Ann. Inst. Statist. Math.
Vol. 46, No. 2, 279–293 (1994)

NONPARAMETRIC TIME SERIES REGRESSION

YOUNG K. TRUONG*

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514, U.S.A.

(Received February 25, 1992; revised February 2, 1993)

Abstract. Consider a k-times differentiable unknown regression function $\theta(\cdot)$ of a d-dimensional measurement variable. Let $T(\theta)$ denote a derivative of $\theta(\cdot)$ of order $m < k$ and set $r = (k - m)/(2k + d)$. Given a bivariate stationary time series of length n, under some appropriate conditions, a sequence of local polynomial estimators of the function $T(\theta)$ can be chosen to achieve the optimal rate of convergence n^{-r} in L_2 norms restricted to compacts; and the optimal rate $(n^{-1} \log n)^r$ in the L_∞ norms on compacts. These results generalize those by Stone (1982, *Ann. Statist.*, **10**, 1040–1053) which deals with nonparametric regression estimation for random (i.i.d.) samples. Applications of these results to nonlinear time series problems will also be discussed.

Key words and phrases: Nonparametric regression, kernel estimator, local polynomials, optimal rates of convergence.

1. Introduction

Let $\{(\boldsymbol{X}_i, Y_i) : \boldsymbol{X}_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, i = 0, \pm 1, \pm 2, \ldots\}$ be a vector-valued stationary time series and let $\theta(\boldsymbol{x}) = E(Y_0 \mid \boldsymbol{X}_0 = \boldsymbol{x})$ denote the regression function of Y on $\boldsymbol{X}$. For nonnegative integers $\alpha_1, \ldots, \alpha_d$, set $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $[\alpha] = \alpha_1 + \cdots + \alpha_d$. Let k be a positive integer, $k \geq 1$, and suppose $\theta(\cdot)$ is a k-time differentiable function on $\mathbb{R}^d$. Set

$$T(\cdot) \equiv T(\cdot; \theta) = \sum_{[\alpha] < k} q_\alpha D^\alpha \theta(\cdot),$$

where q_α are constants and D^α denotes the differential operator defined by

$$D^\alpha = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

* This work was completed while the author was visiting Mathematical Sciences Research Institute at Berkeley, California. Research was supported in part by NSF Grant DMS-8505550, NC Board of Science and Technology Development Award 90SE06 and UNC Research Council.

Denote the order of T by m; that is, $m = \max\{[\alpha] : 0 \le [\alpha] < k \text{ and } q_\alpha \ne 0\}$. Then the function $T\theta(\cdot) = \theta(\cdot)$ corresponds to $m = 0$, while $T\theta(\cdot) = \partial\theta(\cdot)/\partial x_1$ corresponds to $m = 1$.

Given a set of realizations $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$, the current paper studies the asymptotic behaviors of nonparametric estimators of the smooth regression function $\theta(\cdot)$ and its derivatives $T(\cdot)$. This is responding to the current growing interests in nonlinear time series modellings. A number of specific results on testing and estimation for these nonlinear models have been given in the comprehensive accounts by Tong (1983, 1990), Subba Rao and Gabr (1984) and Priestley (1988). However, the important problem of identification still leaves something for desired. Recently, nonparametric methods for plotting the conditional mean and conditional variance have been suggested and proved to be useful in model selection. See, for example, Györfi *et al.* (1989), Auestad and Tjøstheim (1990) and Härdle (1990).

The paper is also aiming at generalizing Stone's results to time series, which can be described briefly as follows. Set $r = (k-m)/(2k+d)$. For random samples $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n) \sim_{\text{iid}} (\boldsymbol{X}, Y)$, Stone (1980, 1982) established that the minimax bound for the estimation of the function $T(\cdot)$ is $\{n^{-r}\}$ in both pointwise and the L_q $(1 \le q < \infty)$ norms, while it is $\{(n^{-1}\log n)^r\}$ in the L_∞ norm restricted to compacts. Moreover, these optimal bounds can be achieved by nonparametric estimators constructed using local polynomials. Under appropriate conditions to be given in the following sections, it will be shown that these optimal properties continue to hold in the context of time series.

Furthermore, it is worth noting that the approaches considered by Györfi *et al.* (1989), Auestad and Tjøstheim (1990) and Härdle (1990) are kernel methods based on local means or local constant fits, and it is known that these estimators can not possess the optimal rates of convergence $\{n^{(k-m)/(2k+d)}\}$ (with $k > m \ge 1$) without imposing further smoothness conditions on the marginal distribution of $\boldsymbol{X}_0$. These conditions can be avoided by adopting an approach based on local polynomials, which will be shown in this paper.

To make the problems more concrete, note that the vector or multivariate setup considered here has several applications depending on the nature of the problems. These will be illustrated in the following examples.

Example 1 (*Univariate time-series*). Let X_i, $i = 0, \pm 1, \pm 2, \ldots$ be a real-valued stationary time series and let m be an integer. Then the plot of $E(X_{i+d+m} \mid X_{i+1}, \ldots, X_{i+d})$ can provide some insight about the autoregression of $Y_i = X_{i+d+m}$ on $\boldsymbol{X}_i = (X_{i+1}, \ldots, X_{i+d})$.

Example 2 (*Bivariate time-series*). Let (X_i, Z_i), $i = 0, \pm 1, \ldots$ be an $\mathbb{R}^2$-valued stationary time series, let m be a nonnegative integer. Then the relationship between $Y_i = Z_{i+d+m}$ and $\boldsymbol{X}_i = (X_{i+1}, \ldots, X_{i+d})$ can be examined by plotting $E(Y_0 \mid \boldsymbol{X}_0)$. More generally, set

$$Y_i = Z_{i+d+m} \quad \text{and} \quad \boldsymbol{X}_i = (X_{i+1}, \ldots, X_{i+k}, Z_{i+k+1}, \ldots, Z_{i+d}),$$

where k is a positive integer such that $k \le d$. These types of series are often encountered in control theory where feedback plays an important role. Here, nonlinear problems can be studied by regressing Y_i on $\boldsymbol{X}_i$.

The rest of the paper is organized as follows. Nonparametric estimators of $T(\cdot)$ based on local polynomials will be described in Section 2. Conditions required by these estimators for the achievability of the optimal rates of convergence are given in Section 3. Discussions and further extensions of these results are provided in Section 4. Proofs are given in Section 5.

