The Annals of Statistics
1992, Vol. 20, No. 1, 98–120

ASYMPTOTIC NORMALITY OF THE RECURSIVE KERNEL REGRESSION ESTIMATE UNDER DEPENDENCE CONDITIONS

BY GEORGE G. ROUSSAS AND LANH T. TRAN

University of California, Davis, and Indiana University

For $i = 1, 2, \dots,$ let X_i and Y_i be $\mathbb{R}^d$-valued ($d \geq 1$ integer) and $\mathbb{R}$-valued, respectively, random variables, and let $\{(X_i, Y_i)\}$, $i \geq 1$, be a strictly stationary and α-mixing stochastic process. Set $m(x) = \mathscr{E}(Y_1 | X_1 = x)$, $x \in \mathbb{R}^d$, and let $\hat{m}_n(x)$ be a certain recursive kernel estimate of $m(x)$. Under suitable regularity conditions and as $n \to \infty$, it is shown that $\hat{m}_n(x)$, properly normalized, is asymptotically normal with mean 0 and a specified variance. This result is established, first under almost sure boundedness of the Y_i's, and then by replacing boundedness by continuity of certain truncated moments. It is also shown that, for distinct points $x_1, \dots, x_N$ in $\mathbb{R}^d$ ($N \geq 2$ integer), the joint distribution of the random vector, $(\hat{m}_n(x_1), \dots, \hat{m}_n(x_N))$, properly normalized, is asymptotically N-dimensional normal with mean vector 0 and a specified covariance function.

1. Introduction and statement of the problem. Consider the strictly stationary time series $\{Z_i\}$, $i = 1, 2, \dots,$ where the random variables (r.v.'s) Z_i, $i \geq 1$, are defined on the probability space $(\Omega, \mathscr{A}, P)$ and take values in the Euclidean space $\mathbb{R}^t$ with $t \geq 1$ integer, and suppose we are interested in estimating some function of s r.v.'s in the future, $Z_{i+1}, \dots, Z_{i+s}$, given the immediately previous k r.v.'s, $Z_{i-k+1}, \dots, Z_i$, $i \geq k$. To put it differently, let φ: $\mathbb{R}^\nu \to \mathbb{R}$ ($\nu = st$) be a given measurable function for which the conditional expectation

$$\mathscr{E}\left[\varphi(Z_{i+1}, \dots, Z_{i+s}) | Z_{i-k+1} = z_{i-k+1}, \dots, Z_i = z_i\right]$$

is finite. Then the problem is that of estimating this conditional expectation.

This problem is a special case of the following one, where $\{(X_i, Y_i)\}$, $i \geq 1$, is a strictly stationary sequence of $\mathbb{R}^d$-valued ($d \geq 1$ integer) and $\mathbb{R}$-valued r.v.'s, respectively, and the objective is that of estimating the regression function $m(x) = \mathscr{E}(Y_1 | X_1 = x)$, assuming, of course, that it is finite. The quantity $m(x)$ will be estimated by means of the pairs (X_i, Y_i), $i = 1, \dots, n$, which we have at our disposal.

The proposed estimate is a recursive kernel-type estimate in that it is based on a recursive kernel estimate of the probability density function (p.d.f.) involved. To be more precise, let f be the p.d.f. of X_1 with respect to Lebesgue

---

Received November 1989; revised April 1991.
AMS 1980 *subject classifications*. Primary 62G05, 62M09; secondary 62J02, 62E20.
Key words and phrases. Asymptotic normality, recursive kernel regression estimate, dependence, strong mixing, asymptotic joint normality.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

measure, which is assumed to exist, let K be a p.d.f. defined on $\mathbb{R}^d$ and let $\{b_n\}$ be a sequence of positive numbers converging to 0. Here, as well as elsewhere in this paper, limits are taken as $n \to \infty$, unless otherwise specified. Let $\hat{f}_n(x)$ be the usual recursive kernel estimate of $f(x)$; that is,

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b_i^d}K\left(\frac{x-X_i}{b_i}\right). \tag{1.1}$$

Define $\hat{w}_n(x)$ by

$$\hat{w}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b_i^d}Y_iK\left(\frac{x-X_i}{b_i}\right). \tag{1.2}$$

Then the proposed estimate of $m(x)$ is $\hat{m}_n(x)$ defined by

$$\hat{m}_n(x) = \hat{w}_n(x)/\hat{f}_n(x). \tag{1.3}$$

The estimate $\hat{m}_n$ is a recursive counterpart of the regression estimate

$$m_n(x) = \sum_{i=1}^{n} Y_iK\left(\frac{x-X_i}{b_n}\right)\Bigg/\sum_{i=1}^{n} K\left(\frac{x-X_i}{b_n}\right)$$

proposed by Nadaraya (1964, 1970) and Watson (1964). The pointwise convergence of m_n to m is treated in Watson (1964) and Rosenblatt (1969). The uniform convergence of m_n with sharp rates was obtained by Mack and Silverman (1982). Schuster (1972) established the joint asymptotic normality of $m_n(x_1), \dots, m_n(x_N)$ at fixed points $x_1, \dots, x_N$.

