JOURNAL OF MULTIVARIATE ANALYSIS 41, 163–177 (1992)

Robust Nonparametric Regression in Time Series

YOUNG K. TRUONG

University of North Carolina, Chapel Hill

Communicated by the Editors

Consider a stationary time series $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \ldots$ with $\mathbf{X}_t$ being $\mathbb{R}^d$-valued and Y_t real-valued. Let $\psi(\cdot)$ denote a monotone function and let $\theta(\cdot)$ denote the robust conditional location functional so that $E[\psi(Y_0 - \theta(\mathbf{X}_0)) \mid \mathbf{X}_0] = 0$. Given a finite realization $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$, the problem of estimating $\theta(\cdot)$ is considered. Under appropriate regularity conditions, it is shown that a sequence of the robust conditional location functional estimators can be chosen to achieve the optimal rate of convergence $n^{-1/(2+d)}$ both pointwise and in L_q $(1 \leqslant q < \infty)$ norms restricted to a compact; it can also be chosen to achieve the optimal rate of convergence $(n^{-1} \log(n))^{1/(2+d)}$ in L_∞ norm restricted to a compact.

1. INTRODUCTION

Let $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \ldots$ denote a strictly stationary time series with $\mathbf{X}_t$ being $\mathbb{R}^d$-valued and Y_t being real-valued. Let $\theta(\mathbf{x})$ denote the regression function of Y_0 on $\mathbf{X}_0$ so that $\theta(\mathbf{x}) = E(Y_0 \mid \mathbf{X}_0 = \mathbf{x})$. To estimate this function, Nadaraya [13] and Watson [19] proposed

$$\hat{\theta}(\mathbf{x}) = \sum_1^n W_n(\mathbf{x}, \mathbf{X}_i)\, Y_i,$$

where the kernel weights $W_n(x, \mathbf{X}_i)$ satisfy $W_n(\mathbf{x}, \mathbf{X}_i) \geqslant 0$ and $\sum_1^n W_n(\mathbf{x}, \mathbf{X}_i) = 1$. $\hat{\theta}(\cdot)$ is called a kernel estimator based on local average. Alternately, $\hat{\theta}(\mathbf{x})$ can be obtained by minimizing over $t(\mathbf{x})$

$$\sum_1^n W_n(\mathbf{x}, \mathbf{X}_i)(Y_i - t(\mathbf{x}))^2,$$

Received December 7, 1990.

AMS 1980 subject classifications: 62G05, 62J02.

Key words and phrases: kernel estimator, local M-estimator, nonparametric regression, optimal rates of convergence, stationary time series, mixing processes.

0047-259X/92 $5.00

or, by solving

$$\sum_{1}^{n} W_n(\mathbf{x}, \mathbf{X}_i)(Y_i - t(\mathbf{x})) = 0.$$

But the solution to this least squares problem is very sensitive to outliers which are known to occur quite often in practice, even in the context of nonparametric regression analysis. Examples given in Härdle [11] indicated that nonparametric estimators constructed by taking the average over a small neighborhood about x, large Y values may result in peaks and bumps.

To remedy this problem, one may replace the above square function by other functions that put less weight on large value of Y_i. This leads to the estimator $\hat{\theta}(\cdot)$ defined implicitly as the solution to

$$G(\hat{\theta}(\mathbf{x})) = \sum_{1}^{n} W_i(\mathbf{x}, \mathbf{X}_i)\, \psi(Y_i - \hat{\theta}(\mathbf{x})) = 0,$$

where $\psi(\cdot)$ is an increasing and bounded function. Note that the boundedness condition here has the effect of truncating large Y values. For example, in dealing with heavy-tailed symmetric distributions such as the "gross error models," $\psi(\cdot)$ is chosen as $\psi(y) = \max\{-k, \min\{y, k\}\}$, $k > 0$, which has the effect of trimming large Y value to k, and the resulting estimator takes a form of trimmed mean. Such functions are very important and have been studied extensively in the area of robust estimation. See Hampel *et al.* [6] and Huber [12].

More generally, regression analysis can be viewed as follows. Let $\psi(\cdot)$ denote a monotone function and let $\theta(\cdot) \equiv \theta_\psi(\cdot)$ denote a function such that $E[\psi(Y_0 - \theta(\mathbf{X}_0)) \mid \mathbf{X}_0] = 0$ almost surely. The function $\theta(\cdot)$ is called the robust conditional location functional of Y_0 on $\mathbf{X}_0$ by Boente and Fraiman [1]. For example, in regression analysis involving asymmetric conditional distributions such as income distributions and if it desired to estimate the conditional median function, then the function $\psi(\cdot)$ is chosen to be $\psi(y) = \text{sign}(y)$ so that $\theta(\mathbf{x}) = \text{med}(Y_0 \mid \mathbf{X}_0 = \mathbf{x})$.

Given a realization of length n from $(\mathbf{X}_i, Y_i)$, $i = 0, \pm 1, \pm 2, \ldots$, the present paper considers the problem of estimating the function $\theta(\cdot)$. Note that many important time series problems can be analyzed via this setup. Specifically, the applications include problems of estimating the autoregression function of the "present" on its "past" in univariate time series, regression function estimation, and dynamic modellings based on bivariate time series. These are discussed in Examples 1–3 of Truong and Stone [18].

