ELSEVIER Stochastic Processes and their Applications 65 (1996) 81–101

stochastic processes and their applications

Multivariate regression estimation Local polynomial fitting for time series[1]

Elias Masry

Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA

Received May 1995; revised June 1996

Abstract

We consider 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 its partial derivatives, for stationary random processes $\{Y_i, X_i\}$ using local higher-order polynomial fitting. Particular cases of Ψ yield estimation of the conditional mean, conditional moments and conditional distributions. Joint asymptotic normality is established for estimates of the regression function and its partial derivatives for strongly mixing and ρ-mixing processes. Expressions for the bias and variance/covariance matrix (of the asymptotically normal distribution) for these estimators are given.

Keywords: Multivariate regression estimation; Local polynomial fitting; Mixing processes; Joint asymptotic normality

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 (1964)–Watson (1964) estimator in connection with regression estimation. See, for example, Mack and Silverman (1982), and Härdle (1990) and the references therein in the i.i.d. case. For dependent data, see Rosenblatt (1969), Robinson (1983, 1986), Collomb and Härdle (1986), Roussas (1990),

E-mail: masry@ece.ucsd.edu.

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

PII S0304-4149(96)00095-6

Truong and Stone (1992) Roussas and Tran (1992), Fan and Masry (1992) among others.

The local polynomial fitting approach was introduced originally by Stone (1977) and studied by Cleveland (1979), Fan (1992, 1993) and many others (see Fan (1993) for additional references). Local polynomial fitting has significant advantages over the Nadaraya–Watson regression estimator: For local linear fitting in the univariate case it has been shown to reduce the bias (see Chu and Marron, 1991, Fan, 1992); it adapts automatically to the boundary of design points (see Fan and Gijbels, 1992; Hastie and Loader, 1993; Ruppert and Wand, 1994); – no boundary modification is required. It is superior to the Nadaraya–Watson estimator in the context of estimating the derivatives of the regression function (see Fan and Gijbels, 1992; and Ruppert and Wand, 1994); in particular, the differentiability of the kernel is not required. All the above-cited works consider i.i.d. setting. In a recent work, Masry and Fan (1996), established the asymptotic normality of the univariate regression function $m(x)$ and its derivatives ($d = 1$) for dependent data using local higher-order polynomial fit.

The purpose of this paper is to formulate the multivariate regression estimation problem in the general setting given in (1.1), in conjunction with local higher-order polynomial fitting, and establish the joint asymptotic normality of $\hat{m}(x_1, \dots, x_d)$ 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 are given. We remark at this point that in the case of i.i.d. setting $\{Y_i, X_i\}$ with $m(\boldsymbol{x}) = E[Y_1 | X_1 = \boldsymbol{x}]$ where X_i is R^d - valued random variable, Ruppert and Wand (1994) consider local *quadratic* fit and provide leading bias and variance terms for $\hat{m}(\boldsymbol{x})$ and its derivatives. Our setting is completely different from Ruppert and Wand (1994) in that

- We consider regression estimation from a vector of *past* data as in (1.1).
- The processes $\{Y_i, X_i\}$ are individually and jointly dependent.
- Higher-order local polynomial fit (of arbitrary order p) is considered.
- Joint asymptotic normality of the estimates of $m(\boldsymbol{x})$ of (1.1) and its partial derivatives is established along with its implications for the bias and variance/covariance matrix of the estimators.

Before we formulate our problem we discuss potential applications. We first remark that the function Ψ in the definition of the regression function $m(\boldsymbol{x})$ in (1.1) is arbitrary. Some special cases of importance in practice are: (a) $\Psi(y) = y$ corresponds to estimating the condition mean of Y_d and its derivatives from a vector of past data $(X_1, \dots, X_d)$. (b) $\Psi(y) = I\{Y \leqslant y\}$ corresponds to estimating the conditional distribution $m(\boldsymbol{x}) = P[Y_d \leqslant y | X_1 = x_1, \dots, X_d = x_d]$ and its derivatives from past data. (c) $\Psi(y) = y^2$ corresponds to estimating the conditional second moment from past data. Prediction problems are also included in our formulation: Put $Y_i = X_{i+r}$ for some $r \geqslant 1$. Then the regression problem (1.1) reduces to estimating $E[\Psi(X_{d+r}) | X_1 = x_1, \dots, X_d = x_d]$. 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 ARCH time series

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

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

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

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

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 and we can thus establish their asymptotic normality.

We formulate our problem as follows. Let

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

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

We assume throughout the paper that derivatives of total order $p+1$ of $m(\boldsymbol{z})$ exist and are continuous at the point $\mathbf{x}$. 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

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

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

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

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

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

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

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 $(D^{\boldsymbol{k}} m)\hat{}(\boldsymbol{x}) = \boldsymbol{k}!\hat{b}_{\boldsymbol{k}}(\boldsymbol{x})$. The minimization of (1.9) leads to the set of equations

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

where

$$t_{n,j} = \frac{1}{n-d+1} \sum_{i=0}^{n-d} \Psi(Y_{d+i}) \left[\frac{X_i - x}{h}\right]^j K_h(X_i - x), \tag{1.11}$$

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

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

The organization of the paper is as follows. Section 2 establishes the quadratic-mean convergence of $s_{n,j}$, the centering of $t_{n,j}$, the bias of the estimates $\hat{b}_j$, and the asymptotic covariance structure of the centered $t_{n,j}$. Section 3 derives the joint asymptotic normality of the centered $t_{n,j}$ and of the regression function's estimate $\hat{m}(x)$ and its partial derivatives and provides expression for their bias and variance/covariance matrix (of the asymptotic distribution). We remark that a companion paper (Masry, 1996) provides uniform rates of almost sure convergence of $\hat{m}(x)$ and its derivatives.

2. Mean-square convergence

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 j with $|j| = i$. 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 $N_{|j|}$ values of $t_{n,j}$ in a column vector $\tau_{n,|j|}$ according to the above order. Then

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

Define

$$\tau_n = \begin{bmatrix} \tau_{n,0} \\ \tau_{n,1} \\ \vdots \\ \tau_{n,p} \end{bmatrix} \tag{2.2}$$

and note that the column vector τ_n is of dimension $N = \sum_{i=0}^{p} N_i \times 1$. Similarly arrange the distinct values of $h^{|k|}\hat{b}_k, 0 \leqslant |k| \leqslant p$, as a column vector of dimension

$N \times 1$ in the form

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

Finally, arrange the possible values of $s_{n,j+k}$ by a matrix $\boldsymbol{S}_{n,|j|,|k|}$ in a lexicographical order with the (ℓ, m) element of $\boldsymbol{S}_{n,|j|,|k|}$ given by

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

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 $\boldsymbol{\tau}_n = \boldsymbol{S}_n \hat{\boldsymbol{\beta}}_n$. Because of the functional form of $s_{n,j+k}$ we have

$$\sum_{0 \leqslant |j| \leqslant p} \sum_{0 \leqslant |k| \leqslant p} c_j c_k s_{n,j+k}$$

$$= \frac{1}{n-d+1} \sum_{i=0}^{n-d} \left[\sum_{0 \leqslant |j| \leqslant p} c_j (\boldsymbol{X}_i - \boldsymbol{x})/h)^j \right]^2 K_h(\boldsymbol{X}_i - \boldsymbol{x}) \geqslant 0.$$

It follows that the $N \times N$ matrix whose components are $\{s_{n,j,k}: 0 \leqslant |\boldsymbol{j}| \leqslant p, 0 \leqslant |\boldsymbol{k}| \leqslant p\}$ is positive-semidefinite. It would be positive-definite if there are sufficient number of data points $\{\boldsymbol{X}_i: (\boldsymbol{X}_i - \boldsymbol{x})/h \in \text{support of } K\}$. In any case, we assume henceforth that the matrix $\boldsymbol{S}_n$ is positive-definite and we write $\hat{\boldsymbol{\beta}}_n = \boldsymbol{S}_n^{-1} \boldsymbol{\tau}_n$, as the solution of the set of equations (1.10).