In the independent case, weak conditions for various forms of consistency of $\hat{m}_n$ have been obtained by Ahmad and Lin (1976) and Devroye and Wagner (1980). Under dependence, m_n has been investigated by Robinson (1983, 1986), Collomb (1984), Collomb and Härdle (1986) and Roussas (1990). Again in the independent case, Devroye and Wagner (1980), Krzyżak and Pawlak (1984) and Greblicki and Pawlak (1987) have considered another recursive estimate, $\tilde{m}_n(x)$, of $m(x)$ defined by

$$\tilde{m}_n(x) = \sum_{i=1}^{n} Y_iK\left(\frac{x-X_i}{b_i}\right)\Bigg/\sum_{i=1}^{n} K\left(\frac{x-X_i}{b_i}\right).$$

This paper, however, concerns itself with $\hat{m}_n$ and establishes its asymptotic normality under a certain set of regularity conditions.

A problem closely related to the one discussed here is that of estimating $f(x)$ by $\hat{f}_n(x)$. For relevant papers, the reader is referred to Masry (1986, 1987), Masry and Györfi (1987), Pham and Tran (1985), Roussas (1989), Tran (1989, 1990) and the references therein. The methods of proof and the assumptions employed in this paper are reminiscent of those used in Robinson (1983), Masry (1986), Schuster (1972) and Tran (1990).

The recursive computation of $\hat{m}_n$ can be carried out by means of the relations $\hat{f}_0(x) = \hat{w}_0(x) = \hat{m}_0(x) = 0$, and for $n \geq 1$,

$$\hat{f}_n(x) = \frac{n-1}{n}\hat{f}_{n-1}(x) + \frac{1}{nb_n^d}K((x - X_n)/b_n),$$

$$\hat{w}_n(x) = \frac{n-1}{n}\hat{w}_{n-1}(x) + \frac{1}{nb_n^d}Y_n K((x - X_n)/b_n),$$

so that

$$\hat{m}_n(x) = \frac{(n-1)b_n^d \hat{f}_{n-1}(x)\hat{m}_{n-1}(x) + Y_n K((x - X_n)/b_n)}{(n-1)b_n^d \hat{f}_{n-1}(x) + K((x - X_n)/b_n)}.$$

This recursive property is particularly useful in large sample sizes since $\hat{m}_n$ can be easily updated with each additional observation. This is especially relevant in a time series context, where recently there has been an interest in the use of nonparametric estimates in very long financial time series. Also, under certain circumstances, the recursive estimate is more efficient than its nonrecursive counterpart m_n, when efficiency is measured in terms of the variance of an appropriate asymptotic (normal) distribution. Although recently recursive estimation of p.d.f.'s and regression functions has been given some attention, it is felt that its full potential has not been appreciated as yet. This paper may also be looked upon as a contribution toward this end. At this point, it should be mentioned that stationarity of the underlying process is extensively used throughout the paper. Nonstationary time series are certainly of great interest. The study of such stochastic processes is an undertaking of major proportions, which cannot be accommodated by minor modifications here.

The paper is organized as follows: In Section 2 the assumptions used throughout the paper are gathered together for easy reference, followed by some brief comments. The main results of the paper, Theorems 2.1–2.3, are also formulated in the same section. The following three sections are devoted to the proofs of the theorems. In the course of these proofs, a substantial amount of auxiliary results are needed. Their justification is deferred to an appendix in order not to disrupt the continuity of the arguments. The letter C will be used throughout to indicate constants whose values are unimportant and may vary.

2. Assumptions and statement of main results. One of the basic assumptions made in this paper is that of α-mixing. For the definition of the strong mixing property and relevant literature, the reader is referred to Rosenblatt (1956); see also Roussas and Ioannides (1987).

ASSUMPTION A.1.

(i) For $i \geq 1$, the sequence $\{(X_i, Y_i)\}$, $i \geq 1$, is strictly stationary.
(ii) The sequence $\{(X_i, Y_i)\}$, $i \geq 1$, is α-mixing (or strongly mixing).
(iii) $\mathscr{E}|Y_1|^{2+\delta} < \infty$ for some $\delta > 0$.
(iv) $\mathscr{E}(Y_1|X_1 = x)$ is finite for all $x \in \mathbb{R}^d$.
(v) $|Y_i(\omega)| \leq C$ for all $i \geq 1$ and almost all $[P]$, $\omega \in \Omega$.

ASSUMPTION A.2.