To estimate the function $\theta(\cdot)$, the parametric approach starts with specific assumptions about the relationship between $\mathbf{X}_0$ and Y_0 and about the variation in the Y-sequence that may or may not be accounted for by the $\mathbf{X}$-sequence. For instance, the standard autoregressive method in time series starts with an a priori model for the conditional mean function $\theta(\mathbf{x}) = E(Y_0 \mid \mathbf{X}_0 = \mathbf{x})$ which, by assumption or prior knowledge, is a linear function that contains finitely many unknown parameters. Under the assumption that the joint distribution is Gaussian, it is an optimal prediction rule; if the distribution is non-Gaussian, it is not generally possible to determine such function; so one might settle for the *best* linear predictor. By contrast, in the nonparametric approach, the function will be estimated directly without assuming such an a priori model for $\theta(\cdot)$. In recent years, nonparametric estimation has become an active area in statistics because of its flexibility in fitting data. Hence, this approach is adopted in this paper.

Robust parametric estimation for i.i.d. data in the context of linear models is discussed at length in Huber [12] and recently in Hampel *et al.* [6]. The nonparametric approach in regression analysis has been considered by Boente and Fraiman [1], Härdle [7], Härdle and Gasser [8], Härdle and Luckhaus [9], Härdle and Tsybakov [10], and Härdle [11]. However, the parallel studies in time series analysis appear to be much less developed. Recently, Collomb and Härdle [4] established a uniform consistency result based on strong assumptions on the correlated structure. Robinson [14] obtained a central limit theorem under weaker mixing conditions. Boente and Fraiman [2, 3] addressed issues on robust nonparametric estimation for dependent observations. Results on rates of convergence are also presented in the above papers. However, the results for dependent observations are not optimal in the minimax sense according to Stone [16, 17]. The present paper focuses on these optimal properties and it will be shown that under rather weak conditions, optimal rates of convergence for the robust location functional estimators can be achieved.

The rest of this paper is organized as follows. A class of robust conditional location functional will be defined in Section 2, along with results on optimal rates of convergence. Section 3 contains proofs of these results.

2. Statement of Results

Let U be a nonempty open subset of the origin of $\mathbb{R}^d$. The following smoothness condition is imposed on the conditional M-functional.

Condition 1. There is a positive constant M_0 such that

$$|\theta(\mathbf{x}) - \theta(\mathbf{x}')| \leqslant M_0 \, \|\mathbf{x} - \mathbf{x}'\| \qquad \text{for} \quad \mathbf{x}, \mathbf{x}' \in U,$$

where $\|\mathbf{x}\| = (x_1^2 + \cdots + x_d^2)^{1/2}$ for $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$.

Condition 2. The distribution of $\mathbf{X}_0$ is absolutely continuous and its density $f(\cdot)$ is bounded away from zero and infinity. That is, there is a positive constant M_1 such that $M_1^{-1} \leqslant f(\mathbf{x}) \leqslant M_1$ for $\mathbf{x} \in U$.

The following technical condition is required for bounding the variance of various terms in the proof.

Condition 3. The conditional distribution of $\mathbf{X}_j$ given $\mathbf{X}_0$ is absolutely continuous and its density $h(\cdot|\cdot)$ is bounded away from zero and infinity on U. That is $M_1^{-1} \leqslant h(\mathbf{x}_j \mid \mathbf{x}_0) \leqslant M_1$ for $\mathbf{x}_0$ and $\mathbf{x}_j \in U$, $j = 1, 2, \ldots$.

Conditions on the function $\psi(\cdot)$ are required to guarantee the uniqueness of the robust conditional location functional $\theta(\cdot)$ (uniqueness will ensure consistency) and also the achievability of the desired rate of convergence. (The same condition is required in order to obtain the usual asymptotic result about the M-estimate in the univariate case.) The proof of the existence and the uniqueness of the function $\theta(\cdot)$ can be found in Boente and Fraiman [1] and Härdle [7].

Condition 4. (i) The function $\psi(\cdot)$ is bounded and increasing with

$$E(\psi(Y - \theta(\mathbf{x})) \mid \mathbf{X} = \mathbf{x}) = 0, \qquad \mathbf{x} \in U.$$

(ii) There exists a positive constant M_2 such that

$$|E(\psi(Y - \theta(\mathbf{x}) + t) \mid \mathbf{X} = \mathbf{x})| > M_2 |t| \qquad \text{for} \quad |t| < M_2^{-1}, \mathbf{x} \in U.$$

(iii) There exists a positive constant M_3 such that

$$|E(\psi(Y - \theta(\mathbf{x}) + t) \mid \mathbf{X} = \mathbf{x})| > M_3 \qquad \text{for} \quad |t| \geqslant M_2^{-1}, \mathbf{x} \in U.$$

Also, there exist positive constants $M_4(\geqslant M_3)$ and $M_5(\geqslant M_2^{-1})$ such that

$$\psi(y) = \begin{cases} -M_4 & \text{if} \quad y \leqslant -M_5; \\ M_4 & \text{if} \quad y \geqslant M_5. \end{cases}$$

See Härdle and Luckhaus [9] for motivations of Conditions 4(i) and (ii). These conditions ensure uniqueness and finite variance in estimation.