2. Nonparametric estimators

The estimator of the function $T(\cdot)$ will now be described. For each given $n \geq 1$, let $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ denote observations obtained from the bivariate series $\{(\boldsymbol{X}_i, Y_i) : i = 0, \pm 1, \pm 2, \ldots\}$ and let δ_n be positive numbers that tend to zero as $n \to \infty$. For a given $\boldsymbol{x} \in \mathbb{R}^d$, set

$$I_n(\boldsymbol{x}) = \{i : 1 \leq i \leq n \text{ and } \|\boldsymbol{X}_i - \boldsymbol{x}\| \leq \delta_n\}$$

and let $N_n(\boldsymbol{x}) = \#I_n(\boldsymbol{x})$ denote the number of points in $I_n(\boldsymbol{x})$, where $\|\boldsymbol{x}\| = (x_1^2 + \cdots + x_d^2)^{1/2}$ for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

For nonnegative integers $\alpha_1, \ldots, \alpha_d$, set $\alpha = (\alpha_1, \ldots, \alpha_d)$, $\alpha! = (\alpha_1)! \cdots (\alpha_d)!$, $[\alpha] = \alpha_1 + \cdots + \alpha_d$, and $\boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. Given an integer $k \geq 1$, let $P(\boldsymbol{x}; k)$ denote a polynomial in $\boldsymbol{x}^\alpha$ of degree $(k-1)$. That is, $P(\boldsymbol{x}; k) = \sum_{[\alpha]<k} c_\alpha \boldsymbol{x}^\alpha$, where c_α are the coefficients.

Denote $\hat{P}_n(\cdot; \boldsymbol{x})$ the polynomial such that it minimizes $\sum_{i \in I_n(\boldsymbol{x})} \{Y_i - P(\boldsymbol{X}_i - \boldsymbol{x}; k)\}^2$ over the class of polynomials $P(\cdot; k)$ of degree $(k-1)$. (It will be shown in Section 5 that the existence of $\hat{P}_n(\cdot; \boldsymbol{x})$ follows from Lemmas 5.2 and 5.3.) Define the estimator of $\theta(\boldsymbol{x})$ by $\hat{\theta}_n(\boldsymbol{x}) = \hat{P}_n(\boldsymbol{x}; \boldsymbol{x})$. Moreover, the estimator of $T(\boldsymbol{x}) = T(\boldsymbol{x}; \theta)$ based on $\hat{P}_n(\cdot; \boldsymbol{x})$ is defined by

$$\hat{T}_n(\boldsymbol{x}) = T(\boldsymbol{x}; \hat{\theta}_n) = \sum_{[\alpha]<k} q_\alpha D^\alpha \hat{\theta}_n(\boldsymbol{x}).$$

As an example, suppose $d = 1$ and $k = 2$. Then $\hat{P}_n(X; x) = \hat{a} + \hat{b}(X - x)$ with $\hat{a}$ and $\hat{b}$ minimizing $\sum_{i \in I_n(x)} \{Y_i - a - b(X_i - x)\}^2$. Here $\hat{a}$ and $\hat{b}$ are called local linear estimators of $\theta(x) = E(Y \mid X = x)$ and $\theta'(x)$, respectively.

3. Optimal rates of convergence

Conditions for the nonparametric estimators described in Section 2 to achieve the optimal rates of convergence are given in this section.

Let U be a nonempty open subset (of $\mathbb{R}^d$) containing the origin. The following smoothness condition is imposed on the function $\theta(\cdot)$ so that the bias can be estimated.

Condition 1. $\theta(\cdot)$ has bounded partial derivatives of order k on U.

The following two conditions are required to show the existence of $\hat{P}_n(\cdot; \boldsymbol{x})$. Consequently, it also guarantees the existence of the nonparametric estimator $\hat{T}_n(\cdot)$. These conditions are similar to those used in ordinary least square theory for random effect models.

CONDITION 2. The distribution of $\boldsymbol{X}_1$ is absolutely continuous and its density $f(\cdot)$ is bounded away from zero and infinity on U; that is, there is a positive constant M_1 such that $M_1^{-1} \leq f(\boldsymbol{x}) \leq M_1$ for $\boldsymbol{x} \in U$.

CONDITION 3. For $j \geq 1$, the conditional distribution of $\boldsymbol{X}_j$ given $\boldsymbol{X}_0 = \boldsymbol{x}$ has a density $f_j(\cdot \mid \boldsymbol{x})$; there is a positive constant M_2 such that

$$M_2^{-1} \leq f_j(\boldsymbol{x}' \mid \boldsymbol{x}) \leq M_2 \quad \text{for} \quad \boldsymbol{x}, \boldsymbol{x}' \in U \quad \text{and} \quad j \geq 1.$$

A moment condition is required:

CONDITION 4. (1) There is a positive constant $q > 2(2k+d)/k$ such that $E(|Y|^q) < \infty$.

(2) There is a positive constant $s > 3$ such that

$$\sup_{\boldsymbol{x} \in U} E(|Y|^s \mid \boldsymbol{X} = \boldsymbol{x}) < \infty.$$

Let $\mathcal{F}_j$ and $\mathcal{F}^j$ denote the σ-fields generated respectively by $(\boldsymbol{X}_i, Y_i)$, $-\infty < i \leq j$, and $(\boldsymbol{X}_i, Y_i)$, $j \leq i < \infty$. Given a positive integer u set

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

The stationary sequence is said to be α-mixing or strongly mixing if $\alpha(u) \to 0$ as $u \to \infty$. In this paper, the stationary time series $\{(\boldsymbol{X}_i, Y_i) : i = 0, \pm 1, \pm 2, \ldots\}$ is assumed to be α-mixing and it satisfies one of the following conditions.

CONDITION 5. (1) For some $s > 2$, $\sum_{i \geq N} \alpha^{1-2/s}(i) = O(N^{-1})$ as $N \to \infty$.
(2) For some $0 < \rho < 1$, $\alpha(u) = O(\rho^u)$ as $u \to \infty$.

Sufficient conditions for linear processes to be α-mixing are studied by Gorodetskii (1977) and Withers (1981). See also Auestad and Tjøstheim (1990) for an illuminating discussion on the role of α-mixing (or geometric ergodicity) for model identification in nonlinear time series analysis. Condition 5(2) can be weakened to algebraic rates. However, it is used here to simplify our presentation. See Tran (1993).

