The Annals of Statistics
1993, Vol. 21, No. 2, 924–942

AN IMPROVED MONOTONE CONDITIONAL QUANTILE ESTIMATOR

BY HARI MUKERJEE

Wichita State University

Suppose that $(X_1, Y_1), \dots, (X_n, Y_n)$ are i.i.d. bivariate random vectors and that $\xi_p(x)$ is the p-quantile of Y_1 given $X_1 = x$ for $0 < p < 1$. Estimation of $\xi_p(x)$, when it is monotone in x, has been studied in the literature. In the nonparametric conditional quantile estimation one uses only some smoothness assumptions. The asymptotic properties are superior in the latter case; however, monotonicity is not guaranteed. We introduce a new estimator that enjoys both of the above properties.

1. Introduction. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random vectors distributed as (X, Y) and let $\xi_p(x)$ be the p-quantile of Y given $X = x$ for $0 < p < 1$. In many applications it is reasonable to assume that $\xi_p(x)$ is nondecreasing in x, $0 < p < 1$; the nonincreasing case is similar. Fix $0 < p < 1$. Let $G(\cdot|x)$ be the conditional d.f. of Y given $X = x$. For $A \subset \mathbb{R}$ let

$$N(n, A) = \#\{1 \le i \le n : X_i \in A\}$$

and, assuming $N(n, A) > 0$, let

$$\hat{G}_n(y|A) = \frac{1}{N(n, A)} \sum I[Y_i \le y,\ X_i \in A],$$

the e.d.f. of $\{Y_i : X_i \in A\}$ evaluated at y,

$$\hat{\xi}_{pn}(A) = \text{the}[N(n, A)p]\text{th order statistic of } \{Y_i : X_i \in A\}$$
$$= \inf\{t : \hat{G}_n(t|A) \ge [N(n, A)p]/N(n, A)\}.$$

Throughout this paper $N(n, A) \to \infty$ a.s. for every $A \subset \mathbb{R}$ that we consider so that the above quantities are well defined a.s. for all large n. We will assume that n is at least this large without explicit mention. For notational simplicity we will omit the subscripts p and n.

In analogy with the monotone median estimator due to Robertson and Waltman (1968) that was studied further by Cryer, Robertson, Wright and Casady (1972), Casady and Cryer (1976) introduced the nondecreasing estimator $\xi^{**}(x)$ of $\xi(x)$ by

$$\xi^{**}(x) = \max_{r \le x} \min_{s \ge x} \hat{\xi}([r, s]), \tag{1.1}$$

and studied its consistency and convergence rate. Wright (1984) showed that

Received April 1991; revised March 1992.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G30.
Key words and phrases. Monotone conditional quantiles, Bahadur representation.

$n^{1/3}[\xi^{**}(x) - \xi(x)]$ has an asymptotic nondegenerate (nonnormal) distribution when $\xi'(x) > 0$.

Bhattacharya and Gangopadhyay (1990) [to be referred to as BG (1990)] found a Bahadur-type representation of the kernel estimator $\hat{\xi}([x - h/2, x + h/2])$ under some smoothness assumptions on the distributions, where h is the bandwidth. For the optimal bandwidth, $h \propto n^{-1/5}$, this representation yields asymptotic normality of the estimator with the norming of $n^{2/5}$. Moreover, using $h(t) = n^{-1/5}t$ with $c \le t \le d$ for some $0 < c < d < \infty$, this representation also yields weak convergence of the stochastic process $\{\hat{\xi}([x - h(t)/2, x + h(t)/2])\colon c \le t \le d\}$, which in turn yields an asymptotic linear model for the estimator as a function of t that could be used to estimate the optimal $t \in [c, d]$; see BG (1990) for details. Unfortunately, this kernel estimator does not guarantee monotonicity.

The monotone estimator (1.1) is actually a kernel estimator with a random bandwidth of $O_p(n^{-1/3})$ when $\xi'(x) > 0$ [Wright (1984)]. As in the case of a fixed bandwidth of the same order used by Cheng (1983) and Stute (1986), this yields an asymptotic distribution with the norming of $n^{1/3}$ only. The improved convergence rate in BG (1990) is obtained by enlarging the bandwidth to a multiple of $n^{-1/5}$. To obtain a similar improvement in the case of monotone estimation we introduce an estimator that adds on an additional nonrandom band of width $\propto n^{-1/5}$. Fix $0 < c < d < \infty$ and let $J = J(c, d) = [cn^{-1/5}, dn^{-1/5}]$. We now define our estimators of $\xi(x)$ by

$$\xi^*(x) = \max_{r \le x - h/2} \min_{s \ge x + h/2} \hat{\xi}([r, s]), \qquad h \in J, \tag{1.2}$$

the h-dependence being suppressed in ξ^*. The monotonicity of the estimators follows from the fact that as x increases, the set of maximization is enlarged and the set of minimization is reduced. It will be shown that asymptotically these estimators behave exactly the same way as those of BG (1990):

$$\tilde{\xi}(x) = \hat{\xi}([x - h/2, x + h/2]), \qquad h \in J. \tag{1.3}$$

In fact, we will show that $\xi^*(x)$ has the same Bahadur-type representation as $\tilde{\xi}(x)$, but the remainder term is $O((\log\log n/n)^{1/2})$ a.s. instead of $O(n^{-3/5}\log n)$ a.s. due to the random bias (conditional on $\{X_i\}$) introduced by the "max–min" operation [see (2.13) and Remark 4 in Section 3]. However, the convergence rate is fast enough for the entire machinery developed for the unconstrained estimators to be applied to our monotone estimators.

The asymptotic equivalence of the fixed and the random bandwidth estimators stems from the following key observation. As mentioned earlier, Wright (1984) showed that the random bandwidth employed by the estimator (1.1), which is the estimator (1.2) with $h = 0$, is $O_p(n^{-1/3})$ when $\xi'(x) > 0$. This was obtained by a direct translation of the lemma in Wright (1981) for monotone regression estimates that uses the Hájek–Rényi inequality assuming only bounded conditional second moments. Using Hoeffding's (1963) inequality for 0-1 variables the result could be strengthened to show that the random bandwidth is $O((\log n/n)^{1/3})$ a.s. (see Section 3). Now, the estimator (1.2) adds a random, and typically asymmetric, "skin" to the fixed band of width h. When the band is symmetric the first order term in the bias vanishes [BG (1990)], but this is not the case if the band is asymmetric. If the skinwidth remained as large as in the case of $h = 0$, the extra (random) bias would have been too large to establish the above asymptotic equivalence. Fortunately, as the order of magnitude of the fixed bandwidth increases, the skinwidth decreases until it reaches the asymptotic value of $O((\log\log n/n)^{1/2})$ a.s., which turns out to be sufficient for our purposes. The convergence rate of $n^{-1/3}$ is typical in many other isotonic estimators, for example, estimators of the conditional mean, the hazard rate and the density, under monotonicity restrictions. The discussion above points out that these monotone estimators of functionals of distributions or conditional distributions could be improved upon (i) to obtain asymptotic normality and (ii) to use the norming of $n^{2/5}$ instead of $n^{1/3}$, by smoothing over wider intervals as in (1.2) when the smoothness assumptions are valid. In Section 4 we present a kernel-smoothed version of ξ^* as was done in Mukerjee (1988) for isotonic conditional mean estimators. This estimator is differentiable, and we discuss strong uniform consistency of the estimator and its derivative on compact intervals and the asymptotic normality of the estimator. The estimator $\tilde{\xi}$, of course, can be smoothed further in the same way.

