The Annals of Statistics
1991, Vol. 19, No. 2, 760–777

NONPARAMETRIC ESTIMATES OF REGRESSION QUANTILES AND THEIR LOCAL BAHADUR REPRESENTATION[1]

By Probal Chaudhuri

University of Wisconsin, Madison and Indian Statistical Institute, Calcutta

Let (X, Y) be a random vector such that X is d-dimensional, Y is real valued and $Y = \theta(X) + \varepsilon$, where X and ε are independent and the αth quantile of ε is 0 (α is fixed 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 a pointwise 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} under appropriate regularity conditions. Further, a local Bahadur type representation is shown to hold for the estimate $\hat{T}_n$ and this is used to obtain some useful asymptotic results.

1. Introduction. Suppose that we have i.i.d. observations (X_1, Y_1), $(X_2, Y_2), \ldots, (X_n, Y_n)$, where the Y_i's are real valued and the X_i's are d-dimensional satisfying the model

$$Y_i = \theta(X_i) + \varepsilon_i, \qquad 1 \le i \le n.$$

Here θ is an unknown function to be estimated from the data, and the ε_i's are i.i.d. unobservable random variables, which are assumed to be independent of the X_i's. In the usual regression problems, one assumes that ε_i has expected value 0 and tries to construct estimates for θ using the method of least squares. The least squares estimates certainly have many optimal properties, particularly when the random error ε_i follows normal distribution. However, it is well known [see, for example, Harter (1974–1975), Huber (1973), etc.] that the method of least squares does not perform very well when the ε_i has a heavy-tailed distribution because the method is highly sensitive to extreme values or outliers among the ε_i's. The method of least absolute deviations, which is one of the competitors to the method of least squares, has been neglected in the past because of the computational difficulties associated with it and the complexity of the distribution of the resulting statistical estimates. The possible inefficiency of the method of least squares when the random errors in the data follow nonnormal probability laws and recent developments in the robustness studies for statistical procedures have motivated the search

Received August, 1988; revised June 1990.
[1]Research partially supported by NSF Grant DMS-86-00409 and a Wisconsin Alumni Research Foundation Grant.
AMS 1980 *subject classifications*. Primary 62G05, 62G35; secondary 62G20, 62E20.
Key words and phrases. Regression quantiles, Bahadur representation, optimal nonparametric rates of convergence.

for more robust methods than the method of least squares—the method of least absolute deviations being one potential candidate. Further, recent developments in efficient algorithms for obtaining least absolute deviation fits [Barradole and Roberts (1973, 1974), Bartels, Conn and Sinclair (1978), Bloomfield and Steiger (1980), Gonin and Money (1989), etc.] and rapid advancements in high-speed computing facilities have played major roles in motivating people to investigate the method of least absolute deviations as a way of robustifying the least squares regression.

Koenker and Bassett (1977) [see Stone (1977), Ferguson (1967), etc.] suggested the use of the loss function

$$H_\alpha(t) = |t| + (2\alpha - 1)t$$

instead of the usual squared error loss to estimate parameters in linear models. Here α is fixed such that $0 < \alpha < 1$. Note that, for a random variable Z, $E\{H_\alpha(Z - t)\}$ is minimized by taking $t =$ the αth quantile of Z [just as $E\{(Z - t)^2\}$ is minimized by taking $t =$ the mean of Z]. Particularly important is the case $\alpha = 1/2$, when the function $H_\alpha(t)$ becomes the absolute loss function, which is an appropriate thing to use if θ is the conditional median function of Y given X. Koenker and Bassett (1978) were motivated in part by Huber's (1973) observation that outliers are very difficult to identify in the regression context. Further, it is worth noting here that one can construct analogs of L-estimates (like the trimmed mean, or other linear combinations of order statistics in one-sample location problems) for the parameters in a linear model using the regression quantile estimates [see Ruppert and Carroll (1980), Koenker and Portnoy (1987), etc.]. Algorithms to compute regression quantiles have been discussed in Koenker and D'Orey (1987), Narula and Wellington (1986), etc.

Several people [Koenker and Bassett (1978), Bloomfield and Steiger (1983), Ruppert and Carroll (1980), Koenker and Portnoy (1987), etc.] have studied the asymptotic statistical behavior of the estimates constructed using the loss function $H_\alpha(t)$, establishing the $\sqrt{n}$-consistency and the asymptotic normality of the estimates under appropriate regularity conditions. However, all these people have studied the problem from a linear parametric approach. In other words, they assumed that the function θ belongs to a fixed finite-dimensional linear space of functions, so that the form of θ, as it depends on the regressor X, is known except for a fixed and finite number (which does not depend on the sample size) of unknown parameters. So, in their cases, it becomes a problem of estimating a finite-dimensional Euclidean parameter. Recent developments in nonparametric regression provide a strong stimulus to investigate the asymptotic behavior of nonparametric estimates of θ, which are constructed using the loss function $H_\alpha(t)$, under the assumption that θ is a suitably smooth function instead of assuming a finite-dimensional linear parametric model for θ.

Under appropriate regularity conditions, Stone (1980, 1982) proved the attainability of the optimal nonparametric rates of convergence in the asymptotic minimax sense by constructing estimates for the regression function and its derivatives using local polynomial fits minimizing the squared error loss. With a desire to robustify nonparametric regression, in his 1982 paper he asked *if the same rates of convergence are achievable in the estimation of the conditional median function of Y given X*. It is worth mentioning here that in his 1980 and 1982 papers he obtained the lower bounds for the rates of convergence for nonparametric estimates. Although he was mainly talking there about estimates of the regression function, it is clear that the lower bounds obtained in these papers also apply to estimates of the conditional median function. As a matter of fact, the lower bounds are determined by the local behavior of the Kullback–Leibler divergence of the conditional distribution [Yatracos (1988)] or, more broadly speaking, by the "geometry" of the problem [Donoho and Liu (1988)]. Truong (1989) provided a partial answer to Stone's question using local median. Earlier efforts to robustify nonparametric regression were made by Härdle (1984), Härdle and Gasser (1984), Härdle and Luckhaus (1984), Härdle and Tsybakov (1988), etc. In the present work, a complete affirmative answer to this question is provided by constructing pointwise estimates for conditional quantile functions and their derivatives using local polynomial fits that minimize the loss function $H_\alpha(t)$ and showing that such estimates achieve the optimal nonparametric rates of convergence in the asymptotic minimax sense under mild regularity conditions. One of the main objectives here is to gain theoretical insights into the asymptotic behavior of nonparametric estimates of regression quantiles constructed through local polynomial fits. A major advantage of working with the squared error loss is that it yields estimates that are linear functions of $Y_1, Y_2, \dots, Y_n$. When one replaces the squared error loss by the loss function $H_\alpha(t)$ this advantage is lost. As a result, the standard techniques of asymptotic theory, which are capable of dealing with sums of independent random variables, cannot be used directly any more. The *local Bahadur type representation* obtained in this paper can be used as an elegant tool to cope with this situation. In addition to providing several useful asymptotic results, this asymptotic linear representation helps to get a very good insight into the asymptotic behavior of nonparametric estimates of regression quantiles by making apparent the critical role played here by the density f of the random error ε. In particular, it explains why the inference about regression quantiles is intrinsically different from that about the usual regression function.