(i) For any i and j with $1 \leq i < j$, the r.v.'s X_i, Y_i, X_j, Y_j have a joint p.d.f. $f_{X_i, Y_i, X_j, Y_j}(\cdot, \cdot, \cdot, \cdot)$ with respect to Lebesgue measure.
(ii) The following moment of the p.d.f. $f_{X_1, Y_1}(\cdot, \cdot)$ is finite for all $x \in \mathbb{R}^d$: $h(x) = \int |y|^{2+\delta} f_{X_1, Y_1}(x, y)\, dy$, where δ is as in Assumption A.1(iii). Also set

$$w(x) = \int y f_{X_1, Y_1}(x, y)\, dy,$$

$$w^*(x) = \int |y| f_{X_1, Y_1}(x, y)\, dy,$$

$$v(x) = \int y^2 f_{X_1, Y_1}(x, y)\, dy.$$

For all $i \neq j$ and all values of the arguments involved,
(iii) $|f_{X_i, X_j}(x_i, x_j) - f_{X_i}(x_i) f_{X_j}(x_j)| \leq C$.
(iv) $|f_{(X_i, X_j)|Y_j}(x_i, x_j|y_j) - f_{X_j|Y_j}(x_j|y_j) f_{X_i}(x_i)| \leq C$.
(v) $f_{(X_i, X_j)|Y_i, Y_j}(x_i, x_j|y_i, y_j) \leq C$.
(vi) The p.d.f. $f_{X_1}(\cdot)$, to be denoted by $f(\cdot)$, has continuous second-order partial derivatives which are bounded in $\mathbb{R}^d$.
(vii) The function $w(\cdot)$ defined in Assumption A.2(ii) has continuous second-order partial derivatives which are bounded in $\mathbb{R}^d$.
(viii) For each $L > 0$, the following truncated moments

$$w_L(x) = \int_{(|y| \leq L)} y f_{X_1, Y_1}(x, y)\, dy,$$

$$v_L(x) = \int_{(|y| \leq L)} y^2 f_{X_1, Y_1}(x, y)\, dy,$$

$$h_L(x) = \int_{(|y| \leq L)} |y|^{2+\delta} f_{X_1, Y_1}(x, y)\, dy,$$

$$w_L^*(x) = \int_{(|y| \leq L)} y^2 f_{X_1, Y_1}(x, y)\, dy$$

are continuous, and so are the quantities $\overline{w}_L(\cdot)$, $\overline{v}_L(\cdot)$, $\overline{h}_L(\cdot)$ and $\overline{w}_L^*(\cdot)$, where $\overline{w}_L(\cdot) = w(\cdot) - w_L(\cdot)$, and similarly for the others.

ASSUMPTION A.3.

(i) The kernel K is a bounded p.d.f. defined on $\mathbb{R}^d$.
(ii) $\|x\|^d K(x) \to 0$ as $\|x\| \to \infty$, where $\|\cdot\|$ is the usual norm in $\mathbb{R}^d$.
(iii) $\int u_j K(u)\,du = 0$, $j = 1, \dots, d$, and $\int \|u\|^2 K(u)\,du < \infty$.

ASSUMPTION A.4.

(i) $b_n \downarrow 0$.
(ii) $(1/n)\sum_{i=1}^n (b_n/b_i)^{\lambda d} \to \theta_{\lambda d}$ for some $\lambda \in [1, 3)$.
(iii) $nb_n^{d+4} = o(1)$ and $\sum_{i=1}^n (b_i/b_n)^2 = O(n)$.

ASSUMPTION A.5.

(i) There exists a selection of positive numbers $c_n \to \infty$ such that

$$c_n = o\left(b_n^{-d}\right) \quad \text{and} \quad b_n^{-(1-\gamma)d} \sum_{l=c_n}^{\infty} \alpha^{1-\gamma}(l) \to 0,$$

where $\gamma = 2/(2+\delta)$ and δ is as in Assumption A.1(iii).
(ii) There is a choice of $p = p(n)$, $q = q(n)$ and $r = r(n)$ such that

$$qrn^{-1} \to 0, \qquad r\alpha(q) \to 0, \qquad p^2 n^{-1} b_n^{-d} \to 0, \qquad b_n^{-(1-\gamma)d} \sum_{l=q}^{\infty} p^{1-\gamma}(l) \to 0.$$

REMARK 2.1. Let the α-mixing coefficient be given by $\alpha(n) = O(n^{-\rho})$, $\rho > 0$, and consider the popular choice for the bandwidth $b(n) = n^{-\theta}$, $\theta > 0$. Then there always exist $p = p(n)$, $q = q(n)$ and $r = r(n)$ for which Assumptions A.4 and A.5 are satisfied. Detailed discussion may be found in Roussas and Tran (1989).