It should be pointed out that this method of improvement of convergence rates of isotonic estimators cannot be extended to the case of partial orders in general.

2. Main results and proofs. For the random vector (X, Y) let f be the marginal p.d.f. of X. We wish to estimate $\xi(x_0)$ for some $x_0 \in \mathbb{R}$. We now make the following assumptions:

ASSUMPTION 1. The derivative $\xi'(x)$ exists continuously in a neighborhood of x_0 and $\xi'(x_0) > 0$.

ASSUMPTION 2. (i) $f(x_0) > 0$.

(ii) $f''(x)$ exists in a neighborhood of x_0, and there exists $A_1 < \infty$ such that $|f''(x) - f''(y)| \le A_1|x - y|$ for x and y in the neighborhood.

ASSUMPTION 3. There exists a neighborhood of $(x_0, \xi(x_0))$ such that:

(i) The derivatives $g(y|x) = G_y(y|x)$, $g_y(y|x)$, $g_x(y|x)$, $g_{xx}(y|x)$, $G_x(y|x)$ and $G_{xx}(y|x)$ exist for (x, y) in the neighborhood.

(ii). There exists $A_2 < \infty$ such that (x, y) and (t, y) in the neighborhood implies

$$g(\xi(x)|x) > 0, \qquad |g_y(y|x)| \le A_2, \qquad |g_x(y|x)| \le A_2, \qquad |g_{xx}(y|x)| \le A_2,$$

$$|g_{xx}(y|x) - g_{xx}(y|t)| \le A_2|x - t|, \qquad |G_{xx}(y|x) - G_{xx}(y|t)| \le A_2|x - t|.$$

Note that Assumptions 1 and 3 imply the uniqueness of $\xi(x)$ as the solution of $G(y|x) = p$ for $|x - x_0|$ sufficiently small. Also note that Assumptions 1–3 hold uniformly for all x in a neighborhood of x_0. The results in BG (1990) have been derived under Assumption 2 and 3, but with some of the conditions weakened to hold pointwise at x_0 only. Since we need to consider their representation of $\tilde{\xi}(x)$ for all x in a neighborhood of x_0, the uniformity conditions have been added.

We assume that $x_0 = 0$ w.l.o.g., and, to simplify the notation, we write $G(\cdot)$, $g(\cdot)$, $G_i(\cdot)$, ξ, ξ^* and $\tilde{\xi}$ for $G(\cdot|0)$, $g(\cdot|0)$, $G(\cdot|X_i)$, $\xi(0)$, $\xi^*(0)$ and $\tilde{\xi}(0)$, respectively.

THEOREM 2.1 [(BG (1990)]. *Under Assumptions* 2 *and* 3,

$$\tilde{\xi} - \xi = \beta(\xi) f^2(0) h^2 + \frac{1}{[nhf(0)]g(\xi)} \sum_{1}^{[nhf(0)]} [I(Z_i > \xi) - (1-p)] + R_n,$$

where

$$\beta(\xi) = -[f(0)G_{xx}(\xi) + 2f'(0)G_x(\xi)]/[24 f^3(0) g(\xi)],$$

$\{Z_i\}$ *are i.i.d. random variables with the d.f.* $G(\cdot)$,

and

$$\sup_{h \in J} |R_n| = O(n^{-3/5} \log n) \quad a.s. \tag{2.1}$$

THEOREM 2.2. *Under Assumptions* 1–3, $\tilde{\xi}$ *may be replaced by* ξ^* *in the conclusion of Theorem* 1 *with* (2.1) *replaced by*

$$\sup_{h \in J} |R_n| = O\big((\log\log n/n)^{1/2}\big) \quad a.s. \tag{2.2}$$

We first prove a lemma about the a.s. bound on the random bandwidth used by ξ^*. For $b > 0$ let $b_n = b(\log\log n/n)^{1/2}$ and define

$$\xi^{*b} = \max_r \min_s \{\hat{\xi}([r,s]) \colon -b_n - h/2 \le r \le -h/2,$$
$$h/2 \le s \le h/2 + b_n\},\ h \in J.$$

LEMMA 2.3. *Under Assumptions* 1–3, *there exists* $B < \infty$ *such that*

$$P\Big[\bigcup_{h \in J} \{\xi^{*B} \ne \xi^* \quad i.o.\}\Big] = 0.$$

PROOF. We first argue that $P[\xi^{*b} \neq \xi^*]$ is bounded above by the sum of

$$P\Big[\max_{r \le -b_n - h/2} \hat{\xi}([r, -h/2]) > \min_{s \ge h/2} \hat{\xi}((-h/2, s])\Big] \tag{2.3}$$

and

$$P\Big[\max_{r \le -h/2} \hat{\xi}([r, h/2)) > \min_{s \ge h/2 + b_n} \hat{\xi}([h/2, s])\Big]. \tag{2.4}$$

This can be seen by taking complements of the events above. Suppose that $r_0 \le -h/2$, $s_0 \ge h/2 + b_n$ and $\hat{\xi}([r_0, h/2)) \le \hat{\xi}([h/2, s_0])$. Then by the Cauchy mean value property of $\hat{\xi}$ [Robertson and Wright (1974)], $\hat{\xi}([r_0, s_0]) \le \hat{\xi}([h/2, s_0])$ so that

$$\xi^* = \max_r \min_s \{\hat{\xi}([r, s]): r \le -h/2, h/2 \le s \le h/2 + b_n\}.$$

Using a similar argument for the event in (2.3), it can now be seen that the event $\{\xi^{*b} = \xi^*\}$ is contained in the intersection of the complements of the two events in (2.3) and (2.4), which completes the argument.