We now introduce the mixing coefficients. Let $\mathscr{F}_a^b$ be the σ-algebra of events generated by the random variables $\{Y_j, X_j, a \leqslant j \leqslant b\}$ and L_2 $(\mathscr{F}_a^b)$ denote the collection of all second-order random variables which are $\mathscr{F}_a^b$-measurable. The stationary processes $\{Y_j, X_j\}$ are called strongly mixing (Rosenblatt, 1956) if

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

and are called ρ-mixing (Kolmogorov and Rozanov, (1960) if

$$\sup_{\substack{U \in L_2(\mathscr{F}_{-\infty}^0) \\ V \in L_2(\mathscr{F}_k^\infty)}} \frac{|\operatorname{cov}\{U, V\}|}{\operatorname{var}^{1/2}[U] \operatorname{var}^{1/2}[V]} = \rho(k) \to 0 \quad \text{as } k \to \infty.$$

$\alpha(k)$ is the strong mixing coefficient and $\rho(k)$ is the maximal correlation coefficient. It is well known that $\alpha(k) \leqslant \frac{1}{4}\rho(k)$.

2.1. Main results

We make the following assumptions on the kernel function K and on the random processes $\{Y_i, X_i\}$.

Condition 1. (a) The kernel $K \in L_1$ is bounded and $\|\boldsymbol{u}\|^{4p} K(\boldsymbol{u}) \in L_1$ and $\|\boldsymbol{u}\|^{4p+d} K(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\| \to \infty$. (b) $|f_{x_0,x_\ell}(\mathbf{u},\boldsymbol{v};\ell) - f_{x_0}(\boldsymbol{u})f_{x_\ell}(\boldsymbol{v})| \leqslant A_1 < \infty$ for all $\ell \geqslant 1$ where $f(\boldsymbol{u})$ and $f(\boldsymbol{u},\boldsymbol{v};\ell)$ denote, respectively, the probability density of $\boldsymbol{X}_0$ and $(\boldsymbol{X}_0,\boldsymbol{X}_\ell)$.

(c) Either the processes $\{Y_i, X_i\}$ are ρ-mixing with $\sum_{j=1}^{\infty}\rho(j) < \infty$; or are strongly mixing with $\sum_{j=1}^{\infty} j^a[\alpha(j)]^{1-2/v} < \infty$ for some $v > 2$ and $a > 1 - 2/v$. In the latter case we assume further that $\|\boldsymbol{u}\|^{2vp+d} K(\boldsymbol{u}) \to 0$ as $\|\boldsymbol{u}\| \to \infty$.

Remark 1. Note that for $1 \leqslant \ell \leqslant d-1$ the components of $\boldsymbol{X}_0$ and $\boldsymbol{X}_\ell$ overlap. The joint density f_{X_0,X_ℓ} in Condition 1(b) is then the density $(X_1, \ldots, X_{d+\ell})$.

Theorem 1. *Under Condition* 1 *and the assumption that* $h_n \to 0$, $nh_n^d \to \infty$ *as* $n \to \infty$, *we have at every point of continuity of* f

$$E[s_{n,\boldsymbol{j}}] \to f(\boldsymbol{x})\mu_{\boldsymbol{j}}, \quad nh_n^d \operatorname{var}[s_{n,\boldsymbol{j}}] \to f(\boldsymbol{x})\gamma_{2\boldsymbol{j}} \tag{2.6}$$

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

$$\mu_{\boldsymbol{j}} = \int_{\mathbb{R}^d} \boldsymbol{u}^{\boldsymbol{j}} K(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \quad \gamma_{\boldsymbol{j}} = \int_{\mathbb{R}^d} \boldsymbol{u}^{\boldsymbol{j}} K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}. \tag{2.7}$$

Define the $N \times N$ dimensional matrices $\boldsymbol{M}$ and Γ by

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{0,0} & \boldsymbol{M}_{0,1} & \cdots & \boldsymbol{M}_{0,p} \\ \boldsymbol{M}_{1,0} & \boldsymbol{M}_{1,1} & \cdots & \boldsymbol{M}_{1,p} \\ \vdots & & & \vdots \\ \boldsymbol{M}_{p,0} & \boldsymbol{M}_{p,1} & \cdots & \boldsymbol{M}_{p,p} \end{bmatrix} \quad \Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & & & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{bmatrix} \tag{2.8}$$

where $\boldsymbol{M}_{i,j}$ and $\Gamma_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (ℓ,m) element are, respectively, $\mu_{g_i(\ell)+g_j(m)}$ and $\gamma_{g_i(\ell)+g_j(m)}$. Note that the elements of the matrices $\boldsymbol{M}$ and Γ are simply multivariate moments of the kernel K and K^2, respectively.

Corollary 1. *Under the conditions of Theorem* 1 *we have*

$$\boldsymbol{S}_n \xrightarrow{m.s} \boldsymbol{M}f(\boldsymbol{x}) \text{ as } n \to \infty$$

at points of continuity of f *in the sense that each element of the matrix* $\boldsymbol{S}_n$ *converges in the mean-square sense to a constant multiple of* $f(\boldsymbol{x})$.

We now center $t_{n,\boldsymbol{j}}$ of (1.11) as follows. Let

$$t^*_{n,\boldsymbol{j}} = \frac{1}{n-d+1}\sum_{i=0}^{n-d}[\Psi(Y_{d+i}) - m(\boldsymbol{X}_i)]\left[\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right]^{\boldsymbol{j}} K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{2.9}$$

With the help of Theorem 1 we can now determine the bias contribution of the coefficients' estimates $\{\hat{b}_k\}$. We have

$$t_{n,j} - t_{n,j}^* = \frac{1}{n-d+1} \sum_{i=0}^{n-d} m(X_i) \left[\frac{X_i - x}{h}\right]^j K_h(X_i - x). \tag{2.10}$$

Expanding $m(X_i)$ in a Taylor series around x for $\|X_i - x\| \leqslant h$ and since $m(x)$ has continuous derivatives of total order $p+1$, we have

$$m(X_i) = \sum_{0 \leqslant |k| \leqslant p+1} \frac{1}{k!}(D^k m)(x)\,(X_i - x)^k + o_p(h^{p+1}). \tag{2.11}$$

Substituting in (2.10) and using (1.12) we find

$$t_{n,j} - t_{n,j}^* = \sum_{0 \leqslant |k| \leqslant p+1} \frac{1}{k!} h^{|k|} (D^k m)(x)\, s_{n,j+k} + o_p(h^{p+1}) s_{n,j+0}.$$

Using (1.10) and $D^k m = k! b_k$ we obtain

$$\begin{aligned} t_{n,j}^* = &\sum_{0 \leqslant |k| \leqslant p} h^{|k|} [\hat{b}_k(x) - b_k(x)] s_{n,j+k} \\ &- h^{p+1} \sum_{|k| = p+1} \frac{1}{k!}(D^k m)(x)\, s_{n,j+k} + o_p(h^{p+1}) s_{n,j+0}. \end{aligned} \tag{2.12}$$

By Theorem 1, the last term on the right-hand side is $o_p(h^{p+1})$ since $s_{n,j+0}$ converges in quadratic mean to $f(x)\mu_j$. Now arrange the N_{p+1} elements of the derivatives $(1/j!)(D^j m)(x)$ for $|j| = p+1$ as a column vector $m_{p+1}(x)$ using the lexicographical order introduced earlier. Similarly let the $N \times N_{p+1}$ matrix B_n be defined by

$$B_n = \begin{bmatrix} S_{n,0,p+1} \\ S_{n,1,p+1} \\ \vdots \\ S_{n,p,p+1} \end{bmatrix} \tag{2.13}$$

where the matrix $S_{n,i,p+1}$ is defined in (2.4). Then we can write (2.12), using (2.3) and the centered version of (2.2), in the matrix form

$$\tau_n^* = S_n(\hat{\beta}_n - \beta) - h^{p+1} B_n m_{p+1}(x) + o_p(h^{p+1}).$$

Thus

$$\hat{\beta}_n - \beta = S_n^{-1} \tau_n^* + h^{p+1} S_n^{-1} B_n m_{p+1}(x) + o_p(h^{p+1}).$$

By Corollary 1 we have S_n converges in mean square to $f(x)M$ and, similarly, B_n converges in mean square to $f(x)B$ where the matrix B is given by

$$B = \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}. \tag{2.14}$$

It follows that

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \boldsymbol{S}_n^{-1}\boldsymbol{\tau}_n^* + h^{p+1}\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x}) + o_p(h^{p+1}). \tag{2.15}$$

It is seen from (2.15) that the bias term of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ is of order h^{p+1} and is proportional to a linear combinations of the derivatives of $m(\boldsymbol{x})$ of total order $p+1$. Also note that the i-th element of $\hat{\boldsymbol{\beta}}_n$ represents an estimate of the derivative of $m(\boldsymbol{x})$ via the relationship

$$(\hat{\boldsymbol{\beta}}_n)_i = \frac{h^{|j|}(D^j m)\hat{}(\boldsymbol{x})}{j!}, \quad i = g_{|j|}^{-1}(j) + \sum_{k=0}^{|j|-1} N_k. \tag{2.16}$$

