JOURNAL OF MULTIVARIATE ANALYSIS 39, 246–269 (1991)

Global Nonparametric Estimation of Conditional Quantile Functions and Their Derivatives*

PROBAL CHAUDHURI

University of Wisconsin-Madison and Indian Statistical Institute, Calcutta, India

Communicated by C. R. Rao

Let (X, Y) be a random vector such that X is d-dimensional, Y is real valued, and $\theta(X)$ is the conditional αth quantile of Y given X, where α is a fixed number such that $0<\alpha<1$. Assume that θ is a smooth function with order of smoothness $p>0$, and set $r=(p-m)/(2p+d)$, where m is a nonnegative integer smaller than p. Let $T(\theta)$ denote a derivative of θ of order m. It is proved that there exists estimate $\hat{T}_n$ of $T(\theta)$, based on a set of i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, that achieves the optimal nonparametric rate of convergence n^{-r} in L_q-norms $(1 \leqslant q < \infty)$ restricted to compacts under appropriate regularity conditions. Further, it has been shown that there exists estimate $\hat{T}_n$ of $T(\theta)$ that achieves the optimal rate $(n/\log n)^{-r}$ in L_∞-norm restricted to compacts.

1. INTRODUCTION

Consider a regression setup with a d-dimensional random regressor X and a real valued response variable Y, which satisfy $Y=\theta(X)+\varepsilon$. Here, ε is an unobservable random variable assumed to be independent of X, and θ is an unknown function to be estimated from a set of i.i.d. observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. In usual regression problems, ε is assumed to have mean 0, and that makes $\theta(X)$ the conditional mean of Y given X. On the other hand, in quantile regression [21, 25], the αth quantile (α is a fixed number such that $0<\alpha<1$) of ε is assumed to be at 0, and $\theta(X)$ becomes the conditional αth quantile of Y given X. Of particular importance is the case when $\alpha=\frac{1}{2}$ in which case $\theta(X)$ becomes the conditional median of Y given X. It is well known that the estimates constructed by using the

Received January 24, 1990; revised April 10, 1991.

AMS 1980 subject classifications: primary 62G05, 62G20; secondary 62G35.

Key words and phrases: regression quantiles, nonparametric estimates, bin smoothers, optimal rates of convergence.

* The research presented here was supported in part by the NSF Grant DMS 86-00409 and a Wisconsin Alumni Research Foundation Grant.

0047-259X/91 $3.00

method of least squares in usual regression problems are not very robust and are very badly affected by the presence of outlying observations in the data (see [20, 22, 23, 25, 33] for detailed discussions). Also, the least squares estimates often turn out to be quite inefficient when the random error ε follows a non-normal probability distribution. One of the major motivations behind looking at the regression quantiles is to make the usual regression robust and to extend the techniques of L-estimation to the regression context from location problems with univariate data (see [4, 31, 41, 42]).

In the case of a finite dimensional linear parametric model for θ, it is assumed that θ is a member of a fixed finite dimensional vector space of real valued functions. In that case, the problem of estimating θ boils down to the problem of estimating a finite dimensional Euclidean parameter. The $\sqrt{n}$-consistency and the asymptotic normality of the quantile regression estimates in linear parametric models have already been established by several people [25, 31, 27, 41, 42]. Others [1, 10] have investigated the almost sure behavior and strong consistency of minimum L_1-norm and other related estimates in the context of linear regression. Stone in his 1982 paper on "Optimal Global Rates of Convergence for Nonparametric Regression" raised the question whether the optimal nonparametric rates of convergence are achievable in the estimation of the conditional median function of Y given X. He was motivated by the idea of making the nonparametric regression robust. In this paper, we will provide an affirmative answer to the above question (see also [40]) by constructing nonparametric estimates for the conditional quantile function (which is assumed to be suitably smooth) and its derivatives and showing that, under mild regularity conditions, such estimates achieve the optimal nonparametric rates of convergence in L_q-norms ($1 \leqslant q \leqslant \infty$) restricted to compact sets. Stone [36] (see also [35]) obtained the lower bounds for the global rates of convergence for nonparametric estimates of a regression function and its derivatives. It is clear from his work that these lower bounds apply also to the nonparametric estimates of conditional quantile functions. In fact, the lower bounds are determined by the local behavior of the Kullback–Leibler divergence for the family of conditional distributions [43] or, more broadly speaking, by the "geometry" of the problem [14]. Results presented in this paper do not require any moment condition on the random error ε_i, whereas Stone [36] did use such a condition. It implies that the estimates discussed here will perform well even if the ε_i's have a distribution with heavy tail(s) and they will be resistant to outliers among $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ (see also [3, 11, 12, 16, 17, 18, 19, 24, 28, 29, 39]).

The estimates discussed in this paper are polynomial smoothers constructed following the idea of bin smoothers or the histogram type estimates. After obtaining the data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, we first split the compact set, on which the estimate for the conditional quantile function is to be constructed, into an appropriate number of bins of equal size. Then a polynomial in X of suitable degree (depending on the assumed amount of smoothness of the function being estimated) is fitted to the data points that fall inside a bin using the loss function $H_\alpha(t) = |t| + (2\alpha - 1)\, t$ (see [25, 34, 15]) with each bin being treated separately. Note that in the case of estimating the location parameter based on univariate data, the estimate constructed by minimizing the loss H_α is nothing but the sample αth quantile (just as the estimate constructed by minimizing the squared error loss is the sample mean, and the estimate constructed by minimizing the absolute error loss is the sample median). The number of bins chosen will depend on the sample size, the dimension of X, and the assumed amount of smoothness of the unknown conditional quantile function. As it is typical in nonparametric function estimation, this involves a subtle and complex form of "bias versus variance game." One of the main objectives here is to gain theoretical insights into the asymptotic behavior of nonparametric estimates of regression quantiles constructed through piecewise polynomial fits. The piecewise polynomial nature of the estimates makes them discontinuous at the boundaries of the chosen bins. One resolution of this problem is to smooth out these estimates by weighted averaging with appropriately chosen weights without affecting their optimal asymptotic properties. Such estimates have been constructed and thoroughly discussed in Chaudhuri, Huang, Loh, and Yao [9] in the context of regression function estimation via recursive partitioning.

2. Description of the Estimates for Conditional Quantile Function and Its Derivatives

Suppose that (X_1, Y_1), $(X_2, Y_2), \ldots, (X_n, Y_n)$ are i.i.d. observations, where the Y_i's are real valued and the X_i's are d-dimensional. We assume that $Y_i = \theta(X_i) + \varepsilon_i$, where ε_i is independent of X_i and has a distribution with 0 as the αth quantile. Note that the assumption that ε_i has the αth quantile at 0 is only a centering assumption to fix the idea and simplify notations. One can always make 0 to be the αth quantile of the common distribution of the ε_i's by addition of a common constant to the ε_i's and subtraction of the same constant from θ, if necessary.

For $u = (u_1, \ldots, u_d)$, a d-dimensional vector of nonnegative integers, let D^u denote the differential operator $\partial^{[u]}/(\partial x_1^{u_1} \cdots \partial x_d^{u_d})$, where $[u] = u_1 + \cdots + u_d$.

Let $\mathbf{V}$ be a nonempty open subset of R^d and $|\ |$ denote the usual Euclidean norm. Let C be a fixed compact subset of $\mathbf{V}$. For a fixed nonnegative integer k and real numbers c and γ such that $c>0$ and $0<\gamma\leqslant 1$, let $\Theta(c,k,\gamma)$ be the collection of all functions θ on $\mathbf{V}$ satisfying

(i) $D^u\theta(x)$ exists and is continuous in x for all $x\in\mathbf{V}$ and $[u]\leqslant k$,

and

(ii) $|D^u\theta(x_1)-D^u\theta(x_2)|\leqslant c|x_1-x_2|^\gamma$ for all $x_1,x_2\in\mathbf{V}$ and $[u]=k$.

So, the functions in $\Theta(c,k,\gamma)$ are continuously differentiable up to order k on $\mathbf{V}$, and their kth derivatives are uniformly Holder continuous with exponent γ on $\mathbf{V}$. We refer to $p=k+\gamma$ as the order of smoothness of the functions in $\Theta(c,k,\gamma)$. We assume that the conditional quantile function of Y given X is a member of the family $\Theta(c,k,\gamma)$ for some fixed c, k, and γ (see [35, 36, 13]). Let $T(\theta)=D^u\theta$, where $[u]=m\leqslant k$. An estimate $\hat{T}_n(x)$ of $T(\theta)(x)$ will now be described for $x\in C$.

From now on, assume that $C=[-0.5,0.5]^d$. Note that there is no loss of generality in assuming $C=[-0.5,0.5]^d$. Any compact subset of $\mathbf{V}$ can be covered by a union of finitely many d-dimensional rectangles (i.e., d-fold products of compact intervals of real line) in $\mathbf{V}$ such that any two of these rectangles will either be disjoint or intersect only at their boundaries. The purpose of assuming $C=[-0.5,0.5]^d$ is to simplify notations and thereby making the proofs more readable. For general C, one only needs to do some trivial modifications of the arguments given here. Let J_n be a sequence of positive integers, which tend to ∞ as n tends to ∞, and set $\delta_n=1/J_n$. The choice of J_n will be described later. Split the cube $C=[-0.5,0.5]^d$ into J_n^d smaller subcubes of equal size each with side length δ_n and having the boundaries parallel to the standard coordinate hyperplanes in R^d. Let $C_{n,r}$, where $1\leqslant r\leqslant J_n^d$, be a typical such subcube with center at $x_{n,r}$. Define random sets $S_{n,r}$ in terms of the data as $S_{n,r}=\{i\colon 1\leqslant i\leqslant n,\ X_i\in C_{n,r}\}$ and set $N_{n,r}=\#(S_{n,r})$.