Given positive numbers b_n and c_n, $n \geq 1$, let $b_n \sim c_n$ mean that b_n/c_n is bounded away from zero and infinity. Given random variables V_n, $n \geq 1$, let $V_n = O_p(b_n)$ mean that the random variables $b_n^{-1}V_n$, $n \geq 1$, are bounded in probability; that is,

$$\lim_{c \to \infty} \limsup_n P(|V_n| > cb_n) = 0.$$

THEOREM 3.1. *Suppose Conditions* 1–5(1) *hold and that* $\delta_n \sim n^{-1/(2k+d)}$. *Then*

$$|\hat{T}_n(\boldsymbol{x}) - T(\boldsymbol{x};\theta)| = O_p(n^{-(k-m)/(2k+d)}), \qquad \boldsymbol{x} \in U.$$

Let C be a fixed compact subset of U having a nonempty interior. Given a real-valued function $g(\cdot)$ on C, set

$$\|g\|_2 = \left\{ \int_C |g(\boldsymbol{x})|^2 d\boldsymbol{x} \right\}^{1/2} \qquad \text{and} \qquad \|g\|_\infty = \sup_{\boldsymbol{x}\in C} |g(\boldsymbol{x})|.$$

The L_2 and L_∞ rates of convergence are given in the following results.

THEOREM 3.2. *Suppose Conditions* 1–5(2) *hold and that* $\delta_n \sim n^{-1/(2k+d)}$. *Then*

$$\|\hat{T}_n(\cdot) - T(\cdot;\theta)\|_2 = O_p(n^{-(k-m)/(2k+d)}).$$

THEOREM 3.3. *Suppose Conditions* 1–5(2) *hold and that* $\delta_n \sim (n^{-1} \cdot \log n)^{1/(2k+d)}$. *Then there is a positive constant* c *such that*

$$\lim_n P(\|\hat{T}_n(\cdot) - T(\cdot;\theta)\|_\infty \geq c(n^{-1}\log n)^{(k-m)/(2k+d)}) = 0.$$

Proofs of Theorems 3.1–3.3 will be given in Section 5.

4. Discussions

Given a univariate or bivariate time series, nonparametric estimates of the regression function based on the vector-valued setup in Section 1 can be useful for gaining insight toward model identification. See Auestad and Tjøstheim (1990). The current paper contributes to this approach by establishing optimal sampling properties for these nonparametric estimators.

For a class of k-time differentiable regression function $\theta(\cdot)$, optimal rates of convergence are achieved by using local polynomial estimators instead of the ordinary kernel method based on local mean, this is because optimal rates of convergence of the latter approach require extra smoothness conditions on the marginal distributions.

This paper also answers a question raised by Truong and Stone (1992): "Can the local constant fits be generalized to the local polynomials?" By taking a semiparametric approach, which did not include the univariate time series, Truong and Stone (1993) obtained optimal rates of convergence (with $k \geq 2$) for the local polynomial estimators. The current results generalize our previous results, and they are established by using a truncation argument and a theorem of Bradley (1983), which was first exploited by Tran (1989). Truong and Stone (1992) also considered an approach using local constant fit based on median. Followings are related open questions.

1. Set $\theta(\boldsymbol{x}) = \text{median}(Y_0 \mid \boldsymbol{X}_0 = \boldsymbol{x})$ and denote its estimate by $\hat{\theta}_n(\boldsymbol{x}) = \hat{P}_n(\boldsymbol{x};\boldsymbol{x})$, where $\hat{P}_n(\cdot;\boldsymbol{x})$ is a polynomial that minimizes $\sum_{i\in I_n(\boldsymbol{x})} |Y_i - P(\boldsymbol{X}_i - \boldsymbol{x};k)|$

over the class of polynomials $P(\cdot;k)$ of degree $(k-1)$. (See Section 2.) Moreover, define the estimator of $T(\boldsymbol{x}) = T(\boldsymbol{x};\theta) = \sum_{[\alpha]<k} q_\alpha D^\alpha \theta(\boldsymbol{x})$ based on $\hat{P}_n(\cdot;\boldsymbol{x})$ by

$$\hat{T}_n(\boldsymbol{x}) = T(\boldsymbol{x};\hat{\theta}_n) = \sum_{[\alpha]<k} q_\alpha D^\alpha \hat{\theta}_n(\boldsymbol{x}).$$

Then the question is: Will $\hat{T}_n(\boldsymbol{x})$ achieve the optimal rates of convergence given in Theorems 3.1–3.3?

2. For kernel estimators based on local constants, the "smoothing parameter" δ_n can be chosen by minimizing

$$\mathrm{CV}(\delta_n) = n^{-1}\sum (Y_i - \hat{\theta}_{n,i}(\boldsymbol{X}_i))^2 w(\boldsymbol{X}_i),$$

where $\hat{\theta}_{n,i}(\cdot)$ is computed by leaving $(\boldsymbol{X}_i, Y_i)$ case out, and $w(\cdot)$ is a weight function. See Auestad and Tjøstheim (1990), Härdle and Vieu (1992) and the references given therein. This cross-validatory technique can be generalized similarly to local polynomial fits in choosing δ_n. An interesting question would be: Is (4.1) of Härdle and Vieu (1992) still valid for local polynomial fits? What can be gained here?

3. Suppose the regression function for an univariate time series is given by

$$E(X_{i+1} \mid X_{i_1}, X_{i_2}, \ldots, X_{i_q}),$$

where $i \geq i_1 \geq \cdots \geq i_q \geq i-d+1$. How would the final prediction error (FPE) (Akaike (1969)) and Akaike information criteria (AIC) (Akaike (1973)) for linear autoregressive process be further developed to handle the nonparametric case in selecting the lags?