2. Description of the estimates for the conditional quantile function and its derivatives. We assume that, in the model $Y_i = \theta(X_i) + \varepsilon_i$, ε_i is independent of X_i and has a distribution with 0 as the αth quantile. Here α is a fixed number such that $0 < \alpha < 1$. In other words, the conditional αth quantile of Y_i given $X_i = x_i$ is $\theta(x_i)$. In this paper, we will discuss estimates of the function θ or its derivatives evaluated at a particular point, which will be assumed without loss of generality to be $0 \in R^d$.

For $u = (u_1, \dots, u_d)$, 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 V be some fixed open neighborhood of 0 in R^d. We will write $|\ |$ to denote the usual Euclidean norm. For a fixed nonnegative integer

k and real numbers c and γ such that $c > 0$ and $0 < \gamma \leq 1$, let $\Theta(c, k, \gamma)$ be the collection of all real valued functions θ on V such that

(i) $D^u\theta(x)$ exists and is continuous in x for all $x \in V$ and $[u] \leq k$,
(ii) $|D^u\theta(x) - D^u\theta(0)| \leq c|x|^\gamma$ for all $x \in V$ and $[u] = k$.

Thus the functions $\Theta(c, k, \gamma)$ are continuously differentiable up to order k on V and their kth derivatives are uniformly Hölder continuous at 0 with exponent γ. We will refer to $p = k + \gamma$ as the order of smoothness of the functions in $\Theta(c, k, \gamma)$ at 0. We assume that the conditional quantile function of Y given X is an element of $\Theta(c, k, \gamma)$ for some fixed c, k and γ [see Stone (1980, 1982), Devroye and Györfi (1985), etc.]. For u, a d-dimensional vector of nonnegative integers such that $[u] = m \leq k$, set $T(\theta) = D^u\theta(0)$. An estimate $\hat{T}_n$ of $T(\theta)$ will now be described.

Construction of $\hat{T}_n$. Consider a sequence of positive real numbers $\delta_n = an^{-1/(2p+d)}$, where a is a positive constant. Let C_n denote the cube $[-\delta_n, \delta_n]^d$ in R^d. So, for n sufficiently large, this cube will be completely contained in the open set V. *From now on, for the rest of this paper, we will assume that n is such that* $C_n \subseteq V$. Let S_n be the collection of all the subscripts of the X_i's (for $1 \leq i \leq n$) that fall in the cube C_n and let N_n be the number of such subscripts. So, S_n is a random set and N_n is a random variable defined in terms of the data as

$$S_n = \{i: 1 \leq i \leq n,\ X_i \in C_n\} \quad \text{and} \quad N_n = \#(S_n).$$

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

$$\sum_{i \in S_n} H_\alpha(Y_i - P_n(\beta, X_i)),$$

where $H_\alpha(t) = |t| + (2\alpha - 1)t$. Note here that since $0 < \alpha < 1$, $H_\alpha(t) \to \infty$ as $|t| \to \infty$. So, the above minimization problem always has a solution in view of the fact that, for any fixed value of N_n, this minimization problem becomes the problem of minimizing a continuous function over a bounded and closed subset of a linear subspace of R^{N_n}. Further, under the conditions of Theorem 3.1 and for $k \geq 1$, asymptotically (as $n \to \infty$) the above minimization problem will *almost surely* have a unique solution. On the other hand, when $k = 0$, the polynomial $P_n(\beta, x)$ 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 $\hat{\beta}_n$ can be defined to be the right endpoint of that interval (see the arguments at the end of the proof of Theorem 3.1).

Step 2: Set $\hat{T}_n = D^u P_n(\hat{\beta}_n, 0) = \delta_n^{-[u]} \hat{\beta}_{n,u}\{u!\}$. Here, for $u \in A$, we define $u! = \prod_{i=1}^d u_i!$ with the convention that $0! = 1$.

3. Main theorems. In order to derive asymptotic results about the estimate $\hat{T}_n$, we need some conditions to be satisfied.

CONDITION 3.1. The distribution of X_i is absolutely continuous on V with a density w, which is positive on V and continuous at 0.

CONDITION 3.2. ε_i has density f that is positive and Hölder continuous with a positive exponent η in an open neighborhood around 0.

Condition 3.1 [see Stone (1980, 1982)] ensures that, as $n \to \infty$, asymptotically there will be sufficiently many X_i's that will fall in the cube C_n. Condition 3.2 amounts to assuming that the distribution of the random error ε_i has a density that is positive and Hölder continuous in an open neighborhood of its αth quantile. This is satisfied by all classical examples of probability density functions.

THEOREM 3.1. *Assume that ε_i has a continuous distribution and that condition* 3.1 *is satisfied. Let E_n denote the event that the minimization problem in step* 2 *in the construction of $\hat{T}_n$ has a unique solution. Then, there is a set A_n, defined in terms of $X_1, X_2, \dots, X_n$, such that*

(i) $\Pr(\liminf A_n) = 1$,

(ii) *for $k \geq 1$, there is a positive integer N such that, for $n \geq N$, the conditional probability of E_n given that A_n has occurred is* 1.

THEOREM 3.2. *Let $r = (p - m)/(2p + d)$ and assume that the density f of ε_i is positive and continuous* (*note that this is weaker than Condition* 3.2) *in a neighborhood of* 0. *Then, under Condition* 3.1 *and whenever $\theta \in \Theta(c, k, \gamma)$, $|\hat{T}_n - T(\theta)|$ is almost surely of order $O(n^{-r}\sqrt{\log n})$ as $n \to \infty$.*

The next theorem in this section is on *a local Bahadur type representation* [Bahadur (1966), Kiefer (1967), Ruppert and Carroll (1980), Koenker and Portnoy (1987), Bhattacharya and Gangopadhyay (1990) and Dabrowska (1988)] of the estimate $\hat{\beta}_n$. But before we state the theorem we need to introduce some notation. Let $w_{\delta_n}(x)$ denote the conditional density of the vector $\delta_n^{-1}X_i$ given that X_i falls in C_n so that, for $x \in [-1, 1]^d$, $w_{\delta_n}(x) = w(\delta_n x)(\int_{[-1,1]^d} w(\delta_n x)\, dx)^{-1}$. Given a d-dimensional random vector X_i and a d-dimensional vector $u \in A$ (A is as in step 1 in the construction of $\hat{T}_n$), write $X_i(\delta_n, u) = \delta_n^{-[u]}(X_i)^u$ and define an $s(A)$-dimensional random vector $X_i(\delta_n, A) = (X_i(\delta_n, u))_{u \in A}$. For $u, v \in A$, let $q_n^{u,v}$ denote the conditional expectation of the product $X_i(\delta_n, u) X_i(\delta_n, v)$ given that $X_i \in C_n$. In other words, $q_n^{u,v} = \int_{[-1,1]^d} x^u x^v w_{\delta_n}(x)\, dx$. If Q_n denotes the $s(A) \times s(A)$ symmetric matrix with $q_n^{u,v}$ as a typical entry, Q_n is invertible in view of Condition 3.1. In fact, Condition 3.1 implies that $w_{\delta_n}(x)$ converges to the uniform probability density on $[-1,1]^d$ as $n \to \infty$. This convergence is uniform *in* $x \in [-1,1]^d$, and $q_n^{u,v}$ converges to the corresponding integral with respect to the uniform probability measure on $[-1,1]^d$.