Let $\mathscr{F}_t$ and $\mathscr{F}^t$ denote the σ-fields generated respectively by $(\mathbf{X}_i, Y_i)$, $-\infty < i \leqslant t$, and $(\mathbf{X}_i, Y_i)$, $t \leqslant i < \infty$. Given a positive integer k set

$$\alpha(k) = \sup\{|P(A \cap B) - P(A)\, P(B)| : A \in \mathscr{F}_t \text{ and } D \in \mathscr{F}^{t+k}\}.$$

According to Rosenblatt [15], a stationary sequence is said to be α-mixing or strongly mixing if $\alpha(k) \to 0$ as $k \to \infty$. This mixing condition is known to be very weak and hence it has been used in different forms by many authors to study nonparametric regression for dependent observations. See, for example, Boente and Fraiman [2, 3], Fan [5], Robinson [14], and Truong and Stone [18]. The following formulation is from Truong and Stone [18].

Condition 5. The process $(\mathbf{X}_i, Y_i)$, $i = 0, \pm 1, \pm 2, \ldots$, is α-mixing with

(i) $\sum_{i \geqslant N} \alpha(i) = O(N^{-1})$ as $N \to \infty$.

(ii) $\alpha(N) = O(\rho^N)$ as $N \to \infty$ for some ρ with $0 < \rho < 1$.

In nonparametric regression estimation based on kernel method, it is necessary to assume that the marginal has a smooth distribution (see, for example, Condition H4 of Boente and Fraiman [1]). We adopt an approach by Stone [16, 17] that avoids this problem. Let δ_n, $n \geqslant 1$, be positive numbers that tend to zero as $n \to \infty$. For $\mathbf{x} \in \mathbb{R}^d$ and $n \geqslant 1$, set

$$I_n(\mathbf{x}) = \{i\colon 1 \leqslant i \leqslant n \text{ and } \|\mathbf{X}_i - \mathbf{x}\| \leqslant \delta_n\}$$

and let $N_n(\mathbf{x}) = \# I_n(\mathbf{x})$ denote the number of points in I_n. Given $\mathbf{x} \in \mathbb{R}^d$, the robust conditional location functional estimator (also referred to as local M-estimators or M-smoother) is defined as the solution $\hat{\theta}_n(\mathbf{x})$ of the equation

$$\frac{1}{N_n(\mathbf{x})} \sum_{I_n(\mathbf{x})} \psi(Y_i - \hat{\theta}_n(\mathbf{x})) = 0.$$

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

$$\lim_{c \to \infty} \overline{\lim_{n}}\, P(|V_n| > c b_n) = 0.$$

Set $r = 1/(2 + d)$. The local (pointwise) rate of convergence of $\hat{\theta}_n(\cdot)$ is given in the following result.

THEOREM 1. *Suppose* $\delta_n \sim n^{-r}$ *and that Conditions* 1–3, 4(i), *and* 5(i) *hold. Then*

$$|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| = O_p(n^{-r}), \qquad \mathbf{x} \in U.$$

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

$$\|g(\cdot)\|_q = \left\{\int_C |g(\mathbf{x})|^q \, d\mathbf{x}\right\}^{1/q}, \ 1 \leqslant q < \infty, \qquad \text{and} \qquad \|g(\cdot)\|_\infty = \sup_{\mathbf{x} \in C} |g(\mathbf{x})|.$$

The L_∞ rate of convergence is given in the following result.

THEOREM 2. *Suppose* $\delta_n \sim (n^{-1} \log n)^r$ *and that Conditions* 1–3, 4(i), (ii), *and* 5(ii) *hold. Then there is a positive constant c such that*

$$\lim_n P(\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty \geqslant c(n^{-1} \log(n))^r) = 0.$$

The L_q rate of convergence is given in the following result.

THEOREM 3. *Suppose* $\delta_n \sim n^{-r}$ *and that Conditions* 1–4 *and* 5(ii) *hold. Then*

$$\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_q = O_p(n^{-r}), \qquad 1 \leqslant q < \infty.$$

Proofs of Theorems 1–3 will be given in Section 3.

Remark 1. According to Stone [16, 17] and since i.i.d. is a special case of stationary sequences, the rates presented in Theorems 1–3 are optimal rates of convergence.

Remark 2. Collomb and Härdle [4], Boente and Fraiman [2, 3] addressed the uniform consistency of local M-estimators corresponding to smooth $\psi(\cdot)$'s. The asymptotic independence used by Collomb and Härdle [4] is formulated in terms of ϕ-mixing, which is stronger than the α-mixing adopted in this paper. Moreover, the L_∞ rates of convergence presented in [2–4] are slower than the optimal rate $(n^{-1} \log n)^r$ given in Theorem 2. The L_q $(1 \leqslant q < \infty)$ rates of convergence in Theorem 3 are more difficult to obtain and were not considered by [2–4]. A central limit theorem (a pointwise result) for local M-estimators is given in Robinson [14].

Remark 3. Examples of the function $\psi(\cdot)$ that satisfy Condition 4 are $\psi(y) = \text{sign}(y)$ and Huber's $\psi(y) = \max\{-k, \min\{y, k\}\}$. Thus the above results unify approaches based on local medians and local M-estimators.

Remark 4. Truong and Stone [18] obtained the L_∞ rates of convergence for conditional mean function estimators by assuming the sequence Y_i to be bounded. Theorem 2 suggests a way to remove this boundedness condition by linearizing the tail behavior of Y through the function $\psi(\cdot)$.