5. Proofs

We recall the following notations. For $\alpha = (\alpha_1, \ldots, \alpha_d)$, $[\alpha] = \alpha_1 + \cdots + \alpha_d$, $\alpha! = (\alpha_1)! \cdots (\alpha_d)!$ and $\boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. Also, set $\mathcal{A} = \{\alpha : [\alpha] < k\}$. Given $\boldsymbol{x} \in C$, let $\boldsymbol{Y}_n(\boldsymbol{x})$, $\boldsymbol{X}_n(\boldsymbol{x})$ and $\boldsymbol{A}_n(\boldsymbol{x})$ be defined as follows: $\boldsymbol{Y}_n(\boldsymbol{x}) = (Y_{ni}(\boldsymbol{x}))$ is the n-dimensional column vector given by $Y_{ni}(\boldsymbol{x}) = Y_i$ if $i \in I_n(\boldsymbol{x})$ and $Y_{ni}(\boldsymbol{x}) = 0$ otherwise; $\boldsymbol{X}_n(\boldsymbol{x}) = (X_{ni\alpha}(\boldsymbol{x}))$ is the $n \times \#(\mathcal{A})$ matrix given by $X_{ni\alpha}(\boldsymbol{x}) = (\boldsymbol{X}_i - \boldsymbol{x})^\alpha / \delta_n^{[\alpha]}$ if $i \in I_n(\boldsymbol{x})$ and $\alpha \in \mathcal{A}$ and $X_{ni\alpha}(\boldsymbol{x}) = 0$ otherwise; $\boldsymbol{A}_n(\boldsymbol{x}) = (A_{n\alpha\beta}(\boldsymbol{x})) = \boldsymbol{X}_n'(\boldsymbol{x})\boldsymbol{X}_n(\boldsymbol{x})$, which is a $\#(\mathcal{A}) \times \#(\mathcal{A})$ matrix. Since $\hat{P}_n(\cdot;\boldsymbol{x})$ is the polynomial of degree $k-1$ minimizing

$$\sum_{I_n(\boldsymbol{x})} \{Y_i - P(\boldsymbol{X}_i - \boldsymbol{x}; k)\}^2,$$

we have that

$$\hat{P}_n(\cdot;\boldsymbol{x}) = \sum_{[\alpha]<k} \hat{b}_{n\alpha}(\boldsymbol{x}) \frac{(\cdot - \boldsymbol{x})^\alpha}{\delta_n^{[\alpha]}},$$

where

$$\hat{b}_{n\alpha}(\boldsymbol{x}) = (\boldsymbol{A}_n^{-1}(\boldsymbol{x})\boldsymbol{X}_n'(\boldsymbol{x})\,\boldsymbol{Y}_n(\boldsymbol{x}))_\alpha.$$

Let $\boldsymbol{Q}_n$ denote the $\#(\mathcal{A})$-dimensional column vector defined by $\boldsymbol{Q}_n = (\alpha! q_\alpha / \delta_n^{[\alpha]})_{\alpha \in \mathcal{A}}$. Then

$$\hat{T}_n(\boldsymbol{x}) = T(\boldsymbol{x}; \hat{\theta}_n) = \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) \boldsymbol{Y}_n(\boldsymbol{x}).$$

Let $\theta_k(\cdot; \boldsymbol{x})$ denote the k-th order Taylor polynomial of $\theta(\cdot)$ about $\boldsymbol{x}$ defined by

$$\theta_k(\cdot; \boldsymbol{x}) = \sum_{[\alpha] < k} \frac{D^\alpha \theta(\boldsymbol{x})}{\alpha!} (\cdot - \boldsymbol{x})^\alpha, \qquad \boldsymbol{x} \in C.$$

Define the n-dimensional column vectors $\boldsymbol{T}_n(\boldsymbol{x}) = (T_{ni}(\boldsymbol{x}))$ and $\boldsymbol{T}_{kn}(\boldsymbol{x}) = (T_{kni}(\boldsymbol{x}))$ by

$$\begin{aligned} T_{ni}(\boldsymbol{x}) &= \theta(\boldsymbol{X}_i), \qquad & i \in I_n(\boldsymbol{x}), \\ T_{kni}(\boldsymbol{x}) &= \theta_k(\boldsymbol{X}_i; \boldsymbol{x}), \qquad & i \in I_n(\boldsymbol{x}), \end{aligned}$$

and $T_{ni}(\boldsymbol{x}) = T_{kni}(\boldsymbol{x}) = 0$, otherwise. Then

$$T_{kni}(\boldsymbol{x}) = \sum_{[\alpha] < k} \frac{(\boldsymbol{X}_i - \boldsymbol{x})^\alpha}{\delta_n^{[\alpha]}} \frac{\delta_n^{[\alpha]} D^\alpha \theta(\boldsymbol{x})}{\alpha!}, \qquad i \in I_n(\boldsymbol{x}),$$

so

$$\begin{aligned} &(\boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) \boldsymbol{T}_{kn}(\boldsymbol{x}))_\alpha = \frac{\delta_n^{[\alpha]} D^\alpha \theta(\boldsymbol{x})}{\alpha!} \qquad \text{and} \\ &T(\boldsymbol{x}; \theta) = \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) \boldsymbol{T}_{kn}(\boldsymbol{x}). \end{aligned}$$

Consequently,

$$\begin{aligned} (5.1) \quad \hat{T}_n(\boldsymbol{x}) - T(\boldsymbol{x}; \theta) &= \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) [\boldsymbol{Y}_n(\boldsymbol{x}) - \boldsymbol{T}_n(\boldsymbol{x})] \\ &\quad + \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) [\boldsymbol{T}_n(\boldsymbol{x}) - \boldsymbol{T}_{kn}(\boldsymbol{x})], \qquad \boldsymbol{x} \in C. \end{aligned}$$

Moreover,

$$\begin{aligned} (5.2) \qquad & \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) [\boldsymbol{Y}_n(\boldsymbol{x}) - \boldsymbol{T}_n(\boldsymbol{x})] \\ &= \sum_{\alpha, \beta} \frac{\alpha! q_\alpha}{\delta_n^{[\alpha]}} [N_n(\boldsymbol{x}) (\boldsymbol{A}_n^{-1}(\boldsymbol{x}))_{\alpha\beta}] \\ &\quad \cdot \left(N_n(\boldsymbol{x})^{-1} \sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x}) (Y_i - \theta(\boldsymbol{X}_i)) \right), \end{aligned}$$

where $M_i^\beta(\boldsymbol{x}) = (\boldsymbol{X}_i - \boldsymbol{x})^\beta / \delta_n^{[\beta]}$. Also, recall that

$$N_n(\boldsymbol{x}) = \#\{i : 1 \le i \le n \text{ and } \|\boldsymbol{X}_i - \boldsymbol{x}\| \le \delta_n\}, \qquad \boldsymbol{x} \in C.$$

By Conditions 2, 3, 5(1) and Lemma 5 of Truong and Stone (1992), there are positive constants c_1 and c_2 such that

$$\lim_n P(c_1 n\delta_n^d \le N_n(\boldsymbol{x}) \le c_2 n\delta_n^d) = 1, \qquad \boldsymbol{x} \in C. \tag{5.3}$$