We next derive the asymptotic covariance of the centered $t_{n,j}^*$. Consider an arbitrary linear combination of the $t_{n,j}^*$,

$$Q_n = \sum_{0 \leqslant |j| \leqslant p} c_j t_{n,j}^* = \frac{1}{n-d+1}\sum_{i=0}^{n-d}[\Psi(Y_{d+i}) - m(\boldsymbol{X}_i)]C_h(\boldsymbol{X}_i - \boldsymbol{x}) \tag{2.17}$$

where

$$C_h(\boldsymbol{u}) = \sum_{0 \leqslant |j| \leqslant p} c_j(\boldsymbol{u}/h)^j K_h(\boldsymbol{u}) \equiv \frac{1}{h^d}C(\boldsymbol{u}/h) \tag{2.18a}$$

with

$$C(\boldsymbol{u}) = \sum_{0 \leqslant |j| \leqslant p} c_j \boldsymbol{u}^j K(\boldsymbol{u}). \tag{2.18b}$$

Put

$$Z_i = [\Psi(Y_{d+i}) - m(\boldsymbol{X}_i)]C_h(\boldsymbol{X}_i - \boldsymbol{x}) \tag{2.19}$$

Then

$$Q_n = \frac{1}{n-d+1}\sum_{i=0}^{n-d} Z_i. \tag{2.20}$$

We find the asymptotic variance of Q_n from which the covariance of the $t_{n,j}^*$'s is obtained. We make the following assumption on the kernel K and on the random processes $\{Y_i, \boldsymbol{X}_i\}$.

Condition 2. (a) $K(\boldsymbol{u})$ is bounded with compact support (say $K(\boldsymbol{u}) = 0$ for $||\boldsymbol{u}|| > 1$).

(b) The conditional density $f_{\boldsymbol{X}_0, \boldsymbol{X}_\ell | Y_d, Y_{d+\ell}}(\boldsymbol{u}, \boldsymbol{v} | y_1, y_2) \leqslant A_2 < \infty$ for all $\ell \geqslant 1$.

(c) For ρ-mixing processes we assume

$$E|\Psi(Y_1)|^2 < \infty, \ \sum_{i=1}^{\infty} \rho(i) < \infty.$$

For strongly mixing processes we assume that for some $v > 2$ and $a > 1 - 2/v$,

$$\sum_{j=1}^{\infty} j^a[\alpha(j)]^{1-2/v} < \infty, \quad E[|\Psi(Y_d)|^v | \boldsymbol{X}_0 = \boldsymbol{u}] \leqslant A_3 < \infty \quad \text{for } \mathbf{u} \text{ in a neighbor-}$$

hood of $\boldsymbol{x}$.

Put

$$\sigma^2(\boldsymbol{x}) = \mathrm{var}[\Psi(Y_d)|\boldsymbol{X}_0 = \boldsymbol{x}]. \tag{2.21}$$

Theorem 2. *Under Condition 2 and the assumption that $h_n \to 0$, $nh_n^d \to \infty$ as $n \to \infty$, we have the following convergence results at every continuity point $\boldsymbol{x}$ of $\{\sigma^2, f\}$:*

(a) $h_n^d \ \mathrm{var}[Z_0] \to \sigma^2(\boldsymbol{x})f(\boldsymbol{x})\int_{\mathbb{R}^d} C^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$.

(b) $h_n^d \sum_{\ell=1}^{n-d} \mathrm{cov}(Z_0, Z_\ell) = \mathrm{o}(1)$.

(c) $nh_n^d \mathrm{var}[Q_n] \to \sigma^2(\boldsymbol{x})f(\boldsymbol{x}) \int_{\mathbb{R}^d} C^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$.

(d) $\mathrm{cov}\{(n h_n^d)^{1/2} t_{n,j}^*, (nh_n^d)^{1/2} t_{n,k}^*\} \to \sigma^2(\boldsymbol{x})f(\boldsymbol{x})\gamma_{j+k}$, where γ_j is defined in (2.7).

2.2. Proofs

Proof of Theorem 1. By (1.12) we have

$$E[s_{n,j}] = \int_{\mathbb{R}^d} ((\boldsymbol{u} - \boldsymbol{x})/h)^j K_h(\boldsymbol{u} - \boldsymbol{x})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}.$$

Under Condition 1, the function $K_{j,h} = (1/h^d)K_j\ (\boldsymbol{u}/h)$ with $K_j(\boldsymbol{u}) = \boldsymbol{u}^j K(\boldsymbol{u})$ is an approximation of the identity as $h \to 0$ (note that $|\boldsymbol{u}^j| \leqslant ||\boldsymbol{u}||^{|j|}$). Hence by Bochner's lemma (Wheeden and Zygmund, 1977, Theorem 9.9) we have $E[s_{n,j}] \to f(\boldsymbol{x})\mu_j$ at continuity points of f. For the variance, let

$$U_{i,j} = \left[\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right]^j K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{2.22}$$

Then by stationarity,

$$\begin{aligned}\mathrm{var}[s_{n,j}] &= \frac{1}{n-d+1}\left[\mathrm{var}[U_{0,j}] + 2\sum_{\ell=1}^{n-d}\left(1 - \frac{\ell}{n-d+1}\right)\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}\right] \\ &= J_1 + J_2.\end{aligned} \tag{2.23}$$

For J_1 we have $(n-d+1)J_1 = E[U_{0,j}^2] + \mathrm{O}(1)$ so that

$$(n-d+1)h_n^d J_1 = \int_{\mathbb{R}^d} ((\boldsymbol{u} - \boldsymbol{x})/h)^{2j} \left[\frac{1}{\mathrm{h}^d} K^2((\boldsymbol{u} - \boldsymbol{x})/h)\right] f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} + \mathrm{O}(h_n^d).$$

By Bochner's lemma we then have

$$nh_n^d J_1 \to f(\boldsymbol{x})\int_{\mathbb{R}^d} \boldsymbol{u}^{2j} K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = f(\boldsymbol{x})\gamma_{2j} \tag{2.24}$$

at continuity points of f. It remains to show that $nh_n^d J_2 = \mathrm{o}(1)$. We decompose the sum in J_2 in three sums

$$\frac{1}{n-d+1}\sum_{\ell=1}^{n-d} |\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}|$$

$$= \frac{1}{n-d+1}\sum_{\ell=1}^{d-1} + \frac{1}{n-d+1}\sum_{\ell=d}^{\pi_n} + \frac{1}{n-d+1}\sum_{\ell=\pi_n+1}^{n-d}$$

$$= J_{21} + J_{22} + J_{23} \tag{2.25}$$

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 $\boldsymbol{X}_0$ and $\boldsymbol{X}_\ell$. Let $f(\boldsymbol{u}', \boldsymbol{u}'', \boldsymbol{u}''')$ be the joint density of the $d + \ell$ distinct random variables in $(\boldsymbol{X}_0, \boldsymbol{X}_\ell)$, where $\boldsymbol{u}'$, $\boldsymbol{u}''$, and $\boldsymbol{u}'''$ are of dimensions ℓ, $d - \ell$, and ℓ respectively. Then

$$\mathrm{cov}\{U_{0,j}, U_{\ell,j}\} = \frac{1}{h^{d-\ell}}\int_{\mathbb{R}^{d+\ell}} (\boldsymbol{u}')^{j'}(\boldsymbol{u}'')^{j''}(\boldsymbol{u}''')^{j'''} K(\boldsymbol{u}',\boldsymbol{u}'')K(\boldsymbol{u}'',\boldsymbol{u}''')$$
$$\times [f(\boldsymbol{x}' - h\boldsymbol{u}', \boldsymbol{x}'' - h\boldsymbol{u}'', \boldsymbol{x}''' - h\boldsymbol{u}''')$$
$$- f(\boldsymbol{x}' - h\boldsymbol{u}', \boldsymbol{x}'' - h\boldsymbol{u}'')f(\boldsymbol{x}'' - h\boldsymbol{u}'', \boldsymbol{x}''' - h\boldsymbol{u}''')]$$
$$\mathrm{d}\boldsymbol{u}'\,\mathrm{d}\boldsymbol{u}''\,\mathrm{d}\boldsymbol{u}'''.$$