3. PROOFS

The proofs depend on the general properties of the function $\psi(\cdot)$ given in Condition 4 and are refinements of the arguments in Truong and Stone [18].

Proof of Theorem 1. By symmetry, it suffices to show that

$$\lim_{c\to\infty} \overline{\lim_{n}}\, P(\hat{\theta}_n(\mathbf{x}) > \theta(\mathbf{x}) + cn^{-r}) = 0. \tag{3.1}$$

Set $I_n = I_n(\mathbf{x})$. It follows from Condition 1 that $\theta(\mathbf{X}_i) \leqslant \theta(\mathbf{x}) + M_0\delta_n$ for $i \in I_n$. Let $c > M_0$; then by Condition 4(ii),

$$\begin{aligned} E[\psi(Y_i - \theta(\mathbf{x}) - c\delta_n) \mid \mathbf{X}_i] &= E[\psi(Y_i - \theta(\mathbf{X}_i) + (\theta(\mathbf{X}_i) - \theta(\mathbf{x})) - c\delta_n) \mid \mathbf{X}_i] \\ &\leqslant -(c - M_0)\, M_2\delta_n, \qquad n \gg 1,\ i \in I_n. \end{aligned} \tag{3.2}$$

Set

$$Z_i = \psi(Y_i - \theta(\mathbf{x}) - c\delta_n) - E[\psi(Y_i - \theta(\mathbf{x}) - c\delta_n) \mid \mathbf{X}_i].$$

Then

$$E\left[\sum_{I_n} Z_i\right] = 0$$

and, by Conditions 2–4(i) and 5(i) (see Lemma 6 of Truong and Stone [18]),

$$\operatorname{Var}\left(\sum_{I_n} Z_i\right) = O(n\delta_n^d).$$

Let $c > M_0$. Then by (3.2),

$$N_n^{-1} \sum_{I_n} E[\psi(Y_i - \theta(\mathbf{x}) - c\delta_n) \mid \mathbf{X}_i] \leqslant -(c - M_0)\, M_2\delta_n, \qquad n \gg 1.$$

It now follows from (3.2) and Lemma 5 of Truong and Stone [18] that, for some $c_2 > 0$ and $n \gg 1$,

$$\begin{aligned} P(\hat{\theta}_n(\mathbf{x}) \geqslant \theta(\mathbf{x}) + c\delta_n) &\leqslant P\left(N_n^{-1} \sum_{I_n} \psi(Y_i - \theta(\mathbf{x}) - c\delta_n) \geqslant 0\right) \\ &= P\left(N_n^{-1} \sum_{I_n} Z_i \geqslant -N_n^{-1} \sum_{I_n} E[\psi(Y_i - \theta(\mathbf{x}) - c\delta_n) \mid \mathbf{X}_i]\right) \\ &\leqslant P\left(N_n^{-1} \sum_{I_n} Z_i \geqslant (c - M_0)\, M_2\delta_n\right) \\ &\leqslant P\left(N_n^{-1} \sum_{I_n} Z_i \geqslant (c - M_0)\, M_2\delta_n;\ N_n \geqslant c_2 n\delta_n^d\right) + P(N_n < c_2 n\delta_n^d) \\ &\leqslant P\left(\sum_{I_n} Z_i \geqslant (c - M_0)\, M_2 c_2 n\delta_n^{d+1}\right) + o(1). \end{aligned}$$

Since $n\delta_n^{d+2} \sim 1$, (3.1) now follows from Chebyshev's inequality. This completes the proof of Theorem 1.

Proof of Theorem 2. We can assume that $C = [-\frac{1}{2}, \frac{1}{2}]^d \subset U$. Set $L_n = [n^{2r}]$. Let W_n be the collection of $(2L_n + 1)^d$ points in C each of whose coordinates is of the form $j/(2L_n)$ for some integer j such that $|j| \leqslant L_n$. Then C can be written as the union of $(2L_n)^d$ subcubes, each having length $2\lambda_n = (2L_n)^{-1}$ and all of its vertices in W_n. For each $\mathbf{x} \in C$, there is a subcube $Q_\mathbf{w}$ with center $\mathbf{w}$ such that $\mathbf{x} \in Q_\mathbf{w}$. Let C_n denote the collection of the centers of these subcubes. Then

$$P\left(\sup_{\mathbf{x} \in C} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| \geqslant c(n^{-1} \log n)^r\right)$$

$$= P\left(\max_{\mathbf{w} \in C_n} \sup_{\mathbf{x} \in Q_\mathbf{w}} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| \geqslant c(n^{-1} \log n)^r\right).$$

It follws from $\lambda_n \sim n^{-2r}$ and Condition 1 that (for n sufficiently large)

$$|\theta(\mathbf{x}) - \theta(\mathbf{w})| \leqslant M_0 \|\mathbf{x} - \mathbf{w}\| \leqslant M_0 \delta_n \qquad \text{for} \quad \mathbf{x} \in Q_\mathbf{w}, \mathbf{w} \in C_n.$$

Therefore, to prove the theorem, it is sufficient to show that there is a positive constant c such that

$$\lim_n P\left(\max_{\mathbf{w} \in C_n} \sup_{\mathbf{x} \in Q_\mathbf{w}} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})| \geqslant c(n^{-1} \log n)^r\right) = 0. \tag{3.3}$$