The existence of the polynomial $\hat{P}_n$ follows from (see Lemma 5.3 below)

$$N_n(\boldsymbol{x})(\boldsymbol{A}_n^{-1}(\boldsymbol{x}))_{\alpha\beta} = O_p(1), \qquad \boldsymbol{x} \in C, \quad \alpha, \beta \in \mathcal{A}. \tag{5.4}$$

LEMMA 5.1. *Suppose Conditions* 2–5(1) *hold and that* $\delta_n \sim n^{-1/(2k+d)}$. *Then*

$$N_n(\boldsymbol{x})^{-1} \sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i - \theta(\boldsymbol{X}_i)) = O_p(\delta_n^k), \qquad \boldsymbol{x} \in C, \quad \beta \in \mathcal{A}.$$

PROOF. According to Lemma 6 of Truong and Stone (1992),

$$E\left|\sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i - \theta(\boldsymbol{X}_i))\right|^2 = O\left(n\delta_n^d\right).$$

It follows from Markov's inequality and (5.3) that

$$\begin{aligned}
&P\left(N_n(\boldsymbol{x})^{-1} \sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i - \theta(\boldsymbol{X}_i)) > c\delta_n^k\right) \\
&\le \frac{E|\sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i - \theta(\boldsymbol{X}_i))|^2}{(cc_1 n\delta_n^{d+k})^2} + P(N_n(\boldsymbol{x}) \le c_1 n\delta_n^d) \\
&= O\left(\frac{1}{c^2 c_1^2 n\delta_n^{d+2k}}\right) + o(1) \\
&= o(1) \quad \text{as} \quad c \to \infty.
\end{aligned}$$

□

5.1 *Proof of Theorem* 3.1

According to Condition 1,

$$|T_{ni}(\boldsymbol{x}) - T_{kni}(\boldsymbol{x})| = O(\delta_n^k), \qquad i \in I_n(\boldsymbol{x}), \quad \boldsymbol{x} \in C.$$

It follows from (5.4) that

$$\boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x})[\boldsymbol{T}_n(\boldsymbol{x}) - \boldsymbol{T}_{kn}(\boldsymbol{x})] = O_p(\delta_n^{k-m}), \qquad \boldsymbol{x} \in C. \tag{5.5}$$

By (5.2), (5.4) and Lemma 5.1

$$\boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x})[\boldsymbol{Y}_n(\boldsymbol{x}) - \boldsymbol{T}_n(\boldsymbol{x})] = O_p(\delta_n^{k-m}). \tag{5.6}$$

The desired result follows from (5.1), (5.5) and (5.6). □

The proofs of Theorems 3.2–3.3 depend on a sequence of lemmas. First of all, (5.3) and (5.4) are required to hold uniformly over $\boldsymbol{x} \in C$.

LEMMA 5.2. *Suppose Conditions* 2, 3, 5(2) *hold and that* $\delta_n \sim n^{-r}$ *or* $\delta_n \sim (n^{-1}\log n)^r$. *Then there are positive constants* c_3 *and* c_4 *such that* $\lim_n P(\Omega_n) = 1$, *where* $\Omega_n = \{c_3 n\delta_n^d \le N_n(\boldsymbol{x}) \le c_4 n\delta_n^d$ *for* $\boldsymbol{x} \in C\}$.

PROOF. See Lemma 7 of Truong and Stone (1992). □

The following lemma will be needed to show the existence of the estimators $\hat{\theta}_n(\cdot)$ and $\hat{T}_n(\cdot)$.

LEMMA 5.3. *Suppose Conditions* 2, 3, 5(2) *hold. Then there is a positive constant* c_5 *such that* $\lim_n P(\Phi_n) = 1$, *where* $\Phi_n = \{N_n(\boldsymbol{x})(\boldsymbol{A}_n^{-1}(\boldsymbol{x}))_{\alpha\beta} \le c_5$ *for all* $\boldsymbol{x} \in C$ *and* $\alpha, \beta \in \mathcal{A}\}$.

PROOF. See Lemma 1 of Truong and Stone (1993). □

5.2 *Proof of Theorem* 3.2

By Lemmas 5.2 and 5.3, there is a positive constant c_6 such that

$$
\begin{aligned}
P&\left(\int_C |\boldsymbol{Q}_n'\boldsymbol{A}_n^{-1}(\boldsymbol{x})\boldsymbol{X}_n'(\boldsymbol{x})[\boldsymbol{Y}_n(\boldsymbol{x}) - \boldsymbol{T}_n(\boldsymbol{x})]|^2 d\boldsymbol{x} \ge c^2\delta_n^{2(k-m)}\right) \\
&\le P\left(c_6\delta_n^{-2m}\max_{\beta}\int_C \left|\sum_{I_n(\boldsymbol{x})} M_i^{\beta}(\boldsymbol{x})[Y_i - \theta(\boldsymbol{X}_i)]\right|^2 d\boldsymbol{x} \ge c^2\delta_n^{2(k-m)}(c_3 n\delta_n^d)^2\right) \\
&\quad + P(\Omega_n^c) \\
&\le \frac{c_6^2\int_C E|\sum_{I_n(\boldsymbol{x})} M_i^{\beta}(\boldsymbol{x})[Y_i - \theta(\boldsymbol{X}_i)]|^2 d\boldsymbol{x}}{c^2 c_3^2\delta_n^{2k}(n\delta_n^d)^2} + P(\Omega_n^c) \\
&= O\left(\frac{1}{c^2 n\delta_n^{2k+d}}\right) + P(\Omega_n^c).
\end{aligned}
$$

It follows from $n\delta_n^{2k+d} \sim 1$ and Lemma 5.2 that

$$
\int_C |\boldsymbol{Q}_n'\boldsymbol{A}_n^{-1}(\boldsymbol{x})\boldsymbol{X}_n'(\boldsymbol{x})[\boldsymbol{Y}_n(\boldsymbol{x}) - \boldsymbol{T}_n(\boldsymbol{x})]|^2 d\boldsymbol{x} = O_p(\delta_n^{2(k-m)}). \tag{5.7}
$$

The conclusion follows from (5.5) and (5.7). □

To prepare for the proof of Theorem 3.3, we need a truncation argument. Let

$$
B_n = \{n(\log n)^a\}^{1/\nu}, \qquad \nu = 2(2k+d)/k, \quad 1 < a < (4k+d)/(2k). \tag{5.8}
$$

LEMMA 5.4. *Suppose Condition* 4(1) *holds. Then* $P(|Y_n| > B_n \text{ i.o.}) = 0$.