By Condition 1(b) we have

$$h^{d-\ell}|\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}|$$
$$\leqslant A_1 \int_{\mathbb{R}^{d+\ell}} \prod_{i=1}^{\ell} |u_i|^{j_i} \prod_{i=\ell+1}^{d} |u_i|^{2j_i} \prod_{i=d+1}^{d+\ell} |u_i|^{j_i} K(\boldsymbol{u}',\boldsymbol{u}'')K(\boldsymbol{u}'',\boldsymbol{u}''')\mathrm{d}\boldsymbol{u}'\,\mathrm{d}\boldsymbol{u}''\,\mathrm{d}\boldsymbol{u}'''.$$

Hence

$$nh_n^d |J_{21}| \leqslant \text{const.} \sum_{\ell=1}^{d-1} h_n^\ell = \mathrm{O}(h_n) \to 0. \tag{2.26a}$$

For J_{22}, there is no overlap between the components of $\boldsymbol{X}_0$ and $\boldsymbol{X}_\ell$ so that by Condition 1(b) we have

$$|\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}| \leqslant A_1\left[\int_{\mathbb{R}^d} \prod_{i=1}^{d} |u_i|^{j_i} K(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]^2 < \infty.$$

Hence

$$nh_n^d |J_{22}| \leqslant \text{const.} \sum_{\ell=d}^{\pi_n} h_n^d = \mathrm{O}(\pi_n h_n^d) \to 0 \tag{2.26b}$$

by the choice of π_n. For J_{23} we distinguish between ρ-mixing and strongly mixing processes. For ρ-mixing processes we have $|\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}| \leqslant \rho(\ell - d + 1)\mathrm{var}[U_{0,j}]$ and by (2.24)

$$nh_n^d |J_{23}| \leqslant f(\boldsymbol{x})\gamma_{2j}(1 + \mathrm{o}(1)) \sum_{\ell=\pi_n+1}^{\infty} \rho(\ell - d + 1) \to 0 \tag{2.26c}$$

by Condition 1(c). For strongly mixing processes, we have by Davydov's lemma (Hall and Heyde, 1980, Corollary A2)

$$\mathrm{cov}\{U_{0,j}, U_{\ell,j}\}| \leqslant 8[\alpha(\ell - d + 1)]^{1-2/v}[E|U_{0,j}|^v]^{2/v}.$$

It is easily seen that

$$E|U_{0,j}|^v \leqslant \int_{\mathbb{R}^d} ||(\boldsymbol{u} - \boldsymbol{x})/h||^{v|j|}\frac{1}{h_n^{vd}}K^v((\boldsymbol{u} - \boldsymbol{x})/h)f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

so that

$$h_n^{(v-1)d}E|U_{0,j}|^v \leqslant \int_{\mathbb{R}^d}\left[\frac{1}{h_n^d}||\boldsymbol{u}/h||^{v|j|}K^v(\boldsymbol{u}/h)\right]f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \to f(\boldsymbol{x})\int_{\mathbb{R}^d}||\boldsymbol{u}||^{v|j|}K^v(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

By Bochner's lemma. Thus,

$$nh_n^d|J_{23}| \leqslant \frac{\text{const.}}{h_n^{d(1-2/v)}}\sum_{\ell=\pi_n}^{\infty}[\alpha(\ell - d + 1)]^{1-2/v} \leqslant \frac{\text{const.}}{h_n^{d(1-2/v)}\pi_n^a}\sum_{\ell=\pi_n}^{\infty}\ell^a[\alpha(\ell)]^{1-2/v}.$$

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

$$nh_n^d|J_{23}| \leqslant \text{const.}\sum_{\ell=\pi_n}^{\infty}\ell^a[\alpha(\ell)]^{1-2/v} \to 0 \tag{2.26d}$$

as $n \to \infty$ by Condition 1(c). Thus be (2.23), (2.25). and (2.26) we have $nh_n^d J_2 \to 0$ as $n \to \infty$ and the result follows by (2.24). □

Proof of Theorem 2. We provide an outline of the proof taking into account the fact that some steps are similar to those in the proof of Theorem 1. First note that $E[Z_i] = 0$. Next by conditioning on $\boldsymbol{X}_0$,

$$\mathrm{var}[Z_0\} = E[Z_0^2] = E[\sigma^2(\boldsymbol{X}_0)C_h^2(\boldsymbol{X}_0 - \boldsymbol{x})] = \frac{1}{h^{2d}}\int_{\mathbb{R}^d}\sigma^2(\boldsymbol{u})C^2((\boldsymbol{u} - \boldsymbol{x})/h)f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

and by Condition 2(a) and Bochner's lemma

$$h_n^d\,\mathrm{var}[Z_0] \to \sigma^2(\boldsymbol{x})f(\boldsymbol{x})\int_{\mathbb{R}^d}C^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \tag{2.27}$$

at continuity points of $\sigma^2 f$. Next

$$\mathrm{var}[Q_n] = \frac{1}{n-d+1}\mathrm{var}[Z_0] + \frac{2}{n-d+1}\sum_{\ell=1}^{n-d}\left[1 - \frac{\ell}{n-d+1}\right]\mathrm{cov}\{Z_0, Z_\ell\}. \tag{2.28}$$

Let $\pi_n \to \infty$ such that $\pi_n h_n^d \to 0$. Write

$$J = \sum_{\ell=1}^{n-d}|\mathrm{cov}\{Z_0, Z_\ell\}| = \sum_{\ell=1}^{d-1} + \sum_{\ell=d}^{\pi_n} + \sum_{\ell=\pi_n+1}^{n-d} = J_1 + J_2 + J_3. \tag{2.29}$$

We show that $J = o(h_n^{-d})$ from which Part (c) of the theorem follows. Part (d) follows from Part (c) and (2.18b). It remains to prove Part (b) of the theorem. For J_1 we note that since the kernel K has compact support, $m(\boldsymbol{X}_i)$ is bounded in the neighborhood of $\|\boldsymbol{X}_i - \boldsymbol{x}\| \leqslant h$. Let $A_4 = \sup_{\|\boldsymbol{X}-\boldsymbol{x}\| \leqslant h} |m(\boldsymbol{X})|$. Conditioning on $(Y_d, Y_{d+\ell})$ and using Condition 2(b) we have for $1 \leqslant \ell \leqslant d-1$,

$$|\mathrm{cov}\{Z_0, Z_\ell\}| \leqslant A_2 E\{[|\Psi(Y_d)| + A_4][|\Psi(Y_{d+\ell})| + A_4]\}$$

$$\times \int_{\mathbb{R}^{d+\ell}} |C_h(\boldsymbol{u}' - \boldsymbol{x}', \boldsymbol{u}'' - \boldsymbol{x}'')||C_h(\boldsymbol{u}'' - \boldsymbol{x}'', \boldsymbol{u}''' - \boldsymbol{x}''')| \mathrm{d}\boldsymbol{u}' \mathrm{d}\boldsymbol{u}'' \mathrm{d}\boldsymbol{u}'''$$

as in the proof of Theorem 1. Thus

$$|\mathrm{cov}\{Z_0, Z_\ell\}| \leqslant \text{const.}\, h^{-(d-\ell)}$$

and

$$h_n^d |J_1| \leqslant \text{const} \sum_{\ell=1}^{d-1} h_n^\ell = \mathrm{O}(h_n) \to 0 \tag{2.30}$$

as $n \to \infty$. For J_2 we find similarly

$$|\mathrm{cov}\{Z_0, Z_\ell\}| \leqslant A_2 E\{[|\Psi(Y_d)| + A_4][|\Psi(Y_{d+\ell})| + A_4]\} \left[\int_{\mathbb{R}^d} |C_h(\boldsymbol{u} - \boldsymbol{x})| \mathrm{d}\boldsymbol{u}\right]^2$$

$$= \text{const. for all } \ell \geqslant \mathrm{d}.$$

Thus

$$h_n^d J_2 = \mathrm{O}(h_n^d \pi_n) \to 0 \tag{2.31}$$

by the choice of π_n. For J_3 under ρ-mixing we proceed as in the proof of Theorem 1 to obtain

$$|J_3| \leqslant \mathrm{var}[Z_0] \sum_{\ell=\pi_n}^{\infty} \rho(\ell)$$

and by (2.27) and Condition 2(c)

$$h_n^d |J_3| \to 0 \quad \text{as } n \to \infty. \tag{2.32a}$$

For strongly mixing processes, using Davydov's lemma, we have

$$|\mathrm{cov}\{Z_0, Z_\ell\}| \leqslant 8[\alpha(\ell - d + 1)]^{1-2/\nu}[E|Z_0|^\nu]^{2/\nu}.$$