We first consider (2.4). For fixed $b > 0$, (2.4) is bounded above by the sum of

$$P\Big[\max_r \{\hat{\xi}([r, h/2)): r \le -h/2\} \ge \xi(h/4)\Big] \tag{2.5}$$

and

$$P\Big[\min_s \{\hat{\xi}([h/2, s]): s \ge h/2 + b_n\} \le \xi(h/4)\Big]. \tag{2.6}$$

We derive an upper bound for (2.6). Crude estimates will be sufficient. By Assumption 2 and the LIL for the e.d.f. there exists $B < \infty$ such that

$$P\Big[\inf_{h \in J} N(n, [h/2, h/2 + B_n]) \le (n \log\log n)^{1/2} \text{ i.o.}\Big] = 0, \tag{2.7}$$

where $B_n = B(\log\log n/n)^{1/2}$. For $s \ge h/2 + B_n$ let $\overline{G}_s(\cdot) = \sum G_i(\cdot) I(h/2 \le X_i \le s)/N(n, [h/2, s])$, and note that, for all $h \in J$ and $s \ge h/2 + B_n$, we have $\overline{G}_s(\xi(h/4)) \le \overline{G}_{h/2+B_n}(\xi(h/4)) \le p - Cn^{-1/5}$ for some $C > 0$, if n is sufficiently large, by Assumptions 1, 3 and the monotonicity of the conditional quantiles. Now, $\hat{\xi}([h/2, s]) \le \xi(h/4)$ with $s \ge h/2 + B_n$ implies

$$\begin{aligned} &\frac{1}{N(n, [h/2, s])} \sum \{I[Y_i \le \xi(h/4)] - G_i(\xi(h/4))\} I[h/2 \le X_i \le s] \\ &\qquad \ge [N(n, [h/2, s])p]/N(n, [h/2, s]) - \overline{G}_s(\xi(h/4)) \ge Dn^{-1/5} \end{aligned}$$

for some $D > 0$, not depending on h or $s \ge h/2 + B_n$, if n is sufficiently large, by Assumptions 1, 3 and the above. Let $\mathscr{A} = \sigma(X_1, X_2, \dots)$, $U_i = I[Y_i \le \xi(h/4)]$ and $\mu_i = G_i(\xi(h/4)) = E[U_i | \mathscr{A}]$. Since $\{Y_i\}$ are independent given $\mathscr{A}$, by Hoeffding's (1963) inequality and (2.7), for all n sufficiently large,

$$\sup_{h\in J}\max_{s\geq h/2+B_n} P\left[\frac{1}{N(n,[h/2,s])}\sum(U_i-\mu_i)I[h/2\leq X_i\leq s]\right.$$

$$(2.8)\qquad \left.\geq Dn^{-1/5}|\mathscr{A}\right]$$

$$\leq \sup_{h\in J}\max_{s\geq h/2+B_n}\exp(-2N(n,[h/2,s])D^2n^{-2/5})$$

$$\leq \exp\left[-Mn^{1/10}(\log\log n)^{1/2}\right]\quad \text{a.s. for some } M>0.$$

We now have

$$P\left[\bigcup_{h\in J}\left\{\min_{s\geq h/2+B_n}\hat{\xi}([h/2,s])\leq \xi(h/4)\right\}\right]$$

$$(2.9)\qquad \leq n^2\sup_{h\in J}\max_{s\geq h/2+B_n}P\left[\hat{\xi}([h/2,s])\leq\xi(h/4)\right]$$

$$\leq n^2\exp\left[-Mn^{1/10}(\log\log n)^{1/2}\right],\quad \text{which is summable.}$$

By a similar argument it can be seen that

$$(2.10)\qquad P\left[\bigcup_{h\in J}\left\{\max_{r\leq -h/2}\hat{\xi}([r,h/2))\geq\xi(h/4)\right\}\right]$$

$$\leq n^2\exp\left[-M'n^{4/5}(n^{-1/5})^2\right]=n^2\exp(-M'n^{2/5})$$

for some $M'>0$, and hence is summable. The derivation of an upper bound of (2.5) is similar, and is omitted. The use of the Borel–Cantelli lemma now completes the proof. □

PROOF OF THEOREM 2.2. By Theorem 2.1 and Lemma 2.3 it is sufficient to show that there exists a finite M, not depending on h, such that

$$P\left[|\xi^{*B}-\tilde{\xi}|\geq M(\log\log n/n)^{1/2}\text{ i.o.}\right]=0,\qquad h\in J,$$

with B as in Lemma 2.3. We note that for any h

$$\min_{h/2\leq s\leq h/2+B_n}\hat{\xi}([-h/2,s])\leq\xi^{*B}\leq\max_{-h/2-B_n\leq r\leq -h/2}\hat{\xi}([r,h/2]).$$

We will show that for an M as above

$$(2.11)\qquad P\left[\min_{h/2\leq s\leq h/2+B_n}\hat{\xi}([-h/2,s])\right.$$

$$\left.\leq\tilde{\xi}-M(\log\log n/n)^{1/2}\text{ i.o.}\right]=0,\qquad h\in J.$$

The proof of the inequality going the other way is similar, and will be omitted.

To prove (2.11) it is sufficient to show that, for $h \in J$,

$$(2.12)\quad \sum n \max_{h/2 \le s \le h/2 + B_n} P\Big[\hat{\xi}([-h/2, s]) \le \tilde{\xi} - M(\log\log n/n)^{1/2}\Big] < \infty.$$

The easiest way to prove (2.12) is to use the representation of $\hat{\xi}([-h/2, s])$ as an estimator of $\xi(s/2 - h/4)$ as given by Theorem 2.1. This is permissible since the representation holds for all $h \in J(c, d)$, c and d are arbitrary with $0 < c < d < \infty$ and $B_n = o(h)$, and Assumptions 1–3 hold uniformly for $x_0 = s/2 - h/4$ in place of $x_0 = 0$ for $|s/2 - h/4|$ sufficiently small.

Let $t = s - h/2$. For $0 \le t \le B_n$ let

$$\overline{G}_t(\cdot) = \frac{1}{N(n, [-h/2, h/2 + t])} \sum G_i(\cdot) I(-h/2 \le X_i \le h/2 + t),$$

$\tilde{\xi}_t = \hat{\xi}([-h/2, h/2 + t])$, and let $\bar{\xi}_t$ be the p-quantile of $\overline{G}_t(\cdot)$, which is defined by the unique solution of $\overline{G}_t(\bar{\xi}_t) = p$ for all n large enough by our assumptions; note that this $\overline{G}_t$ is not the same as the one defined in the proof of Lemma 2.3. In the language of Theorem 2.1, $\tilde{\xi}_t$ is the kernel estimator of $\xi(t/2)$, with Assumptions 2 and 3 being valid for $x_0 = t/2$ when n is large. Now, one may think of $\tilde{\xi}_t$ as an estimator of $\bar{\xi}_t$, which gives rise to a bias, $\bar{\xi}_t - \xi(t/2)$, as well as a random error in estimating $\bar{\xi}_t$ by $\tilde{\xi}_t$.

We will be quoting results in BG (1990) that were derived for their nearest neighbor estimator, but the corresponding results for the kernel estimator are also valid as shown in Section 8 of that paper.