PROOF. This follows from Condition 4(1) and Markov's inequality

$$P(|Y_n| > B_n) \le B_n^{-\nu} E|Y_n|^\nu$$

via Borel-Cantelli Lemma. □

We also need a lemma on approximating α-mixing sequence of random variables by independent random variables.

LEMMA 5.5. *Let $\mathcal{B}$ be a Borel space and let V be a random element taking value in $\mathcal{B}$. Let U and S denote random variables so that $U \sim \mathrm{unif}(0,1)$ and U is independent of (V, S). Suppose further that ζ and q are positive constants so that $\zeta \le \|S\|_q = (E|S|^q)^{1/q} < \infty$. Then there exists a real-valued random variable $W = \varphi(V, S, U)$, where $\varphi : \mathcal{B} \times \mathbb{R} \times (0,1) \to \mathbb{R}$ is measurable, so that*
1. *W is independent of V,*
2. *W and S have the same distribution, and*

$$P(|W - S| \ge \zeta) \le 18(\|S\|_q/\zeta)^{q/(2q+1)} \cdot \alpha(\sigma(V), \sigma(S))^{2q/(2q+1)},$$

where $\alpha(\sigma(V), \sigma(S)) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \sigma(V), B \in \sigma(S)\}$.

PROOF. See Theorem 3 of Bradley (1983). □

Set

$$Y_i^B = Y_i 1_{\{|Y_i| \le B_n\}}, \qquad i = 1, 2, \ldots, n. \tag{5.9}$$

LEMMA 5.6. *Suppose Conditions* 2–5 *hold and that* $\delta_n \sim (n^{-1}\log n)^{1/(2k+d)}$. *Then there are positive constants c_7 and c_8 such that, for $\xi > 0$, and n is sufficiently large,*

$$P\left(\left|N_n(\boldsymbol{x})^{-1} \sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i^B - E(Y_i^B|\boldsymbol{X}_i))\right| \ge \xi\sqrt{(\log n)/n\delta_n^d}\right) \le c_7 n^{-c_8\xi},$$
$$\boldsymbol{x} \in C, \quad \beta \in \mathcal{A}.$$

PROOF. We may suppose that $n = 2n_1n_2$ and set

$$n_1 = [\delta_n^{-k}(n(\log n)^b)^{-1/\nu}], \qquad b > a,$$

where a and ν are given in (5.8). Note that $n_1 \to \infty$ because $\nu > (2k+d)/k$.
Write

$$\sum_{i \in I_n(\boldsymbol{x})} M_i^\beta(\boldsymbol{x})(Y_i^B - E(Y_i^B \mid \boldsymbol{X}_i)) = \sum_{j=1}^{n_2} V_{1j} + \sum_{j=1}^{n_2} V_{2j}, \tag{5.10}$$

where

$$V_{1j} = \sum_{i=(j-1)n_1+1}^{jn_1} K_{i,\beta}(Y_i^B - E(Y_i^B \mid \boldsymbol{X}_i)),$$
$$V_{2j} = \sum_{i=jn_1+1}^{(j+1)n_1} K_{i,\beta}(Y_i^B - E(Y_i^B \mid \boldsymbol{X}_i)),$$

and

$$K_{i,\beta} = 1_{\{\|\boldsymbol{X}_i - \boldsymbol{x}\| \le \delta_n\}} M_i^\beta(\boldsymbol{x}).$$

By Lemma 5.5, there exists a sequence of independent random variables $W_1, \ldots, W_{n_2}$ so that $W_j \stackrel{d}{=} V_{1j}$, and

$$P(|W_j - V_{1j}| \ge \zeta) \le 18(\|V_{1j}\|_q/\zeta)^{q/(2q+1)} \cdot \alpha(n_1)^{2q/(2q+1)}.$$

Set $z_n = \sqrt{n\delta_n^d \log n}$. If $(\xi/2n_2)z_n \le \|V_{1j}\|_q$, then

$$\begin{aligned} P(|W_j - V_{1j}| \ge (\xi/2n_2)z_n) &\le 18\left(\frac{\|V_{1j}\|_q \cdot \alpha(n_1)^2}{(\xi/2n_2)z_n}\right)^{q/(2q+1)} \\ &\le 18\left(\frac{nB_n\alpha(n_1)^2}{\xi z_n}\right)^{q/(2q+1)}. \end{aligned} \tag{5.11}$$

If $(\xi/2n_2)z_n > \|V_{1j}\|_q$, then

$$\begin{aligned} P(|W_j - V_{1j}| \ge (\xi/2n_2)z_n) &\le P(|W_j - V_{1j}| \ge \|V_{1j}\|_q) \\ &\le 18\alpha(n_1)^{2q/(2q+1)}. \end{aligned} \tag{5.12}$$

By Condition 5(2), (5.11) and (5.12), there exists a constant c_9 so that

$$P(|W_j - V_{1j}| \ge (\xi/2n_2)z_n) = O(\rho^{n^{c_9}}). \tag{5.13}$$

By Lemma 6 of Truong and Stone (1992), there exists a constant c_{10} so that

$$\sum_1^{n_2} E|W_j|^2 = \sum E|V_{1j}|^2 \le c_{10} n\delta_n^d.$$

Set $t = \delta_n^k$. Then $t|W_j| = O((\log n)^{-(b-a)/s}) \to 0$. Thus, $e^{tW_j} \le 1 + tW_j + t^2W_j^2$. By Markov's inequality,

$$\begin{aligned} P\left(\left|\sum_1^{n_2} W_j\right| \ge (\xi/2)z_n\right) &\le \frac{E(\exp(t\sum_1^{n_2} W_j))}{\exp(t(\xi/2)z_n)} \\ &\le \frac{\exp(t^2 \sum EW_j^2)}{\exp(t(\xi/2)z_n)} \\ &= \exp(c_{10}\log n - (\xi/2)\log n). \end{aligned} \tag{5.14}$$

Note that

$$P\left(\left|\sum_1^{n_2} V_{1j}\right| \geq \xi z_n\right) \leq P\left(\left|\sum_1^{n_2} W_j\right| \geq (\xi/2) z_n\right) + P\left(\left|\sum_1^{n_2}(V_{1j} - W_j)\right| \geq (\xi/2) z_n\right). \tag{5.15}$$