The notation $C(g)$ is used to indicate the set of continuity points of the function g.

THEOREM 2.1. *Let $\hat{m}_n(x)$ be defined by* (1.3) *and suppose that Assumptions* A.1, A.2(i)–(vii) *and* A.3–A.5 *are satisfied. Then, for every $x \in C(w^*) \cap C(v) \cap C(h)$ with $f(x) > 0$,*

$$\left(nb_n^d\right)^{1/2}\left[\hat{m}_n(x) - m(x)\right] \to_d N(0, \sigma^2(x)),$$

where

$$\sigma^2(x) = \left[f(x)v(x) - w^2(x)\right]\theta_d \int K^2(u)\,du \Big/ f^3(x).$$

THEOREM 2.2. *Under Assumptions* A.1(i)–(iv) *and* A.2–A.5 *and for every $x \in C(v) \cap C(h)$ with $f(x) > 0$, the conclusion of the previous theorem holds true.*

Finally, let $N \geq 2$ be an integer and let $x_1, \ldots, x_N$ be distinct points in $\mathbb{R}^d$ arbitrarily chosen but kept fixed. Then we have the following theorem.

THEOREM 2.3. *Under Assumptions* A.1(i)–(iv) *and* A.2–A.5 *and for every* N *distinct continuity points* $x_1, \ldots, x_N$ *of* v *and* h *with* $f(x_1), \ldots, f(x_N) > 0$,

$$\left((nb_n^d)^{1/2}[\hat{m}_n(x_1) - m(x_1)], \ldots, (nb_n^d)^{1/2}[\hat{m}_n(x_N) - m(x_N)]\right)$$
$$\to_d N(0, \Sigma_N(\mathbf{x})),$$

where $\mathbf{x} = (x_1, \ldots, x_N)$, *the covariance* $\Sigma_N(\mathbf{x})$ *is given by* $\Sigma_N(\mathbf{x}) = D(\mathbf{x})\theta_d \int K^2(u)\,du$ *and* $D(\mathbf{x})$ *is the diagonal matrix with (diagonal) elements given by*

$$\sigma_{ii}(x_i) = \begin{vmatrix} f(x_i) & w(x_i) \\ w(x_i) & v(x_i) \end{vmatrix} \Big/ f^3(x_i), \qquad i = 1, \ldots, N.$$

REMARK 2.2. Notice that the variance in Theorem 2.1 and the covariance in Theorem 2.3 are, in general, distinct from the corresponding quantities for the nonrecursive case with the same bandwidths, by the factor θ_d [see, e.g., the theorem on page 85 in Schuster (1972)]. This fact may lead to a reduction of the variance or of the covariance. This is, indeed, the case for the popular selection of the bandwidths $h_n = cn^{-\theta}$, $c > 0$, $\frac{1}{5} < \theta < \frac{1}{2}$. For details, the reader is referred to Example 8.1 in Roussas and Tran (1990).

3. Proof of Theorem 2.1. First, note that $m(x) = w(x)/f(x)$. Second, by (1.1)–(1.3) and the standard Cramér–Wold device [see, e.g., Theorem (xi), page 103, in Rao (1965)], the desired convergence in Theorem 2.1 is equivalent to

$$\xi_n(x) \to_d N(0, \tau_{cd}^2(x)) \quad \text{for every } c_1, d_1 \text{ in } \mathbb{R} \text{ with } c_1^2 + d_1^2 \neq 0, \tag{3.1}$$

where

$$\xi_n(x) = (nb_n^d)^{1/2}\left\{c_1\left[\hat{f}_n(x) - f(x)\right] + d_1[\hat{w}_n(x) - w(x)]\right\}$$

and

$$\tau_{cd}^2(x) = \left[c_1^2 f(x) + 2c_1 d_1 w(x) + d_1^2 v(x)\right]\theta_d \int K^2(u)\,du; \tag{3.2}$$

recall that the quantities $w(x)$, $v(x)$ are defined in Assumption A.2(ii). In turn, (3.1) is equivalent to the following two convergences; namely,

$$\psi_n(x) - \xi_n(x) \to 0, \tag{3.3}$$

$$\psi_n(x) \to_d N(0, \tau_{cd}^2(x)), \tag{3.4}$$

where $\psi_n(x) = (nd_n^d)^{1/2}\{c_1[\hat{f}_n(x) - \mathscr{E}\hat{f}_n(x)] + d_1[\hat{w}_n(x) - \mathscr{E}\hat{w}_n(x)]\}$.

Relations (3.3) and (3.4) are established in Sections 3.1 and 3.2 below. There and subsequently, we restrict ourselves in presenting the main points of the proofs. Details may be found in Roussas and Tran (1989).

