If $\|V_n(2j-1)\|_\infty < \eta/(4q)$ then by (3.40) and (3.30) with $\mu \equiv \|V_n(2j-1)\|_\infty$ we have

$$\sup_{x \in R^d} I_2(x) \leqslant \text{constant} \times q(n)[\alpha\{r(n) - d + 2\}] \tag{3.42ii}$$

which is of smaller order than (3.42i).

Thus by (3.31), (3.29) and (3.42),

$$\sup_{x \in R^d} P\left[|W_n'(x)| > \frac{\eta}{2}\right] \leqslant 2\exp\left(-\frac{\lambda_n \eta}{4} + \frac{\lambda_n^2 A_2}{nh_n^d}\right)$$

$$+ \text{constant} \times q(n)\frac{1}{(\eta h_n^d)^{1/2}}[\alpha\{r(n) - d + 2\}]$$

and similarly for $P[W_n''| > \eta/2]$. Then by (3.29) we have with $\eta = \eta_n = A_3\{\ln n/(nh_n^d)\}^{1/2}$ that

$$P[\max_{1 \leqslant k \leqslant L(n)} |s_{n,j}(x_k) - E\{s_{n,j}(x_k)\}| > \eta_n] \leqslant 4L(n)\exp\left(-\frac{\lambda_n \eta_n}{4} + \frac{A_2 \lambda_n^2}{nh_n^d}\right)$$

$$+ \text{constant} \times \psi(n). \tag{3.43}$$

Now by (3.34)

$$\frac{\lambda_n \eta_n}{4} = \frac{A_3}{16A_1}\ln n \qquad \frac{\lambda_n^2 A_2}{nh_n^d} = \frac{A_2}{16A_1^2}\ln n.$$

The first term on the right-hand side of (3.43) is therefore of the form

$$\frac{4L(n)}{n^a}; \qquad a = \frac{1}{16A_1}\left(A_3 - \frac{A_2}{A_1}\right)$$

and by selecting A_3 large enough we can ensure that $L(n)/n^a$ is summable. Because $\{\psi(n)\}$ is summable by assumption, it follows by (3.43) and the Borel–Cantelli lemma that

$$Q_2 = \mathrm{O}(\eta_n) = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \text{ almost surely} \tag{3.44}$$

and Theorem 2 now follows from (3.20), (3.22), (3.23) and (3.44). □

Proof of Proposition 2. We have

$$E\{G_{n,j+k}(x)\} = \int_{R^d} u^{j+k}\rho(h_n u; x)K(u)f(x+h_n u)\,du. \tag{3.45}$$

By (3.12) and Condition 4(b),

$$|\rho(h_n u; x)| \leqslant \int_0^1 \omega(wh_n\|u\|)(1-w)^p\,dw$$

where $\omega(\eta)$ is the modulus of continuity of $D^k m$. Hence by the boundedness of f

$$|E\{G_{n,j+k}(x)\}| \leqslant C_1 \int_{R^d}\int_0^1 |u^{j+k}|\omega(wh_n\|u\|)K(u)\,dw\,du$$

and the result follows by dominated convergence. □

Proof of Theorem 3. The proof is carried out in the manner of the proof of Theorem 1. Here put

$$U_{i,j} = \left(\frac{X_i - x}{h}\right)^j \rho(X_i - x; x)K_h(X_i - x)$$

(compare with (3.3)). Then, by stationarity,

$$\begin{aligned}\operatorname{var}\{G_{n,j+k}(x)\} &= \frac{1}{n-d+1}\left\{\operatorname{var}(U_{0,j}) + 2\sum_{l=1}^{n-d}\left(1 - \frac{l}{n-d+1}\right)\operatorname{cov}(U_{0,j}, U_{l,j})\right\}\\ &\equiv J_1(x) + J_2(x).\end{aligned}$$

The main difference here (compared with the proof of Theorem 1) is the presence of the factor $\rho(X_i - x; x)$ in $U_{i,j}$. However, by (3.12) and Condition 5(d) we have

$$|\rho(u; x)| \leqslant \frac{2C_5}{p+1}$$

and using this bound the proof follows in the manner of the proof of Theorem 1. □