Given $\mathbf{x} \in Q_\mathbf{w}$, set

$$N_n = N_n(\mathbf{x}),$$

$$I_n = I_n(\mathbf{x}),$$

$$\bar{I}_n = \bar{I}_n(\mathbf{w}) = \{i\colon 1 \leqslant i \leqslant n \text{ and } \|\mathbf{X}_i - \mathbf{w}\| \leqslant \delta_n + \lambda_n \sqrt{d}\},$$

$$\bar{N}_n = \bar{N}_n(\mathbf{w}) = \# \bar{I}_n(\mathbf{w}),$$

$$\underline{N}_n = \underline{N}_n(\mathbf{w}) = \# \{i\colon \|\mathbf{X}_i - \mathbf{w}\| \leqslant \delta_n - \lambda_n \sqrt{d}\}.$$

Then there is a positive constant c_3 such that

$$\lim_n P(\Pi_n) = 1, \tag{3.4}$$

where

$$\Pi_n = \left\{\max_{\mathbf{w} \in C_n} \sup_{\mathbf{x} \in Q_\mathbf{w}} \left|\frac{\sum_{\bar{I}_n} \psi_i}{\bar{N}_n} - \frac{\sum_{I_n} \psi_i}{N_n}\right| \leqslant c_3 \delta_n\right\},$$

$$\psi_i = \psi(Y_i - \theta(\mathbf{w}) - c\delta_n).$$

In fact, by Conditions 2, 3, and 5(ii), there are positive constants c_4 and c_5 such that

$$\lim_n P(\Psi_n) = 1, \tag{3.5}$$

where $\Psi_n = \{\bar{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) \leqslant c_4 n \delta_n^{d-1} \lambda_n$ and $\bar{N}_n(\mathbf{w}) \geqslant c_5 n \delta_n^d$ for all $\mathbf{w} \in C_n\}$. (Proof of (3.5) is given in (2.13) of Truong and Stone [18]). Given $\mathbf{x} \in C$, choose $\mathbf{w}$ such that $\mathbf{x} \in Q_{\mathbf{w}}$. Then $\underline{N}_n \leqslant N_n \leqslant \bar{N}_n$ and

$$\frac{\sum_{\bar{I}_n} \psi_i}{\bar{N}_n} - \frac{\sum_{I_n} \psi_i}{N_n} = \frac{N_n \sum_{\bar{I}_n \setminus I_n} \psi_i - (\bar{N}_n - N_n) \sum_{I_n} \psi_i}{\bar{N}_n N_n}.$$

Thus

$$\left| \frac{\sum_{\bar{I}_n} \psi_i}{\bar{N}_n} - \frac{\sum_{I_n} \psi_i}{N_n} \right| \leqslant \frac{(\bar{N}_n - \underline{N}_n)}{\bar{N}_n} \max_{\bar{I}_n \setminus I_n} |\psi_i| + \frac{(\bar{N}_n - \underline{N}_n)}{\bar{N}_n} \max_{\bar{I}_n} |\psi_i|$$

and, hence,

$$\left| \frac{\sum_{\bar{I}_n} \psi_i}{\bar{N}_n} - \frac{\sum_{I_n} \psi_i}{N_n} \right| \leqslant 2 \frac{(\bar{N}_n - \underline{N}_n)}{\bar{N}_n} \max_{\bar{I}_n} |\psi_i|.$$

Consequently, (3.4) follows from (3.5) and the boundedness of $\psi(\cdot)$.
By (3.4),

$$\begin{aligned} \{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \geqslant c\delta_n\} &\subseteq \left\{ N_n^{-1} \sum_{I_n} \psi(Y_i - \theta(\mathbf{w}) - c\delta_n) \geqslant 0 \right\} \\ &\subseteq \left\{ \bar{N}_n^{-1} \sum_{\bar{I}_n} \psi(Y_i - \theta(\mathbf{w}) - c\delta_n) \geqslant -c_3 \delta_n \right\} \\ &\qquad \text{on} \quad \Pi_n, n \gg 1. \end{aligned}$$

Thus

$$\begin{aligned} &\bigcup_{Q_{\mathbf{w}}} \{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \geqslant c\delta_n\} \\ &\quad \subseteq \left\{ \bar{N}_n^{-1} \sum_{\bar{I}_n} \psi(Y_i - \theta(\mathbf{w}) - c\delta_n) \geqslant -c_3 \delta_n \right\} \qquad \text{on} \quad \Pi_n, n \gg 1. \end{aligned}$$

Hence

$$\begin{aligned} &P\left(\max_{\mathbf{w} \in C_n} \sup_{\mathbf{x} \in Q_{\mathbf{w}}} [\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})] \geqslant c\delta_n \right) \\ &\quad \leqslant P\left(\bigcup_{C_n} \bigcup_{Q_{\mathbf{w}}} \{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \geqslant c\delta_n\} \right) \\ &\quad \leqslant P\left(\bigcup_{C_n} \left\{ \bar{N}_n^{-1} \sum_{\bar{I}_n} \psi(Y_i - \theta(\mathbf{w}) - c\delta_n) \geqslant -c_3 \delta_n \right\} \right) + P(\Pi_n^c). \end{aligned} \tag{3.6}$$

According to Condition 1,

$$|\theta(\mathbf{X}_i)-\theta(\mathbf{w})| \leqslant M_0(\delta_n+\lambda_n\sqrt{d}), \text{ whenever } \|\mathbf{X}_i-\mathbf{w}\| \leqslant \delta_n+\lambda_n\sqrt{d}.$$