Let β_n be the $s(A)$-dimensional vector $(\beta_{n,u})_{u \in A}$, where $\beta_{n,u} = D^u\theta(0)\delta_n^{[u]}\{u!\}^{-1}$. Let $\theta_n^*(x) = \sum_{u \in A}\beta_{n,u}x^u\delta_n^{-[u]}$ denote the Taylor polynomial of θ around 0 containing terms up to the kth order. So, we can write $\theta(x) = \theta_n^*(x) + r_n(x)$, where, in view of the choice of δ_n, $r_n(x)$ is uniformly of order $O(n^{-p/(2p+d)})$ for $x \in C_n$, $\theta \in \Theta(c,k,\gamma)$ and $n \to \infty$.

THEOREM 3.3 (Local Bahadur type representation of $\hat{\beta}_n$). *Under Conditions* 3.1 *and* 3.2 *and for* $\theta \in \Theta(c,k,\gamma)$, *one has the following asymptotic representation*:

$$\hat{\beta}_n - \beta_n = [N_n]^{-1}[f(0)]^{-1}[Q_n]^{-1}\sum_{i \in S_n} X_i(\delta_n, A)\left[\alpha - I(Y_i \le \theta_n^*(X_i))\right] + R_n,$$

where I is the standard zero–one valued indicator function and the remainder term R_n *satisfies*

$$\text{(i)}\quad R_n = O\left([\log n]^{(1+\eta)/2}n^{-p(1+\eta)/(2p+d)}\right) \quad \textit{almost surely, as } n \to \infty, \quad \textit{if } 0 < \eta < 1/2,$$

and

$$\text{(ii)}\quad R_n = O\left([\log n]^{3/4}n^{-3p/2(2p+d)}\right) \quad \textit{almost surely, as } n \to \infty, \quad \textit{if } 1/2 \le \eta \le 1.$$

[Note that when $N_n = 0$, one can define the first term on the right in the above representation to be 0. It does not matter, since $\Pr(N_n = 0) \to 0$ as $n \to \infty$.]

THEOREM 3.4. *Let r be as in Theorem* 3.2. *Then, under Conditions* 3.1 *and* 3.2,

$$\lim_{K \to \infty}\limsup_n \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta\left[\left|\hat{T}_n - T(\theta)\right| > Kn^{-r}\right] = 0.$$

4. Some useful asymptotic results. Several useful asymptotic results follow from the linear representation in Theorem 3.3. The results presented below are quite helpful in understanding the *bias–variance trade off* in the estimation of conditional quantiles via local polynomial fits. First note that the linear term in the linear representation in Theorem 3.3 can be written as

$$
\begin{aligned}
&[N_n]^{-1}[f(0)]^{-1}[Q_n]^{-1} \sum_{i \in S_n} X_i(\delta_n, A)[\alpha - I(Y_i \le \theta_n^*(X_i))] \\
&\quad = [N_n]^{-1}[f(0)]^{-1}[Q_n]^{-1} \\
&\qquad \times \sum_{i \in S_n} X_i(\delta_n, A)[F(\theta_n^*(X_i) - \theta(X_i)) - I(Y_i \le \theta_n^*(X_i))] \\
&\qquad + [N_n]^{-1}[f(0)]^{-1}[Q_n]^{-1} \sum_{i \in S_n} X_i(\delta_n, A)[\alpha - F(\theta_n^*(X_i) - \theta(X_i))],
\end{aligned} \tag{4.1}
$$

where F is the distribution function corresponding to the density f of ε_i. The second term on the right-hand side in (4.1), which we will denote by B_n, may be thought of as a *bias term* arising due to the polynomial approximation of θ in a neighborhood of 0.

PROPOSITION 4.1. *Under Conditions* 3.1 *and* 3.2, B_n *is uniformly of order* $O(n^{-p/(2p+d)})$ *as* $n \to \infty$ *and* $\theta \in \Theta(c, k, \gamma)$. *Note that here we have a deterministic bound for the random object* B_n. (*Recall that* B_n *is equal to* 0 *when* $N_n = 0$.)

PROOF. First note that $|X_i(\delta_n, A)| \le [s(A)]^{1/2}$ for $X_i \in C_n$. Also, by Condition 3.1, Q_n tends to a positive definite matrix Q as $n \to \infty$. So, the Euclidean norm of the matrix Q_n^{-1} remains bounded as $n \to \infty$. Also, in view of Condition 3.2, we have

$$
|\alpha - F(\theta_n^*(X_i) - \theta(X_i))| = |F(0) - F(-r_n(X_i))| = O(n^{-p/(2p+d)})
$$

as $n \to \infty$, uniformly for $\theta \in \Theta(c, k, \gamma)$ and $X_i \in C_n$. □

PROPOSITION 4.2. *Let* V_n *denote the first term on the right of* (4.1). *Then, under Conditions* 3.1 *and* 3.2, $[N_n]^{1/2}V_n$ *converges weakly to an* $s(A)$-*dimensional Gaussian random vector with zero mean and* $\alpha(1-\alpha)[f(0)]^{-2}Q^{-1}$ *as the dispersion matrix. Here* Q *is the limit of* Q_n *as* $n \to \infty$, *as mentioned in the proof of Proposition* 4.1.

PROOF. First note that V_n can be written as

$$
\begin{aligned}
&[N_n]^{-1}[f(0)]^{-1}[Q_n]^{-1} \\
&\quad \times \sum_{i=1}^{n} X_i(\delta_n, A)[F(\theta_n^*(X_i) - \theta(X_i)) - I(Y_i \le \theta_n^*(X_i))] I(X_i \in C_n).
\end{aligned} \tag{4.2}
$$

Now, N_n is a binomial random variable with the number of trials $= n$ and the probability parameter $= P(X_i \in C_n)$. In view of Condition 3.1, $(2\delta_n)^{-d}\Pr(X_i \in C_n) \to w(0)$ as $n \to \infty$. Hence, an application of Bernstein's inequality [see Pollard (1984), page 193, or Shorack and Wellner (1986), page 855] implies that $[n(2\delta_n)^d]^{-1}N_n$ converges *almost surely* to $w(0)$ as $n \to \infty$. On the other hand, (4.2) is a sum of n i.i.d. random vectors, each of which is bounded and has mean 0. The desired result now follows from a simple calculation of the dispersion matrix of the sum, Conditions 3.1 and 3.2, and an application of Lindeberg's theorem. □

5. Discussion and remarks. 1. The rate of convergence obtained in Theorem 3.4 is optimal by Stone (1980) and provides an affirmative answer to question 4 raised by Stone (1982) as far as the pointwise estimation of conditional quantile functions or their derivatives is concerned. The attainability of the optimal nonparametric rates of convergence in the global sense (i.e., in L_q-norms restricted to compacts, with $1 \leq q \leq \infty$) will be considered separately in another paper [see Stone (1982) and Truong (1989)].