PROOF OF THEOREM 4. The proof is carried out in the manner of the proof of Theorem 2. There are some differences which are discussed below. We decompose the error as in (3.20):

$$\sup_{x \in D} |G_{n,j+k}(x) - E\{G_{n,j+k}(x)\}| \leqslant Q_1 + Q_2 + Q_3. \tag{3.46}$$

Here we have

$$G_{n,j+k}(x) = \frac{1}{(n-d+1)h_n^d} \sum_{i=0}^{n-d} H_{j+k}\left(\frac{X_i - x}{h_n}\right) \rho(X_i - x; x) \tag{3.47}$$

and there is the additional factor ρ. In bounding Q_1 in the proof of Theorem 2, we used the Lipschitz condition on H_j (cf. (3.21)). Here we note that

$$H(u)\rho(u; x) - H(v)\rho(v; x) = \{H(u) - H(v)\}\rho(u; x) + H(u)\{\rho(v; x) - \rho(u; x)\}.$$

Now by (3.12) and Condition 5(d)

$$\sup_{u \in R^d} |\rho(u; x)| \leqslant \frac{2C_5}{p+1} \tag{3.48}$$

and by Condition 6

$$\begin{aligned} &|\rho(X_i - x; x) - \rho(X_i - x_k; x_k)| \\ &= \int_0^1 ([(D^k m)\{x + w(X_i - x)\} - (D^k m)\{x_k + w(X_i - x_k)\}] \\ &\quad + \{(D^k m)(x_k) - (D^k m)(x)\})(1 - w)^p \, dw \\ &\leqslant \left(C_6 \int_0^1 (2 - w)(1 - w)^p \, dw\right) \|x - x_k\|. \end{aligned}$$

It then follows with $A_1 = \sup_{u \in R^d} |H_{j+k}(u)|$ that

$$\begin{aligned} &|G_{n,j+k}(x) - G_{n,j+k}(x_k)| \\ &\leqslant \frac{2C_4 C_5/(p+1)}{h_n^{d+1}} \|x - x_k\| + \frac{C_6 A_1 \int_0^1 (2 - w)(1 - w)^p \, dw}{h_n^d} \|x - x_k\| \\ &\leqslant \frac{\text{constant}}{h_n^{d+1}} \|x - x_k\| \{1 + \mathrm{O}(h_n)\}. \end{aligned} \tag{3.49}$$

We now obtain from (3.49) that

$$Q_1 = \mathrm{O}\left\{\left(\frac{\ln n}{n h_n^d}\right)^{1/2}\right\} \text{ almost surely.} \tag{3.50}$$

From (3.49) we immediately have

$$Q_3 = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\}. \tag{3.51}$$

The bounding of Q_2 proceeds along the same lines as in the proof of Theorem 2. The only point to be made is that $Z_{n,i}$ of (3.25) is now replaced by

$$Z_{n,i} = H_{j+k}\left(\frac{X_i - x}{h_n}\right)\rho(X_i - x; x) - E\left\{H_{j+k}\left(\frac{X_i - x}{h_n}\right)\rho(X_i - x; x)\right\} \tag{3.52}$$

and (3.48) ensures the boundedness of $\rho(X_i - x; x)$ so that $Z_{n,i}$ above satisfies an inequality similar to (3.32) (with a different constant): $|Z_{n,i}| \leqslant \{4A_1C_5/(p+1)\}/h_n^d$. □

4. UNIFORM STRONG CONSISTENCY OF THE REGRESSION FUNCTION $m(x)$ AND ITS DERIVATIVES—RATES OF CONVERGENCE

4.1. *Main results*

We first establish uniform rates of almost sure convergence for the centered $t_{n,j}^*(x)$ of (2.7). Since the function ψ is not necessarily bounded, we use a well-known truncation argument (e.g. Masry and Tjostheim, 1995). Let $\{T_n\}$ be a sequence of positive numbers specified in (4.4) below. Define the truncated regression function $m^{\mathrm{T}}(x)$ by

$$m^{\mathrm{T}}(x) = E[\psi(Y_d)I\{|\psi(Y_d)| \leqslant T_n\}|X_0 = x] \tag{4.1}$$

where I is the indicator function, and define the truncated variable $t_{n,j}^{*\mathrm{T}}(x)$ by

$$t_{n,j}^{*\mathrm{T}}(x) = \frac{1}{n-d+1}\sum_{i=0}^{n-d}[\psi(Y_{d+i})I\{|\psi(Y_{d+i})| \leqslant T_n\} - m^{\mathrm{T}}(X_i)]$$
$$\times\left(\frac{X_i - x}{h}\right)^j K_h(X_i - x). \tag{4.2}$$