Let A be the set of all d-dimensional vectors u with nonnegative integral components such that $[u]\leqslant k$ and set $s(A)=\#(A)$. Let $\beta=(\beta_u)_{u\in A}$ be a vector of dimension $s(A)$. Also, given $x_1,x_2\in R^d$, define $P_n(\beta,x_1,x_2)$ to be the polynomial $\sum_{u\in A}\beta_u[(x_1-x_2)/\delta_n]^u$. Here, if $z=(z_1,\dots,z_d)$ is an element of R^d and $u=(u_1,\dots,u_d)$ is a vector of nonnegative integral components, we set $z^u=\prod_{i=1}^d z_i^{u_i}$ with the convention that $0^0=1$. Let $\hat{\beta}_{n,r}$ be a minimizer of

$$\sum_{i\in S_{n,r}} H_\alpha[Y_i-P_n(\beta,X_i,x_{n,r})],$$

where $H_\alpha(t)=|t|+(2\alpha-1)\,t$. Since $0<\alpha<1$, $H_\alpha(t)$ tends to ∞ as $|t|$ tends to ∞. Therefore, the above minimization problem always has a solution. In view of Theorem 3.1 below, one does not need to worry about the uniqueness of $\hat{\beta}_{n,r}$ as long as one is interested in asymptotic results and

$k \geqslant 1$. On the other hand, when $k=0$ the polynomial $P_n(\beta, x_1, x_2)$ reduces to a constant and β becomes a real number. In this case, as n increases, the solution set for the above minimization problem will eventually turn out to be a compact interval and $\hat{\beta}_{n,r}$ can be defined to be the right or the left end point of that interval, so that $\hat{\beta}_{n,r}$ becomes a local quantile of the Y values for which the X values fall in the cube $C_{n,r}$. The existence and the uniqueness of regression quantile estimates in linear parametric models have been discussed in Koenker and Bassett [25]. Algorithms to compute regression quantiles exploiting simplex method have been developed and studied in Wellington and Narula [44], Narula and Wellington [30], Koenker and D'Orey [26] (see also [5]).

For x in the interior of $C_{n,r}$, we will set $\hat{T}_n(x) = D^u P_n(\hat{\beta}_{n,r}, x, x_{n,r})$. Here the differential operator D^u acts by differentiating with respect to the second argument x. For a point, which lies on the boundary of several subcubes, one can define $\hat{T}_n$ as the simple average of the different values that arise from different subcubes having the point in question as a common boundary point.

3. Main Results

From now on, for two sequences $\{a_n\}$ and $\{b_n\}$ of positive real numbers, the notation $a_n \sim b_n$ will mean that a_n/b_n tends to 1 as n tends to ∞. Given a real valued function g on $\mathbf{V}$, define $\|g\|_\infty = \sup_{x \in C} |g(x)|$. Also, for $1 \leqslant q < \infty$, we set $\|g\|_q = [\int_C |g(x)|^q \, dx]^{1/q}$.

In order to derive asymptotic results about the estimate $\hat{T}_n$, we need to impose some conditions on the distributions of X_i and ε_i.

Condition 3.1. The distribution of X_i is absolutely continuous on $\mathbf{V}$ with a density w that is continuous and positive on $\mathbf{V}$.

Condition 3.2. ε_i has a density f that is continuous and positive in an open neighborhood around 0.

THEOREM 3.1. *Let J_n be a sequence of positive integers such that either $J_n \sim n^\lambda$ or $J_n \sim (n/\log n)^\lambda$, where $0 < \lambda < 1/d$. Also, assume that ε_i has a continuous distribution and Condition* 3.1 *is satisfied. Then, there exists event C_n defined in terms of $X_1, ..., X_n$ such that*

(i) $P(\liminf C_n) = 1$.

(ii) *Let E_n be the event that $\hat{\beta}_{n,r}$ defined in the construction of $\hat{T}_n$ exists uniquely for each r with $1 \leqslant r \leqslant J_n^d$. Then, if $k \geqslant 1$, there is a positive integer N_0 depending on k, d, and λ such that for $n \geqslant N_0$, the conditional probability of E_n given that C_n has occurred is* 1.

THEOREM 3.2. *Set* $\lambda = 1/(2p+d)$ *and* $r = (p-m)/(2p+d)$, *and let* $J_n \sim (n/\log n)^{\lambda}$. *Then, under Conditions* 3.1 *and* 3.2, *we have the following.*

(i) *There exists a constant* $K_1 > 0$ (K_1 *depends on* f, w, *and* c) *such that*

$$\lim_{n} \sup_{\theta \in (c,k,\gamma)} P_\theta[\|\hat{T}_n - T(\theta)\|_\infty > K_1(n/\log n)^{-r}] = 0.$$

(ii) *For* $\theta \in \Theta(c, k, \gamma)$, $\|\hat{T}_n - T(\theta)\|_\infty$ *is almost surely* $O[(n/\log n)^{-r}]$ *as n tends to* ∞.

THEOREM 3.3. *Let* λ *and* r *be as in Theorem* 3.2 *and suppose that* $J_n \sim n^{\lambda}$. *Then, under Conditions* 3.1 *and* 3.2, *we have the following.*

(i) *There exists a constant* $K_2 > 0$ (K_2 *depends on* f, w, *and* c) *such that*

$$\lim_{n} \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta[\|\hat{T}_n - T(\theta)\|_q > K_2 n^{-r}] = 0 \qquad (1 \leqslant q < \infty).$$

(ii) *For* $\theta \in \Theta(c, k, \gamma)$, $\|\hat{T}_n - T(\theta)\|_\infty$ *is almost surely* $O(n^{-r}\sqrt{\log n})$ *as n tends to* ∞.

4. DISCUSSION

(1) The rates of convergence obtained in Theorem 3.2 and in (i) in Theorem 3.3 are optimal by Stone [36]. In the special case when $\alpha = \frac{1}{2}$, these two theorems answer question 4 raised by Stone [36] regarding the attainability of the optimal global rates of convergence in the non-parametric estimation of conditional median. Troung [40] provided a partial answer to this question considering the special case $p = 1$ and using local median type estimators. Asymptotic behavior of point-wise estimates of a conditional quantile function and its derivatives are explored in Chaudhuri [6].

(2) Condition 3.2 amounts to assuming that the distribution of the random error ε_i has a density which is continuous and positive in a neighborhood of its αth quantile. This is satisfied by all classical examples of probability density functions. As it will be clear from the proof of Theorem 3.1, one needs Condition 3.1 (see Stone [36]) to ensure sufficiently many observations (at least asymptotically) in each of the cubes $C_{n,r}$ defined in the construction of $\hat{T}_n$. This is an important requirement for $\hat{T}_n$ to achieve the optimal global rates.

(3) In order to understand some intuitive ideas behind the proofs of the theorems stated in Section 3, consider Theorem 3.3, where J_n has been chosen to be of order $n^{1/(2p+d)}$. Any function θ in $\Theta(c, k, \gamma)$ can be expanded in $C_{n,r}$ using Taylor series around $x_{n,r}$ up to terms involving the kth order derivatives, and the remainder term in that expansion will be of order $n^{-p/(2p+d)}$. The construction of $\hat{T}_n$ uses approximation of θ by a polynomial in each of the cubes $C_{n,r}$, and the remainder term mentioned above may be viewed as a cause of error ("bias") in this approximation procedure. On the other hand, under Condition 3.1, one would expect that $N_{n,r}$ defined in the construction of $\hat{T}_n$ will be of order $n^{2p/(2p+d)}$ (see [36]). Hence, in view of the way $\hat{\beta}_{n,r}$ is defined in the construction of $\hat{T}_n$, it is natural to expect that under appropriate conditions, this vector will estimate the coefficients (scaled by δ_n) in the Taylor expansion (in the cube $C_{n,r}$ around the point $x_{n,r}$) of θ with an error ("variance": caused by the random noise ε present in the data) of order $n^{-p/(2p+d)}$. So, one can expect that $\hat{T}_n$ will achieve the optimal rate of convergence globally. The idea used here is close to that of Stone [35, 36]. However, the "bias-variance trade-off" is more complicated here than in his work on non-parametric regression. A major source of complication is the non-linear nature of the estimates obtained by using the loss function $H_\alpha(t)$ instead of the method of least squares. These ideas will be more clear from the proofs in Section 5.

(4) A serious problem in nonparametric function estimation is data sparseness in high dimension (the "curse of dimensionality"). This is reflected in the slow rates of convergence for nonparametric estimates when the dimension d of the regressor X is high. Stone's work on "dimensionality reduction principle" [37, 38] motivates the following question. Suppose that the conditional quantile function $\theta(x)$ of Y given $X=x$ can be written as $\theta(x)=\theta_1(x_1)+\cdots+\theta_d(x_d)$, where $x=(x_1, ..., x_d)\in R^d$ and each θ_i $(1\leqslant i\leqslant d)$ is a real valued smooth function of a single real variable. Then, is it possible to construct estimates of θ or its derivatives that will achieve the same rates of convergence as the optimal non-parametric rates of convergence when $d=1$?

5. Proofs of the Theorems in Section 3

Before going into the proofs of the theorems in Section 3 we need to introduce some notations and prove a few preliminary results. Let $A(d, k)$ be the set of all d-dimensional vectors u with nonnegative integral components such that $0<[u]\leqslant k$. Note that $A(d, k)=A-\{(0, 0, ..., 0)\}$, where A is as defined in the construction of $\hat{T}_n$ in Section 2. Let $s(d, k)=\#A(d, k)=\#(A)-1=s(A)-1$. For $x\in R^d$, write $x(d, k)$ for the $s(d, k)$-dimensional vector $(x^u)_{u\in A(d,k)}$.