Thus by Condition 4(ii), there is a positive constant c_6 such that $c \geqslant 2M_0$,

$$\begin{aligned} E(\psi(Y_i-\theta(\mathbf{w})-c\delta_n)\,|\,\mathbf{X}_i) &= E(\psi(Y_i-\theta(\mathbf{X}_i)+\theta(\mathbf{X}_i) \\ &\quad -\theta(\mathbf{w})-c\delta_n+M_0\lambda_n\sqrt{d})\,|\,\mathbf{X}_i) \\ &\leqslant -cc_6\delta_n, \quad n \gg 1,\, i \in \bar{I}_n. \end{aligned}$$

Hence

$$\bar{N}_n^{-1}\sum_{\bar{I}_n} E(\psi(Y_i-\theta(\mathbf{w})-c\delta_n)\,|\,\mathbf{X}_i) \leqslant -cc_6\delta_n. \tag{3.7}$$

Set $Z_i=\psi(Y_i-\theta(\mathbf{w})-c\delta_n)-E(\psi(Y_i-\theta(\mathbf{w})-c\delta_n)\,|\,\mathbf{X}_i)$. By (3.4)–(3.7), there is a positive constant κ such that for $n \gg 1$,

$$\begin{aligned} &P\left(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_{\mathbf{w}}}(\hat{\theta}_n(\mathbf{x})-\theta(\mathbf{w})) \geqslant c\delta_n\right) \\ &\quad \leqslant P\left(\bigcup_{C_n}\left\{\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \geqslant (cc_6-c_3)\,\delta_n\right\}\right)+o(1) \\ &\quad \leqslant n^{\kappa}\max_{C_n} P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \geqslant (cc_6-c_3)\,\delta_n\right)+o(1). \end{aligned} \tag{3.8}$$

Set $p_n=p_n(\mathbf{w})=P(\|\mathbf{X}_i-\mathbf{w}\| \leqslant \delta_n+\lambda_n\sqrt{d})$. Then $p_n \sim \delta_n^d$. Note that $\sum_{\bar{I}_n} Z_i=\sum_i K_iZ_i$ and $E(K_iZ_i)=0$. By Lemma 6 of Truong and Stone [18], $\mathrm{Var}(\sum_i K_iZ_i)=O(n\delta_n^d)$. It follows from $\alpha(n)=O(\rho^n)$ and a double application of Lemma 8 of Truong and Stone [18] that there are positive constants c_7 and c_8 such that

$$\begin{aligned} P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \geqslant (cc_6-c_3)\,\delta_n\right) &\leqslant P\left(\bar{N}_n \leqslant \frac{1}{2}np_n\right) \\ &\quad + P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \geqslant cc_6\delta_n;\, \bar{N}_n \geqslant \frac{1}{2}np_n\right) \\ &\leqslant \exp(-c_7n\delta_n^d/n^{2\gamma})+\exp(-c^2c_8n\delta_n^{d+2}) \\ &\quad + O\left(n\rho^{[n^\gamma]}\left(\frac{1}{\delta_n^{2d}}+\frac{1}{\delta_n^d\log n}\right)\right) \quad \text{for } \mathbf{w}\in C_n. \end{aligned}$$

(The first term on the right is bounded by applying Lemma 8 of Truong and Stone [18] with $\gamma<r$, $\sigma^2 \sim \delta_n^d$, and $R^2=M_1^{-1}n\delta_n^{2d}$. The second term is bounded by using that Lemma again with $R^2=M_1^{-1}(cc_6-c_3)^2\,n\delta_n^{2d+2}$.)

Now it follows from $n\delta_n^{d+2} \sim \log n$ that there is a positive constant c such that

$$n^{\kappa} \max_{C_n} P\left(\bar{N}_n^{-1} \sum_{\bar{I}_n} Z_i \geqslant (cc_6 - c_3)\,\delta_n\right) \to 0 \qquad \text{as} \quad n \to \infty. \tag{3.9}$$

Hence by (3.8) and (3.9),

$$\lim_n P\left(\max_{C_n} \sup_{\mathbf{x} \in Q_{\mathbf{w}}} (\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})) \geqslant c\delta_n\right) = 0 \qquad \text{for} \quad c > 0. \tag{3.10}$$

Similarly,

$$\lim_n P\left(\max_{C_n} \sup_{\mathbf{x} \in Q_{\mathbf{w}}} (\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})) \leqslant -c\delta_n\right) = 0 \qquad \text{for} \quad c > 0. \tag{3.11}$$

It follows from (3.10) and (3.11) that (3.3) is valid. This completes the proof of Theorem 2.