Note that $t_{n,j}^{*\mathrm{T}}(x)$ has zero mean. Write

$$t_{n,j}^*(x) = \{t_{n,j}^*(x) - t_{n,j}^{*\mathrm{T}}(x)\} + t_{n,j}^{*\mathrm{T}}(x). \tag{4.3}$$

The main task is to establish uniform rates of almost sure convergence for $t_{n,j}^{*\mathrm{T}}(x)$ as the first term on the right-hand side of (4.3) is eventually zero for large n. We make the following assumption.

CONDITION 7.

(a) $E|\psi(Y_1)|^\sigma < \infty$ for some $\sigma > 2$.

(b) The conditional density $f_{X_0|Y_d}(x|y)$ of X_0 given Y_d exists and is bounded, $f_{X_0|Y_d}(x|y) \leqslant C_7 < \infty$.

(c) The conditional density $f_{(X_0, X_l)|(Y_d, Y_{d+l})}$ of (X_0, X_l) given (Y_d, Y_{d+l}) exists and is bounded,

$$f_{(X_0, X_l)|(Y_d, Y_{d+l})}\{(u, v)|(y_1, y_2)\} \leqslant C_8 < \infty$$

for all $l \geqslant 1$.

(d) The processes $\{Y_i, X_i\}_{i=-\infty}^{\infty}$ are strongly mixing with the mixing coefficient $\alpha(k)$ satisfying

$$\sum_{j=1}^{\infty} j^a \{\alpha(j)\}^{1-2/\nu} < \infty$$

for some $\nu > 2$ and $a > 1 - 2/\nu$.

Define the truncation sequence $\{T_n\}$ by

$$T_n = \{n \ln n (\ln \ln n)^{1+\delta}\}^{1/\sigma} \text{ for some } 0 < \delta < 1. \tag{4.4}$$

LEMMA 1. *Under Condition 7(a) we have*

$$t^*_{n,j}(x) - t^{*\mathrm{T}}_{n,j}(x) = 0 \textit{ almost surely for large } n.$$

Next we obtain a uniform bound on the variance of $t^{*\mathrm{T}}_{n,j}(x)$.

LEMMA 2. *Let D be any compact subset of R^d. Let Condition 3(a) and Condition 7 (with $\nu \leqslant \sigma$) hold. Assume the bandwidth $h_n \to 0$ such that $nh_n^d \to \infty$ as $n \to \infty$. Then for all j with $0 \leqslant |j| \leqslant p$ we have*

$$nh_n^d \sup_{x \in D} \operatorname{var}(t^{*\mathrm{T}}_{n,j}) \leqslant A_4\{1 + o(1)\}$$

where the constant A_4 is specified in (4.16) (later).

Next we obtain a uniform rate of almost sure convergence for $t^{*\mathrm{T}}_{n,j}(x)$.

LEMMA 3. *Let D be any compact subset of R^d. Let Condition 3 and Condition 7 (with $\nu \leqslant \sigma$) hold. Assume the bandwidth $h_n \to 0$ slowly enough such that*

$$\phi_n \triangleq \frac{n^{1-2/\sigma} h_n^d}{(\ln n)\{(\ln n)(\ln \ln n)^{1+\delta}\}^{2/\sigma}} \to \infty \textit{ as } n \to \infty. \tag{4.5}$$

Define

$$r_1(n) \triangleq \phi_n^{1/2} = \frac{(nh_n^d/\ln n)^{1/2}}{T_n} \qquad L_1(n) \triangleq \left(\frac{nT_n^2}{h_n^{d+2} \ln n}\right)^{d/2} \tag{4.6}$$

and

$$\psi_1(n) \triangleq \frac{nL_1(n)}{r_1(n)}\left(\frac{nT_n^2}{h_n^d \ln n}\right)^{1/4} \alpha\{r_1(n)\}. \tag{4.7a}$$