PROPOSITION 5.1 (An extended version of the Riemann–Lebesgue lemma). *Let g be a complex-value function on R^d such that $\int_{R^d} |g(x)|\, dx < \infty$ and $\phi = (\phi_u)_{u \in A(d,k)}$ be an element of $R^{s(d,k)}$. Then*

$$\lim_{|\phi| \to \infty} \int_{R^d} \exp\{i\langle \phi, x(d, k)\rangle\}\, g(x)\, dx = 0.$$

Here $i^2 = -1$, $|\ |$ is the usual Euclidean norm, and $\langle\ ,\ \rangle$ is the usual Euclidean inner product.

Proof. First note that it is enough to prove the assertion when g is a probability density function on R^d. The proof will be given by induction on d and will make use of the standard Riemann–Lebesgue lemma. Assume that $d = 1$ and in this case, $x(d, k) = x(1, k) = (x, x^2, ..., x^k)$, $A(d, k) = A(1, k) = \{(1), (2), ..., (k)\}$ and $s(d, k) = s(1, k) = k$. Also, for $\phi \in R^k$, $\int_R \exp\{i\langle \phi, x(1, k)\rangle\}\, g(x)\, dx$ is the characteristic function of the random vector $(X, X^2, ..., X^k)$, where X is a random variable with probability density function g. Hence, if $k = 1$, the assertion in the proposition actually reduces to the ordinary Riemann–Lebesgue lemma. For $k > 1$, let $X_1, X_2, ..., X_k$ be k i.i.d. random variables each with probability density function g and form the k-dimensional random vector $Y = (\sum_{i=1}^k X_i, \sum_{i=1}^k X_i^2, ..., \sum_{i=1}^k X_i^k)$. Then, since the transformation from R^k into R^k that maps $(x_1, x_2, ..., x_k)$ into $(\sum_{i=1}^k x_i, \sum_{i=1}^k x_i^2, ..., \sum_{i=1}^k x_i^k)$ is smooth and has a Jacobian that vanishes only on a set of Lebesgue measure 0, the random vector Y must have a probability density h (say). Hence, again by the standard Riemann–Lebesgue lemma, we have that

$$\lim_{|\phi| \to \infty} \int_{R^k} \exp\{\langle i\phi, y\rangle\}\, h(y)\, dy = 0.$$

The proof for the case $d = 1$ is now complete by noting that

$$\int_{R^k} \exp\{\langle i\phi, y\rangle\}\, h(y)\, dy = \left[\int_R \exp\{\langle i\phi, x(1, k)\rangle\}\, g(x)\, dx\right]^k.$$

Assume now that the assertion in the proposition is true for dimension $= d - 1$. Define g_d, a function on R, as

$$g_d(x_d) = \int_{R^{d-1}} g(x_1, x_2, ..., x_{d-1}, x_d) \prod_{l=1}^{d-1} dx_l.$$

So, g_d is the marginal probability density of the dth coordinate obtained from the joint probability density g. Set $g^* = g/g_d$ on the set in R^d on which the dth coordinate is such that $g_d \neq 0$. Then g^* is the conditional density of the first $d-1$ coordinates after fixing the dth coordinate. Define $A^{**}(d,k)$ to be the set $\{u_1, u_2, ..., u_k\}$, where u_j is the d-dimensional vector $(0, 0, ..., 0, j)$. Then consider the set $A^*(d,k) = A(d,k) - A^{**}(d,k)$, and let $s^*(d,k) = \# A^*(d,k) = s(d,k) - k$. For $\phi \in R^{s(d,k)}$ and $x \in R^d$, define $\phi^* = (\phi_u)_{u \in A^*(d,k)}$ and $x^*(d,k) = (x^u)_{u \in A^*(d,k)}$. So, ϕ^* and $x^*(d,k)$ are elements of $R^{s^*(d,k)}$. With these notations in hand, we can write

$$\int_{R^d} \exp\{i\langle \phi, x(d,k)\rangle\}\, g(x)\, dx$$

$$= \int_R \left[\exp\left\{i \sum_{j=1}^{k} x_d^j \phi_{u_j}\right\}\right] \Psi(\phi^*, x_d)\, g_d(x_d)\, dx_d, \tag{5.1}$$

where

$$\Psi(\phi^*, x_d) = \int_{R^{d-1}} \exp\{i\langle \phi^*, x^*(d,k)\rangle\}\, g^*(x_1, x_2, ..., x_{d-1}, x_d) \prod_{l=1}^{d-1} dx_l. \tag{5.2}$$

Suppose now that $\{\phi_n\}$ is a sequence in $R^{s(d,k)}$ such that $|\phi_n|$ tends to ∞ as n tends to ∞. There are two cases that may occur.

Case 1. Assume that $|\phi_n^*|$ tends to ∞ as n tends to ∞. Then, from (5.2) and using the induction hypothesis, $\Psi(\phi_n^*, x_d)$ must tend to 0 as n tends to ∞ for all non-zero values of x_d for which $g_d(x_d)$ is positive. Hence, an application of Lebesgue's dominated convergence theorem to (5.1) gives

$$\lim_{n\to\infty} \int_{R^d} \exp\{i\langle \phi_n, x(d,k)\rangle\}\, g(x)\, dx = 0.$$

Case 2. Assume now that $|\phi_n^*|$ remains bounded as n varies but $|(\phi_{u_1,n}, \phi_{u_2,n}, ..., \phi_{u_k,n})|$ tends to ∞ as n tends to ∞ (here $\phi_{u_j,n}$ is the component of ϕ_n indexed by u_j). In this case, one can extract a subsequence $\{\phi_{n'}\}$ of $\{\phi_n\}$ such that $\phi_{n'}^*$ tends to ϕ_0^*, an element in $R^{s^*(d,k)}$. Hence, an application of the dominated convergence theorem to (5.2) implies that $\Psi(\phi_{n'}^*, x_d)$ converges to $\Psi(\phi_0^*, x_d)$ for all values of x_d for which g_d is positive. Further, since $|\Psi(\phi_0^*, x_d)| \leqslant 1$, in view of the arguments in the case $d=1$, we have that

$$\lim_{n'\to\infty} \int_R \left[\exp\left\{\sum_{j=1}^{k} x_d^j \phi_{u_j,n'}\right\}\right] \Psi(\phi_0^*, x_d)\, g_d(x_d)\, dx_d = 0. \tag{5.3}$$

Equation (5.3) together with (5.1) implies through an application of the dominated convergence theorem that

$$\lim_{n'\to\infty} \int_{R^d} \exp\{i\langle \phi_{n'}, x(d,k)\rangle\}\, g(x)\, dx = 0.$$

This completes the proof of the proposition. ∎

An immediate and useful consequence of this proposition is the fact that whenever X has a density g on R^d, the random vector $X(d, k)$ defined above has a characteristic function that satisfies "Cramer's condition."

For $y \in D_n = [0.5(-1+\delta_n), 0.5(1-\delta_n)]^d$, where $\{\delta_n\}$ is a sequence of positive numbers (smaller than 1) such that δ_n tends to 0 as n tends to ∞, let us define

$$w_{\delta_n}(x, y) = \frac{w(y+\delta_n x)}{\int_{[-0.5, 0.5]^d} w(y+\delta_n x)\, dx},$$

where $x \in [-0.5, 0.5]^d$. Recall that w is the density of X_i in $\mathbf{V}$ as mentioned in Condition 3.1. So, $w_{\delta_n}(x, y)$ is the conditional density of $(X_i - y)\,\delta_n^{-1}$ given that X_i falls in the cube with center y and side length δ_n. Let $u(x)$ be the uniform density on $[-0.5, 0.5]^d$, which is identically 1 on the set $[-0.5, 0.5]^d$ and identically 0 outside this set. In view of the continuity of w assumed in Condition 3.1, w_{δ_n} converges to u as n tends to ∞, and this convergence is uniform in both x and y. This leads to the following fact as an immediate consequence of Proposition 5.1.

Fact 5.1. Under Condition 3.1, for any η such that $0<\eta<1$, there is $M_1>0$ and an integer $N_1>0$ (which may depend on η as well as on w and the sequence $\{\delta_n\}$) such that

$$\sup_{n \geqslant N_1} \sup_{|\phi| \geqslant M_1} \sup_{y \in D_n} \left| \int_{[-0.5, 0.5]^d} \exp\{i\langle \phi, x(d, k)\rangle\}\, w_{\delta_n}(x, y)\, dx \right| < \eta.$$

The following fact will be used in the proofs of Theorem 3.2 and (ii) in Theorem 3.3.

Fact 5.2. Let F be a distribution function on the real line with the property that $F(0)=\alpha$, where $0<\alpha<1$. Assume that F has a density f that satisfies Condition 3.2. For $x \in R^d$, denote by $x(A)$ the $s(A)$-dimensional vector $(x^u)_{u \in A}$, where A is as defined in the construction of $\hat{T}_n$ in Section 2 and $s(A) = \#(A)$. Let $\Delta = (\Delta_u)_{u \in A}$ be a vector in $R^{s(A)}$, and for $0<\delta \leqslant 1$, let $R(\delta, x)$ be a real valued function with the property that there is $M_2>0$ such that $|R(\delta, x)| \leqslant M_2\, \delta$ for all $x \in [-0.5, 0.5]^d$. Define an $s(A)$-dimensional vector valued function $G_{n,y}(\Delta, \delta)$ as

$$G_{n,y}(\Delta, \delta) = \int_{[-0.5, 0.5]^d} [F\{\langle \Delta, x(A)\rangle + R(\delta, x)\} - \alpha]\, x(A)\, w_{\delta_n}(x, y)\, dx.$$

Then, under Condition 3.1, there exist positive constants c_1, M_3, ε_1, ε_2 and a positive integer N_2 such that we have $|G_{n,y}(\Delta, \delta)| \geqslant \min(\varepsilon_1 \text{ and } c_1|\Delta|)$ for all $y \in D_n$, $n \geqslant N_2$, $\delta \leqslant \varepsilon_2$ and $|\Delta| \geqslant M\delta$, where $M \geqslant M_3$.