Conditioning on $\boldsymbol{X}_0$ and using the second part of Condition 2(b) we have

$$E|Z_0|^\nu \leqslant \left[\sup_{\|\boldsymbol{X}_0 - \boldsymbol{x}\| \leqslant h} E[|\Psi(Y_d) + A_4|^\nu |\boldsymbol{X}_0]\right] E[|C_h(\boldsymbol{X}_0 - \boldsymbol{x})|^\nu]$$

$$\leqslant 2^{\nu-1}[A_3 + A_4^\nu]\left[\sup_{\|\boldsymbol{u} - \boldsymbol{x}\| \leqslant h} f_{\boldsymbol{X}_0}(\boldsymbol{u})\right] \int_{\mathbb{R}^d} |C_h(\boldsymbol{u})|^\nu \mathrm{d}\boldsymbol{u} \leqslant \text{const.}\, h^{-d(\nu-1)}.$$

It follows that

$$h_n^d|J_3| \leqslant \frac{\text{const.}}{h_n^{d(1-2/v)}} \sum_{\ell=\pi_n}^{\infty} [\alpha(\ell)]^{1-2/v}$$

and this tends to zero as $n \to \infty$ under Condition 2(c) in the manner of the proof of Theorem 1 (for the term J_{23}). □

3. Joint asymptotic normality

3.1. Main results

We first obtain the asymptotic normality of Q_n of (2.17). Recall that Q_n is an arbitrary linear combination of $t_{n,j}^*$'s *and* Q_n can be written in the form (2.20). Let

$$\theta^2(\boldsymbol{x}) = \sigma^2(\boldsymbol{x})f(\boldsymbol{x})\int_{\mathbb{R}^d} C^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \tag{3.1}$$

where $C(\boldsymbol{u})$ is defined in (2.18b). We make the following assumption on the mixing coefficients.

Condition 3. Let $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$. For ρ-mixing and strongly mixing processes, we assume that there exists a sequence $\{v_n\}$ of positive integers satisfying $v_n \to \infty$ and $v_n = o((nh_n^d)^{1/2})$ such that

$$(n/h_n^d)^{1/2}\rho(v_n) \to 0, \quad (n/h_n^d)^{1/2}\alpha(v_n) \to 0, \quad \text{as } n \to \infty.$$

Condition 4. The conditional distribution $G(y|\boldsymbol{u})$ of Y_d given $\boldsymbol{X}_0 = \boldsymbol{u}$ is continuous at the point $\boldsymbol{u} = \boldsymbol{x}$.

By dominated convergence, Condition 4 implies that for each $L > 0$, the functions $E[\Psi(Y_d)I(|\Psi(Y_d)| < L)|(\boldsymbol{X}_0 = \boldsymbol{u}]$, $E[\Psi(Y_d)^2 I(|\Psi(Y_d)| < L)|\boldsymbol{X}_0 = \boldsymbol{u}]$, are continuous at the point $\boldsymbol{x}$. Hence for each $L > 0$, $\tilde{\sigma}_L^2(\boldsymbol{u}) = \operatorname{var}[\Psi(Y_d)I(|\Psi(Y_d)| > L)|\boldsymbol{X}_0 = \boldsymbol{u}]$ is continuous at the point $\boldsymbol{x}$ provided $m(\cdot)$ and $\sigma(\cdot)$ are continuous at the point $\boldsymbol{x}$. Condition 4 is needed in the proof of Theorem 3 where a truncation argument is employed and the continuity of $\tilde{\sigma}_L^2(\boldsymbol{u})$ at $\boldsymbol{u} = \boldsymbol{x}$ is required in (3.31).

Theorem 3. *Under Conditions* 1–4, *we have the following asymptotic normality as* $n \to \infty$:

$$(nh_n^d)^{1/2} Q_n \xrightarrow{L} \mathrm{N}(0, \theta^2(\boldsymbol{x}))$$

at continuity points $\boldsymbol{x}$ *of* $\{\sigma^2, f\}$.

Proof of Theorem 3 is given in Section 3.2. We first remark on the mixing conditions required by Conditions 1–3.

Remark 2. If $h_n = An^{-\lambda}$ $(0 < \lambda < 1/d,\ A > 0)$, a sufficient condition for the mixing coefficients in Condition 3 is $\alpha(j) = O(j^{-\bar{a}})$ and $\rho(j) = O(j^{-\bar{a}})$ with $\bar{a} > (1 + \lambda d)/(1 - \lambda d)$ [with $v_n = (nh_n^d)^{1/2}/\log n$]. A sufficient condition for the mixing coefficients in Conditions 1(c) and 2(c) is $\alpha(j) = O(j^{-\bar{a}})$ with $\bar{a} > (2v - 2)/(v - 2)$ and $\rho(j) = O(j^{-\bar{a}})$ *with* $\bar{a} > 1$. Thus for ρ-mixing processes it suffices that $\rho(j) = O(j^{-\bar{a}})$ with $\bar{a} > (1 + \lambda d)/(1 - \lambda d)$ and $E|\Psi(Y_1)|^2 < \infty$. For strongly mixing processes with $E|\Psi(Y_1)|^v < \infty$ for some $v > 2$ it suffices that

$$\alpha(j) = O(j^{-\bar{a}}); \bar{a} > \max[(1 + \lambda d)/(1 - \lambda d), (2v - 2)/(v - 2)]$$

and it is seen that there is a trade-off between the order of the moment of $\Psi(Y_1)$ and the decay rate of the strong mixing coefficient: the larger the order v, the weaker is the decay rate of $\alpha(j)$.

Corollary 2. *Under the assumptions of Theorem* 3 *we have that the* $\{(nh_n^d)^{1/2} t_{n,j}^*\}$, *for distinct* $\boldsymbol{j}$*'s, are jointly asymptotically normal with zero means and covariance*

$$\operatorname{cov}[(nh_n^d)^{1/2} t_{n,j}^*, (nh_n^d)^{1/2} t_{n,k}^*] \to \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \gamma_{j+k}$$

at continuity points $\boldsymbol{x}$ *of* $\{\sigma^2, f\}$ *where* γ_j *is given in* (2.7).

Since Theorem 3 holds for all linear combinations of $t_{n,j}^*$, we immediately have for the $N \times 1$ lexicographically ordered vector $\boldsymbol{\tau}_n^*$ that $(nh_n^d)^{1/2}\boldsymbol{\tau}_n^* \xrightarrow{L} N(0, \sigma^2(\boldsymbol{x})\ f(\boldsymbol{x})\Gamma)$ at continuity points $\boldsymbol{x}$ of $\{\sigma^2,\ f\}$ where the $N \times N$ matrix Γ is given in (2.8). It now follows from Theorem 1 that

$$(nh_n^d)^{1/2}\boldsymbol{S}_n^{-1}\boldsymbol{\tau}_n^* \xrightarrow{L} N(0, \sigma^2(\boldsymbol{x})\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1}/f(\boldsymbol{x})] \tag{3.2}$$

at continuity points $\mathbf{x}$ of $\{\sigma^2, f\}$ whenever $f(\boldsymbol{x}) > 0$, where the $N \times N$ matrix $\boldsymbol{M}$ is given by (2.8). By (2.15) we have

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

This and (3.2) gives the following asymptotic normality result for the estimate $\hat{\boldsymbol{\beta}}_n$.

Theorem 4. *Under Conditions* 1–4 *and* $h_n = O(n^{-1/(d+2p+2)})$ *we have*

$$(nh_n^d)^{1/2}[\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}] - h^{p+1}\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x})] \xrightarrow{L} N[0, \sigma^2(\boldsymbol{x})\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1}/f(\boldsymbol{x})]$$

at continuity points $\boldsymbol{x}$ *of* $\{\sigma^2, f\}$ *whenever* $f(\boldsymbol{x}) > 0$.