If the strong mixing coefficient $\alpha(k)$ *satisfies*

$$\sum_{n=1}^{\infty} \psi_1(n) < \infty \tag{4.7b}$$

then for each $\boldsymbol{j}$ *with* $0 \leqslant |\boldsymbol{j}| \leqslant p$ *we have*

$$\sup_{x \in D} |t_{n,j}^{*\mathrm{T}}(\boldsymbol{x})| = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \textit{ almost surely.}$$

Combining Lemmas 1 and 3 we obtain the following.

THEOREM 5. *Let D be any compact subset of* R^d. *Let Condition 3 and Condition 7 (with* $v \leqslant \sigma$*) hold. Assume the bandwidth* $h_n \to 0$ *slowly enough such that (4.5) is satisfied. Assume that the strong mixing coefficient* $\alpha(k)$ *satisfies the summability condition (4.7). Then, for each* $\boldsymbol{j}$ *with* $0 \leqslant |\boldsymbol{j}| \leqslant p$ *we have*

$$\sup_{x \in D} |t_{n,j}^{*}(\boldsymbol{x})| = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \textit{ almost surely}$$

and thus

$$\sup_{x \in D} |\boldsymbol{\tau}_n^{*}(\boldsymbol{x})| = \mathrm{O}\left\{\left(\frac{\ln n}{nh_n^d}\right)^{1/2}\right\} \textit{ almost surely.}$$

We can now state the principal result regarding the uniform strong convergence and rates for the regression function $m(\boldsymbol{x})$ and its partial derivatives. We make the following additional assumption on the probability density f of $\boldsymbol{X}_0$.

CONDITION 8. Let D be a compact subset of R^d. Assume that the probability density f of $\boldsymbol{X}_0$ satisfies

$$\inf_{x \in D} f(\boldsymbol{x}) = C_9 > 0. \tag{4.8}$$

Using the fundamental relationship (2.13), Corollaries 2 and 3 and Theorem 5 we obtain the following.

THEOREM 6. *Let D be a compact subset of* R^d. *Let Conditions 1(a), 2(b), 3, 4(b), 5(d), 6, Condition 7 (with* $v \leqslant \sigma$*) and Condition 8 hold. Assume the bandwidth* $h_n \to 0$ *slowly enough such that (4.5) is satisfied. Assume that the strong mixing coefficient* $\alpha(k)$ *satisfies the summability condition (4.7). Then*

$$\sup_{x \in D} |\hat{\boldsymbol{\beta}}_n(x) - \boldsymbol{\beta}(x)| = \mathrm{O}\left\{ \left(\frac{\ln n}{nh_n^d} \right)^{1/2} \right\} + \mathrm{O}(h_n^{p+1}) \textit{ almost surely} \tag{4.9}$$

and for each j *with* $0 \leqslant |j| \leqslant p$ *we have*

$$\sup_{x \in D} |(D^j m)\hat{}(x) - (D^j m)(x)| = \mathrm{O}\left\{ \left(\frac{\ln n}{nh_n^{d+2|j|}} \right)^{1/2} \right\} + \mathrm{O}(h_n^{p-|j|+1}) \textit{ almost surely.} \tag{4.10}$$

Note that we have pruned the conditions in Theorem 6 because of redundancy. Since the kernel K is bounded with compact support (Condition 3), all moments conditions on K are automatically satisfied. Also, condition (4.7) on the strong mixing coefficient is the most stringent so we removed the other assumptions on $\alpha(k)$. The rates of convergence given in Theorem 6 for local polynomial regression estimates coincide with the optimal rates established in Stone (1982) for regression.