For the special case when w is the uniform density implying that w_{δ_n} is also the uniform density on $[-0.5, 0.5]^d$, a proof of this fact can be found in Chaudhuri [7, Proposition 6.1]. Since w_{δ_n} converges uniformly to the uniform density as n tends to ∞, the proof of the above fact can be obtained by minor modifications of this special case. Technical details can be found in Chaudhuri [6, 8].

In several places of the proofs of the theorems stated in Section 3, the following simple facts will be used:

Fact 5.3. Let X be a random vector in R^d with a distribution, which is absolutely continuous with respect to the Lebesgue measure, and let $p(X)$ be a nonzero polynomial in X. Then the probability of the event $\{p(X)=0\}$ is 0.

Fact 5.4. Let $\mathbf{X}$ be a random matrix of dimension $m \times m$ where m is a positive integer. Denote by $X^{(i)}$ the ith row of $\mathbf{X}$ $(1 \leqslant i \leqslant m)$. Assume that $X^{(1)}, X^{(2)}, ..., X^{(m)}$ are independent random vectors, and that each $X^{(i)}$ has a distribution with the property that for any fixed vector subspace H of R^m such that $\dim(H) \leqslant m-1$, $P(X^{(i)} \in H)=0$. Then $\mathbf{X}$ has full rank $=m$ with probability 1.

Unlike the problem of minimizing the squared error loss, the solution(s) to the minimization problem in the construction of $\hat{T}_n$ does (do) not have a nice closed form. We will be using Theorems 3.1 and 3.3 in Koenker and Bassett [25] several times in the proofs that follow. These two theorems enable us to exploit some fundamental algebraic properties of the elements in the solution set for the minimization problem that arise in the construction of the estimate $\hat{T}_n$. For the rest of the paper, we will refer to this minimization problem as problem (P).

Suppose that we have a matrix (vector) $\mathbf{X}$ with rows (components) indexed by the elements of a nonempty finite set $\mathbf{S}$ (e.g., a nonempty finite subset of the set of integers). Then, for any nonempty subset $\mathbf{h}$ of $\mathbf{S}$, denote by $\mathbf{X}(\mathbf{h})$ the submatrix (subvector) of $\mathbf{X}$ with rows (components) which are indexed by the elements of $\mathbf{h}$ [25]. Write $X_{n,r}$ for the matrix with $N_{n,r}$ rows and $s(A)=\#(A)$ (A is as defined in the construction of $\hat{T}_n$) columns, where the rows of $X_{n,r}$ are the vectors $X_i(x_{n,r}, \delta_n, A)=((X_i-x_{n,r})^u\, \delta_n^{-[u]})_{u \in A}$ with $i \in S_{n,r}$. So, we may naturally assume that the rows of $X_{n,r}$ are indexed by the elements of $S_{n,r}$, and its columns are indexed by the elements of A. Also, denote by $Y_{n,r}$ the vector of Y_i's for which $i \in S_{n,r}$. Let $H_{n,r}$ be the collection of all subsets of $S_{n,r}$ of size $s(A)$. For h in $H_{n,r}$, write h^c for the set theoretic complement of h in $S_{n,r}$. Let $\beta_{n,r}$ be the $s(A)$-dimensional vector $(\beta_{n,r,u})_{u \in A}$, where $\beta_{n,r,u}=D^u\theta(x_{n,r})\,\delta_n^{[u]}(u!)^{-1}$. Here, for $u=(u_1, u_2, ..., u_d) \in A$, we define $u!=\prod_{i=1}^d u_i!$ with the convention that $0!=1$. Set $\theta^*_{n,r}(x)=\sum_{u \in A} \beta_{n,r,u}(x-x_{n,r})^u\, \delta_n^{-[u]}=$ the

Taylor polynomial of θ around $x_{n,r}$ up to the kth order terms. Also, set $\theta(x)=\theta^*_{n,r}(x)+R_{n,r}(x)$. In view of the definition of $\Theta(c,k,\gamma)$, we can find $c_+>0$ such that $|R_{n,r}(x)|\leqslant c_+\,\delta_n^p$ for all $\theta\in\Theta(c,k,\gamma)$ and $x\in C_{n,r}$.

Proof of Theorem 3.1. First note that, under Condition 3.1, w is bounded away from 0 and ∞ on the compact set $[-0.5,0.5]^d$. Recall from the construction of $\hat{T}_n$ in Section 2 that $N_{n,r}$ is a binomial random variable that counts the number of X_i's that fall in the cube $C_{n,r}$. Now, for the present choices of $\delta_n=J_n^{-1}$, we can choose positive constants λ_1 and λ_2 (which may depend on w) such that, in view of Condition 3.1, we have $\lambda_1\,\delta_n^d\leqslant P(X_i\in C_{n,r})\leqslant\lambda_2\,\delta_n^d$ for all n and $1\leqslant r\leqslant J_n^d$. So, the expected value of $N_{n,r}$ is going to lie between $\lambda_1 n\,\delta_n^d$ and $\lambda_2 n\,\delta_n^d$. Hence, we can choose positive constants c_2, c_3, c_4, c_5 (which may again depend on w) such that an application of Bernstein's inequality (see e.g., [32]) gives

$$\begin{aligned}P(\{c_2 n\,\delta_n^d\leqslant N_{n,r}\leqslant c_3 n\,\delta_n^d \text{ for all } r \text{ with } 1\leqslant r\leqslant J_n^d\})\\ \geqslant 1-c_4 J_n^d\exp(-c_5 n\delta_n^d)\end{aligned}$$

for all n. The assertion (i) in the theorem now follows by defining C_n to be the event enclosed in $\{\ \}$ above and using Borel–Cantelli lemma as, for the present choices of J_n, $\sum_{n=1}^{\infty} J_n^d\exp(-c_5 n\,\delta_n^d)<\infty$.

Next, fix an r such that $1\leqslant r\leqslant J_n^d$. Recall that the conditional distribution of $(X_i-x_{n,r})\,\delta_n^{-1}$ given that $X_i\in C_{n,r}$ is absolutely continuous with density $w_{\delta_n}(x,x_{n,r})$. Also, note that given the set $S_{n,r}$ (i.e., given the indices of the X_i's which fall in $C_{n,r}$), the vectors $(X_i-x_{n,r})\,\delta_n^{-1}$ for $i\in S_{n,r}$ are conditionally independently distributed. So, in view of Facts 5.3 and 5.4, any $s(A)\times s(A)$ submatrix of the matrix $X_{n,r}$ is going to have rank $=s(A)$ with conditional probability 1 given the set $S_{n,r}$. Theorem 3.1 in Koenker and Bassett [25] now implies that if $N_{n,r}\geqslant s(A)$, problem (P) has at least one solution of the form $\hat{\beta}_{n,r}=[X_{n,r}(h)]^{-1}\,Y_{n,r}(h)$ for some $h\in H_{n,r}$ with conditional probability 1 given the set $S_{n,r}$.

For $h\in H_{n,r}$ and $\hat{\beta}_{n,r}=[X_{n,r}(h)]^{-1}\,Y_{n,r}(h)$, set

$$\begin{aligned}L_{n,r}(h)=\sum_{i\in h^c}\,&[1/2-1/2\operatorname{sgn}\{Y_i-\langle X_i(x_{n,r},\delta_n,A),\hat{\beta}_{n,r}\rangle\}-\alpha]\\ &\times[X_{n,r}(h)]^{-1}\,X_i(x_{n,r},\delta_n,A),\end{aligned}$$

where $\operatorname{sgn}(x)$ is $+1$ or -1 depending on whether x is positive or negative respectively. So, $L_{n,r}(h)$ is an $s(A)$-dimensional random vector. It is now immediate from Theorem 3.3 in Koenker and Bassett [25] using the continuity of the distribution of ε_i and its independence from X_i that for some $h\in H_{n,r}$, $\hat{\beta}_{n,r}=[X_{n,r}(h)]^{-1}\,Y_{n,r}(h)$ is a unique solution to problem (P) if and only if $L_{n,r}(h)\in(\alpha-1,\alpha)^{s(A)}$. Further, it follows from the arguments in the proof of Theorem 3.3 in Koenker and Bassett [25] that if, for some $h \in H_{n,r}$, $\hat{\beta}_{n,r} = [X_{n,r}(h)]^{-1} Y_{n,r}(h)$ is a solution (not necessarily unique) to problem (P), we must have $L_{n,r}(h) \in [\alpha - 1, \alpha]^{s(A)}$. Here $(\alpha - 1, \alpha)^{s(A)}$ and $[\alpha - 1, \alpha]^{s(A)}$ are $s(A)$-dimensional intervals in $R^{s(A)}$. Hence, if we choose a positive integer N_0 in such a way that $n \geqslant N_0$ implies that $c_2 n \delta_n^d \geqslant s(A)$, the assertion (ii) in the theorem follows from Fact 5.3 in view of the absolute continuity of the conditional distribution of $(X_i - x_{n,r}) \delta_n^{-1}$ given that $X_i \in C_{n,r}$. ∎

Proof of Theorem 3.2. Assume that $n \geqslant N_0$. Recall that $\delta_n = J_n^{-1}$, where J_n is as defined in the statement of the theorem, so that $\delta_n \sim (\log n/n)^{1(2p+d)}$. Fix a positive integer r such that $1 \leqslant r \leqslant J_n^d$ and for some constant $K_1^* > 0$, let U_n be the event defined as

$$U_n = \{ \max_{1 \leqslant r \leqslant J_n^d} |\hat{\beta}_{n,r} - \beta_{n,r}| > K_1^* \delta_n^p \}.$$