Proof of Theorem 3. By Condition 1, $\theta(\cdot)$ is bounded on C (compact). Thus it follows from Theorem 2 that there is a positive constant $T \geqslant 1$ such that $\|\theta(\cdot)\| \leqslant T$ and

$$\lim_n P(\Phi_n) = 1, \tag{3.12}$$

where $\Phi_n = \{\|\hat{\theta}_n(\cdot)\|_\infty \leqslant T\}$. For $i = 1, ..., n$, set

$$Y_i' = \begin{cases} -(T + M_5) & \text{if} \quad Y_i < -(T + M_5); \\ Y_i & \text{if} \quad |Y_i| \leqslant T + M_5; \\ T + M_5 & \text{if} \quad Y_i > T + M_5. \end{cases}$$

Let $\bar{\theta}_n(\mathbf{x})$ denote the solution of the equation

$$\frac{1}{N_n(\mathbf{x})} \sum_{I_n(\mathbf{x})} \psi(Y_i' - \bar{\theta}_n(\mathbf{x})) = 0.$$

Note that $\bar{\theta}_n(\mathbf{x}) = \hat{\theta}_n(\mathbf{x})$ for $\mathbf{x} \in C$ except on Φ_n^c. Thus by (3.12), in order to prove the theorem, it is sufficient to show that

$$\lim_{c \to \infty} \lim_n P(\|\bar{\theta}_n - \theta\|_q \geqslant cn^{-r}) = 0. \tag{3.13}$$

To verify (3.13), we may assume that $C = [-\frac{1}{2}, \frac{1}{2}]^d \subset U$. According to $\alpha(n) = O(\rho^n)$, there is a positive constant c_9 such that

$$\lim_n P(\Omega_n) = 1, \tag{3.14}$$

where $\Omega_n = \{N_n(\mathbf{x}) \geqslant c_9 n \delta_n^d \text{ for } \mathbf{x} \in C\}$. (See Lemma 7 of Truong and Stone [18]).

Write $P_{\Omega_n}(\cdot) = P(\cdot; \Omega_n) = P(\cdot \cap \Omega_n)$ and $E_{\Omega_n}(W) = E(W 1_{\Omega_n})$, where W is a real-valued random variable. By (3.14), there is a sequence of positive numbers $\varepsilon_n \to 0$ such that

$$\begin{aligned} P & \left(\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q \, d\mathbf{x} \geqslant (cn^{-r})^q \right) \\ & \leqslant P \left(\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q \, d\mathbf{x} \geqslant (cn^{-r})^q; \Omega_n \right) + \varepsilon_n \\ & \leqslant \frac{E_{\Omega_n}[\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q \, d\mathbf{x}]}{(cn^{-r})^q} + \varepsilon_n. \end{aligned} \tag{3.15}$$

Set $U_n(\mathbf{x}) = |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|$. By Condition 1, $U_n(\mathbf{x})$ is bounded by $2T$ for $\mathbf{x} \in C$. Thus there is a positive constant c_{10} such that

$$\begin{aligned} E_{\Omega_n}[U_n^q(\mathbf{x})] &= \int_0^{2T} q t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t) \, dt \\ &= \int_0^{2M_0 \delta_n} q t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t) \, dt \\ &\quad + \int_{2M_0 \delta_n}^{2T} q t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t) \, dt \\ &\leqslant c_{10} \delta_n^q + \int_{2M_0 \delta_n}^{2T} q t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t) \, dt. \end{aligned} \tag{3.16}$$

By Conditions 1–3, 4(iii), and 5(ii), there is a positive number c_{11} such that

$$\int_{2M_0 \delta_n}^{2T} q t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t) \, dt \leqslant c_{11} \, \delta_n^q \qquad \text{for} \quad \mathbf{x} \in C. \tag{3.17}$$

(The proof of (3.17) will be given shortly.) It follows from (3.16) and (3.17) that there is a positive constant c_{12} such that

$$E_{\Omega_n}[U_n^q(\mathbf{x})] \leqslant c_{12} \, \delta_n^q.$$

Thus there is a positive constant c_{13} such that

$$E_{\Omega_n} \left[\int_C U_n^q(\mathbf{x}) \, d\mathbf{x} \right] = \int_C E_{\Omega_n}[U_n^q(\mathbf{x})] \, d\mathbf{x} \leqslant c_{13} \delta_n^q. \tag{3.18}$$

The conclusion of Theorem 3 follows from (3.15) and (3.18).

Finally, (3.17) will be proven. Let $\mathbf{x} \in C$ be fixed. By Conditions 4(ii), (iii),

$$-N_n^{-1} \sum_{I_n} E[\psi(Y_i' - \theta(\mathbf{x}) - t) \mid \mathbf{X}_i] \geqslant M_2 t, \qquad 2M_0 \delta_n \leqslant t \leqslant M_2^{-1};$$

$$-N_n^{-1} \sum_{I_n} E[\psi(Y_i' - \theta(\mathbf{x}) - t) \mid \mathbf{X}_i] \geqslant M_3, \qquad M_2^{-1} \leqslant t \leqslant 2T.$$

Set

$$Z_i = \psi(Y_i' - \theta(\mathbf{x}) - t) - E[\psi(Y_i' - \theta(\mathbf{x}) - t) \mid \mathbf{X}_i].$$

Then

$$\begin{aligned} &P_{\Omega_n}(\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) > t) \\ &\quad \leqslant P_{\Omega_n}\left(N_n^{-1} \sum_{I_n} \psi(Y_i' - \theta(\mathbf{x}) - t) \geqslant 0\right) \\ &\quad \leqslant P_{\Omega_n}\left(N_n^{-1} \sum_{I_n} Z_i \geqslant -N_n^{-1} \sum_{I_n} E[\psi(Y_i' - \theta(\mathbf{x}) - t) \mid \mathbf{X}_i]\right) \\ &\quad \leqslant P\left(\sum_{I_n} Z_i \geqslant c_9 M_2 tn\, \delta_n^d\right) \qquad \text{for} \quad 2M_0 \delta_n \leqslant t < M_2^{-1}. \end{aligned} \tag{3.19}$$