REMARK 1. It is of practical interest to provide an explicit rate of decay for the strong mixing coefficient $\alpha(k)$ of the form $\alpha(k) = \mathrm{O}(1/k^c)$ for some $c > 0$ (to be determined) under which Theorem 6 holds. It is easily seen that, among all the conditions imposed on $\alpha(k)$, the summability condition (4.7) is the most restrictive. We assume that

$$h_n \sim \left(\frac{\ln n}{n} \right)^{\bar{a}} \text{for some } 0 < \bar{a} < \frac{1}{d}\left(1 - \frac{2}{\sigma}\right)$$

so that the condition (4.5) and $[(\ln n)/(nh_n^d)] \to 0$ on h_n are satisfied. Algebraic calculations show that the summability condition (4.7) is satisfied provided

$$\alpha(k) = \mathrm{O}\left(\frac{1}{k^c} \right) \text{ with } c > \frac{(3 + 2d) + \sigma(7/2 + d)(1 + \bar{a}d)}{\sigma(1 - \bar{a}d) - 2} \equiv C(\sigma, d, \bar{a}). \tag{4.11}$$

It can be seen that $C(\sigma, d, \bar{a})$ is monotonically decreasing in σ and therefore there is a tradeoff between the order σ of the moment $E|\psi(Y_1)|^\sigma < \infty$ in Condition 7(a) and the decay rate of the strong mixing coefficient $\alpha(k)$; the existence of high order moments allows for weaker conditions on $\alpha(k)$. In particular, when $E|\psi(Y_1)|^\sigma < \infty$ for all $\sigma \geqslant 1$ (e.g. when ψ is bounded as when $\psi(Y_1) = I(Y_1 \leqslant y)$ corresponding to estimating conditional distributions) it suffices that $\alpha(k) = \mathrm{O}(1/k^c)$ with $c > (7 + 2d)/(1 - \bar{a}d)$ with $0 < \bar{a} < 1/d$.

REMARK 2. Note that the right-hand side of (4.10) is the sum of two terms corresponding to the 'variance' and 'bias' terms. The optimal choice of the bandwidth parameter h_n which minimizes the right-hand side of (4.10) is *uniform* for all derivatives of total order up to p. We find

$$h_n \sim \left(\frac{\ln n}{n}\right)^{1/\{d+2(p+1)\}}$$

in which case we have for each derivative of order j with $0 \leqslant |j| \leqslant p$ that

$$\sup_{x \in D} |(D^j m)\hat{}(x) - (D^j m)(x)| = \mathrm{O}\left\{\left(\frac{\ln n}{n}\right)^{(p-|j|+1)/\{d+2(p+1)\}}\right\}.$$

4.2. *Proofs*

PROOF OF LEMMA 1. We have

$$t^*_{n,j} - t^{*\mathrm{T}}_{n,j} = \frac{1}{n-d+1}\sum_{i=0}^{n-d}(\psi(Y_{d+i})I\{|\psi(Y_{d+i})| > T_n\}$$
$$- E[\psi(Y_d)I\{|\psi(Y_d)| > T_n\}|X_i])\left(\frac{X_i - x}{h_n}\right)^j K_h(X_i - x). \tag{4.12}$$

But $P[|\psi(Y_{d+n})| > T_n] \leqslant T_n^{-\sigma} E|\psi(Y_{d+n})|^\sigma$ and, by the summability of the right-hand side and the Borel–Cantelli lemma, $|\psi(Y_{d+n})| \leqslant T_n$ almost surely for all sufficiently large n. Since T_n is increasing we have $|\psi(Y_{d+i})| \leqslant T_n$ almost surely for all $i \leqslant n$. This implies that $t^*_{n,j} - t^{*\mathrm{T}}_{n,j}$ is eventually zero, almost surely. □

PROOF OF LEMMA 2. The proof is carried out in the manner of the proof of Theorem 1 with the following differences. Here we define

$$U_{i,j} = [\psi(Y_{d+i})I\{|\psi(Y_{d+i})| \leqslant T_n\} - m^{\mathrm{T}}(X_i)]\left(\frac{X_i - x}{h}\right)^j K_h(X_i - x) \tag{4.13}$$

which has zero mean, and we write, by stationarity,

$$(n-d+1)\operatorname{var}\{t^{*\mathrm{T}}_{n,j}(x)\} = \operatorname{var}(U_{0,j}) + 2\sum_{l=1}^{n-d}\left(1 - \frac{l}{n-d+1}\right)\operatorname{cov}(U_{0,j}, U_{l,j})$$
$$= J_1(x) + J_2(x). \tag{4.14}$$