By Lemma 5 of BG (1990),

$$\bar{\xi}_t - \xi(t/2) = \beta_t(\xi(t/2)) f^2(t/2)(h + t)^2 + O(n^{-3/5}) \quad \text{a.s.},$$

where $\xi(t/2) = \xi + O(t)$, $f^2(t/2) = f^2(0) + O(t)$, $(h + t)^2 = h^2 + O(ht)$, and

$$\begin{aligned}\beta_t(\xi(t/2)) &= -\big[f(t/2) G_{xx}(\xi(t/2)|t/2) \\ &\qquad + 2f'(t/2) G_x(\xi(t/2)|t/2)\big] / \big[24 f^3(t/2) g(\xi(t/2)|t/2)\big] \\ &= \beta(\xi) + O(t) \quad \text{as } t \to 0 \text{ by our assumptions.}\end{aligned}$$

Hence,

$$(2.13)\quad \bar{\xi}_t = \xi + \beta(\xi) f^2(0) h^2 + O\big((\log\log n/n)^{1/2}\big) \quad \text{a.s. for } 0 \le t \le B_n.$$

Let $V_i = [I(Y_i > \xi) - \{1 - G_i(\xi)\}]$. Assume that $0 \le t \le B_n$. By the representation (22b) of BG (1990),

$$(2.14)\quad \begin{aligned}\tilde{\xi}_t - \bar{\xi}_t &= \big[N(n, [-h/2, h/2 + t]) g(\xi(t/2)|t/2)\big]^{-1} \\ &\qquad \times \sum V_i I(-h/2 \le X_i \le h/2 + t) + R_{nt},\end{aligned}$$

where $\sup_{h \in J} \max_t |R_{nt}| = O(n^{-3/5} \log n)$ a.s. Then, noting that $\tilde{\xi}_0 = \tilde{\xi}$ and $\bar{\xi}_0 = \beta(\xi) f^2(0) h^2 + O(n^{-3/5})$ a.s. from BG (1990), $(\tilde{\xi}_t - \bar{\xi}_t) - (\tilde{\xi} - \bar{\xi}_0)$ is almost surely equal to

$$\Big\{[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1}$$
$$-[N(n,[-h/2,h/2])g(\xi)]^{-1}\Big\}\sum V_i I(-h/2 \le X_i \le h/2)$$
$$+[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1}\sum V_i I(h/2 < X_i \le h/2+t)$$
$$+R'_{nt},$$

where $\sup_{h \in J} \max_t |R'_{nt}| = O(n^{-3/5} \log n)$ a.s.

Now, $g(\xi(t/2)|t/2) = g(\xi) + O((\log\log n/n)^{1/2})$ and, by Lemma 10 of BG (1990) and Assumption 2,

$$N(n,[-h/2,h/2+t]) = n(h+t)f(t/2) + \Delta_{nt}$$
$$= nhf(0) + O\big((n \log\log n)^{1/2}\big) + \Delta_{nt},$$

where $\sup_{h \in J} \max_t |\Delta_{nt}| = O(n^{2/5} \log n)$ a.s. Thus, for some positive C_1, C_2, C_3 and C_4,

$$\big|[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1} - N(n,[-h/2,h/2])g(\xi)]^{-1}\big|$$
$$\le C_1 n^{-11/10}(\log\log n)^{1/2} \quad \text{a.s.},$$
$$N(n,[-h/2,h/2]) \le C_2 n^{4/5} \quad \text{a.s.},$$
$$[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1} \le C_3 n^{-4/5} \quad \text{a.s.} \quad \text{and}$$
$$N(n,[-h/2,h/2+t]) - N(n,[-h/2,h/2]) \le C_4 (n \log\log n)^{1/2} \quad \text{a.s.},$$

where we have used the LIL for the e.d.f. and Assumption 2 in the last inequality. Since $\{V_i\}$ are independent given $\mathscr{A} = \sigma(\{X_i\})$ with $E[V_i|\mathscr{A}] \equiv 0$ a.s., for any $\varepsilon > 0$, we have, using Hoeffding's (1963) inequality,

$$
\begin{aligned}
&P\Big[\Big|\Big\{[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1} \\
&\qquad -[N(n,[-h/2,h/2])g(\xi)]^{-1}\Big\} \\
&\qquad \times \sum V_i I(-h/2 \le X_i \le h/2)\Big| \ge \varepsilon(\log\log n/n)^{1/2}\Big|\mathscr{A}\Big] \\
&\le 2\exp\Big\{-2C_5 N(n,[-h/2,h/2]) \\
&\qquad \times \Big\{\big|[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1} \\
&\qquad\qquad -[N(n,[-h/2,h/2])g(\xi)]^{-1}\big|^{-1} \\
&\qquad \times (\log\log n/n)^{1/2}/N(n,[-h/2,h/2])\Big\}^2\Big\} \\
&\le 2\exp\big(-C_6 n^{2/5}\big) \quad \text{a.s.}
\end{aligned}
\tag{2.15}
$$

and

$$P\Big[\Big|[N(n,[-h/2,h/2+t])g(\xi(t/2)|t/2)]^{-1}$$
$$\times \sum V_i I(h/2 < X_i \le h/2+t)\Big| \ge \varepsilon(\log\log n/n)^{1/2}|\mathscr{A}\Big]$$
$$\le 2\exp\Big\{-2C_7 N(n,(h/2,h/2+t])\big(N(n,[-h/2,h/2+t]) \times(\log\log n/n)^{1/2}/N(n,(h/2,h/2+t])\big)^2\Big\} \tag{2.16}$$
$$\le 2\exp\big\{-C_8 n^{1/10}(\log\log n)^{1/2}\big\} \quad \text{a.s.}$$

for some positive C_5, C_6, C_7 and C_8. The inequality (2.12) now follows from the inequalities (2.13)–(2.16). This completes the proof of Theorem 2.2. □

3. Remarks. As mentioned in the introduction, the key to the proof of Theorem 2.1 is Lemma 2.3. We can make the following additional observations.

Remark 1. It may be noted that all of the a.s. order of magnitude bounds on random variables in this paper as well as in BG (1990) were obtained by exponential probability bounds and subsequent use of Borel–Cantelli lemma. Using this fact it can be seen that Lemma 2.3 can be extended to hold uniformly for all points for which Assumptions 1–3 hold uniformly; this entails a change in the multipliers of the exponential probability bounds in (2.9) and (2.10) from n^2 to n^3. It is also easy to verify that the representation of ξ^* given by Theorem 2.2 holds for these points with the remainder term bounded by (2.2) uniformly in these points. For example, Assumptions 1–3 as stated imply this uniform representation of $\xi^*(x)$ for all x in a neighborhood of x_0.

Remark 2. The summability of the probability bound in (2.9) simply requires that $N(n,[h/2,s])h^2 \ge C\log n$ and that in (2.10) requires $nh^3 \ge D\log n$ for sufficiently large C and D. Thus, B_n will still be $O((\log\log n/n)^{1/2})$ a.s. if h is reduced to const.$\times(\log n)^{1/2}/(n\log\log n)^{1/4}$. If the order of magnitude of h is reduced further, the random skindepth will be increased.