2. The usefulness of conditional quantile functions as good descriptive statistics has been discussed by Hogg (1975), who called them percentile regression lines. Janssen and Veraverbeke (1987), Lejeune and Sarda (1988) and Nahm (1989) explored some nonparametric estimates of regression using the idea of quantile regression. Koenker and Portnoy (1987) obtained a uniform Bahadur type representation for regression quantiles in a linear parametric setup under rather strong assumptions on the density of the error distribution. In Theorem 3.3, a weaker assumption on the density of the error distribution has been used and results have been proved in a nonparametric setup. We have obtained results on the *almost sure* behavior of the remainder term R_n whereas Koenker and Portnoy (1987) obtained an asymptotic linear representation that is valid *in probability*. The arguments used in the proof of Theorem 3.3 are based in part on considerable simplifications of some of the ideas in Koenker and Portnoy (1987). Ruppert and Carroll (1987) also obtained a similar representation in a parametric setup. Both of these two papers may be viewed as generalizations of earlier work by Bahadur (1966) and Ghosh (1971).

3. The asymptotic normality of estimators of conditional quantiles has been proved by Cheng (1983, 1984), who considered kernel estimators in the fixed design case, and by Stute (1986), who considered nearest-neighbor type estimators in the random design case. The linear representation in Theorem 3.3 can be used as an elegant toll to derive results on the joint asymptotic behavior of nonparametric estimates of several regression quantiles for different values of α. One has to carry out calculations similar to those in the proofs of Propositions 4.1 and 4.2 to get such results. Koenker and Bassett (1978) looked at a similar problem from the parametric point of view.

4. Bhattacharya and Gangopadhyay (1990) obtained a Bahadur type representation of local quantiles under a setup different from but related to the present one. They restricted their attention to a real valued regressor X and to kernel and nearest-neighbor estimates. In the special case $p = 2$ and $d = 1$, the rates obtained in Theorem 3.3 are comparable with those obtained by them. Actually the rates obtained in Theorem 3.3 are faster than those obtained by Bhattacharya and Gangopadhyay (1990). Dabrowska (1987) obtained a Bahadur type representation for nonparametric estimates of conditional quantiles with censored data. However, she worked under a setup that is very much different from the present one. As a result, the estimates studied by

her are quite different in nature and so are the rates of convergence obtained by her.

5. One can prove a weaker version of Theorem 3.3 (with a slower rate of convergence for the remainder term) under weaker conditions on the density f. However, then the arguments become quite tedious and complicated [see Chaudhuri (1988)].

6. Proofs of the theorems in Section 3. In several places in the proofs that follow, the following simple facts will be used.

FACT 6.1. Let x be a vector in R^m and $p(x)$ be a nonzero polynomial in x. Then the Lebesgue measure of the set $\{x | p(x) = 0\}$ is 0.

FACT 6.2. Let $X^{(1)}, X^{(2)}, \dots, X^{(m)}$ be independent random vectors in R^m with the property that $\Pr(X^{(i)} \in H) = 0$ for any i and any given linear subspace H of R^m with $\dim(H) \le m - 1$. Then the collection $\{X^{(1)}, X^{(2)}, \dots, X^{(m)}\}$ is *almost surely* linearly independent.

We need to introduce some additional notation at this point. 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 $\mathbf{s}$ of $\mathbf{S}$, we will denote by $\mathbf{X}(\mathbf{s})$ the submatrix (subvector) of $\mathbf{X}$ with rows (components) that are indexed by the elements of $\mathbf{s}$ [Koenker and Bassett (1978)]. Let DX_n be the matrix of dimension $N_n \times s(A)$, whose rows are the vectors $X_i(\delta_n, A)$, where $i \in S_n$. So, it is natural to assume that the rows of DX_n are indexed by the elements of S_n and its columns are indexed by the elements of A. Also, we will write VY_n to denote the vector whose components are Y_i for $i \in S_n$. Let H_n be the collection of all subsets h of S_n such that $\#(h) = s(A) = \#(A)$.

The following two facts, which are essentially restatements of two theorems from Koenker and Bassett (1978) [see also Theorem 1 in page 7 in Bloomfield and Steiger (1983)], play crucial roles in the proofs of the theorems stated in Section 3. *In the remaining part of the paper, the minimization problem in step* 1 *in the construction of* $\hat{T}_n$ *will be referred to as problem* (P).

FACT 6.3 [Theorem 3.1 in Koenker and Bassett (1978)]. Suppose that the matrix DX_n has rank $= s(A)$. Then, there is a subset h of S_n with $\#(h) = s(A)$ such that problem (P) has at least one solution of the form $\hat{\beta}_n = [DX_n(h)]^{-1} VY_n(h)$. So, for such a $\hat{\beta}_n$, we have $Y_i = P_n(\hat{\beta}_n, X_i)$ for all $i \in h$.

FACT 6.4 [Theorem 3.3 in Koenker and Bassett (1978)]. Suppose that the matrix DX_n has rank $= s(A)$, and, for $h \in H_n$, define

$$L_n(h) = \sum_{i \in h^c} \left[\tfrac{1}{2} - \tfrac{1}{2}\operatorname{sgn}\left(Y_i - \langle X_i(\delta_n, A), \hat{\beta}_n \rangle\right) - \alpha\right] \times [DX_n(h)]^{-1} X_i(\delta_n, A),$$

where $\operatorname{sgn}(x)$ is $+1$ or -1 depending on whether x is positive or negative respectively, $\langle , \rangle$ is the usual Euclidean inner product and h^c is the set theoretic complement of h in S_n. Then, under the assumption that ε_i has a continuous distribution and is independent of X_i, $\hat{\beta}_n = [DX_n(h)]^{-1}VY_n(h)$ is a unique solution to problem (P) if and only if $L_n(h) \in (\alpha - 1, \alpha)^{s(A)}$. Further, if $\hat{\beta}_n = [DX_n(h)]^{-1}VY_n(h)$ is a solution (not necessarily unique) to problem (P), we must have $L_n(h) \in [\alpha - 1, \alpha]^{s(A)}$. Here $(\alpha - 1, \alpha)^{s(A)}$ and $[\alpha - 1, \alpha]^{s(A)}$ are $s(A)$-dimensional cubes in $R^{s(A)}$—the former being an open one and the latter being a closed one.