3.1. *Proof of the convergence in* (3.3). It would be convenient to write $K_i = b_i^{-d}K((x - X_i)/b_i))$ and observe that

$$\mathscr{E}K_i = \int K(u)f(x - b_i u)\,du. \tag{3.5}$$

Then, from (1.1) and (3.5), we obtain

$$\mathscr{E}\hat{f}_n(x) - f(x) = n^{-1}\sum_{i=1}^{n}\int K(u)[f(x - b_i u) - f(x)]\,du. \tag{3.6}$$

Next, use Assumptions A.2(vi), A.3(iii) and A.4(iii), in conjunction with the dominated convergence theorem and relation (3.6), in order to obtain

$$(nb_n^d)^{1/2}[\mathscr{E}\hat{f}_n(x) - f(x)] \to 0. \tag{3.7}$$

Utilizing Assumptions A.2(vii), A.3(iii), A.4(iii) and working similarly with $\hat{w}_n(x)$, we obtain

$$(nd_n^d)^{1/2}[\mathscr{E}\hat{w}_n(x) - w(x)] \to 0. \tag{3.8}$$

Relations (3.7) and (3.8) establish (3.3).

3.2. *Proof of the convergence in* (3.4). Set

$$Z_i = Z_{in}(x) = b_n^{d/2}\{c_1(K_i - \mathscr{E}K_i) + d_1[(Y_iK_i) - \mathscr{E}(Y_iK_i)]\}, \tag{3.9}$$

and $S_n = S_n(x) = \sum_{i=1}^{n} Z_i$. Then $\psi_n = n^{-1/2}S_n$, and it will be shown that

$$n^{-1/2}S_n \to_d N(0, \tau_{cd}^2(x)). \tag{3.10}$$

In discussing the convergence in (3.10), we use the familiar technique of Doob (1953) (see Theorem 7.5, pages 228–231) of big blocks separated by small blocks; see also Masry (1986), page 262. More precisely, let $p = p(n)$, $q = q(n)$ be positive integers tending to ∞ and such that $p + q \le n$. Let $r = r(n)$ be the largest positive integer such that $r(p + q) \le n$. The quantities p, q and r are also selected, so that the first convergence in Assumption A.5(ii) is satisfied. Partition the set $\Delta_n = \{1, \dots, n\}$ into $2r + 1$ subsets $\Delta'_{mn}, \Delta''_{mn}$, $m = 1, \dots, r$, and Δ'''_{rn}, where

$$\Delta'_{mn} = \{(m - 1)(p + q) + 1, \dots, (m - 1)(p + q) + p,\ m = 1, \dots, r\},$$

$$\Delta''_{mn} = \{(m - 1)(p + q) + p + 1, \dots, m(p + q),\ m = 1, \dots, r\},$$

$$\Delta'''_{rn} = \{r(p + q) + 1, \dots, n\},$$

so that $\Delta_n = \Sigma_{m=1}^r \Delta'_{mn} + \Sigma_{m=1}^r \Delta''_{rn} + \Delta'''_{rn}$. Ignore dependence on x, and for $m = 1, \ldots, r$, set

$$y_{mn} = \sum_{i \in \Delta'_{mn}} Z_{in} = \sum_{i=k_m}^{k_m+p-1} Z_{in}, \qquad k_m = (m-1)(p+q)+1,$$

$$(3.11)\quad y'_{mn} = \sum_{j \in \Delta''_{mn}} Z_{jn} = \sum_{j=l_m}^{l_m+q-1} Z_{jn}, \qquad l_m = (m-1)(p+q)+p+1,$$

$$y'_{r+1} = \sum_{k \in \Delta'''_{rm}} Z_{kn} = \sum_{k=r(p+q)+1}^{n} Z_{kn}.$$

Also, set

$$(3.12)\qquad S'_n = \sum_{m=1}^r y_{mn}, \qquad S''_n = \sum_{m=1}^r y'_{mn}, \qquad S'''_n = y'_{r+1},$$

so that $S_n = S'_n + S''_n + S'''_n$. It will be shown (see Section A.1) that

$$(3.13)\qquad n^{-1}\left[\mathscr{E}(S''_n)^2 + \mathscr{E}(S'''_n)^2\right] \to 0,$$

$$(3.14)\qquad n^{-1/2} S'_n \to_d N(0, \tau^2_{cd}(x)).$$

Then the desired result in (3.10) will follow. In establishing (3.14), the following convergence is also needed (see Section A.2); namely,

$$(3.15)\qquad n^{-1} \sum_{m=1}^r \operatorname{var}(y_{mn}) \to \tau^2_{cd}(x),$$
$$x \in C(f) \cap C(w) \cap C(w^*) \cap C(v) \cap C(h).$$