Since K has compact support and the set D is compact, there is a constant A_5 such that

$$\sup_{x \in D} \sup_{\{u:\|u-x\| \leqslant A_0 h\}} |m^{\mathrm{T}}(u)| = A_5 < \infty. \tag{4.15}$$

Then, conditioning on Y_d and using Condition 7(b),

$$\sup_{x \in D} J_1(x) \leqslant \frac{C_7}{h_n^d} E[\{|\psi(Y_1)| + A_5\}^2]\int_{R^d} u^{2j} K^2(u)\, du \equiv A_4/h_n^d. \tag{4.16}$$

Now decompose $J_2(x)$ as in (3.18). For J_{21} conditioning on (Y_d, Y_{d+l}) and using

Condition 7(c) and (4.15) yields

$$h^{d-l}|\operatorname{cov}(U_{0,j}, U_{l,j})| \leqslant C_8 E[\{|\psi(Y_d)| + A_5\}\{|\psi(Y_{d+l})| + A_5\}] \int_{R^{d+l}} \prod_{i=1}^{l} |u_i|^{j_i} \times \prod_{i=l+1}^{d} |u_i|^{2j_i} \prod_{i=d+1}^{d+l} |u_i|^{j_i} K(\boldsymbol{u}', \boldsymbol{u}'')K(\boldsymbol{u}'', \boldsymbol{u}''')\, d\boldsymbol{u}'\, d\boldsymbol{u}''\, d\boldsymbol{u}'''.$$

Hence

$$h_n^d \sup_{\boldsymbol{x}\in D} |J_{21}(\boldsymbol{x})| \leqslant \text{constant} \times \sum_{l=1}^{d-1} h_n^l = \mathrm{O}(h_n) \to 0 \text{ as } n \to \infty. \tag{4.17}$$

For J_{22} one finds similarly

$$|\operatorname{cov}(U_{0,j}, U_{l,j})| \leqslant C_8 E[\{|\psi(Y_d)| + A_5\}\{|\psi(Y_{d+l})| + A_5\}] \left\{ \int_{R^d} \prod_{i=1}^{d} |u_i|^{j_i} K(\boldsymbol{u})\, d\boldsymbol{u} \right\}^2 < \infty.$$

Hence

$$nh_n^d \sup_{\boldsymbol{x}\in D} |J_{22}(\boldsymbol{x})| \leqslant \text{constant} \times \sum_{l=d}^{\pi_n} h_n^d = \mathrm{O}(\pi_n h_n^d) \to 0 \text{ as } n \to \infty. \tag{4.18}$$

For J_{23} we use Davydov's lemma as in the proof of Theorem 1 but employ conditioning on Y_d and (4.15) to bound the moment

$$h_n^{(\nu-1)d} E|U_{0,j}|^\nu \leqslant C_7 E|\psi(Y_d) + A_5|^\nu \int_{R^d} \|\boldsymbol{u}\|^{\nu|j|} K^\nu(\boldsymbol{u})\, d\boldsymbol{u}$$

which, as in the proof of Theorem 1, leads to

$$nh_n^d \sup_{\boldsymbol{x}\in D} |J_{23}(\boldsymbol{x})| \leqslant \text{constant} \times \sum_{l=\pi_n}^{\infty} l^a \{\alpha(l)\}^{1-2/\nu} \to 0 \text{ as } n \to \infty. \tag{4.19}$$

□

PROOF OF LEMMA 3. The proof is carried out in the manner of the proof of Theorem 2 with $r_1(n)$ and $L_1(n)$ replacing $r(n)$ and $L(n)$, respectively, and the following differences. Here we have

$$\begin{aligned} \sup_{\boldsymbol{x}\in D} |t_{n,j}^{*\mathrm{T}}(\boldsymbol{x})| &= \max_{1\leqslant k\leqslant L_1(n)} \sup_{\boldsymbol{x}\in D\cap I_k} |t_{n,j}^{*\mathrm{T}}(\boldsymbol{x})| \\ &\leqslant \max_{1\leqslant k\leqslant L_1(n)} \sup_{\boldsymbol{x}\in D\cap I_k} |t_{n,j}^{*\mathrm{T}}(\boldsymbol{x}) - t_{n,j}^{*\mathrm{T}}(\boldsymbol{x}_k)| + \max_{1\leqslant k\leqslant L_1(n)} |t_{n,j}^{*\mathrm{T}}(\boldsymbol{x}_k)| \\ &\equiv Q_1 + Q_2. \end{aligned} \tag{4.20}$$