Remark 3. When $h=0$, corresponding to the estimator (1.1), the random bandwidth used by the estimator is $O((\log n/n)^{1/3})$ a.s. This can be seen by noting that Lemma 2.3 still holds if $h=0$ and $b_n = b(\log n/n)^{1/3}$, $b>0$. The following modifications of (2.3), (2.4), (2.6) and (2.8) should make it clear how to prove the result:

$$(2.3)' \qquad P\left[\max_{r \le -b_n} \hat{\xi}([r, -b_n/2]) > \min_{s \ge 0} \hat{\xi}((-b_n/2, s])\right],$$

$$(2.4)' \qquad P\left[\max_{r \le 0} \hat{\xi}([r, b_n/2)) > \min_{s \ge b_n} \hat{\xi}([b_n/2, s])\right],$$

$$(2.6)' \qquad P\left[\min_{s \ge b_n} \hat{\xi}([b_n/2, s]) \le \xi(b_n/4)\right],$$

$$(2.8)' \qquad \begin{aligned} &\max_{s \ge b_n} P\left[\frac{1}{N(n, [h/2, s])} \sum (U_i - \mu_i) I(b_n/2 \le X_i \le s)\right. \\ &\qquad\qquad \left. \ge D(\log n/n)^{1/3} | \mathscr{A}\right] \\ &\le \max_{s \ge b_n} \exp\left[-2N(n, [h/2 + s]) D^2 (\log n/n)^{2/3}\right] \\ &\le \exp\left[-Mn(\log n/n)^{1/3} (\log n/n)^{2/3}\right] \quad \text{a.s.} \end{aligned}$$

REMARK 4. For the kernel estimator $\tilde{\xi}$, the conditional mean, $E(\tilde{\xi}|\mathscr{A})$, depends only on $\{X_i\}$. However, $E(\xi^*|\mathscr{A})$ adds a random component of $O((\log\log n/n)^{1/2})$ a.s. due to the added random bandwidth depending on $\{Y_i\}$, resulting from the isotonization process as shown in Lemma 2.3. This causes the convergence rate in the remainder term in (2.2) to go down from that in (2.1). This rate is still fast enough for the Bahadur-type representation to go through. By Remark 3, if we smoothed the Casady–Cryer estimator (1.1) by a kernel, the resulting random bias would have been $O((\log n/n)^{1/3})$ a.s., and this representation would not have been possible.

REMARK 5. If $\xi'(x_0) = 0$ then this representation does not necessarily hold. For monotone conditional mean estimators Parsons (1979) has shown that the random bandwidth is $O_p(1)$ if $\xi'(x) = 0$ for all x in an open interval containing x_0. Wright (1981) has studied the means case when $\xi'(x) = 0$, but $|\xi(x) - \xi(x_0)| = A|x - x_0|^{\alpha}(1 + o(1))$ as $x \to x_0$, with $A > 0$ and $\alpha > 1$. In this case the random bandwidth is $O_p(n^{-1/(2\alpha+1)})$. It would appear that something similar to Lemma 2.3 will hold if α is sufficiently close to 1, but the uniform representation of $\tilde{\xi}$ near x_0 given by Theorem 2.1 will not hold, and other methods need to be used to prove the equivalent of Theorem 2.2.

4. A smooth estimator. The estimator ξ^* improves on the convergence properties of ξ^{**} in (1.1), but it is not smooth. In this section we propose a differentiable version of ξ^* using a kernel smoother the same way the monotone conditional mean of the type ξ^{**} was smoothed in Mukerjee (1988). In the process we lose the Bahadur-type representation, but the representation of ξ^* will turn out to be very useful in analyzing the asymptotic properties of the smoothed version. We define

$$\xi_s^*(x) = \frac{1}{na}\sum_{i=1}^{n}\delta\left[\frac{(x - X_i)}{a}\right]\frac{\xi^*(X_i)}{f_n(x)},$$

where $f_n(x) = (1/na)\Sigma_{i=1}^n \delta[(x - X_i)/a]$, a is the band width and δ is a kernel function. By our assumptions (to be stated) $f_n(x) \to f(x)$ uniformly in x in a set of interest and thus the above quantities are well defined for all x in the set a.s. for all large n, and we will assume that n is at least this large without explicit mention. We now make the following assumptions.

ASSUMPTION 4. (i) $\int v\delta(v)\,dv = 0$;
(ii) δ'' exists and is continuous;
(iii) $a = O(h)$ and $na/\log(1/a) \to \infty$;
(iv) ξ'' exists and is bounded in a neighborhood of x; and
(v) δ is a log-concave density kernel with compact support.

We estimate $\xi(x)$ by $\xi_s^*(x)$ and $\xi'(x)$ by $\xi_s^{*\prime}(x)$. Differentiability of δ implies ξ_s^* is differentiable. Log-concave kernels are continuous and bounded, and guarantee monotonicity of ξ_s^*, if ξ^* is, as argued in Mukerjee [(1988) page 743]. Note that a kernel that is not nonnegative does not guarantee this. We restrict our attention to compact kernels only since those with infinite support are generally suboptimal.

We first consider the strong uniform consistency of ξ_s^* and $\xi_s^{*\prime}$ on a fixed interval I.

THEOREM 4.1. *Assume that Assumptions* 1–3 *hold uniformly in a neighborhood of* I. *If* $a \to 0$ *and* $na/\log(1/a) \to \infty$, *then* $\sup_{x\in I}|\xi_s^*(x) - \xi(x)| \to 0$ *a.s. If* δ'' *exists continuously,* $a \to 0$ *and* $na^3/\log(1/a) \to \infty$, *then* $\sup_{x\in I}|\xi_s^{*\prime}(x) - \xi'(x)| \to 0$ *a.s.*

INDICATION OF PROOF. By Remark 1 of Section 3, $|\xi^*(x) - \xi(x)| \to 0$ a.s., uniformly in a neighborhood of I. Using this result the proof of the theorem exactly parallels the proofs of Theorem 3.3 and Theorem 3.5 in Mukerjee (1988) for the kernel-smoothed monotone conditional *mean* estimator using kernels with compact support after we note the following. Theorem 3.4, that proves a.s. globally uniform convergence of estimators of f and f' under Silverman's (1978) conditions [(3.11) in Mukerjee (1988)], was used in proving Theorem 3.5. These conditions are very hard to verify. For this reason we have used a slightly stronger analytic assumption on δ for the second half of the theorem. Under our assumptions, the conditions of Theorem 3.4 in the above paper are satisfied in a neighborhood of I and the conclusions of that theorem hold in a neighborhood of I. □

We now derive the asymptotic distribution of $\xi_s^*(x_0)$ for an x_0 for which Assumptions 1–3 hold. From the indication of proof of Theorem 4.1 and

Theorem A of Silverman (1978), under Assumptions 1–3, Assumption 4(ii) and (iii),

$$f_n(x) \to f(x) \quad \text{uniformly in } x \text{ in a neighborhood of } x_0. \tag{4.1}$$

Let $\alpha = h/(2a)$ and let F_δ be the d.f. corresponding to the density δ. Define

$$\psi(\alpha, \delta) = \frac{2\alpha}{\alpha + 1} \Big/ \int [F_\delta(s(\alpha + 1) + \alpha) - F_\delta(s(\alpha + 1) - \alpha)]^2 \, ds.$$

Since $a = O(h)$ and δ has bounded support, $\psi(\alpha, \delta)$ is bounded away from 0. Denote the support of δ by $[-s_1, s_2]$, where s_1 and s_2 are some positive numbers. Then the integrand in the denominator of $\psi(\alpha, \delta)$ is 0 if $s > (\alpha + s_2)/(\alpha + 1)$ or $s < -(\alpha + s_1)/(\alpha + 1)$. If $a = o(h)$, that is, if $\alpha \to \infty$, then $\psi(\alpha, \delta) \to 1$. To see this we note that $-(\alpha - s_2)/(\alpha + 1) \leq s \leq (\alpha - s_1)/(\alpha + 1)$ implies that $F_\delta(s(\alpha + 1) + \alpha) = 1$ and $F_\delta(s(\alpha + 1) - \alpha) = 0$, and thus the integral becomes $(2\alpha - s_1 - s_2)/(\alpha + 1) + \gamma(s_1 + s_2)/(\alpha + 1)$ for some $0 < \gamma < 1$.