Now we proceed as follows. Let Y_{mn}, $m = 1, \ldots, r$, be independent r.v.'s with Y_{mn} distributed as $y_{mn} n^{-1/2}$, so that $\mathscr{E} Y_{mn} = 0$ for all m. Let Φ_{mn} be the characteristic function (ch.f.) of y_{mn}, so that the ch.f. of $y_{mn} n^{-1/2}$ is $\Phi_{mn}(tn^{-1/2})$ and that of $\Sigma_{m=1}^r Y_{mn}$ is $\Pi_{m=1}^r \Phi_{mn}(tn^{-1/2})$. It will be shown that $n^{-1/2}\Sigma_{m=1}^r y_{mn} = n^{-1/2} S'_n$ and $\Sigma_{m=1}^r Y_{mn}$ have the same asymptotic distribution and that this distribution is $N(0, \tau^2_{cd}(x))$. This will establish the desired result. Now, by Lemma 1.1 in Volkonskii and Rozanov (1959) [see also Theorem 7.2 in Roussas and Ioannides (1987)],

$$(3.16)\qquad \left| \mathscr{E}\left[\prod_{m=1}^r \exp(itn^{-1/2} y_{mn})\right] - \prod_{m=1}^r \mathscr{E}\left[\exp(itn^{-1/2} y_{mn})\right] \right|$$
$$\le 16(r-1)\alpha(q).$$

At this point, suppose that p, q and r have been chosen as specified so far and also such that the second convergence in Assumption A.5(ii) is satisfied. Then the left-hand side of (3.16) tends to 0; or

$$\left| \mathscr{E}\left[\prod_{m=1}^r \exp(itn^{-1/2} y_{mn})\right] - \prod_{m=1}^r \Phi_{mn}(tn^{-1/2}) \right| \to 0.$$

It remains to be shown that $\Pi_{m=1}^{r}\Phi_{mn}(tn^{-1/2})$ converges to the ch.f. of $N(0, \tau_c^2(x))$. To this end, set

$$s_n^2 = \sum_{m=1}^{r} \operatorname{var}(Y_{mn}) \left(= n^{-1} \sum_{m=1}^{r} \operatorname{var}(y_{mn}) \right),$$

and let $X_{mn} = Y_{mn}/s_n$. Then the r.v.'s X_{mn}, $m = 1, \ldots, r$, are independent with $\mathscr{E}X_{mn} = 0$ and $\Sigma_{m=1}^{r} \operatorname{var}(X_{mn}) = 1$. So, by the normal convergence criterion [Loève (1963), page 295], $\Sigma_{m=1}^{r} X_{mn} \to_d N(0, 1)$, provided, for every $\varepsilon > 0$, $g_n(\varepsilon) = \Sigma_{m=1}^{r} \int_{(|x| \geq \varepsilon)} x^2 \, dF_{mn} \to 0$, where F_{mn} is the distribution function of X_{mn}. But

$$\int_{(|x| \geq \varepsilon)} x^2 \, dF_{mn} = \mathscr{E}\left[X_{mn}^2 I_{(|X_{mn}| \geq \varepsilon)} \right] = s_n^{-2} n^{-1} \mathscr{E}\left[y_{mn}^2 I_{(|y_{mn}| \geq \varepsilon s_n n^{-1/2})} \right].$$

From the definition of y_{mn}, the boundedness of the p.d.f. K, and the (almost sure) boundedness of the r.v.'s Y_i, we have

$$|y_{mn}| \leq C b_n^{d/2} \sum_{i=k_m}^{k_m+p-1} b_i^{-d} = C b_n^{-d/2} \sum_{i=k_m}^{k_m+p-1} \left(\frac{b_n}{b_i} \right)^d \leq C p b_n^{-d/2}.$$

Therefore,

$$\int_{(|x| \geq \varepsilon)} x^2 \, dF_{mn} \leq C \varepsilon^{-2} \left(s_n^2\right)^{-2} n^{-2} p^2 b_n^{-d} \operatorname{var}(y_{mn}),$$

so that

$$g_n(\varepsilon) \leq C \varepsilon^{-2} \left(s_n^2\right)^{-2} n^{-1} p^2 b_n^{-d} n^{-1} \sum_{m=1}^{r} \operatorname{var}(y_{mn}) = C \varepsilon^{-2} \left(s_n^2\right)^{-1} \left(\frac{p^2}{n b_n^d} \right).$$

Since $s_n^2 \to \tau_{cd}^2(x)$ [by (3.15)], the right-hand side above converges to 0, on account of the third convergence in Assumption A.5(ii). This completes the proof of (3.4) and therefore that of Theorem 2.1. □

REMARK 3.1. It is to be noticed that after relation (3.16) the arguments proceed as in the independent case. On the basis of such an observation, a referee has suggested that the lack of any influence of the dependence structure may suggest a very weak kind of dependence, hidden under the conditions imposed.