Now by (4.12) and Condition 3,

$$|t^{*\mathrm{T}}_{n,j}(x) - t^{*\mathrm{T}}_{n,j}(x_k)| \leqslant \frac{2T_n}{h_n^{d+1}} C_4 \|x - x_k\|$$

and thus

$$Q_1 \leqslant 2T_n \frac{C_4 l_n}{h_n^{d+1}} = \mathrm{O}\left\{ \left(\frac{\ln n}{n h_n^d} \right)^{1/2} \right\} \text{ almost surely.} \tag{4.21}$$

For Q_2 define

$$W_n(x) = t^{*\mathrm{T}}_{n,j}(x) = \frac{1}{n-d+1} \sum_{i=0}^{n-d} Z_{n,i}$$

with

$$Z_{n,i} = \frac{1}{h_n^d} [\psi(Y_{d+i}) I\{|\psi(Y_{d+i})| \leqslant T_n\} - m^{\mathrm{T}}(X_i)] H_j\left(\frac{X_i - x}{h_n} \right) \tag{4.22}$$

(compare with (3.25)). Define $V_n(j)$, W'_n, W''_n, W'''_n as in (3.26) and (3.27). Then, for any $\eta > 0$,

$$P[Q_2 > \eta] \leqslant 2L_1(n) \max_{1 \leqslant k \leqslant L_1(n)} P\left[|W'_n(x_k)| > \frac{\eta}{2} \right].$$

Note that since the number of D-covering cubes $L_1(n) \to \infty$, the size of each cube tends to zero so that the centers $\{x_k\}$ of the cubes fall in a compact subset $D_1 \supset D$. Thus

$$P[Q_2 > \eta] \leqslant 2L_1(n) \sup_{x \in D_1} P\left[|W'_n(x)| > \frac{\eta}{2} \right]. \tag{4.23}$$

(In view of Lemma 2, we are unable to take the supremum over $x \in R^d$ as we did in the proof of Theorem 2.) Now proceeding as in the proof of Theorem 2 we note that by (4.22)

$$|Z_{n,i}| \leqslant \frac{2T_n}{h_n^d} A_1 \qquad i = 1, 2, \ldots, n \tag{4.24}$$

(compare with (3.32)). Then $V_n(j)$ satisfies (3.35) and defining λ_n as in (3.34) and as in (3.35)–(3.39) we conclude that

$$\sup_{x \in D_1} I_1(x) \leqslant \exp\left(-\frac{\lambda_n \eta}{4} + \frac{\lambda_n^2 A_4}{n h_n^d} \right) \tag{4.25}$$

where we have used Lemma 2 (instead of Theorem 1), i.e.

$$\sup_{x \in D_1} \operatorname{var}\{t^{*\mathrm{T}}_{n,j}(x)\} \leqslant \frac{A_4}{n h_n^d} \{1 + \mathrm{o}(1)\}$$

since Lemma 2 is valid over any compact set. For I_2 we proceed as in the proof of Theorem 2 noting that here we have, in view of (4.24),

$$|V_n(j)| \leqslant \frac{2A_1 r_1(n)}{nh_n^d} T_n$$

(compare with (3.33)). As a consequence we obtain

$$\sup_{x \in R^d} I_2(x) \leqslant \text{constant} \times q(n) \left\{ \frac{q(n)}{\eta} \right\}^{1/2} \left\{ \frac{r_1(n) T_n}{nh_n^d} \right\}^{1/2} \alpha\{r_1(n) - d + 2\}. \quad (4.26)$$