Note that any derivative $T(\theta)$ of θ of order $m \leqslant k$ can be expanded locally in Taylor series around the center $x_{n,r}$ of the cube $C_{n,r}$ (here r is such that the argument of the function θ falls in $C_{n,r}$) up to degree $k - m$ with a remainder term which is uniformly (with respect to the family $\Theta(c, k, \gamma)$) of order $O(\delta_n^{p-m})$. Further, since $\sum_{n=1}^{\infty} P(C_n^c) < \infty$, it is enough (in view of the definition of δ_n and the construction of $\hat{T}_n$) to prove that there is a $K_1^* > 0$ such that

$$\sum_{n=1}^{\infty} \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n \cap C_n) < \infty. \tag{5.4}$$

Define now $V_{n,r}$ to be the event $\{|\hat{\beta}_{n,r} - \beta_{n,r}| > K_1^* \delta_n^p\}$, so that

$$P_\theta(U_n \cap C_n) \leqslant \sum_{r=1}^{J_n^d} P_\theta(V_{n,r} \cap C_n). \tag{5.5}$$

We will try to get an upper bound for $P_\theta(V_{n,r} \cap C_n)$, where $\theta \in \Theta(c, k, \gamma)$. For $\Delta_{n,r} \in R^{s(A)}$, consider the $s(A)$-dimensional random vector

$$\begin{aligned} Z_{n,r,i} = [\tfrac{1}{2} - \tfrac{1}{2} \operatorname{sgn}\{\varepsilon_i - \langle \Delta_{n,r}, X_i(x_{n,r}, \delta_n, A) \rangle \\ + R_{n,t}(X_i)\} - \alpha] \, X_i(x_{n,r}, \delta_n, A). \end{aligned}$$

The norm of this random vector $Z_{n,r,i}$ is bounded by $\{s(A)\}^{1/2}$ whenever $i \in S_{n,r}$. Also, for fixed $\Delta_{n,r}$, the conditional expectation of $Z_{n,r,i}$ given that $i \in S_{n,r}$ is equal to $G_{n,x_{n,r}}(\Delta_{n,r}, \delta_n^p)$ (take $y = x_{n,r}$, $\Delta = \Delta_{n,r}$, $\delta = \delta_n^p$, $R(\delta, x) = -R_{n,r}(x)$ and $M_2 = c_+$ in the statement of Fact 5.2). Using Theorems 3.1 and 3.3 in Koenker and Bassett [25], we can choose a positive constant c^* depending on $s(A)$ such that the event $V_{n,r} \cap C_n$ is contained in the event

$$\Big\{\text{For some } h \in H_{n,r}, \Big|\sum_{i \in h^c} Z_{n,r,i}\Big| \leqslant c^*, \text{ where } \Delta_{n,r} = \hat{\beta}_{n,r} - \beta_{n,r},$$

$$\hat{\beta}_{n,r} = [X_{n,r}(h)]^{-1}\, Y(h), \text{ and } |\Delta_{n,r}| \geqslant K_1^*\, \delta_n^p\Big\} \cap C_n.$$

Given the set $S_{n,r}$ (i.e., the indices of the X_i's which fall in the cube $C_{n,r}$) and the X_i's and Y_i's for $i \in h$ (h is some fixed element of $H_{n,r}$) the vectors $Z_{n,r,i}$ for $i \in h^c$ are conditionally independently and identically distributed each with conditional mean $G_{n,x_{n,r}}(\Delta_{n,r}, \delta_n^p)$, where we can have $\Delta_{n,r} = [X_{n,r}(h)]^{-1}\, Y(h) - \beta_{n,r}$. So, in view of Bernstein's inequality (see [32]) and Fact 5.2, we can choose K_1^* appropriately large to get hold of constants $c_6 > 0$ and $c_7 > 0$ and a positive integer $N_3 \geqslant N_0$ such that

$$P_\theta(V_{n,r} \cap C_n) \leqslant c_6 (n \delta_n^d)^{s(A)} \exp(-c_7 n\, \delta_n^{d+2p}), \tag{5.6}$$

whenever $n \geqslant N_3$ and $\theta \in \Theta(c, k, \gamma)$. Inequalities (5.5) and (5.6) together imply

$$\sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n \cap C_n) \leqslant c_6 J_n^d (n\, \delta_n^d)^{s(A)} \exp(-c_7 n\, \delta_n^{d+2p}).$$

Further, Fact 5.2 implies that by choosing K_1^* appropriately large, c_7 can be chosen as large as desired. Finally, since $\delta_n \sim (\log n/n)^{1/(2p+d)}$, a suitable choice of K_1^* giving an appropriate value of c_7 depending on p and d ensures (5.4). ∎

Proof of Theorem 3.3. We begin by proving assertion (ii). The proof of this assertion is very similar to the proof of Theorem 3.2. Assume again, as in the proof of Theorem 3.2, that $n \geqslant N_0$ (note that this N_0 may be different from that in Theorem 3.2 depending on the choice of J_n). Also, recall that $\delta_n = J_n^{-1}$, where J_n is as defined in the statement of the theorem (so that $\delta_n \sim n^{-1(2p+d)}$), and assume that r is a positive integer such that $1 \leqslant r \leqslant J_n^d$. For some constant $K_2^* > 0$, let U_n^* be the event defined as

$$U_n^* = \{\max_{1 \leqslant r \leqslant J_n^d} |\hat{\beta}_{n,r} - \beta_{n,r}| \geqslant K_2^*\, \delta_n^p \sqrt{\log n}\}.$$

In view of the arguments preceding (5.4) in the proof of Theorem 3.2, it is enough to prove that there is a $K_2^* > 0$ such that

$$\sum_{n=1}^{\infty} \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n^* \cap C_n) < \infty, \tag{5.7}$$

where C_n is as defined in the proof of Theorem 3.1. Arguments, which are very similar to that used in the proof of Theorem 3.2 making use of Fact 5.2, Theorems 3.1 and 3.3 in Koenker and Bassett [25], and Bernstein's inequality (see [32]), now imply that we can choose a positive integer $N_4 \geqslant N_0$ and two positive constants c_8 and c_9 (depending on K_2^*) such that for $n \geqslant N_4$, we have

$$\sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n^* \cap C_n) \leqslant c_8 J_n^d (n\,\delta_n^d)^{s(A)} \exp(-c_9 n\,\delta_n^{d+2} \log n). \quad (5.8)$$

Also, as in the case of Theorem 3.2, c_9 can be chosen as large as desired by choosing K_2^* appropriately large. Finally, since $\delta_n \sim n^{-1/(2p+d)}$, (5.8) implies (5.7) by an appropriate choice of K_2^* depending on p and d, and this completes the proof of assertion (ii).

In order to prove assertion (i), assume as before that $n \geqslant N_0$, and once again, in view of the arguments preceding (5.4) in the proof of Theorem 3.2, assertion (i) will follow if we can show that there is a positive constant K_3^* such that

$$\lim_{n \to \infty} \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta \left\{ J_n^{-d} \sum_{r=1}^{J_n^d} |\hat{\beta}_{n,r} - \beta_{n,r}| \geqslant K_3^* \delta_n^p \right\} = 0. \quad (5.9)$$

Define random variables $\Psi_{n,r}$ for $1 \leqslant r \leqslant J_n^d$ as

$$\begin{aligned} \Psi_{n,r} &= |\hat{\beta}_{n,r} - \beta_{n,r}|\, \delta_n^{-p} && \text{if } |\hat{\beta}_{n,r} - \beta_{n,r}| \leqslant K_2^* \delta_n^p \sqrt{\log n} \\ &= 0 && \text{otherwise,} \end{aligned}$$

where K_2^* is as chosen in the proof above of assertion (ii) of the present theorem, so that (5.7) is ensured. So, $\Psi_{n,r}$'s are obtained from $|\hat{\beta}_{n,r} - \beta_{n,r}|\, \delta_n^{-p}$ by truncation. For a positive constant K_4^*, denote the event $\{J_n^{-d} \sum_{r=1}^{J_n^d} \Psi_{n,r} \geqslant K_4^*\}$ by V_n^*. Hence, in view of (5.7), to prove (5.9) it is enough to prove that there is a positive constant K_4^* such that

$$\lim_{n \to \infty} \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(V_n^* \cap C_n) = 0. \quad (5.10)$$

Now, given the sets $S_{n,r}$'s for $1 \leqslant r \leqslant J_n^d$ (i.e., given the indices of the X_i's which fall in the cubes $C_{n,r}$'s), the random variables $\hat{\beta}_{n,r}$'s (and hence the random variables $\Psi_{n,r}$'s) are conditionally independently distributed. Also, for fixed r, the conditional distribution of $\hat{\beta}_{n,r}$ (and hence that of $\Psi_{n,r}$) given the sets $S_{n,r}$'s with $1 \leqslant r \leqslant J_n^d$ depends only on $S_{n,r}$ (in fact, it depends only on $\#(S_{n,r}) = N_{n,r}$). Let $E_{n,r}^*(\Psi_{n,r}^2)$ denote the conditional expectation of $\Psi_{n,r}^2$ given the set $S_{n,r}$. So, (5.10) will follow if we can show that there is a positive integer $N^{\#}$ and a positive constant M_3 such that for all $n \geq N^{\#}$, $1 \leq r \leq J_n^d$ and $\theta \in \Theta(c, k, \gamma)$, we have

$$E_{n,r}^{*}(\Psi_{n,r}^{2}) \leq M_3 \tag{5.11}$$

whenever $c_2 n\, \delta_n^d \leq N_{n,r} \leq c_3 n\, \delta_n^d$, where c_2 and c_3 are as defined in the proof of Theorem 3.1. Since the proof of (5.11) is complicated and tedious, it will be given in the appendix. ∎

APPENDIX

From now on, r will denote a positive integer such that $1 \leq r \leq J_n^d$ and θ will denote a member of the class $\Theta(c, k, \gamma)$. Recall at this point that $J_n \sim n^{1/(2p+d)}$ as mentioned in the statement of Theorem 3.3 and $\delta_n = J_n^{-1}$. Assume, as before, that $n \geq N_0$. Since the proof of (5.11) is complicated, it will be split into several steps.