THEOREM 4.2. *Under Assumptions* 1–4,

$$\sqrt{nh\psi(\alpha, \delta)} \Big\{ \xi_s^*(x_0) - \xi(x_0) - \beta(\xi(x_0)) f^2(x_0) h^2 - [f'(x_0)\xi'(x_0)/f(x_0) + \xi''(x_0)/2] a^2 \int v^2 \delta(v) \, dv \Big\} \to N(0, \sigma^2(x_0)) \quad \textit{in distribution},$$

where $\beta(\xi(x_0))$ *is defined the same way as* $\beta(\xi) = \beta(\xi(0))$, *with* 0 *replaced by* x_0, *and*

$$\sigma^2(x_0) = p(1 - p)/[f(x_0) g^2(\xi(x_0)|x_0)].$$

If $a = o(h)$ *then* $\psi(\alpha, \delta)$ *may be replaced by* 1 *and the bias term* $\propto a^2$ *may be omitted.*

PROOF. From our definition of $\xi_s^*(x_0)$ and the representation of $\xi^*(x)$ in a neighborhood of x_0 as discussed in Remark 1 of Section 3,

$$\xi_s^*(x_0) = \xi(x_0) + \beta(\xi(x_0)) f^2(x_0) h^2 + \left[U_{n1}(x_0) + U_{n2}(x_0) + O\left(\sqrt{\log\log n/n}\right)\right]/f_n(x_0) \quad \text{a.s.,} \tag{4.2}$$

where

$$U_{n1}(x_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a} \delta\left[\frac{(x_0 - X_i)}{a}\right] \left[\xi'(x_0)(X_i - x_0) + \frac{\xi''(x_0)(X_i - x_0)^2}{2} + o\left((X_i - x_0)^2\right) + O\left(h^2 |X_i - x_0|\right)\right]$$

and

$$U_{n2}(x_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{a}\delta[(x_0 - X_i)/a][nhf(X_i)g(\xi(X_i))|X_i)]^{-1}$$
$$\times \sum_{j=1}^{[nhf(X_i)]}\left[I\big(Z_j(X_i) > \xi(X_i)\big) - (1-p)\right],$$

where $\{Z_j(X_i)\}$ are i.i.d. random variables with the (random) d.f. $G(\cdot|X_i)$. Now,

$$E[U_{n1}(x_0)] = \frac{1}{a}\int\delta\left(\frac{x_0 - u}{a}\right)\left[\xi'(x_0)(u - x_0) + \frac{\xi''(x_0)(u - x_0)^2}{2}\right.$$
$$\left. + o\big((u - x_0)^2\big) + O\big(h^2|u - x_0|\big)\right]f(u)\,du$$
$$= \int\delta(v)\left[-\xi'(x_0)av + \frac{\xi''(x_0)a^2v^2}{2} + o(a^2v^2) + O(h^2a|v|)\right]$$
$$\times[f(x_0) - f'(x_0)av + o(a|v|)]\,dv$$
$$= a^2\left[f'(x_0)\xi'(x_0) + \frac{f(x_0)\xi''(x_0)}{2}\right]\int v^2\delta(v)\,dv + o(a^2)$$

and

$$E\left\{\frac{1}{a}\delta\left[\frac{(x_0 - X_i)}{a}\right]\left[\xi'(x_0)(X_i - x_0) + \frac{\xi''(x_0)(X_i - x_0)^2}{2}\right.\right.$$
$$\left.\left. + o\big((X_i - x_0)^2\big) + O\big(h^2|X_i - x_0|\big)\right]\right\}^2$$
$$= \frac{1}{a}\int\delta^2(v)[-\xi'(x_0)av + O(a^2)]^2[f(x_0) + O(a)]\,dv = O(a).$$

Thus $\mathrm{Var}[\sqrt{nh\psi(\alpha,\delta)}\,U_{n1}(x_0)] = O(ah)$. Since $o(a^2)\sqrt{nh\psi(\alpha,\delta)} = o((nh^5)^{1/2}) = o(1)$, using (4.1) we now have

$$\sqrt{nh\psi(\alpha,\delta)}\left\{U_{n1}(x_0)/f_n(x_0) - a^2[f'(x_0)\xi'(x_0)/f(x_0)\right.$$
$$\left. + \xi''(x_0)/2]\int v^2\delta(v)\,dv\right\} \to 0 \quad \text{in probability.} \tag{4.3}$$

Next we consider the term involving $U_{n2}(x_0)$. We start with a new representation of $\xi^*(x)$ for x sufficiently close to x_0 by replacing

$$\frac{1}{nhf(x)} \sum_{j=1}^{[nhf(x)]} \left[I(Z_j(x) > \xi(x)) - (1-p) \right] \tag{4.4}$$

by

$$\frac{1}{nhf(x)} \sum_{j=1}^{n} \left[I(Y_j > \xi(X_j)) - (1-p) \right] \left[I\left(|x - X_j| \le \frac{h}{2} \right) \right]. \tag{4.5}$$

[Note that $p = G(\xi(X_j)|X_j)$ for X_j in a neighborhood of x_0.] Comparing the kernel versions of (22c) (given by Theorem 2.1 here) and (22b) in BG (1990), the expression in (4.5) above differs from

$$\frac{1}{nhf(x)} \sum_{j=1}^{n} \left\{ \left[I(Y_j > \xi(x)) - \left[1 - G(\xi(x)|X_j) \right] \right\} \left[I\left(|x - X_j| \le \frac{h}{2} \right) \right], \tag{4.6}$$