Proof of Theorem 3.1. N_n is a binomial random variable, which counts the number of X_i's that fall in the cube C_n. So, in view of Condition 3.1 and Bernstein's inequality [see Pollard (1984)], we can choose positive constants c_1, c_2, c_3, c_4 such that, for all n, we have

$$\Pr\left(\{c_1 n \delta_n^d \le N_n \le c_2 n \delta_n^d\}\right) \ge 1 - c_3 \exp\left(-c_4 n \delta_n^d\right).$$

Now, define A_n to be the event enclosed in { } above. Then, the present choice of δ_n ensures that $\sum_n \Pr(A_n^c) < \infty$, and, by an application of the Borel–Cantelli lemma, we have $\Pr(\liminf A_n) = 1$.

Recall at this point that, for n suitably large, the conditional distribution of $\delta_n^{-1}X_i$, given that $X_i \in C_n$ is absolutely continuous with respect to the Lebesgue measure. Also, note that, given the set S_n (i.e., given the subscripts of those X's which fall in C_n), the vectors $\delta_n^{-1}X_i$ for $i \in S_n$ are conditionally independently distributed. So, in view of Facts 6.1 and 6.2, the matrix DX_n is going to have rank $= s(A)$ with conditional probability 1 given the set S_n, provided that $N_n \ge s(A)$. Further, in the case $k \ge 1$, under Condition 3.1 and in view of Fact 6.1, the absolute continuity of the conditional distribution of $\delta_n^{-1}X_i$ (given that $X_i \in C_n$) prevents $L_n(h)$ from sitting on the boundary of the cube $[\alpha - 1, \alpha]^{s(A)}$ by making the conditional probability of such an event 0 (given the set S_n) whenever $N_n \ge s(A)$.

The proof of the theorem is now complete in view of Facts 6.3 and 6.4, if we choose N appropriately large so that $n \ge N$ implies that $c_1 n \delta_n^d \ge s(A)$.

When $k = 0$ [which implies $s(A) = 1$], problem (P) may not in general have a unique solution (in this case β_n and $\hat{\beta}_n$ become real numbers and the matrix DX_n becomes a vector of 1's). However, in this case, if $c_1 n \delta_n^d \ge 1$, the occurrence of the event A_n implies that the solution set for problem (P) is a compact interval and we can define $\hat{\beta}_n = Y_i$ for some appropriate $i \in S_n$ so that $\hat{\beta}_n$ is the right endpoint of that interval. So, here $\hat{\beta}_n$ is defined as an αth quantile (*local quantile*) of those Y_i's for which $X_i \in C_n$. □

We will now derive some results that will be used in the proof of Theorem 3.2.

Proposition 6.1. *Let F be a distribution function on the real line such that $F(0) = \alpha$, $0 < \alpha < 1$, and F has a density that is continuous and positive in a neighborhood of* 0. *For $x \in R^d$, let $x(A)$ denote the $s(A)$-dimensional vector defined as $x(A) = (x^u)_{u \in A}$ (see step* 1 *in the construction of $\hat{T}_n$). Also, for*

$0 \le \delta \le 1$, let $R(\delta, x)$ be a real valued function with the property that there is $M_1 > 0$ such that $|R(\delta, x)| \le M_1\delta$ for all $x \in [-1, 1]^d$. For $\Delta \in R^{s(A)}$, define an $s(A)$-dimensional vector valued function $G(\Delta, \delta)$ as

$$G(\Delta, \delta) = \int_{[-1,1]^d} \{F(\langle \Delta, x(A) \rangle + R(\delta, x)) - \alpha\} x(A)\, dx.$$

Then, there exist $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $c_5 > 0$ and $M_2 > 0$ such that we have $|G(\Delta, \delta)| \ge$ minimum of (ε_1 and $c_5|\Delta|$) whenever $\delta \le \varepsilon_2$ and $|\Delta| \ge M\delta$, where $M \ge M_2$.

Proof. The proof is somewhat long and therefore we will split it into several steps.

Step 1: We begin by proving that there exist $M_3 > 0$ and $\varepsilon_1 > 0$ such that $|G(\Delta, \delta)| \ge \varepsilon_1$ whenever $|\Delta| \ge M_3$. Assume that this assertion is false. Then, we can construct a sequence $\{\Delta_n\}$ in $R^{s(A)}$ with the property that, as $n \to \infty$, $|\Delta_n| \to \infty$, $\Delta_n/|\Delta_n| \to \Delta^* \in R^{s(A)}$ such that $|\Delta^*| = 1$, and $|G(\Delta_n, \delta)| \to 0$. Since $|R(\delta, x)| \le M_1\delta \le M_1$ for all $x \in [-1, 1]^d$ and $0 \le \delta \le 1$, as $n \to \infty$, $F(\langle \Delta_n, x(A) \rangle + R(\delta, x))$ must tend to 1 or 0 depending on whether $\langle \Delta^*, x(A) \rangle$ is positive or negative respectively. Further, in view of Fact 6.1, the region $[-1, 1]^d \cap \{\langle \Delta^*, x(A) \rangle = 0\}$ must have Lebesgue measure 0. Hence, if $|G(\Delta_n, \delta)| \to 0$ as $n \to \infty$, by a simple application of the dominated convergence theorem, we get

$$(1-\alpha)\int_{[-1,1]^d \cap \{\langle \Delta^*, x(A)\rangle > 0\}} x(A)\, dx = \alpha \int_{[-1,1]^d \cap \{\langle \Delta^*, x(A)\rangle < 0\}} x(A)\, dx. \tag{6.1}$$

By taking the inner product of each side of (6.1) with Δ^*, we get

$$(1-\alpha)\int_{[-1,1]^d \cap \{\langle \Delta^*, x(A)\rangle > 0\}} \langle \Delta^*, x(A) \rangle\, dx = \alpha \int_{[-1,1]^d \cap \{\langle \Delta^*, x(A)\rangle < 0\}} \langle \Delta^*, x(A) \rangle\, dx.$$

Since $0 < \alpha < 1$, the above implies that each of the regions $[-1, 1]^d \cap \{\langle \Delta^*, x(A) \rangle > 0\}$ and $[-1, 1]^d \cap \{\langle \Delta^*, x(A) \rangle < 0\}$ must have Lebesgue measure 0. This is a direct contradiction to Fact 6.1.