Recall from (2.3) that $\hat{\boldsymbol{\beta}}_n$ is the $N \times 1$ lexicographically ordered vector of the scaled partial derivatives estimates of the regression function $m(\boldsymbol{x})$ of all orders up to a total order p: Specifically, the ith element of $\hat{\boldsymbol{\beta}}_n$ is equal to

$$(\hat{\boldsymbol{\beta}}_n)_i = \frac{h_n^{|\boldsymbol{j}|}(D^{\boldsymbol{j}}m)\hat{}(\boldsymbol{x})}{\boldsymbol{j}!} \quad \text{with } i = g_{|\boldsymbol{j}|}^{-1}(\boldsymbol{j}) + \sum_{k=0}^{|\boldsymbol{j}|-1} N_k.$$

Thus Theorem 4 establishes the joint asymptotic normality of the estimates of the regression function $m(\boldsymbol{x})$ and its derivatives up to a total order p. In addition, the bias and variance/covariance matrix (of the asymptotic normal distribution) of these estimators can be read from the theorem. For the individual partial derivatives of the regression function $m(\boldsymbol{x})$ we have, in particular,

Theorem 5. *Under Conditions* 1–4 *and* $h_n = \mathrm{O}(n^{-1/(d+2p+2)})$ *we have*

$$(nh_n^{d+2|\boldsymbol{j}|})^{1/2}\left[[(D^{\boldsymbol{j}}m)\hat{}(\boldsymbol{x}) - (D^{\boldsymbol{j}}m)(\boldsymbol{x})] - \boldsymbol{j}!(\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x}))_i h_n^{p+1-|\boldsymbol{j}|}\right] \xrightarrow{L} \mathrm{N}\left[0, \frac{\sigma^2(\boldsymbol{x})(\boldsymbol{j}!)^2}{f(\boldsymbol{x})}(\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1})_{i,i}\right]$$

at continuity points $\boldsymbol{x}$ *of* $\{\sigma^2, f\}$ *whenever* $f(\boldsymbol{x}) > 0$. Here the relationship between i and $\boldsymbol{j}$ *is given by*

$$i = g_{|\boldsymbol{j}|}^{-1}(\boldsymbol{j}) + \sum_{k=0}^{|\boldsymbol{j}|-1} N_k,$$

$(\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1})_{i,i}$ *is the* (i,i) *diagonal element of the matrix* $\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1}$ *and* $(\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x}))_i$ *is the ith element of the vector* $\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x})$).

Theorem 5 shows that the local higher-order polynomial fit of the partial derivative $(D^{\boldsymbol{j}}m)(\boldsymbol{x})$ has the following expressions for the bias and "variance":

$$\mathrm{bias}[D^{\boldsymbol{j}}m)\hat{}(\boldsymbol{x})] = \boldsymbol{j}!\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x}))_i h_n^{p+1-|\boldsymbol{j}|},$$

$$\mathrm{variance}[D^{\boldsymbol{j}}m)\hat{}(\boldsymbol{x})] = \frac{\sigma^2(\boldsymbol{x})(\boldsymbol{j}!)^2}{f(\boldsymbol{x})}(\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1})_{i,i}\frac{1}{nh_n^{d+2|\boldsymbol{j}|}}.$$

The optimal bandwidth for estimating the $\boldsymbol{j}$th derivative $(D^{\boldsymbol{j}}m)(\boldsymbol{x})$ can be defined as the one which minimizes the sum of the squared bias and "variance" above. One finds

$$h_{n,\boldsymbol{j}} = \left[\frac{\sigma^2(\boldsymbol{x})(d+2|\boldsymbol{j}|)(\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1})_{i,i}/f(\boldsymbol{x})}{2(p+1-|\boldsymbol{j}|)[(\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{m}_{p+1}(\boldsymbol{x}))_i]^2}\right]^{1/(d+2(p+1))}\frac{1}{n^{1/(d+2(p+1))}}.$$

With $h_{n,\boldsymbol{j}} = \mathrm{O}(n^{-1/(d+2(p+1))})$ and using the above expressions for the bias and "variance" for the estimate of $D^{\boldsymbol{j}}m(\boldsymbol{x})$ it is seen that the rate of "mean-square convergence" is $\mathrm{O}(n^{-2(p+1-|\boldsymbol{j}|)/(d+2(p+1))})$ which matches the optimal rate given by Stone (1982) in the vector-valued i.i.d. regression setting.

Remark 3. We note that, in the context of local polynomial regression estimation, the issue of selecting the bandwidth in a data-driven fashion has recently been addressed in the literature (see Fan and Gijbels, 1995; Ruppert et al., (1995)). In particular, in the context of one-dimensional regression with i.i.d. data, Fan and Gijbels (1995) proposed a variable bandwidth selection procedure (a two-stage approach) and showed via examples that the results are comparable to those based on nonlinear wavelet estimators.

Remark 4. It is possible to extend this work by employing a symmetric positive-definite $d \times d$ bandwidth matrix H_n instead of the scalar bandwidth parameter h_n. In this case, the kernel function $(1/h)K(\boldsymbol{u}/h)$ of this paper is replaced by $|H|^{-1}\mathrm{K}(\mathrm{H}^{-1}\boldsymbol{u})$ where $|H|$ is the determinant of H. This makes the notation slightly more complex but the analysis will go through. See Robinson (1983) and Ruppert and Wand (1994) where a bandwidth matrix is utilized.

Remark 5. The proof of Theorem 3 employs the big block–small block procedure. It was suggested by a referee to use instead a recent central limit theorem for stationary strongly mixing processes $\{X_i\}$ (e.g. Theorem 4 of Doukhan et al., (1994)). Unfortunately this does not appear to be feasible for two reasons: In Theorem 4 of Doukhan et al. a central limit theorem is established for $\sum_{i=1}^{n} X_i$ under the weak assumption

$$\int_0^1 \alpha^{-1}(u)[Q_{X_0}(u)]^2\,\mathrm{d}u < \infty \tag{$*$}$$

where Q_{X_0} is the quantile function of $|X_0|$ and $\alpha^{-1}(u)$ is the inverse of the mixing function $\alpha(u) = \alpha_{[u]}$. In our context we seek to establish the asymptotic normality of the partial sums $\sum_{i=1}^{n} Z_{n,i}$ of a triangular array $\{Z_{n,i}\}$ where

$$Z_{n,i} = (h_n^{d/2})[\Psi(Y_{d+i}) - m(\boldsymbol{X}_i)]C_{h_n}(\boldsymbol{X}_i - \boldsymbol{x}).$$

Direct application of a central limit theorem for a single sequence $\{X_i\}$ to the triangular array $\{Z_{n,i}\}$ is not feasible. Moreover, even if it were possible, one needs to impose an integrability condition like $(*)$ on the quantile function $Q_{Z_{n,i}}$ of $|Z_{n,i}|$. Such a condition is utterly unverifiable given the dependence of $Z_{n,i}$ on the bandwidth h_n, the regression function $m(\cdot)$, and the kernel $K(\cdot)$. Ideally, one would like to impose a condition on the quantile functions of the *underlying* processes $\{Y_i, X_i\}$ but we see no clear way of translating it to a condition on the quantile function of $|Z_{n,i}|$.

3.2. Proofs

Proof of Theorem 3. We employ the big block–small block procedure. Put

$$Z_{n,i} = (h_n^d)^{1/2} Z_i, \quad i = 0, 1, \ldots, n-d \tag{3.3a}$$

$$W_n = \sum_{i=0}^{n-1} Z_{n,i}. \tag{3.3b}$$

Then

$$(nh_n^d)^{1/2} Q_n = \left[\frac{n}{n-d+1}\right]^{1/2} \frac{1}{(n-d+1)^{1/2}} W_{n-d+1}.$$

If suffices to show that

$$\frac{1}{\sqrt{\mathrm{n}}} W_n \xrightarrow{L} \mathrm{N}(0, \theta^2(\boldsymbol{x})). \tag{3.4}$$

Note that by Theorem 2 we have

$$\operatorname{var}[Z_{n,0}] \to \theta^2(\boldsymbol{x}); \quad \sum_{\ell=1}^{n-1} |\operatorname{cov}\{Z_{n,0}, Z_{n,\ell}\}| = \mathrm{o}(1). \tag{3.5}$$

Partition the set $\{0, 1, \ldots, n-1\}$ into $2k+1$ subsets with large blocks of size $u = u_n$ and small blocks of size $v = v_n$ where

$$k = k_n = \left\lfloor \frac{n}{u_n + v_n} \right\rfloor. \tag{3.6}$$

Define the random variables

$$\eta_j = \sum_{i=j(u+v)}^{j(u+v)+u-1} Z_{n,i}, \quad 0 \leqslant j \leqslant k-1, \tag{3.7}$$

$$\xi_j = \sum_{i=j(u+v)+u}^{(j+1)(u+v)-1} Z_{n,i}, \quad 0 \leqslant j \leqslant k-1, \tag{3.8}$$

$$\zeta_k = \sum_{i=k(u+v)}^{n-1} Z_{n,i}. \tag{3.9}$$

Write

$$\begin{aligned} W_n &= \sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \zeta_k \\ &\equiv W_n' + W_n'' + W_n'''. \end{aligned} \tag{3.10}$$