It follows from (4.23), (4.25) and (4.26) that

$$P[Q_2 > \eta] \leqslant 4L_1(n) \exp\left(-\frac{\lambda_n \eta_n}{4} + \frac{A_4 \lambda_n^2}{nh_n^d} \right) + \text{constant} \times \psi_1(n)$$

and the proof is now completed in the manner of the proof of Theorem 2. □

PROOF OF THEOREM 6. By Corollary 2 and Condition 8 we have, uniformly in $\boldsymbol{x} \in D$,

$$\boldsymbol{S}_n^{-1}(\boldsymbol{x}) \to \frac{\boldsymbol{M}^{-1}}{f(\boldsymbol{x})} \text{ almost surely.}$$

Combining this with Theorem 5 we have

$$\sup_{\boldsymbol{x} \in D} |\boldsymbol{S}_n^{-1}(\boldsymbol{x}) \boldsymbol{\tau}_n^*(\boldsymbol{x})| = \mathrm{O}\left\{ \left(\frac{\ln n}{nh_n^d} \right)^{1/2} \right\} \text{ almost surely.}$$

By Corollary 2 and Condition 8 we also have, using Condition 5(d), that

$$h_n^{p+1} \sup_{\boldsymbol{x} \in D} |\boldsymbol{S}_n^{-1}(\boldsymbol{x}) \boldsymbol{B}_n(\boldsymbol{x}) \boldsymbol{m}_{p+1}(\boldsymbol{x})| = \mathrm{O}(h_n^{p+1}) \text{ almost surely.}$$

By Corollaries 2 and 3 and Condition 8, we have

$$\sup_{\boldsymbol{x} \in D} |\boldsymbol{S}_n^{-1}(\boldsymbol{x}) \boldsymbol{e}_n(\boldsymbol{x})| = \mathrm{o}(h_n^{p+1}) \text{ almost surely.}$$

The result (4.9) now follows from (2.13). □

NOTE

This work was supported by the Office of Naval Research under Grant N00014-90-J-1175.

REFERENCES

BRADLEY, R. C. (1983) Approximation theorems for strongly mixing random variables. *Michigan Math. J.* 30, 69–81.

COLLOMB, G. and HÄRDLE, W. (1986) Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Processes Appl.* 23, 77–89.

FAN, J. (1992) Design-adaptive nonparametric regression. *J. Am. Stat. Assoc.* 87, 998–1004.

— (1993) Local linear regression smoothers and their minimax efficiency. *Ann. Stat.* 21, 196–216.
— and GIJBELS, I. (1992) Variable bandwidth and local linear regression smoothers. *Ann. Stat.* 20, 2008–36.
— and MASRY, E. (1992) Multivariate regresssion estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* 43, 237–71.
HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.
MASRY, E. (1995) Multivariate regression estimation: local polynomial fitting for time series. Submitted for publication.
— and TJOSTHEIM, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Economet. Theory* 11, 258–89.
NADARAYA, E. A. (1964) On estimating regression. *Theory Probab. Appl.* 9, 141–42.
ROBINSON, P. M. (1983) Nonparametric estimators for time series. *J. Time Ser. Anal.* 4, 185–97.
— (1986) On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression, and density estimators. *Ann. Inst. Stat. Math.* 38, 539–49.
ROSENBLATT, M. (1956) A central limit theorem and strong mixing conditions. *Proc. Natl. Acad. Sci.* 4, 43–47.
— (1969) Conditional probability density and regression estimates. In *Multivariate Analysis II* (ed. Krishnaiah, P. R.). New York: Academic Press, pp. 25–31.
ROUSSAS, G. G. (1990) Nonparametric regression estimation under mixing conditions. *Stochastic Processes Appl.* 36, 107–16.
— and TRAN, L. T. (1992) Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Ann. Stat.* 20, 98–120.
RUPPERT, D. and WAND, M. P. (1994) Multivariate weighted least squares regression. *Ann. Stat.* 22, 1346–70.
STONE, C. J. (1977) Consistent nonparametric regression. *Ann. Stat.* 5, 595–645.
— (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* 10, 1040–53.
TJOSTHEIM, D. (1994) Non-linear time series: a selective review. *Scand. J. Stat.* 21, 97–130.
WATSON, G. S. (1964) Smooth regression analysis. *Sankhya Ser. A*, 26, 359–372.