Step 2: Write $G(\Delta, \delta)$ as

$$G(\Delta, \delta) = \int_{[-1,1]^d} \{F(\langle \Delta, x(A) \rangle + R(\delta, x)) - F(R(\delta, x))\} x(A)\, dx + \int_{[-1,1]^d} \{F(R(\delta, x)) - F(0)\} x(A)\, dx. \tag{6.2}$$

Then, for $x \in [-1, 1]^d$, $0 < \delta < 1$ and Δ a nonzero vector in $R^{s(A)}$, define a real valued function $g(\Delta, \delta, x)$ as

$$g(\Delta, \delta, x) = \frac{F(\langle \Delta, x(A)\rangle + R(\delta, x)) - F(R(\delta, x))}{\langle \Delta, x(A)\rangle}.$$

When $\langle \Delta, x(A)\rangle = 0$, $g(\Delta, \delta, x)$ can be defined arbitrarily because, for nonzero Δ, the set $[-1, 1]^d \cap \{\langle \Delta, x(A)\rangle = 0\}$ has Lebesgue measure 0. Now, in view of Condition 3.2, there exists $b > 0$ such that the density f is continuous and positive in the interval $[-b, b]$. Let $\varepsilon_2 > 0$ be chosen in such a way that $\delta \le \varepsilon_2$ implies that $|R(\delta, x)| \le M_1\delta < b$. Then, using the function $g(\Delta, \delta, x)$ and applying the mean value theorem of differential calculus to the second term on the right of (6.2), we get

$$\begin{aligned} G(\Delta, \delta) &= \left[\int_{[-1,1]^d} g(\Delta, \delta, x)x(A)x(A)^{\mathrm{T}}\, dx\right]\Delta \\ &\quad + \int_{[-1,1]^d} \{f(\xi_1 R(\delta, x))R(\delta, x)\}x(A)\, dx, \end{aligned} \tag{6.3}$$

where ξ_1 is a number lying between 0 and 1 and may depend on x and δ. We have assumed here that $x(A)$ and Δ are column vectors, and T denotes transpose.

Step 3: Note at this point that each component of the $s(A)$-dimensional vector $x(A)$ is bounded by 1 for $x \in [-1, 1]^d$. Let $\sup_{-b \le t \le b} f(t) = \lambda_1$. Then, for the second term on the right of (6.3), we have

$$\begin{aligned} &\left|\int_{[-1,1]^d} \{f(\xi_1 R(\delta, x))R(\delta, x)\}x(A)\, dx\right|^2 \\ &\quad \le \lambda_1^2[s(A)]2^d \int_{[-1,1]^d} |R(\delta, x)|^2\, dx \le \lambda_1^2[s(A)]2^{2d}M_1^2\delta^2. \end{aligned} \tag{6.4}$$

Step 4: Since the density f is bounded away from 0 on the compact interval $[-b, b]$, we can determine a positive constant λ_2 such that $g(\Delta, \delta, x) \ge \lambda_2 > 0$ for all $\delta \le \varepsilon_2$, $|\Delta| \le M_3$ and $x \in [-1, 1]^d$ such that $\langle \Delta, x(A)\rangle$ is nonzero. Let $\rho > 0$ be the smallest eigenvalue of the $s(A) \times s(A)$ matrix $\int_{[-1,1]^d} x(A)x(A)^{\mathrm{T}}\, dx$. Then, for the first term on the right of (6.3), we have

$$\left|\left[\int_{[-1,1]^d} g(\Delta, \delta, x)x(A)x(A)^{\mathrm{T}}\, dx\right]\Delta\right| \ge \lambda_2\rho|\Delta| \tag{6.5}$$

whenever $\delta \le \varepsilon_2$ and $|\Delta| \le M_3$.

Now, using (6.4) and (6.5), the assertion in the proposition follows by setting $c_5 = \lambda_2\rho/2$ and choosing $M_2 > 2\lambda_1[s(A)]^{1/2}2^d M_1/\lambda_2\rho$. □

For $x \in [-1, 1]^d$, recall from Section 3 that $w_{\delta_n}(x)$ denotes the conditional density of $\delta_n^{-1}X_i$ given that $X_i \in C_n$ and it uniformly converges to the uniform probability density on $[-1, 1]^d$. Hence, some minor modifications of the arguments in the proof above lead to the following fact.

FACT 6.5. Let $G_n(\Delta, \delta)$ be defined as

$$G_n(\Delta, \delta) = \int_{[-1,1]^d} \{F(\langle \Delta, x(A)\rangle + R(\delta, x)) - \alpha\} x(A) w_{\delta_n}(x)\, dx,$$

where F, Δ, δ, R and $x(A)$ are exactly the same as in Proposition 6.1. Then, there exist positive constants ε_1^*, ε_2^*, c_5^* and M_2^* such that we have $|G_n(\Delta, \delta)| \geq$ minimum of (ε_1^* and $c_5^*|\Delta|$) whenever $\delta \leq \varepsilon_2^*$ and $|\Delta| \geq M\delta$, where $M \geq M_2^*$.

PROOF OF THEOREM 3.2. For some positive constant K_1, let U_n be the event defined as

$$U_n = \left\{ \left| \hat{\beta}_n - \beta_n \right| \geq K_1 \delta_n^p \sqrt{\log n} \right\}.$$

Now, in view of the definition of δ_n and $\hat{T}_n$, the assertion in the theorem will follow (by an application of the Borel–Cantelli lemma) if $K_1 > 0$ can be appropriately chosen so that

$$\sum_n \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n) < \infty. \tag{6.6}$$

In view of the construction of A_n in the proof of Theorem 3.1, (6.6) is equivalent to

$$\sum_n \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(U_n \cap A_n) < \infty. \tag{6.7}$$

We will now try to get an upper bound for $P_\theta(U_n \cap A_n)$. To this end, given $\Delta_n \in R^{s(A)}$, set

$$Z_{n,i} = \left[\tfrac{1}{2} - \tfrac{1}{2}\operatorname{sgn}(\varepsilon_i - \langle \Delta_n, X_i(\delta_n, A)\rangle + r_n(X_i)) - \alpha\right] X_i(\delta_n, A).$$

Obviously, the norm of the random vector $Z_{n,i}$ is bounded by $[s(A)]^{1/2}$ whenever $i \in S_n$. Hence, using Facts 6.3 and 6.4, we get that there is a positive constant ϕ_1 [which depends only on $s(A)$] such that the event $U_n \cap A_n$ is contained in the event

$$\Bigg\{ \text{for some } h \in H_n, \left| \sum_{i \in h^c} Z_{n,i} \right| \leq \phi_1, \text{ where } \Delta_n = \hat{\beta}_n - \beta_n,$$
$$\hat{\beta}_n = [DX_n(h)]^{-1} VY_n(h) \text{ and } |\Delta_n| \geq K_1 \delta_n^p \sqrt{\log n} \Bigg\} \cap A_n.$$