We show that as $n \to \infty$,

$$\frac{1}{n} E[W_n'']^2 \to 0, \quad \frac{1}{n} E[W_n''']^2 \to 0, \tag{3.11a}$$

$$\left| E[\exp(it W_n')] - \prod_{j=0}^{k-1} E[\exp(it \eta_j)] \right| \to 0, \tag{3.11b}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E[\eta_j^2] \to \theta^2(\mathbf{x}), \tag{3.11c}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E[\eta_j^2 I\{|\eta_j| > \varepsilon \theta(\mathbf{x}) \sqrt{n}\}] \to 0 \tag{3.11d}$$

for every $\varepsilon > 0$. Eq. (3.11a) implies that W_n'' and W_n''' are asymptotically negligible, (3.11b) shows that the summands $\{\eta_j\}$ in W_n' are asymptotically independent; and (3.11c) and (3.11d) are the standard Lindeberg–Feller conditions for asymptotic normality of W_n' under independence.

We now prove (3.11a)–(3.11d) focusing on the strongly mixing case (which is more involved) and we remark on the differences for ρ-mixing processes. We first choose the large block size u_n. Condition 3 implies that there exist integers $q_n \to \infty$ such that for strongly mixing processes, we have

$$q_n v_n = o((n h_n^d)^{1/2}), \quad q_n (n/h_n^d)^{1/2} \alpha(v_n) \to 0 \quad \text{as } n \to \infty \tag{3.12}$$

[for ρ-mixing processes, $q_n (n/h_n^d)^{1/2} \rho(v_n) \to 0$ as $n \to \infty$]. Now define the large-block size u_n be $u_n = \lfloor (n h_n^d)^{1/2} / q_n \rfloor$. Then using (3.12) and simple algebra show that the following properties hold as $n \to \infty$:

$$v_n/u_n \to 0, \quad u_n/n \to 0, \quad u_n/(nh_n^d)^{1/2} \to 0, \tag{3.13}$$

$$\frac{n}{u_n}\alpha(v_n) \to 0 \tag{3.14}$$

[for ρ-mixing processes (3.14) is proved via the inequality $\alpha(v_n) \leqslant \rho(v_n)/4$]. We now establish (3.11a).

$$E[W_n'']^2 = \operatorname{var}\left[\sum_{j=0}^{k-1} \xi_j\right] = \sum_{j=0}^{k-1} \operatorname{var}[\xi_j] + \sum_{\substack{i=0 \\ i \neq j}}^{k-1}\sum_{j=0}^{k-1} \operatorname{cov}\{\xi_i, \xi_j\} \equiv F_1 + F_2. \tag{3.15}$$

By stationarity and (3.5)

$$\operatorname{var}[\xi_j] = v_n \operatorname{var}[Z_{n,0}] + 2v_n \sum_{i=1}^{v_n - 1}\left(1 - \frac{i}{v_n}\right)\operatorname{cov}\{Z_{n,0}, Z_{n,i}\} = v_n\theta^2(\boldsymbol{x})(1 + \mathrm{o}(1)). \tag{3.16}$$

Thus

$$F_1 = k_n v_n \theta^2(\boldsymbol{x})(1 + \mathrm{o}(1)) \sim \frac{nv_n}{v_n + u_n} \sim \frac{nv_n}{u_n} = \mathrm{o}(n). \tag{3.17}$$

Next consider the term F_2 in (3.15). With $r_j = j(u + v) + u$, we have

$$F_2 = \sum_{\substack{i=0 \\ i \neq j}}^{k-1}\sum_{j=0}^{k-1}\sum_{\ell_1=0}^{v-1}\sum_{\ell_2=0}^{v-1} \operatorname{cov}\{Z_{n,r_i+\ell_1}, Z_{n,r_j+\ell_2}\}. \tag{3.18}$$

since $i \neq j$, $|r_i - r_j + \ell_1 - \ell_2| \geqslant u$ so that

$$|F_2| \leqslant 2\sum_{\ell_1=0}^{n-u-1}\sum_{\ell_2=\ell_1+u}^{n-1} |\operatorname{cov}\{Z_{n,\ell_1}, Z_{n,\ell_2}\}|.$$

Since $u = u_n \to \infty$ we can assume that $u_n > d$ so that the random vectors $\boldsymbol{X}_{\ell_1}$ and $\boldsymbol{X}_{\ell_2}$ (appearing in Z_{n,ℓ_1} and Z_{n,ℓ_2} respectively) do not have common components. By stationarity and (3.5)

$$|F_2| \leqslant 2n\sum_{j=u}^{n-1} |\operatorname{cov}\{Z_{n,0}, Z_{n,j}\}| = \mathrm{o}(n). \tag{3.19}$$

Hence by (3.15), (3.17), and (3.19) we have

$$\frac{1}{n}E[W_n'']^2 \to 0 \quad \text{as } n \to \infty.$$

Using a similar argument, we find together with (3.5) and (3.13)

$$\begin{aligned} \frac{1}{n}E[W_n''']^2 &\leqslant \frac{1}{n}[n - k(u + v)]\operatorname{var}[Z_{n,0}] + 2\sum_{j=1}^{n-1} \operatorname{cov}\{Z_{n,0}, Z_{n,j}\}| \\ &\leqslant \frac{u_n + v_n}{n}\theta^2(\boldsymbol{x}) + \mathrm{o}(1) \to 0 \quad \text{as } n \to \infty. \end{aligned} \tag{3.20}$$

In order to establish (3.11b) we make use of the following lemma due to Volkonskii and Rozanov (1959).

Lemma. *Let* $V_1, \ldots, V_J$ *be random variables measurable with respect to the σ-algebras* $\mathscr{F}_{i_1}^{j_1}, \ldots, \mathscr{F}_{i_J}^{j_J}$ *respectively with* $1 \leqslant i_1 < j_1 < i_2 < \ldots < j_J \leqslant n$, $i_{\ell+1} - j_\ell \geqslant w \geqslant 1$ *and* $|V_j| \leqslant 1$ *for* $j = 1, \ldots, J$. *Then*

$$\left| E\left[\prod_{j=1}^{J} V_j\right] - \prod_{j=1}^{J} E[V_j]\right| \leqslant 16(J-1)\alpha(w).$$

We note that by (2.19), (3.2) and (3.7) η_a is a function of the random variables $\{X_{a(u+v)+1}, \ldots, X_{a(u+v)+u+d-1}; Y_{a(u+v)+d}, \ldots, Y_{a(u+v)+u+d-1}\}$ or η_a is $\mathscr{F}_{i_a}^{j_a}$-measurable with $i_a = a(u+v)+1$, $j_a = a(u+v)+u+d-1$. Also $i_{\ell+1} - j_\ell = v - d + 2$. Hence with $V_j = e^{it\eta_j}$ we have

$$\left| E[\exp(\mathrm{i}tW'_n)] - \sum_{j=0}^{k-1} E[\exp(\mathrm{i}t\eta_j)]\right| \leqslant 16k_n\alpha(v_n - d + 2) \sim \frac{n}{u_n}\alpha(v_n) \tag{3.21}$$

which tends to zero by (3.14).

Next we establish (3.11c). By stationarity and (3.16), with u_n replacing v_n, we have

$$\operatorname{var}[\eta_j] = \operatorname{var}[\eta_0] = u_n\theta^2(\boldsymbol{x})(1 + o(1))$$

so that

$$\frac{1}{n}\sum_{j=0}^{k-1} E[\eta_j^2] = \frac{k_n u_n}{n}\theta^2(\boldsymbol{x})(1 + o(1)) \to \theta^2(\boldsymbol{x}) \tag{3.22}$$

since $v_n/u_n \to 0$.