4. Proof of Theorem 2.2. For some $L > 0$, define $Y_i' = Y_i I_{(|Y_i| \leq L)}$ and set $Y_i'' = Y_i - Y_i' = Y_i I_{(|Y_i| > L)}$, so that $Y_i = Y_i' + Y_i''$. In (3.9), replace Y_i by Y_i' and let $Z_i' = b_n^{d/2}\{c_1(K_i - \mathscr{E}K_i) + d_1[(Y_i'K_i) - \mathscr{E}(Y_i'K_i)]\}$, so that $Z_i = Z_i' + d_1 b_n^{d/2}[(Y_i''K_i) - \mathscr{E}(Y_i''K_i)]$. By setting

$$T_n' = \sum_{i=1}^{n} Z_i', \qquad T_n'' = d_1 b_n^{d/2} \sum_{i=1}^{n} \left[(Y_i''K_i) - \mathscr{E}(Y_i''K_i) \right],$$

we then have $S_n = T_n' + T_n''$. Next, by taking the limits as $n \to \infty$ first and as $L \to \infty$ next, it is shown (see Section A.3) that, for $x \in C(h)$,

$$\text{var}(n^{-1/2}T_n'') \to 0. \tag{4.1}$$

On the other hand, for each $L > 0$, relation (3.10) implies that

$$n^{-1/2}T_n' \to_d N(0, \tau_{cdL}^2(x)), \tag{4.2}$$

where

$$\tau_{cdL}^2(x) = [c_1^2 f(x) + 2c_1 d_1 w_L(x) + d_1^2 v_L(x)]\theta_d \int K^2(u)\,du \tag{4.3}$$

and

$$w_L(x) = \int_{(|y| \le L)} y f_{X_1, X_2}(x, y)\,dy, \qquad v_L(x) = \int_{(|y| \le L)} y^2 f_{X_1, X_2}(x, y)\,dy.$$

Finally, it is also shown (see Section A.4) that

$$\tau_{cdL}^2(x) \to \tau_{cd}^2(x) \quad \text{as } L \to \infty. \tag{4.4}$$

Combining relations (4.1), (4.2) and (4.4), we obtain

$$n^{-1/2}S_n \to_d N(0, \tau_{cd}^2(x)), \tag{4.5}$$

and, in effect, the proof of Theorem 2.2.

[Relevant with the argument below is also the proof of Theorem 18.5.2, pages 344–346, in Ibragimov and Linnik (1971).] In the derivations below, write τ^2 and τ_L^2 rather than $\tau_{cd}^2(x)$ and $\tau_{cdL}^2(x)$, respectively. We have to show that the ch.f. of $n^{-1/2}S_n$ converges to $\exp(-t^2\tau^2/2)$ for all $t \in \mathbb{R}$. Indeed,

$$\begin{aligned} |\mathscr{E}e^{itn^{-1/2}S_n} - e^{-t^2\tau^2/2}| &= |\mathscr{E}e^{itn^{-1/2}T_n' + itn^{-1/2}T_n''} - e^{-t^2\tau^2/2}| \\ &\le \mathscr{E}|e^{itn^{-1/2}T_n''} - 1| + |\mathscr{E}e^{itn^{-1/2}T_n'} - e^{-t^2\tau_L^2/2}| \\ &\quad + |e^{-t^2\tau_L^2/2} - e^{-t^2\tau_G^2/2}|. \end{aligned}$$

The first term on the right-hand side of the above inequality tends to 0 by (4.1). The second term tends to 0 by (4.3). Next, take the limit as $L \to \infty$. The last term converges to 0 by (4.4). This completes the proof. □

5. Proof of Theorem 2.3. Without loss of generality, it suffices to prove the theorem for $N = 2$. Arguing as in Section 3 and utilizing Theorem (iii) on page 322 in Rao (1965), it suffices to show that, for all c_i, d_i in $\mathbb{R}$ with $c_i^2 + d_i^2 \neq 0$, $i = 1, 2$, $\xi_n(\mathbf{x}) \to_d N(0, \tau_{\mathbf{cd}}^2(\mathbf{x}))$, where

$$\begin{aligned} \xi_n(\mathbf{x}) = (nb_n^d)^{1/2}\{&c_1[\hat{f}_n(x_1) - f(x_1)] + d_1[\hat{w}_n(x_1) - w(x_1)] \\ &+ c_2[\hat{f}_n(x_2) - f(x_2)] + d_2[\hat{w}_n(x_2) - w(x_2)]\}, \end{aligned}$$

$$\begin{aligned} \tau_{\mathbf{cd}}^2(\mathbf{x}) = \{&[c_1^2 f(x_1) + 2c_1 d_1 w(x_1) + d_1^2 v(x_1)] \\ &+ [c_2^2 f(x_2) + 2c_2 d_2 w(x_2) + d_2^2 v(x_2)]\}\theta_d \int K^2(u)\,du. \end{aligned}$$