Now, given the set S_n and the X_i's and Y_i's for $i \in h$ (where h is some fixed element of H_n), the vectors $Z_{n,i}$ for $i \in h^c$ are conditionally independently and identically distributed, each with conditional mean $G_n(\Delta_n, \delta_n^p)$, where we can have $\Delta_n = [DX_n(h)]^{-1}VY_n(h) - \beta_n$ [take $\Delta = \Delta_n$, $\delta = \delta_n^p$ and $R(\delta, x) = -r_n(\delta_n x)$ in the statement of Fact 6.5]. Using Fact 6.5, for K_1 (which behaves like M in the statement of Fact 6.5) and n appropriately large, and for $|\Delta_n| > K_1 \delta_n^p \sqrt{\log n}$, we have (since $\delta_n^p \sqrt{\log n} \to 0$ as $n \to \infty$) that $|G_n(\Delta_n, \delta_n^p)| \geq c_5^* K_1 \delta_n^p \sqrt{\log n}$. Further, note that $\#(h^c) = N_n - \#(h) = N_n - s(A)$, and the occurrence of the event A_n implies that $c_1 n \delta_n^d \leq N_n \leq c_2 n \delta_n^d$

and $\#(H_n) \le (c_2 n \delta_n^d)^{s(A)}$. Hence, we can choose constants $c_6 > 0$, $c_7 > 0$ and an integer $N_1 > 0$ so that, by applying Bernstein's inequality [see Pollard (1984)] to the sum $\sum_{i \in h^c} Z_{n,i}$, it follows that

$$\begin{aligned} P_\theta(U_n \cap A_n) &\le c_6\big(n\delta_n^d\big)^{s(A)} \exp\big(-c_7 n \delta_n^{d+2p} \log n\big) \\ &= c_6\big(n\delta_n^d\big)^{s(A)} \exp(-ac_7 \log n) \end{aligned} \tag{6.8}$$

whenever $n \ge N_1$, $\theta \in \Theta(c, k, \gamma)$ and K_1 is appropriately large. Further, in view of Fact 6.5, by choosing K_1 suitably, c_7 can be chosen as large as desired. Finally, a suitable choice of K_1 giving an appropriate value of c_7 (depending on p and d) ensures (6.7). □

Proof of Theorem 3.3. The main arguments in the proof will be split into several steps.

Step 1: Let use define

$$\begin{aligned} \tilde{H}_n(\delta_n, \beta_n) &= \int_{[-1,1]^d} F(\langle \beta_n, x(A)\rangle - \theta(\delta_n x))x(A) w_{\delta_n}(x)\, dx \\ &= \int_{[-1,1]^d} F(\theta_n^*(\delta_n x) - \theta(\delta_n x))x(A) w_{\delta_n}(x)\, dx \\ &= \int_{[-1,1]^d} F(-r_n(\delta_n x))x(A) w_{\delta_n}(x)\, dx, \\ \tilde{H}_n\big(\delta_n, \hat{\beta}_n\big) &= \int_{[-1,1]^d} F\big(\langle \hat{\beta}_n, x(A)\rangle - \theta(\delta_n x)\big)x(A) w_{\delta_n}(x)\, dx \\ &= \int_{[-1,1]^d} F\big(\hat{\theta}_n(\delta_n x) - \theta(\delta_n x)\big)x(A) w_{\delta_n}(x)\, dx \end{aligned}$$

and

$$R_n^{(1)} = \tilde{H}_n\big(\delta_n, \hat{\beta}_n\big) - \tilde{H}_n(\delta_n, \beta_n) - [f(0)]Q_n\big(\hat{\beta}_n - \beta_n\big).$$

If we assume that all the β's and $x(A)$'s are column vectors, the third term in the expression defining $R_n^{(1)}$ can be written as

$$[f(0)]Q_n\big(\hat{\beta}_n - \beta_n\big) = [f(0)]\int_{[-1,1]^d} [x(A)][x(A)]^{\mathrm{T}}\big(\hat{\beta}_n - \beta_n\big) w_{\delta_n}(x)\, dx.$$

So, $R_n^{(1)}$ can be rewritten as

$$\begin{aligned} R_n^{(1)} = \int_{[-1,1]^d} [x(A)]\Big\{&\big[F\big(\langle \hat{\beta}_n, x(A)\rangle - \theta(\delta_n x)\big)\big] \\ &-[F(\langle \beta_n, x(A)\rangle - \theta(\delta_n x))] \\ &-\langle x(A), \hat{\beta}_n - \beta_n\rangle [f(0)]\Big\} w_{\delta_n}(x)\, dx \\ = \int_{[-1,1]^d} [x(A)]\{&[F(\psi_n(x) - r_n(\delta_n x))] \\ &-[F(-r_n(\delta_n x))] - [\psi_n(x)][f(0)]\} w_{\delta_n}(x)\, dx, \end{aligned}$$

where $\psi_n(x) = \hat{\theta}_n(\delta_n x) - \theta_n^*(\delta_n x)$. Hence, under Conditions 3.1 and 3.2 and using Theorem 3.2, we have

$$R_n^{(1)} = O\big([\log n]^{(1+\eta)/2} n^{-p(1+\eta)/(2p+d)}\big) \tag{6.9}$$

almost surely as $n \to \infty$. Further, in view of the proof of Theorem 3.1 and the definition of $\Theta(c, k, \gamma)$, this rate of convergence must be uniform for $\theta \in \Theta(c, k, \gamma)$.

Step 2: Define an $s(A)$-dimensional random vector χ_n as

$$\chi_n = \sum_{i \in S_n} \big[X_i(\delta_n, A) I\big(Y_i \le \hat{\theta}(X_i)\big) - \tilde{H}_n(\delta_n, \hat{\beta}_n)\big] - \sum_{i \in S_n} \big[X_i(\delta_n, A) I\big(Y_i \le \theta_n^*(X_i)\big) - \tilde{H}_n(\delta_n, \beta_n)\big],$$

where I is the standard zero–one valued indicator function. For some constant $K_2 > 0$, let W_n be the event defined as

$$W_n = \big\{|\chi_n| \ge K_2 n^{p/2(2p+d)}[\log n]^{3/4}\big\}.$$

Also, for $h \in H_n$, define (assuming that n is appropriately large)

$$\hat{\beta}_n^h = [DX_n(h)]^{-1} VY_n(h), \qquad \hat{\theta}_n^h(X_i) = \langle \hat{\beta}_n^h, X_i(\delta_n, A)\rangle$$

and

$$\chi_n^h = \sum_{i \in h^c} \big[X_i(\delta_n, A) I\big(Y_i \le \hat{\theta}_n^h(X_i)\big) - \tilde{H}_n(\delta_n, \hat{\beta}_n^h)\big] - \sum_{i \in h^c} \big[X_i(\delta_n, A) I\big(Y_i \le \theta_n^*(X_i)\big) - \tilde{H}_n(\delta_n, \beta_n)\big].$$