The desired result now follows from Lemma 5.2, (5.10), (5.13)–(5.15). □

5.3 *Proof of Theorem* 3.3

Set $T_n(\boldsymbol{x}) = \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}) \boldsymbol{T}_n(\boldsymbol{x})$. Then

$$T_n(\boldsymbol{x}) - T(\boldsymbol{x};\theta) = \boldsymbol{Q}_n' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x})[\boldsymbol{T}_n(\boldsymbol{x}) - \boldsymbol{T}_{nk}(\boldsymbol{x})].$$

Note that $\boldsymbol{T}_n(\boldsymbol{x}) - \boldsymbol{T}_{nk}(\boldsymbol{x})$ is the remainder term of Taylor expansion. Hence by Condition 1,

$$\sup_{\boldsymbol{x}\in C} |T_n(\boldsymbol{x}) - T(\boldsymbol{x};\theta)| = O_p(\delta_n^{k-m}). \tag{5.16}$$

This gives the bound for the "bias".

To establish the uniform error bound (over the set C) of the "variance" term $(\hat{T}_n(\boldsymbol{x}) - T_n(\boldsymbol{x}))$, we need a truncation and an approximation argument.

The truncation. By Lemma 5.4, for all sufficiently large n,

$$|Y_j| \leq B_n, \qquad j \leq n \qquad \text{w.p. } 1. \tag{5.17}$$

In fact, if $\omega \in \cup_{j\geq 1} \cap_{i\geq j} \{|Y_i| \leq B_i\}$, there exists a j_0 so that $|Y_i| \leq B_i$ for all $i \geq j_0$. Since B_n is increasing, the desired result follows by noting that there is a $j^* \geq j_0$ such that $B_{j^*} \geq \max\{|Y_1(\omega)|, |Y_2(\omega)|, \ldots, |Y_{j_0}(\omega)|, B_{j_0}\}$.

Set $\boldsymbol{U}_n(\boldsymbol{x}) = (U_{ni}(\boldsymbol{x}))$, $\boldsymbol{V}_n(\boldsymbol{x}) = (V_{ni}(\boldsymbol{x}))$, where $U_{ni}(\boldsymbol{x}) = Y_i^B - \theta(\boldsymbol{X}_i)$, $V_{ni}(\boldsymbol{x}) = Y_i 1_{\{|Y_i|>B_n\}}$ for $i \in I_n(\boldsymbol{x})$ and $U_{ni}(\boldsymbol{x}) = 0$, $V_{ni}(\boldsymbol{x}) = 0$ otherwise. Then

$$\hat{T}_n(\boldsymbol{x}) - T_n(\boldsymbol{x}) = \boldsymbol{W}_n(\boldsymbol{x}) \boldsymbol{U}_n(\boldsymbol{x}) + \boldsymbol{W}_n(\boldsymbol{x}) \boldsymbol{V}_n(\boldsymbol{x}) = \boldsymbol{W}_n(\boldsymbol{x}) \boldsymbol{U}_n(\boldsymbol{x}) \qquad \text{w.p. } 1, \tag{5.18}$$

where

$$\boldsymbol{W}_n(\boldsymbol{x}) = (W_i(\boldsymbol{x})) = \boldsymbol{Q}' \boldsymbol{A}_n^{-1}(\boldsymbol{x}) \boldsymbol{X}_n'(\boldsymbol{x}), \qquad \boldsymbol{x} \in C.$$

Note that $W_i(\boldsymbol{x}) = 0$ for $i \notin I_n(\boldsymbol{x})$. By Lemmas 5.2 and 5.3, there is a positive constant c_{11} such that

$$\lim_n P\left(\max_i |W_i(\boldsymbol{x})| \leq c_{11} n^{-1} \delta_n^{-d-m}, \boldsymbol{x} \in C\right) = 1. \tag{5.19}$$

An approximation. We need to approximate C by a set C_n having a finite number of elements. The set C_n can be described as follows. Since C is a compact subset of U with nonempty interior, we may assume that $C = [-1/2, 1/2]^d$. Let

λ be a positive integer so that $L_n = n^\lambda$. Let W_n be the collection of $(2L_n + 1)^d$ points in $[-1/2, 1/2]^d$ each of whose coordinates is of the form $j/(2L_n)$ for some integer j such that $|j| \leq L_n$. Then $[-1/2, 1/2]^d$ can be written as the union of $(2L_n)^d$ subcubes, each having length $1/2L_n$ and all of its vertices in W_n. For each $\boldsymbol{x} \in [-1/2, 1/2]^d$ there is a subcube Q_w with center $\boldsymbol{w}$ such that $\boldsymbol{x} \in Q_{\boldsymbol{w}}$. Let C_n denote the collection of centers of these subcubes.

For each fixed $\lambda > 0$, the uniform error bound restricted to C_n is given by

$$\lim_n P\left(\max_{\boldsymbol{w}\in C_n} |\hat{T}_n(\boldsymbol{w}) - T_n(\boldsymbol{w})| \geq c\delta_n^{-m}\sqrt{(\log n)/n\delta_n^d}\right) = 0, \tag{5.20}$$

for some positive constant c. In fact, this follows from Lemma 5.6, (5.2), (5.18) and Condition 4(2) via

$$\begin{aligned} &E(1_{\{\|\boldsymbol{X}_i-\boldsymbol{x}\|\leq\delta_n\}}E(Y_i 1_{\{|Y_i|>B_n\}} \mid \boldsymbol{X}_i)) \\ &\quad\leq B_n^{1-s}\sup_{\boldsymbol{a}\in U} E(|Y_i|^s \mid \boldsymbol{X}_i = \boldsymbol{a})P(\|\boldsymbol{X}_i - \boldsymbol{x}\| \leq \delta_n) \\ &\quad= O(\delta_n^d B_n^{1-s}) = o(\delta_n^{d+k}) \qquad \text{uniformly in } \boldsymbol{x} \in C. \end{aligned}$$

Moreover, for $\lambda > 0$ sufficiently small,

$$\max_{\boldsymbol{w}\in C_n}\sup_{\boldsymbol{x}\in Q_w} |\hat{T}_n(\boldsymbol{x}) - T_n(\boldsymbol{x}) - \hat{T}_n(\boldsymbol{w}) + T_n(\boldsymbol{w})| = o_p(\delta_n^{-m}\sqrt{1/n\delta_n^d}). \tag{5.21}$$

(Proof of (5.21) will be given shortly.)