Step 1. For $1 \leq i \leq n$, let $Z_{n,r,i}$ be the $s(A)$-dimensional random vector as defined in the proof of Theorem 3.2 in Section 5. So, $Z_{n,r,i} = \phi_{n,r,i}(\Delta_{n,r})\, X_i(x_{n,r}, \delta_n, A)$, where, for $\Delta \in R^{s(A)}$,

$$\phi_{n,r,i}(\Delta) = [(1/2) - (1/2)\operatorname{sgn}\{\varepsilon_i - \langle \Delta, X_i(x_{n,r}, \delta_n, A)\rangle + R_{n,r}(X_i)\} - \alpha].$$

Given X_i, $\phi_{n,r,i}(\Delta)$ is a random variable which takes values $(1-\alpha)$ and $-\alpha$ with conditional probabilities $F\{\langle \Delta, X_i(x_{n,r}, \delta_n, A)\rangle - R_{n,r}(X_i)\}$ and $1 - F\{\langle \Delta, X_i(x_{n,r}, \delta_n, A)\rangle - R_{n,r}(X_i)\}$, respectively. Now, for fixed $h \in H_{n,r}$, let $\Lambda_{n,r}(h, \Delta)$ denote the conditional probability of the event

$$\left\{\left| \sum_{i \in h^c} \phi_{n,r,i}(\Delta)\, X_i(x_{n,r}, \delta_n, A)\right| \leq c^{*}\right\}$$

given the set $S_{n,r}$ (i.e., the indices of the X_i's which fall in the cube $C_{n,r}$). Recall here that h^c is the complement of h in $S_{n,r}$, and c^* is a positive constant, which depends on $s(A)$ and occurs also in the proof of Theorem 3.2. Also, note at this point that given the set $S_{n,r}$, the random vectors (ε_i, X_i) for $i \in S_{n,r}$ are conditionally independently and identically distributed with $\delta_n^{-1}(X_i - x_{n,r})$ having the conditional density $w_{\delta_n}(x_i, x_{n,r})$. Hence, $\Lambda_{n,r}(h, \Delta)$ is the same for all $h \in H_{n,r}$.

Now, in view of the arguments (which make use of Theorems 3.1 and 3.3 in Koenker and Bassett [25]) in the proofs of Theorems 3.1 and 3.2 (also see the proof of Theorem 4.2 in Koenker and Bassett [25]) and the construction of $\hat{\beta}_{n,r}$, we have, for any fixed positive constant K (the choice of which will be specified later),

$$\begin{aligned}
E^*_{n,r}(\Psi^2_{n,r}) &= E^*_{n,r}[\Psi^2_{n,r} I(|\hat{\beta}_{n,r}-\beta_{n,r}|\,\delta_n^{-p}<K)] \\
&\quad + E^*_{n,r}[\Psi^2_{n,r} I(|\hat{\beta}_{n,r}-\beta_{n,r}|\,\delta_n^{-p}\geqslant K)] \\
&\leqslant [K]^2 + \#(H_{n,r}) \int_{[-0.5,0.5]^d\times[-0.5,0.5]^d\times\cdots\times[-0.5,0.5]^d} |\tilde{X}|\,\delta_n^{ps(A)} \\
&\quad \times\Bigg[\int_{\{\Delta\in R^{s(A)};\,K\leqslant|\Delta|\leqslant K_2^*\sqrt{\log n}\}} |\Delta|^2 \prod_{i=1}^{s(A)} f\{\langle\Delta\delta_n^p, x_i(A)\rangle \\
&\quad - R_{n,r}(\delta_n x + x_{n,r})\}\, \Lambda_{n,r}(h, \Delta\delta_n^p)\, d\Delta\Bigg] \\
&\quad \times \prod_{i=1}^{s(A)} w_{\delta_n}(x_i, x_{n,r})\, dx_i, \qquad\qquad (\text{A.1})
\end{aligned}$$

where I is the 0–1 valued indicator function and K_2^* is as in the proof of assertion (ii) in Theorem 3.3, so that (5.7) is ensured. $|\tilde{X}|$, which occurs as a part of the integrand above, is the determinant of the $s(A)\times s(A)$ matrix $\tilde{X}$ whose rows are the vectors $x_i(A)$ for $1\leqslant i\leqslant s(A)$ (here $x_i(A)$ is as defined in the statement of Fact 5.2 with $x_i\in R^d$). Here it is assumed that $N_{n,r}\geqslant s(A)$, which is true if $n\geqslant N_0$ and the condition for the occurrence of the event C_n defined in Theorem 3.1 is satisfied.

Clearly, for $x\in[-0.5, 0.5]^d$ the components of the vector $x(A)$ are each bounded by 1. Recall also from Section 5 that $x\in[-0.5, 0.5]^d$ implies that $|R_{n,r}(\delta_n x + x_{n,r})|\leqslant c_+\,\delta_n^p$. Further, in view of Condition 3.2, the density f is bounded in a neighborhood of 0. Hence, using the fact that $\delta_n \sim n^{-1/(2p+d)}$, one can find an integer $N_5>N_0$ and a positive constant M_4 such that (A.1) gives

$$\begin{aligned}
E^*_{n,r}(\Psi^2_{n,r}) &\leqslant [K]^2 + \delta_n^{-ps(A)} M_4 \\
&\quad \times \int_{\{\Delta\in R^{s(A)};\,K\leqslant|\Delta|\leqslant K_2^*\sqrt{\log n}\}} |\Delta|^2\, \Lambda_{n,r}(h, \Delta\delta_n^p)\, d\Delta \qquad (\text{A.2})
\end{aligned}$$

whenever $n\geqslant N_5$ and $c_2 n\,\delta_n^d\leqslant N_{n,r}=\#(S_{n,r})\leqslant c_3 n\,\delta_n^d$ (which is a condition equivalent to the occurrence of the event C_n and implies that $\#(H_{n,r})\leqslant [c_3 n\,\delta_n^d]^{s(A)}$).

Step 2. For fixed $h\in H_{n,r}$, let $P^*_{n,r}(h, \Delta)$ denote the conditional probability of the event

$$\left\{\left|\sum_{i\in h^c} \phi_{n,r,i}(\Delta\delta_n^p)\right|\leqslant c^*\right\}$$

given the set $S_{n,r}$. Define

$$p_{n,r}(\Delta)=\int_{[-0.5,0.5]^d} F\{\langle \Delta\, \delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\}\, w_{\delta_n}(x, x_{n,r})\, dx.$$

So, given the set $S_{n,r}$, the random variables $\phi_{n,r,i}(\Delta\delta_n^p)$ for $i \in S_{n,r}$ are conditionally independently distributed, and each of them takes values $(1-\alpha)$ and $-\alpha$ with conditional probabilities $p_{n,r}(\Delta)$ and $1-p_{n,r}(\Delta)$ respectively. So, on the event C_n, we can choose an integer $N_6 > N_5$ and a positive constant M_5 such that using Stirling's approximation for factorials and the fact that $\delta_n \sim n^{-1/(2p+d)}$, we have

$$\begin{aligned} P^*_{n,r}(h, \Delta\, \delta_n^p) \leqslant M_5\, \delta_n^p [\alpha^{-1} p_{n,r}(\Delta)]^{\#(h^c)\alpha} \\ \times [(1-\alpha)^{-1}\{1-p_{n,r}(\Delta)\}]^{\#(h^c)(1-\alpha)} \end{aligned} \tag{A.3}$$

whenever $n \geqslant N_6$. Also, when $|R_{n,r}(\delta_n x + x_{n,r})| \leqslant c_+\, \delta_n^p$ and $|\Delta| \leqslant K_2^* \sqrt{\log n}$, we can choose an integer $N_7 > N_6$ such that $n \geqslant N_7$ implies, under Condition 3.2 via the mean value theorem of differential calculus, that

$$\begin{aligned} p_{n,r}(\Delta) = \alpha + \int_{[-0.5,0.5]^d} \{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\} \\ \times f(\xi_{n,r}(x))\, w_{\delta_n}(x, x_{n,r})\, dx, \end{aligned} \tag{A.4}$$

where $\xi_{n,r}(x)$ may depend on Δ and $R_{n,r}$ and satisfies

$$|\xi_{n,r}(x)| \leqslant |\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})|.$$

Hence, we can write

$$p_{n,r}(\Delta) = \alpha + \langle \Delta, \mu_{n,r}(\Delta)\rangle\, \delta_n^p - t_{n,r}(\Delta)\, \delta_n^p, \tag{A.5}$$

where

$$\mu_{n,r}(\Delta) = \int_{[-0.5,0.5]^d} x(A)\, f(\xi_{n,r}(x))\, w_{\delta_n}(x, x_{n,r})\, dx$$

and

$$t_{n,r}(\Delta) = \int_{[-0.5,0.5]^d} \delta_n^{-p} R_{n,r}(\delta_n x + x_{n,r})\, f(\xi_{n,r}(x))\, w_{\delta_n}(x, x_{n,r})\, dx.$$

Once again, in view of Condition 3.2, N_7 can be appropriately chosen so that we can get hold of a positive constant M_6 such that

$$\sup_{n \geqslant N_7}\ \sup_{1 \leqslant r \leqslant J_n^d}\ \sup_{\theta \in \Theta(c,k,\gamma)}\ \sup_{|\Delta| \leqslant K_2^* \sqrt{\log n}} |t_{n,r}(\Delta)| \leqslant M_6. \tag{A.6}$$

Further, in view of Conditions 3.1 and 3.2, we have

$$\lim_{n\to\infty} \sup_{1\leqslant r\leqslant J_n^d} \sup_{\theta\in\Theta(c,k,\gamma)} \sup_{|\Delta|\leqslant K_2^*\sqrt{\log n}} |\mu_{n,r}(\Delta)-\mu| = 0, \tag{A.7}$$

where μ is the $s(A)$-dimensional vector given by $\mu = f(0) \times \int_{[-0.5,0.5]^d} x(A)\,dx$. Finally, (A.3) and (A.5) imply that on the event C_n, we can choose an integer $N_8 > N_7$ and a positive constant M_7 such that

$$P^*_{n,r}(h, \Delta) \leqslant M_7\,\delta_n^p \exp[-\tfrac{1}{2}\,\#(h^c)\,\delta_n^{2p}\alpha^{-1}(1-\alpha)^{-1} \times \{\langle \Delta, \mu_{n,r}(\Delta)\rangle - t_{n,r}(\Delta)\}^2] \tag{A.8}$$

whenever $n \geqslant N_8$, $|\Delta| \leqslant K_2^* \sqrt{\log n}$, and $h \in H_{n,r}$.