It remains to establish (3.11d). We employ a truncation argument since Ψ is not necessarily a bounded function. Let

$$a_L(y) = yI\{|y| \leqslant L\} \tag{3.23}$$

where L is a fixed truncation point. Put

$$m_L(\boldsymbol{x}) = E[a_L(\Psi(Y_d))|\boldsymbol{X}_0 = \boldsymbol{x}], \tag{3.24}$$

and

$$\sigma_L^2(\boldsymbol{x}) = E\left[(a_L(\Psi(Y_d)) - m_L(\boldsymbol{x}))^2 |\boldsymbol{X}_0 = \boldsymbol{x}\right], \tag{3.25}$$

$$\theta_L^2(\boldsymbol{x}) = \sigma_L^2(\boldsymbol{x})\int_{\mathbb{R}^d} C^2(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{3.26}$$

Put

$$Z_i^L = [a_L(\Psi(Y_{d+i})) - m_L(\boldsymbol{X}_i)]C_h(\boldsymbol{X}_i - \boldsymbol{x}); \quad Z_{n,i}^L = h_n^{d/2} Z_i^L \tag{3.27}$$

$$W_n^L = \sum_{i=0}^{n-1} Z_{n,i}^L, \quad \tilde{W}_n^L = \sum_{i=0}^{n-1} (Z_{n,i} - Z_{n,i}^L). \tag{3.28}$$

Using the fact that $C(\boldsymbol{u})$ is bounded (since K is bounded with compact support), we have

$$|Z_{n,i}^L| \leqslant \frac{\text{const.}}{h_n^{d/2}}.$$

This implies by (3.7) that

$$\max_{0 \leqslant j \leqslant k-1} |\eta_j^L|/\sqrt{n} \leqslant \text{const.} \frac{u_n}{(n h_n^d)^{1/2}} \to 0$$

by (3.13). Hence when n is large, the set $\{|\eta_i^L| \geqslant \theta_L(\boldsymbol{x})\varepsilon\sqrt{n}\}$ becomes an empty set and thus (3.11d) holds. Consequently (3.11a)–(3.11d) hold for W_n^L so that

$$(n h_n^d)^{1/2}\, W_n^L \xrightarrow{L} \mathrm{N}(0, \theta_L^2(\boldsymbol{x})). \tag{3.29}$$

In order to complete the proof, namely to establish (3.11d) in the general case, it suffices to show that

$$(n h_n^d)\operatorname{var}[\tilde{W}_n^L] \to 0 \quad \text{as first } n \to \infty \text{ and then } L \to \infty. \tag{3.30}$$

Indeed,

$$\begin{aligned}
&|E\exp(\mathrm{i}t\sqrt{(n h_n^d)}\,W_n) - \exp(-t^2\theta^2(\boldsymbol{x})/2)| \\
&\quad = |E\exp(\mathrm{i}t\sqrt{n h_n^d}(W_n^L + \tilde{W}_n^L)) - \exp(-t^2\theta_L^2(\boldsymbol{x})/2) + \exp(-t^2\theta_L^2(\boldsymbol{x})/2) \\
&\qquad - \exp(-t^2\theta^2(\boldsymbol{x})/2)| \\
&\quad \leqslant |E\exp(\mathrm{i}t\sqrt{n h_n^d}\,W_n^L) - \exp(-t^2\theta_L^2(\boldsymbol{x})/2)| + E|\exp(\mathrm{i}t\sqrt{n_n^d}\,\tilde{W}_n^L) - 1| \\
&\qquad + |\exp(-t^2\theta_L^2(\boldsymbol{x})/2) - \exp(-t^2\theta^2(\boldsymbol{x})/2)|.
\end{aligned}$$

Letting $n \to \infty$, the first term goes to zero by (3.29) for every $L > 0$; the second term converges to zero by (3.30) as first $n \to \infty$ and then $L \to \infty$; the third term goes to zero as $L \to \infty$ by dominated covergence. Therefore, it remains to prove (3.30). Note that $\tilde{W}_n^L$ has the same structure as W_n except that $\Psi(Y_i)$ is replaced by $\Psi(Y_i)I\{|\Psi(Y_i)| > L\}$. Hence, as in Theorem 2,

$$\lim_{n \to \infty} n h_n^d \operatorname{var}[\tilde{W}_n^L] = \operatorname{var}[\Psi(Y_d)I\{|\Psi(Y_d)| > L\}|\boldsymbol{X}_0 = \boldsymbol{x}]f(\boldsymbol{x})\int_{\mathbb{R}^d} C^2(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{3.31}$$

By dominated convergence the right-hand side converges to 0 as $L \to \infty$. This establishes (3.11d) and completes the proof of Theorem 3. $\square$

References

C.K. Chu and J.S. Marron, Choosing a kernel regression estimator (with discussion), Statist. Sci. 6 (1991) 404–436.

W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Amer. Statist. Assoc. 74 (1979) 829–836.

G. Collomb and W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations, Stochastic Process. Appl. 23 (1986) 77–89.

P. Doukhan, P. Massart and E. Rio, The functional central limit theorem for strongly mixing processes, Ann. Inst. H. Poincare 30 (1994) 63–82.

J. Fan, Design-adaptive nonparametric regression, Jour. Amer. Statist. Assoc. 87 (1992) 998–1004.

J. Fan, Local linear regression smoothers and their minimax efficiency, Ann. Statist. 21 (1993) 196–216.

J. Fan and I. Gijbels, Variable bandwidth and local linear regression smoothers, Ann. Statist. 20 (1992) 2008–2036.

J. Fan and I. Gijbels, Data driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation, J. Roy. Statist. Soc. Ser. B. 57 (1995) 371–394.

J. Fan and E. Masry, Multivariate regression estimation with errors-in-variables: Asymptotic normality for mixing processes, J. Multivariate Anal. 43 (1992) 237–271.

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

W. Härdle, Applied Nonparametric Regression (Cambridge University Press, Boston, MA, 1990).

T. Hastie and C. Loader, Local regression: Automatic kernel carpentry (with discussions), Statist. Sci. 8 (1993) 120–143.

A.N. Kolmogorov and Yu, A. Rozanov, On strong mixing conditions for stationary Gaussian processes. Theory Probab. Appl. 52 (1960) 204–207.

Y.P. Mack and B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitsheorie view. Gebiete 61 (1982) 405–415.

E. Masry, Multivariate local polynomial regression for time series: Uniform strong consistency and rates, J. Time Series Analysis, 1996.

E. Masry and J. Fan, Local polynomial estimation of regression functions for mixing processes, Scand. J. Statist., to appear.

E. Masry and D. Tjostheim, Nonparametric estimation and identification of nonlinear ARCH time series, Econometric Theory 11 (1995) 258–289.

E.A. Nadaraya, On estimating regression, Theor. Proabab. Appl. 9 (1964) 141–142.

P.M. Robinson, Nonparametric estimators for time series. J. Time Ser. Anal. 4 (1983) 185–297.

P.M. Robinson, On the consistency and finite sample properties of nonparametric kernel time series regression, auto regression, and density estimators, Ann. Inst. Statist. Math. 38 (1986) 539–549.

M. Rosenblatt, A central limit theorem and strong mixing conditions, Proc. Nat. Acad. Sci. 4 (1956) 43–47.

M. Rosenblatt, Conditional probability density and regression estimates; in: P.R. Krishnaiah, ed., Multivariate Analysis, Vol. II (Academic Press, New York, 1969) pp. 25–31.

G.G. Roussas, Nonparametric regression estimation under mixing conditions, Stochastic Process. Appl. 36 (1990) 107–116.

G.G. Roussas and L.T. Tran, Asymptotic normality of the recursive kernel regression estimate under dependence conditions, Ann. Statist. 20 (1992) 98–120.

D. Ruppert, S.J. Sheather and M.P. Wand, An effective bandwidth selection for local least squares regression, J. Amer. Statist. Assoc. 90 (1995) 1257–1270.

D. Ruppert and M.P. Wand, Multivariate weighted least squares regression, Ann. Statist. 22 (1994), 1346–1370.

C.J. Stone, Consistent Nonparametric Regression, Ann. Statist. 5 (1977) 595–645.

C.J. Stone, Optimal global rates of convergence for nonparametric regression, Ann. Statist. 10 (1982) 1040–1053.

D. Tjostheim, Non-linear time series: A selective review, Scand. J. Statist. 21 (1994) 97–130.

Y.K. Troung and C.J. Stone, Nonparametric function estimation involving time series, Ann. Statist. 20 (1992) 77–98.

V.A. Volkonskii and Yu. A. Rozanov, Some limit theorems for random functions, Theory Probab. Appl. 4 (1959) 178–197.

G.S. Watson, Smooth regression analysis, Sankhya Ser. A. 26 (1964) 359–372.

R.L. Wheeden and A. Zygmund, Measure and Integral (Marcel Dekker, New York, 1977).