by at most $O(n^{-3/5} \log n)$ a.s., uniformly in x in a neighborhood of x_0 by our Assumptions 1–3 and Remark 1 of Section 3. Let $T_j = [I(Y_j > \xi(X_j)) - I(Y_j > \xi(x))]I(|x - X_j| \le h/2)$ and $\tau_j = [(1-p) - [1 - G(\xi(x)|X_j)]I(|x - X_j| \le h/2)$. Then, conditional on $\mathscr{A} = \sigma(\{X_i\})$, $\{T_j\}$ are independent random variables with means $\{\tau_j\}$. By expanding $G(\xi(x)|X_j)$, as a function of X_j, about x, and using Assumption 3, a little computation shows that $\mathrm{Var}[T_j|\mathscr{A}] \le E[|\tau_j|] = O(h^2)$, uniformly in x near x_0. Since $|T_j - \tau_j| \le 1$, using the same method used in proving Lemma 7 in BG (1990), using the order of bound on $E[|\tau_j|]$ and Bernstein's inequality [Bennett (1962)], it can be seen that (4.5) and (4.6) differ by at most $O(n^{-3/5} \log n)$ a.s., uniformly in x in a neighborhood of x_0. Let $V_j = I(Y_j > \xi(X_j)) - (1-p)$ and note that, conditional on $\{X_i\}$, $\{V_j\}$ are i.i.d. Bernoulli random variables with their means subtracted for $\{X_j\}$ sufficiently close to x_0. We can now write

$$\begin{aligned} U_{n2}(x_0) &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a} \delta \left[\frac{(x_0 - X_i)}{a} \right] \frac{1}{nhf(X_i) g(\xi(X_i)|X_i)} \\ &\quad \times \sum_{j=1}^{n} V_j I\left(|X_j - X_i| \le \frac{h}{2} \right) + O(n^{-3/5} \log n) \\ &= \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h} V_j Q_n(x_0, X_j) + O(n^{-3/5} \log n) \quad \text{a.s.,} \end{aligned}$$

where

$$Q_n(x_0, x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a} \delta \left[\frac{(x_0 - X_i)}{a} \right] \frac{1}{f(X_i) g(\xi(X_i)|X_i)} I\left(|x - X_i| \le \frac{h}{2} \right).$$

We will show that $Q_n(x_0, x)$ converges to a constant a.s., uniformly in x in a neighborhood of x_0. Now,

$$\begin{aligned}E[Q_n(x_0,x)] &= \frac{1}{a}\int \delta\left(\frac{x_0-u}{a}\right)\frac{1}{f(u)g(\xi(u)|u)}I\left(|x-u|\le\frac{h}{2}\right)f(u)\,du\\ &= \int\delta(v)\left\{\frac{1}{g(\xi(x_0)|x_0)}+O(a|v|)\right\}I\left(\left|v-\frac{x_0-x}{a}\right|\le\frac{h}{2a}\right)dv\\ &= \frac{1}{g(\xi(x_0)|x_0)}\int\delta(v)I\left(\left|v-\frac{x_0-x}{a}\right|\le\frac{h}{2a}\right)dv\\ &\quad + O(a)I_S(x_0-x),\end{aligned}$$

uniformly in x in a neighborhood of x_0, where $S=[-as_1-h/2,\ as_2+h/2]$, using the fact that the support of δ is $[-s_1, s_2]$. Similarly,

$$\begin{aligned}&E\left\{\frac{1}{a}\delta\left[\frac{(x_0-X_i)}{a}\right]\frac{1}{f(X_i)g(\xi(X_i)|X_i)}I\left(|x-X_i|\le\frac{h}{2}\right)\right\}^2\\ &\quad=\frac{1}{a}\int\delta^2(v)\left\{\frac{1}{f(x_0)g^2(\xi(x_0)|x_0)}+O(a)\right\}I\left(\left|v-\frac{x_0-x}{a}\right|\le\frac{h}{2a}\right)dv\\ &\quad=O\left(\frac{1}{a}\right),\end{aligned}$$

uniformly in x in a neighborhood of x_0. Since each summand in the definition of $Q_n(x_0,x)$ has a bound and a variance of $O(1/a)$, using Bernstein's inequality for bounded variables [equation (8) of Bennett (1962)], we have

$$\begin{aligned}&\sum nP\left(|Q_n(x_0,x)-EQ_n(x_0,x)|\ge M\sqrt{\log n/na}\right)\\ &\quad\le\sum 2n\exp\left\{-(naM^2\log n/na)/\left(2+2M\sqrt{\log n/na}\right)\right\}\\ &\quad\le\sum 2n\exp(-M^2\log n/4)\end{aligned}$$

for all large n by Assumption 4(iii), and hence the sum is finite for all large M. Thus, by Borel–Cantelli lemma and the above,

$$Q_n(x_0,x)=Q(x_0,x)+\left\{O\left(\sqrt{\log n/na}\right)+O(a)\right\}I_S(x_0-x)\quad\text{a.s.},$$

uniformly in x in a neighborhood of x_0, where

$$Q(x_0,x)=\frac{1}{g(\xi(x_0)|x_0)}\int\delta(v)I\left(\left|v-\frac{x_0-x}{a}\right|\le\frac{h}{2a}\right)dv.$$

Note that $Q_n(x_0,x)=0$ if $x_0-x>as_2+h/2$ or $x_0-x<-as_1-h/2$. We may now write

$$\begin{aligned}U_{n2}(x_0)&=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}V_j\Bigg[Q(x_0,X_j)\\ &\qquad+\left\{O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{na}}\right)+O(a)\right\}I_S(x_0-X_j)\Bigg].\end{aligned}\tag{4.7}$$

Observe that

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}V_j\int\delta(v)I\left(\left|v-\frac{x_0-X_j}{a}\right|\le\frac{h}{2a}\right)dv$$

$$\times\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}\int\delta(v)I\left(\left|v-\frac{x_0-X_j}{a}\right|\le\frac{h}{2a}\right)dv\right]^{-1}$$

may be looked upon as a standard kernel estimator of the *mean* regression function that is identically zero at the point x_0, with the kernel K given by

$$K(t)=\int\delta(v)I\left(|av-t|\le\tfrac{1}{2}\right)dv,$$

with the properties

$$\int K(t)\,dt=\int\int\delta(v)I\left(|av-t|\le\tfrac{1}{2}\right)dt\,dv=\int\delta(v)\,dv=1$$

and

$$\int tK(t)\,dt=\int\int t\delta(v)I\left(|av-t|\le\tfrac{1}{2}\right)dt\,dv=\int av\delta(v)\,dv=0.$$

Thus, under our assumptions, Schuster's (1972) central limit theorem is applicable to

$$W_n(x_0)\equiv\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}V_j\int\delta(v)I\left(\left|v-\frac{x_0-X_j}{a}\right|\le\frac{h}{2a}\right)dv \tag{4.8}$$

after we verify the moment conditions. To compute the variance of a summand in (4.8) we note that

$$\begin{aligned}
&E\left\{\frac{1}{\sqrt{h}}\int\delta(v)I\left(\left|v-\frac{x_0-X_j}{a}\right|\le\frac{h}{2a}\right)dv\right\}^2\\
&=\frac{1}{h}\int\left\{\int\delta(v)I\left(\left|v-\frac{x_0-u}{a}\right|\le\frac{h}{2a}\right)dv\right\}^2f(u)\,du\\
&=\frac{1}{h}\int\left\{\int\delta(v)I\left(\left|v-\frac{t}{a}\right|\le\frac{h}{2a}\right)dv\right\}^2f(x_0-t)\,dt\\
&=\frac{1}{h}\int\left[F_\delta\left(\frac{2t+h}{2a}\right)-F_\delta\left(\frac{2t-h}{2a}\right)\right]^2[f(x_0)+O(t)]\,dt\\
&=f(x_0)\frac{\alpha+1}{2\alpha}\int[F_\delta(s(\alpha+1)+\alpha)-F_\delta(s(\alpha+1)-\alpha)]^2\,ds+O(h)\\
&=\frac{f(x_0)}{\psi(\alpha,\delta)}+O(h),
\end{aligned}$$

where the substitution $t=(a+h/2)s$ has been utilized above. Using the fact that $E[V_j^2|\mathscr{A}] = p(1-p)$ and the result above in (4.8) we get

$$\mathrm{Var}\left[\sqrt{nh}\, W_n(x_0)\right] = p(1-p)\, f(x_0)/\psi(\alpha,\delta) + O(h).$$