Step 3. Let $X_i^{\#}(x_{n,r}, \delta_n, A)$ denote the random vector of dimension $s(A)-1$ obtained from the $s(A)$-dimensional random vector $X_i(x_{n,r}, \delta_n, A)$ by dropping the component 1. So,

$$X_i^{\#}(x_{n,r}, \delta_n, A) = (\delta_n^{-[u]}(X_i - x_{n,r})^u)_{u\in A(d,k)},$$

where $A(d, k)$ is as defined at the beginning of Section 5. Now, given the set $S_{n,r}$ and the random variables $\phi_{n,r,i}(\Delta\delta_n^p)$ for $i \in S_{n,r}$, the random vectors $\delta_n^{-1}(X_i - x_{n,r})$ for $i \in S_{n,r}$ are conditionally independently distributed. Further, the conditional distribution of $\delta_n^{-1}(X_i - x_{n,r})$ for $i \in S_{n,r}$ depends only on $\phi_{n,r,i}(\Delta\,\delta_n^p)$. In fact, given the set $S_{n,r}$ and the random variable $\phi_{n,r,i}(\Delta\delta_n^p)$, the conditional density of $\delta_n^{-1}(X_i - x_{n,r})$ for $i \in S_{n,r}$ is given by

$$[p_{n,r}(\Delta)]^{-1}\,[F\{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\}]\, w_{\delta_n}(x, x_{n,r}) \quad \text{if} \quad \phi_{n,r,i}(\Delta\delta_n^p) = (1-\alpha)$$

and

$$[1 - p_{n,r}(\Delta)]^{-1}\,[1 - F\{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\}]\, w_{\delta_n}(x, x_{n,r}) \quad \text{if} \quad \phi_{n,r,i}(\Delta\delta_n^p) = -\alpha,$$

where $p_{n,r}(\Delta)$ is as defined in Step 2 above and $x \in [-0.5, 0.5]^d$. So, Condition 3.2 and the fact that $\delta_n \sim n^{-1/(2p+d)}$ imply that

$$\lim_{n\to\infty} \sup_{1\leqslant r\leqslant J_n^d} \sup_{\theta\in\Theta(c,k,\gamma)} \sup_{x\in[-0.5,0.5]^d} \sup_{|\Delta|\leqslant K_2^*\leqslant\sqrt{\log n}} |F\{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\} - \alpha| = 0. \tag{A.9}$$

For $\omega \in R^{s(A)-1}$, let $\sigma_{n,r}(\omega, \phi_{n,r,i}(\Delta\delta_n^p))$ with $i \in S_{n,r}$ denote the conditional characteristic function of the random vector $\phi_{n,r,i}(\Delta\delta_n^p)\,X_i^{\#}(x_{n,r}, \delta_n, A)$

given the set $S_{n,r}$ and the random variable $\phi_{n,r,i}(\Delta\delta_n^p)$. In view of Condition 3.1, Fact 5.1 and (A.9), for any $\eta^\#$ such that $0<\eta^\#<1$, we can choose an integer $N_9>N_8$ and a positive constant K_0 so that

$$\sup_{n\geq N_9}\ \sup_{|\omega|\geq K_0}\ \sup_{1\leq r\leq J_n^d}\ \sup_{\theta\in\Theta(c,k,\gamma)}\ \sup_{|\Delta|\leq K_2^*\sqrt{\log n}}\ \sup_{\phi_{n,r,i}(\Delta\delta_n^p)\in\{-\alpha,(1-\alpha)\}} |\sigma_{n,r}(\omega,\phi_{n,r,i}(\Delta\delta_n^p))|<\eta^\#. \tag{A.10}$$

Recall that $\phi_{n,r,i}(\Delta\delta_n^p)$ takes only two values, namely $-\alpha$ and $1-\alpha$.

For $i\in S_{n,r}$, let $Q_{n,r}^\#(\Delta,\phi_{n,r,i}(\Delta\delta_n^p))$ denote the conditional dispersion matrix of $X_i^\#(x_{n,r},\delta_n,A)$ given the set $S_{n,r}$ and the random variable $\phi_{n,r,i}(\Delta\delta_n^p)$. Then, using (A.9), Condition 3.1, and the form of the conditional density of $\delta_n^{-1}(X_i-x_{n,r})$, we obtain

$$\lim_{n\to\infty}\ \sup_{1\leq r\leq J_n^d}\ \sup_{\theta\in\Theta(c,k,\gamma)}\ \sup_{|\Delta|\leq K_2^*\sqrt{\log n}}\ \sup_{\phi_{n,r,i}(\Delta\delta_n^p)\in\{-\alpha,(1-\alpha)\}} |Q_{n,r}^\#(\Delta,\phi_{n,r,i}(\Delta\delta_n^p))-Q^\#|=0, \tag{A.11}$$

where $Q^\#$ is a positive definite matrix given by

$$Q^\#=\int_{[-0.5,0.5]^d}[x(d,k)][x(d,k)]^{\mathrm{T}}\,dx - \left[\int_{[-0.5,0.5]^d}x(d,k)\,dx\right]\left[\int_{[-0.5,0.5]^d}x(d,k)\,dx\right]^{\mathrm{T}}.$$

Here $x(d,k)$ is a column vector of dimension $s(A)-1$ as defined at the beginning of Section 5. So, $Q^\#$ is nothing but the dispersion matrix of $X(d,k)$ when X is a d-dimensional random vector uniformly distributed on $[-0.5,0.5]^d$. Define, for fixed $h\in H_{n,r}$, $\Phi_{n,r}(h,\Delta\delta_n^p)$ to be the random vector $(\phi_{n,r,i}(\Delta\delta_n^p))_{i\in h^c}$ with dimension $\#(h^c)$. Let $Q_{n,r}^*(h,\Delta,\Phi_{n,r}(h,\Delta\delta_n^p))$ denote the conditional dispersion matrix of

$$\frac{1}{\sqrt{\#(h^c)}}\sum_{i\in h^c}\phi_{n,r,i}(\Delta\delta_n^p)\,X_i^\#(x_{n,r},\delta_n,A)$$

given the set $S_{n,r}$ and the random vector $\Phi_{n,r}(h,\Delta\delta_n^p)$. Equation (A.11) implies that on the event C_n,

$$\lim_{n\to\infty}\ \sup_{1\leq r\leq J_n^d}\ \sup_{|\Delta|\leq K_2^*\sqrt{\log n}}\ \sup_{h\in H_{n,r}}\ \sup_{|\sum_{i\in h^c}\phi_{n,r,i}(\Delta\delta_n^p)|\leq c^*}\ \sup_{\theta\in\Theta(c,k,\gamma)} |Q_{n,r}^*(h,\Delta,\Theta_{n,r}(h,\Delta\delta_n^p))-\alpha(1-\alpha)\,Q^\#|=0. \tag{A.12}$$

Next, let $\mu^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))$ denote the conditional mean of

$$\frac{1}{\sqrt{\#(h^c)}} \sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)\, X_i^{\#}(x_{n,r}, \delta_n, A)$$

given the set $S_{n,r}$ and the random vector $\Phi_{n,r}(h, \Delta\delta_n^p)$. At this point, recall from (A.4) in Step 2 that for $n \geqslant N_7$, we can write

$$\begin{aligned} F\{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\} \\ = \alpha + \{\langle \Delta\delta_n^p, x(A)\rangle - R_{n,r}(\delta_n x + x_{n,r})\}\, f(\xi_{n,r}(x)). \end{aligned}$$

This and the form of the conditional density of $\delta_n^{-1}(X_i - x_{n,r})$ for $i \in S_{n,r}$ given the set $S_{n,r}$ and the random variable $\phi_{n,r,i}(\Delta\delta_n^p)$ imply that there is an integer $N_{10} > N_9 > N_7$ such that for $n \geqslant N_{10}$ and on the event C_n, we have (assuming Δ to be a column vector)

$$\begin{aligned} \mu^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p)) = \delta_n^p \sqrt{\#(h^c)}\, \{\Omega_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))\, \Delta \\ + t^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))\}, \end{aligned} \tag{A.13}$$

where $t^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))$ is a vector of dimension $(s(A)-1)$ with the property that there is a positive constant M_8 such that

$$\sup_{n \geqslant N_{10}}\ \sup_{1 \leqslant r \leqslant J_n^d}\ \sup_{\theta \in \Theta(c,k,\gamma)}\ \sup_{|\Delta| \leqslant K_2^* \sqrt{\log n}}\ \sup_{h \in H_{n,r}}\ \sup_{|\sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)| \leqslant c^*} |t^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))| \leqslant M_8, \tag{A.14}$$

and $\Omega^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))$ is a matrix of dimension $(s(A)-1) \times s(A)$ with the property that

$$\lim_{n \to \infty}\ \sup_{1 \leqslant r \leqslant J_n^d}\ \sup_{\theta \in \Theta(c,k,\gamma)}\ \sup_{|\Delta| \leqslant K_2^* \sqrt{\log n}}\ \sup_{h \in H_{n,r}}\ \sup_{|\sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)| \leqslant c^*} |\Omega^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p)) - \{f(0)\}\, \Omega| = 0. \tag{A.15}$$

Here Ω is a matrix of dimension $(s(A)-1) \times s(A)$ with all the entries in the first column equal to 0 and the remaining $s(A)-1$ columns identical with the columns of $Q^{\#}$. So, in the partitioned form it looks like $\Omega = (0 : Q^{\#})$.