Similarly,

$$P_{\Omega_n}(\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) > t) \leqslant P\left(\sum_{I_n} Z_i \geqslant c_9 M_3 n\, \delta_n^d\right) \qquad \text{for} \quad M_2^{-1} \leqslant t \leqslant 2T. \tag{3.20}$$

Set $K_i = K_i(\mathbf{x}) = 1_{\{\|\mathbf{X}_i - \mathbf{x}\| \leqslant \delta_n\}}$. Note that $\sum_{I_n} Z_i = \sum_i K_i Z_i$, $E(K_i Z_i) = 0$, $E\,|K_i Z_i| = O(\delta_n^d)$, and $E\,|K_i Z_i K_j Z_j| = O(\delta_n^{2d})$. Since Z_i is bounded, it follows from Lemma 9 of Truong and Stone [18],

$$E\left|\sum_{I_n} Z_i\right|^{2k} = E\left|\sum_i K_i Z_i\right|^{2k} = O(n\delta_n^d)^k \qquad \text{for} \quad k = 1, 2, 3, \ldots.$$

Consequently, by Markov's inequality,

$$\begin{aligned} P\left(\sum_{I_n} Z_i \geqslant c_9 M_2 tn\, \delta_n^d\right) &\leqslant \frac{E\,|\sum_{I_n} Z_i|^{2k}}{(c_9 M_2 tn\, \delta_n^d)^{2k}} \\ &= \frac{O(n\delta_n^d)^k}{(c_9 M_2 tn\, \delta_n^d)^{2k}} \qquad \text{for} \quad 2M_0 \delta_n \leqslant t < M_2^{-1} \end{aligned} \tag{3.21}$$

and

$$P\left(\sum_{I_n} Z_i \geqslant c_9 M_3 n\, \delta_n^d\right) \leqslant \frac{O(n\delta_n^d)^k}{(c_9 M_3 n\, \delta_n^d)^{2k}} \qquad \text{for} \quad M_2^{-1} \leqslant t < 2T.$$

By (3.19)–(3.21), there is a positive constant c_{14} such that (note that $n\delta_n^d \sim \delta_n^{-2}$)

$$\begin{aligned} P_{\Omega_n}(\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) > t) &\leqslant c_{14} t^{-2k}\, \delta_n^{2k} \qquad \text{for} \quad 2M_0\, \delta_n \leqslant t < M_2^{-1}, \\ &\leqslant c_{14}\, \delta_n^{2k} \qquad \text{for} \quad M_2^{-1} \leqslant t \leqslant 2T. \end{aligned} \tag{3.22}$$

Similarly,

$$\begin{aligned} P_{\Omega_n}(\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) < -t) &\leqslant c_{14} t^{-2k}\, \delta_n^{2k} \qquad \text{for} \quad 2M_0 \delta_n \leqslant t < M_2^{-1}, \\ &\leqslant c_{14}\, \delta_n^{2k} \qquad \text{for} \quad M_2^{-1} \leqslant t \leqslant 2T. \end{aligned} \tag{3.23}$$

Note that c_{14} does not depend on $\mathbf{x}$. It now follows from (3.22) and (3.23) by choosing $k > q/2$ that

$$\int_{2M_0\delta_n}^{2T} t^{q-1} P_{\Omega_n}(U_n(\mathbf{x}) > t)\, dt \leqslant 2\delta_n^{2k} c_{14} \int_{2M_0\delta_n}^{M_2^{-1}} t^{q-2k-1}\, dt + O(1)\, \delta_n^{2k} = O(\delta_n^q).$$

ACKNOWLEDGMENTS

I thank R. Beran and C. Stone for their generous guidance and suggestions.

REFERENCES

[1] BOENTE, G., AND FRAIMAN, R. (1989). Robust nonparametric regression estimation. *J. Multivariate Anal.* **29** 180–198.

[2] BOENTE, G., AND FRAIMAN, R. (1989). Robust nonparametric regression estimation for dependent observations. *Ann. Statist.* **17** 1242–1256.

[3] BOENTE, G., AND FRAIMAN, R. (1991). Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. *Internat. Statist. Rev.* **59** 355–372.

[4] COLLOMB, B. G., AND HÄRDLE, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Process. Appl.* **23** 77–89.

[5] FAN, Y. (1990). Consistent nonparametric multiple regression for dependent heterogeneous processes: The fixed design case. *J. Multivariate Anal.* **33** 72–88.

[6] HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J., AND STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[7] HÄRDLE, W. (1984). Robus regression function estimation. *J. Multivariate Anal.* **14** 169–180.

[8] HÄRDLE, W., AND GASSER, T. (1984). Robust non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.

[9] HÄRDLE, W., AND LUCKHAUS, S. (1984). Uniform consistency of a class of regression function estimators. *Ann. Statist.* **12** 612–623.

[10] HÄRDLE, W., AND TSYBAKOV, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16** 120–135.

[11] HÄRDLE, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179.

[12] HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

[13] NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

[14] ROBINSON, P. M. (1984). Robust nonparametric regression. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Härdle, and D. Martin, Eds.). Lecture Notes in Statistics, Vol. 26. Springer-Verlag, New York.

[15] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43–47.

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

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

[18] TRUONG, K. Y., AND STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20**. To appear.

[19] WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.