Then, in view of the definition of the events A_n (proof of Theorem 3.1) and U_n (proof of Theorem 3.2) and using Fact 6.3, the event $W_n \cap A_n \cap U_n^c$ is contained in the event

$$\Big\{\text{for some } h \in H_n, \big|\chi_n^h\big| > K_3[\log n]^{3/4} n^{p/2(2p+d)} \text{ and } \big|\hat{\beta}_n^h - \beta_n\big| \le K_1\big[\sqrt{\log n}\,\big] n^{-p/(2p+d)}\Big\} \cap A_n$$

for n sufficiently large. Here $K_3 = K_2'/2$ and K_1 is as in the proof of Theorem 3.2 ensuring (6.6). Note that here we are using the fact that $[\log n]^{3/4} n^{p/2(2p+d)} \to \infty$ as $n \to \infty$, whereas $\#(h) = s(A)$ stays bounded. Now, given the set S_n and also the X_i's and Y_i's for which $i \in h$ ($h \in H_n$ is fixed), the terms in the sum defining χ_n^h are conditionally independently and identically distributed and each of them has conditional mean 0. Further, in view of Condition 3.2, each term in the sum has a conditional dispersion matrix whose Euclidean norm is of order $O(n^{-p/(2p+d)}\sqrt{\log n}\,)$ provided that $|\hat{\beta}_n^h - \beta_n| \le K_1[\sqrt{\log n}\,]n^{-p/(2p+d)}$ and n is suitably large. Now, since the occurrence of the event A_n implies $c_1 n \delta_n^d \le N_n \le c_2 n \delta_n^d$, we can choose a positive integer N_2

and positive constants c_8 and c_9 so that an application of Bernstein's inequality [see Pollard (1984)] to the sum defining χ_n^h gives

$$\sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(W_n \cap A_n \cap U_n^c) \le c_8 (n\delta_n^d)^{s(A)} \exp(-c_9 \log n)$$

for all $n \ge N_2$. Further, c_9 can be chosen as large as desired by choosing K_2 sufficiently large and thereby making K_3 sufficiently large. Now, by making an appropriate choice of c_9 (depending on p and d) we have

$$\sum_n \sup_{\theta \in \Theta(c,k,\gamma)} P_\theta(W_n \cap A_n \cap U_n^c) \le \infty.$$

This, in view of the definition of A_n and U_n, ensures that

$$\chi_n = O\big([\log n]^{3/4} n^{p/2(2p+d)}\big) \tag{6.10}$$

almost surely as $n \to \infty$. Also, note that this asymptotic order of χ_n is uniform for $\theta \in \Theta(c,k,\gamma)$.

Step 3: Note at this point that for n sufficiently large (as argued in the proof of Theorem 3.2 using Facts 6.3 and 6.4)

$$\left| \sum_{i \in S_n} X_i(\delta_n, A)\big[I\big(Y_i \le \hat{\theta}_n(X_i)\big) - \alpha\big] \right| = \left| \sum_{i \in S_n} Z_{n,i} \right| \le \phi_2,$$

where ϕ_2 is a positive constant that depends only on $s(A)$. Hence, for the summand in the linear term in the linear representation in Theorem 3.3, we can write

$$\begin{aligned} &[N_n]^{-1} \sum_{i \in S_n} X_i(\delta_n, A)\big[\alpha - I(Y_i \le \theta_n^*(X_i))\big] \\ &\quad = [N_n]^{-1}\chi_n + \tilde{H}_n(\delta_n, \hat{\beta}_n) - \tilde{H}_n(\delta_n, \beta_n) + R_n^{(2)}, \end{aligned} \tag{6.11}$$

where on the event A_n (the occurrence of which implies $c_1 n\delta_n^d \le N_n \le c_2 n\delta_n^d$), we have

$$R_n^{(2)} \le \frac{\phi_2}{c_1} n^{-2p/(2p+d)}. \tag{6.12}$$

Finally, the assertion in the theorem follows by using (6.9), (6.10), (6.11), (6.12) and the fact that Q_n tends to a positive definite matrix Q as $n \to \infty$. □

PROOF OF THEOREM 3.4. First note that, in view of the arguments in the proof of Proposition 4.2, the conditional mean of $n^{2p/(2p+d)}|V_n|^2$, given the occurrence of the event A_n (A_n is as in Theorem 3.1) remains uniformly bounded for $\theta \in \Theta(c,k,\gamma)$ and $n \to \infty$. The proof is now complete in view of the proofs of Theorem 3.3 and Proposition 4.1. □

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 Le Cam for their help and many stimulating and thought-provoking discussions. An anonymous Associate Editor spent a lot of time going through various details of an earlier draft of this paper. His constructive criticisms and suggestions have improved the presentation beyond measure.

REFERENCES

BADAHUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BARRADOLE, I. and ROBERTS, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximation. *SIAM J. Numer. Anal.* **10** 839–848.

BARRADOLE, I. and ROBERTS, F. D. K. (1974). Algorithm 478: Solution of an overdetermined system of equations in the L_1 norm. *CACM* **14** 319–320.

BARTELS, H. H., CONN, A. R. and SINCLAIR, J. W. (1978). Minimization techniques for piecewise differentiable functions: The L_1 solution to an overdetermined system. *SIAM J. Numer. Anal.* **15** 224–241.

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

BLOOMFIELD, P. and STEIGER, W. L. (1980). Least absolute deviations curve fitting. *SIAM J. Sci. Statist. Comput.* **1** 290–301.

BLOOMFIELD, P. and STEIGER, W. L. (1983). *Least Absolute Deviations, Theory, Applications and Algorithms*. Birkhäuser, Boston.

CHAUDHURI, P. (1988). Asymptotic theory of nonparametric estimation of conditional quantiles. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

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

CHENG, K. F. (1984). Nonparametric estimation of regression function using linear combinations of sample quantile regression function. *Sankhyā Ser. A* **46** 287–302.

DABROWSKA, D. M. (1987). Nonparametric quantile regression with censored data. Technical Report 93, Dept. Statistics, Univ. California, Berkeley.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

DONOHO, D. L. and LIU, R. C. (1988). Geometrizing rates of convergence, I, II and III. Technical Reports 120, 137 and 138, Dept. Statistics, Univ. California, Berkeley.

FERGUSON, T. (1967). *Mathematical Statistics. A Decision Theoretic Approach*. Academic, New York.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

GONIN, R. and MONEY, A. H. (1989). *Nonlinear L_p-Norm Estimation*. Dekker, New York.

HÄRDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.* **14** 169–180.

HÄRDLE, W. and GASSER, T. (1984). Robust nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.

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

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

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

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

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

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

KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and D'OREY, V. (1987). Computing regression quantiles. *J. Roy. Statist. Soc. Ser. C* **36** 383–393.

KOENKER, R. and PORTNOY, S. (1987). L-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

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

NAHM, J. W. (1989). Nonparametric least absolute deviations estimation. Ph.D. dissertation, Dept. Economics, Univ. Wisconsin, Madison.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

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

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

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

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

STONE, C. J. (1982). Optimal global rates of convergence for non-parametric regression. *Ann. Statist.* **10** 1040–1053.

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

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

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

STATISTICAL MATHEMATICS DIVISION
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA 700035
INDIA