Step 4. Define, for fixed $h \in H_{n,r}$, $\Lambda^*_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))$ to be the conditional probability of the event

$$\left\{\left|\sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)\, X_i^{\#}(x_{n,r}, \delta_n, A)\right| \leqslant c^*\right\}$$

given the set $S_{n,r}$ and the random vector $\Phi_{n,r}(h, \Delta\delta_n^p)$. Now, it follows from (A.10), (A.12), (A.13), (A.14) and (A.15) in Step 3 via arguments along the same line as in the proof of Theorem 20.1 and Corollaries 20.3 and 20.4 (in particular, 20.49 in Corollary 20.4) in Chapter 4 of Bhattacharya and Rao [2] (also see Theorem 20.6 in the same chapter) that there is an integer $N_{11} > N_{10}$ and positive constants M_9 and M_{10} such that assuming $\mu_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))$ to be a column vector, we have

$$\begin{aligned}
&\Lambda_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p)) \\
&\quad \leqslant \delta_n^{p[s(A)-1]} M_9 \exp\{-(\tfrac{1}{2})[\mu_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))]^{\mathrm{T}} \\
&\qquad \times [Q_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))]^{-1} [\mu_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))]\} \\
&\qquad + M_{10}\, \delta_n^{p[s(A)-1]} / \sqrt{\#(h^c)} \qquad\qquad (\mathrm{A}.16)
\end{aligned}$$

whenever $n \geqslant N_{11}$, $|\Delta| \leqslant K_2^* \sqrt{\log n}$, $h \in H_{n,r}$, $|\sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)| \leqslant c^*$ and the event C_n occurs.

Step 5. For fixed $h \in H_{n,r}$, let us define

$$\begin{aligned}
&\Xi_{n,r}(h, \Delta, \Phi_{n,r}(\Delta\delta_n^p)) \\
&\quad = \alpha^{-1}(1-\alpha)^{-1} [\mu_{n,r}(\Delta)][\mu_{n,r}(\Delta)]^{\mathrm{T}} \\
&\qquad + [\Omega_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))]^{\mathrm{T}} [Q_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))]^{-1} \\
&\qquad \times [\Omega_{n,r}^*(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))] - \alpha^{-1}(1-\alpha)^{-1} \{f(0)\}^2 Q,
\end{aligned}$$

where Q is a positive definite matrix of dimension $s(A) \times s(A)$ given by

$$Q = \int_{[-0.5, 0.5]^d} [x(A)][x(A)]^{\mathrm{T}}\, dx.$$

Now, it follows from (A.7), (A.12), and (A.15) that

$$\lim_{n \to \infty} \sup_{1 \leqslant r \leqslant J_n^d} \sup_{\theta \in \Theta(c,k,\gamma)} \sup_{|\Delta| \leqslant K_2^* \sqrt{\log n}} \sup_{h \in H_{n,r}} \sup_{|\sum_{i \in h^c} \phi_{n,r,i}(\Delta\delta_n^p)| \leqslant c^*} |\Xi_{n,r}(h, \Delta, \Phi_{n,r}(h, \Delta\delta_n^p))| = 0.$$

Hence, (A.6), (A.8), (A.13), (A.14) and (A.16) imply that there are positive constants κ, M_{11}, M_{12}, and K and an integer $N^\# > N_{11}$ such that on the event C_n,

$$\Lambda_{n,r}(h, \Delta\delta_n^p) \leqslant \delta_n^{p[s(A)]} \left\{ M_{11} \exp(-\kappa|\Delta|^2) + \frac{M_{12}}{\sqrt{\#(h^c)}} \right\} \qquad (\mathrm{A}.17)$$

whenever $n \geqslant N^\#$, $K \leqslant |\Delta| \leqslant K_2^* \sqrt{\log n}$ and $h \in H_{n,r}$.

Finally, (A.17) above and (A.2) in Step 1 together imply (5.11).

ACKNOWLEDGMENTS

This paper has grown in part out of the research work done by the author as a Ph.D. student at the University of California, Berkeley. The author wishes to thank his Ph.D. supervisor Professor Charles J. Stone and his teacher Professor Lucien LeCam for their help and many stimulating and thought provoking discussions.

REFERENCES

[1] BASSETT, G., AND KOENKER, R. (1986). Strong consistency of regression quantiles and related processes. *Econom. Theory* **2** 191–202.

[2] BHATTACHARAYA, R. N., AND RAO, R. R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[3] BHATTACHARYA, P. K., AND GANGOPADHYAY, A. (1990). Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

[4] BICKEL, P. J. (1973). On some analogues of linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

[5] BLOOMFIELD, P., AND STEIGER, W. L. (1983). Least absolute deviations, theory, applications and algorithms. In *Progr. Probab. Statist.* (P. Huber and M. Rosenblatt, Eds.), Birkhäuser, Boston.

[6] CHAUDHURI, P. (1988). *Asymptotic Theory of Nonparametric Estimation of Conditional Quantiles.* Ph.D. thesis, Department of Statistics, University of California, Berkeley.

[7] CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19**.

[8] CHAUDHURI, P. (1989). *Attainability of the Optimal Global Rates of Convergence in Nonparametric Quantile Regression.* Technical Report No. 860, Department of Statistics, University of Wisconsin-Madison.

[9] CHAUDHURI, P., HUANG, M. C., LOH, W. Y., AND YAO, R. (1990). *Regression Trees and Function Estimation.* Technical Report No. 871, Department of Statistics, University of Wisconsin-Madison.

[10] CHEN, X. R., AND WU, Y. H. (1988). Strong consistency of M-estimates in linear models. *J. Multivariate Anal.* **27** 116–130.

[11] CHENG, K. F. (1983). Nonparametric estimators for percentile regression functions. *Comm. Statist. Theory Methods* **12** 681–692.

[12] DABROWSKA, D. M. (1987). *Nonparametric Quantile Regression with Censored Data.* Technical Report No. 93, Department of Statistics, University of California, Berkeley.

[13] DEVROYE, L., AND GYORFI, L. (1985). *Nonparametric Density Estimate: The L_1 View.* Wiley, New York.

[14] DONOHO, D. L., AND LIU, R. C. (1987, 1988). *Geometrizing Rates of Convergence, I, II, and III.* Technical Reports No. 120, 137, and 138, Department of Statistics, University of California, Berkeley.

[15] FERGUSON, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

[16] HARDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.* **14** 169–180.

[17] HARDLE, W. AND GASSER, T. (1984). Robust nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.

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

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

[20] HARTER, H. L. (1974, 1975). The method of least squares and some alternatives. *Internat. Statist. Rev.* **42** 147–174, 235–264; **43** 1–44, 125–190, 269–278.

[21] HOGG, R. V. (1975). Estimates of percentile regression lines using salary data. *J. Amer. Statist. Assoc.* **70** 56–59.

[22] HUBER, P. J. (1973). Robust regression, asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

[23] HUBER, P. J. (1987). The place of the L_1-norm in the robust estimation. In *Statistical Data Analysis Based on the L_1-norm and Related Methods* (Y. Dodge, Ed.), pp. 23–33. North Holland, Amsterdam.

[24] JANSSEN, P., AND VERAVERBEKE, N. (1987). On nonparametric regression estimators based on regression quantiles. *Comm. Statist. Theory Methods* **16** 383–396.

[25] KOENKER, R., AND BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

[26] KOENKER, R., AND D'OREY, V. (1987). Computing regression quantiles. *Appl. Statist.* **36** 383–393.

[27] KOENKER, R., AND PORTNOY, S. (1987). L-Estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

[28] LEJEUNE, M. G., AND SARDA, P. (1988). Quantile regression: A nonparametric approach. *Comput. Statist. Data Anal.* **6** 229–239.

[29] NAHM, J. W. (1989). *Nonparametric Least Absolute Deviations Estimation.* Ph.D. thesis, Department of Economics, University of Wisconsin–Madison.

[30] NARULA, S. C., AND WELLINGTON, J. F. (1986). An algorithm to find all regression quantiles. *J. Statist. Comput. Simulation* **26** 269–281.

[31] RUPPERT, D., AND CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

[32] SHORACK, G. R., AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[33] STIGLER, S. M. (1973). Simon Newcomb, Percy Daniell and the History of Robust Estimation 1885–1920. *J. Amer. Statist. Assoc.* **68** 872–879.

[34] STONE, C. J. (1977). Consistent non-parametric regression. *Ann. Statist.* **5** 549–645.

[35] STONE, C. J. (1980). Optimal rates of convergence for non-parametric estimators. *Ann. Statist.* **8** 1348–1360.

[36] STONE, C. J. (1982). Optimal global rates of convergence for non-parametric regression. Special Invited Paper, *Ann. Statist.* **4** 1040–1053.

[37] STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **2** 689–705.

[38] STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **2** 590–606.

[39] STUTE, W. (1986). Conditional empirical processes. *Ann. Statist.* **14** 638–647.

[40] TRUONG, Y. K. N. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–617.

[41] WELSH, A. H. (1987). The trimmed mean in the linear model. *Ann. Statist.* **15** 20–45.

[42] WELSH, A. H. (1987). One step L-estimation in linear model. *Ann. Statist.* **15** 626–641.

[43] YATRACOS, Y. G. (1988). A lower bound on the error in nonparametric regression type problems. *Ann. Statist.* **16** 1180–1187.

[44] WELLINGTON, J. F., AND NARULA, S. C. (1984). An algorithm for regression quantiles. *Comm. Statist. Simul. Comput.* **13** 683–704.