The variance and the third absolute moment of each summand in (4.8) are $O(h^{-1})$ and $O(h^{-3})$, respectively, from arguments similar to those above. From this, and the fact that $nh^3 \to \infty$, Schuster's (1972) theorem under Liapounov's condition implies that

$$(4.9)\quad \sqrt{nh}\, W_n(x_0) \to N(0, p(1-p)\, f(x_0)/\psi(\alpha,\delta)) \quad \text{in distribution.}$$

For the large order terms in (4.7), it may be noticed that $I_S(x_0 - X_j)$ may be written as

$$I_S(x_0 - X_j) = I\left(\left|x_0 + a\frac{s_1 - s_2}{2} - X_j\right| \le a\frac{s_1 + s_2}{2} + \frac{h}{2}\right).$$

This uniform kernel could be used to estimate the same identically zero regression function as above at the point $x_0 + a(s_1 - s_2)/2$, yielding a central limit theorem similar to (4.9). Thus the large order terms in (4.7) are $o_p((nh)^{-1/2})$ by Assumption 4(iii) and using (4.9) in (4.7) yields

$$(4.10)\quad \sqrt{nh\psi(\alpha,\delta)}\, U_{n2}(x_0) \to N\big(0, \sigma^2(x_0)\, f^2(x_0)\big) \quad \text{in distribution.}$$

Putting the results (4.3) and (4.10) in conjunction with (4.1), in the representation of $\xi_s^*(x_0)$ in (4.2) now completes the proof of the first half of the theorem, noting that $\sqrt{nh}\sqrt{\log\log n/n} = o(1)$. The second half is a consequence of the facts that $\lim_{\alpha\to\infty} \psi(\alpha,\delta) = 1$ and $\sqrt{nh}\, a^2 = o(1)$ if $a = o(h)$. □

We also note that using the representations in (4.2) and (4.7), Härdle's (1984) Theorem 4 could be used to show the following:

THEOREM 4.3. *Under the assumptions of Theorem* 4.2, *if* $n^{-3/5}\log n = O(a)$, *then*

$$\limsup_n \pm \sqrt{\frac{nh\psi(\alpha,\delta)}{2\log\log n}}\, [\xi_s^*(x_0) - \xi(x)] = \sigma(x_0) \quad a.s.$$

REMARKS 4.4. (i) The case $a = o(h)$ corresponds to the interesting situation where we have smoothing over very narrow bands, but the asymptotics are dictated by those of $\tilde{\xi}$.

(ii) As noted earlier, the same smoothing procedure with the same results can be obtained for the unrestricted estimator $\tilde{\xi}$ of BG (1990).

(iii) The estimator $\tilde{\xi}$ uses $h \propto n^{-1/5}$. The results of BG (1990) (our Theorem 2.1) will still hold for somewhat smaller h, but with a remainder term of $O_{\mathrm{a.s.}}((nh)^{-3/4}\log n)$. If $h = o(a)$ (i.e., $\alpha \to 0$), but large enough for Lemma 2.3 to go through (see Remark 2 in Section 3), and $a = O(n^{-1/5})$, then Theorems 2.2, 4.1 and 4.2 will still hold, but another interesting limiting result comes out of Theorem 4.2. By expanding the integrand in the definition of $\psi(\alpha,\delta)$

around $\alpha = 0$, we note that

$$\psi(\alpha, \delta) = \frac{2\alpha}{\alpha + 1} \Big/ \int \left[F_\delta(s(\alpha + 1) + \alpha) - F_\delta(s(\alpha + 1) - \alpha) \right]^2 ds$$

$$= \frac{2\alpha}{\alpha + 1} \Big/ \left\{ \int [2\alpha\delta(s)]^2 \, ds + O(\alpha) \right\}$$

$$= \frac{1}{2\alpha(\alpha + 1)} \Big/ \left\{ \int \delta^2(s) \, ds + O(\alpha) \right\}.$$

Now $\sqrt{nh/[2\alpha(\alpha + 1)]} = \sqrt{n\alpha(1 + O(\alpha))}$. Thus, Theorem 4.2 in this case could be restated with the norming of $\sqrt{n\alpha}$ instead of $\sqrt{nh\psi(\alpha, \delta)}$, erasing the bias term involving h^2, and redefining

$$\sigma^2(x_0) \quad \text{as} \quad p(1 - p) \int \delta^2(s) \, ds / \left[f(x_0) g^2(\xi(x_0) | x_0) \right],$$

which is exactly the result one gets in a single stage kernel estimator with bandwidth a. If, in addition, $a = o(n^{-1/5})$, but still sufficiently large, then the remaining bias term also disappears. If h is very small then the representation given by Theorem 2.2 does not hold [see Remarks 2–4 in Section 3] and hence Theorem 4.2 does not necessarily hold.

Acknowledgments. The author wishes to thank two referees and an Associate Editor for instigating the research on the smoothed estimator.

REFERENCES

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

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

CASADY, R. J. and CRYER, J. D. (1976). Monotone percentile regression. *Ann. Statist.* **4** 532–541.

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

CRYER, J. D., ROBERTSON, T., WRIGHT, F. T. and CASADY, R. J. (1972). Monotone median regression. *Ann. Math. Statist.* **43** 1459–1469.

HÄRDLE, W. (1984). A law of the iterated logarithm for nonparametric regression function estimators. *Ann. Statist.* **12** 624–635.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

MUKERJEE, H. (1988). Monotone nonparametric regression. *Ann. Statist.* **16** 741–750.

PARSONS, V. L. (1979). The limiting distribution of the isotonic estimator at a point. Unpublished manuscript.

ROBERTSON, T. and WALTMAN, P. (1968). On estimating monotone parameters. *Ann. Math. Statist.* **39** 1030–1039.

ROBERTSON, T. and WRIGHT, F. T. (1974). A norm reducing property for isotonized Cauchy mean value functions. *Ann. Statist.* **2** 528–534.

SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184.

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

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448.

WRIGHT, F. T. (1984). The asymptotic behavior of monotone percentile regression estimates. *Canad. J. Statist.* **12** 229–236.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
1845 FAIRMOUNT
WICHITA STATE UNIVERSITY
WICHITA, KANSAS 67208-1595