We conclude from (5.20) and (5.21) that there is a positive constant c such that

$$\lim_n P\left(\sup_{\boldsymbol{x}\in C} |\hat{T}_n(\boldsymbol{x}) - T_n(\boldsymbol{x})| \geq c\delta_n^{-m}\sqrt{(\log n)/n\delta_n^d}\right) = 0. \tag{5.22}$$

The conclusion of Theorem 3.3 follows from (5.16) and (5.22).

Proof of (5.21). Let C_0 be a compact subset of U containing C in its interior. Since $\delta_n \to 0$, we can assume that if $i \in I_n(\boldsymbol{x})$ for some $\boldsymbol{x} \in C$, then $\boldsymbol{X}_i \in C_0$. Thus, $\max\{|\theta(\boldsymbol{X}_i)| : \boldsymbol{X}_i \in C_0\} < \infty$. Hence, from (5.17),

$$\sup_{\boldsymbol{x}\in C}\max_i |U_{ni}(\boldsymbol{x})| = \sup_{\boldsymbol{x}\in C}\max_i |Y_i^B - \theta(\boldsymbol{X}_i)| = O_p(B_n). \tag{5.23}$$

By Conditions 2 and 5(2) (see the proof of (2.13) of Truong and Stone (1992)), for λ sufficiently large,

$$\max_{\boldsymbol{w}\in C_n}\sup_{\boldsymbol{x}\in Q_w} \#(I_n(\boldsymbol{x})\triangle I_n(\boldsymbol{w})) = O_p(B_n). \tag{5.24}$$

(For sets A and B, $A\triangle B$ is the symmetric difference defined by $A\triangle B = (A\setminus B)\cup(B\setminus A)$.) We conclude from Lemmas 5.2, 5.3, (5.19), (5.23) and (5.24) that, for λ sufficiently large,

$$\max_{\boldsymbol{w}\in C_n}\sup_{\boldsymbol{x}\in Q_w} |\boldsymbol{W}_n(\boldsymbol{x})[\boldsymbol{U}_n(\boldsymbol{x}) - \boldsymbol{U}_n(\boldsymbol{w})]| = O_p(B_n^2 n^{-1}\delta_n^{-d-m}).$$

Consequently, by (5.8),

$$\max_{w\in C_n}\sup_{x\in Q_w}|W_n(x)[U_n(x)-U_n(w)]| = o_p(\delta_n^{-m}\sqrt{1/n\delta_n^d}). \tag{5.25}$$

Observe next that

$$\max_{w\in C_n}\sup_{x\in Q_w}\max_{i\in I_n(x)\triangle I_n(w)}\max_{\alpha\in\mathcal{A}}|X_{ni\alpha}(x)-X_{ni\alpha}(w)| = O(n^{-\lambda}\delta_n^{-1}). \tag{5.26}$$

We conclude from Lemmas 5.2, 5.3, (5.23), (5.24) and (5.26) that, for λ sufficiently large,

$$\max_{w\in C_n}\sup_{x\in Q_w}|Q_n' A_n^{-1}(x)[X_n'(x)-X_n'(w)]\,U_n(w)| = o_p(\delta_n^{-m}\sqrt{1/n\delta_n^d}). \tag{5.27}$$

Recall that $A_n(x)=(A_{n\alpha\beta}(x)) = X_n'(x)X_n(x)$. It follows from (5.24) and (5.26) that, for λ sufficiently large,

$$\max_{w\in C_n}\sup_{x\in Q_w}\max_{\alpha,\beta\in\mathcal{A}}|A_{n\alpha\beta}(x)-A_{n\alpha\beta}(w)| = o_p(B_n). \tag{5.28}$$

We conclude from Lemmas 5.2, 5.3 and (5.28) that, for λ sufficiently large,

$$\max_{w\in C_n}\sup_{x\in Q_w}\max_{\alpha,\beta\in\mathcal{A}}|(A_n^{-1}(x))_{\alpha\beta}-(A_n^{-1}(w))_{\alpha\beta}| = o_p(B_n(n\delta_n^d)^{-2})$$

and hence that

$$\max_{w\in C_n}\sup_{x\in Q_w}|Q_n'[A_n^{-1}(x)-A_n^{-1}(w)]X_n'(w)\,U_n(w)| = o_p(\delta_n^{-m}\sqrt{1/n\delta_n^d}). \tag{5.29}$$

It follows from (5.18), (5.25), (5.27) and (5.29) that (5.21) holds. □

Acknowledgements

The author is grateful to Chuck Stone and the referees for helpful comments.

References

Akaike, H. (1969). Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, **21**, 243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281, Akademiai Kiado, Budapest.

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination, *Biometrika*, **77**, 669–687.

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

Gorodetskii, V. V. (1977). On the strong mixing property for linear sequences, *Theory Probab. Appl.*, **22**, 411–413.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). Nonparametric curve estimation from time series, *Lecture Notes in Statist.*, **60**, Springer, New York.

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

Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series, *J. Time Ser. Anal.*, **13**, 211–232.

Priestley, M. B. (1988). *Non-Linear and Non-Stationary Time Series*, Academic Press, New York.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators, *Ann. Statist.*, **8**, 1348–1360.

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

Subba Rao, T. and Gabr, M. M. (1984). An introduction to bispectral analysis and bilinear time series models, *Lecture Notes in Statist.*, **24**, Springer, New York.

Tong, H. (1983). Threshold models in non-linear time series analysis, *Lecture Notes in Statist.*, **21**, Springer, New York.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

Tran, L. T. (1989). Recursive density estimation under dependence, *IEEE Trans. Inform. Theory*, **35**, 1103–1108.

Tran, L. T. (1993). Nonparametric function estimation for time series by local average estimators, *Ann. Statist.*, **21**, 1040–1057.

Truong, Y. K. and Stone, C. J. (1992). Nonparametric function estimation involving time series, *Ann. Statist.*, **20**, 77–97.

Truong, Y. K. and Stone, C. J. (1993). Semiparametric time series regression, *J. Time Ser. Anal.* (to appear).

Withers, C. S. (1981). Conditions for linear processes to be strong mixing, *Z. Wahrsch. Verw. Gebiete*, **57**, 477–480.