Also, define $\psi_n(\mathbf{x})$ by

$$\psi_n(\mathbf{x}) = (nb_n^d)^{1/2}\{c_1[\hat{f}_n(x_1) - \mathscr{E}\hat{f}_n(x_1)] + d_1[\hat{w}_n(x_1) - \mathscr{E}\hat{w}_n(x_1)] + c_2[\hat{f}_n(x_2) - \mathscr{E}\hat{f}_n(x_2)] + d_2[\hat{w}_n(x_2) - \mathscr{E}\hat{w}_n(x_2)]\}.$$

Then, as in Section 3, the proof of the desired result will be completed by showing that

$$(5.1) \qquad \psi_n(\mathbf{x}) - \xi_n(\mathbf{x}) \to 0,$$

$$(5.2) \qquad \psi_n(\mathbf{x}) \to_d N(0, \tau^2_{\mathbf{cd}}(\mathbf{x})).$$

For $j = 1,2$, set

$$\xi_{nj} = (nb_n^d)^{1/2}\{c_j[\hat{f}_n(x_j) - f(x_j)] + d_j[\hat{w}_n(x_j) - w(x_j)]\},$$

$$\psi_{nj} = (nb_n^d)^{1/2}\{c_j[\hat{f}_n(x_j) - \mathscr{E}\hat{f}_n(x_j)] + d_j[\hat{w}_n(x_j) - \mathscr{E}\hat{w}_n(x_j)]\},$$

so that

$$\psi_n(\mathbf{x}) = \psi_n = \psi_{n1} + \psi_{n2}, \qquad \xi_n(\mathbf{x}) = \xi_n = \xi_{n1} + \xi_{n2}.$$

Thus $\psi_n(\mathbf{x}) - \xi_n(\mathbf{x}) = (\psi_{n1} - \xi_{n1}) + (\psi_{n2} - \xi_{n2})$, and the convergence in (5.1) follows by (3.3). In order to establish (5.2), define

$$Z_i(x_j, c_j, d_j) = Z_{ij} = b_n^{d/2}[c_j[K_i(x_j - X_i) - \mathscr{E}K_i(x_j - X_i)] + d_j\{Y_iK_i(x_j - X_i) - \mathscr{E}[Y_iK_i(x_j - X_i)]\}].$$

Then observe that

$$\begin{aligned} Z_i(\mathbf{x}, \mathbf{c}, \mathbf{d}) = Z_i(\mathbf{x}) = b_n^{d/2}[&c_1[K_i(x_1 - X_i) - \mathscr{E}K_i(x_1 - X_i)] \\ &+d_1\{Y_iK_i(x_1 - X_i) - \mathscr{E}[Y_iK_i(x_1 - X_i)]\} \\ &+c_2[K_i(x_2 - X_i) - \mathscr{E}K_i(x_2 - X_i)] \\ &+d_2\{Y_iK_i(x_2 - X_i) - \mathscr{E}[Y_iK_i(x_2 - X_i)]\}]. \end{aligned}$$

Next, the quantities $y_{mn}(\mathbf{x})$, $y'_{mn}(\mathbf{x})$ and $y'_{r+1}(\mathbf{x})$ are defined as in (3.11) by employing the $Z_i(\mathbf{x})$'s. Thus

$$y_{mn}(\mathbf{x}) = y_{mn}(x_1) + y_{mn}(x_2),$$

$$y'_{mn}(\mathbf{x}) = y'_{mn}(x_1) + y'_{mn}(x_2), \qquad y'_{r+1}(\mathbf{x}) = y'_{r+1}(x_1) + y'_{r+1}(x_2).$$

The definition of the quantities $S'_n(\mathbf{x})$, $S''_n(\mathbf{x})$ and $S'''_n(\mathbf{x})$ is obvious. By (3.12),

$$S''_n(\mathbf{x}) = S''_n(x_1) + S''_n(x_2), \qquad S'''_n(\mathbf{x}) = S'''_n(x_1) + S'''_n(x_2).$$

Then, by (3.13),

$$(5.3) \qquad n^{-1}\{\mathscr{E}[S''_n(\mathbf{x})]^2 + \mathscr{E}[S'''_n(\mathbf{x})]^2\} \to 0.$$

As in (3.15), it is also shown (see Section A.5) that

$$n^{-1} \sum_{m=1}^{r} \operatorname{var}[y_{mn}(\mathbf{x})] \to \tau_{\mathbf{cd}}^2(\mathbf{x}). \tag{5.4}$$

By (5.3) and (5.4), the arguments used, following relation (3.15), establish (5.2) for the bounded case; see Assumption A.1(v).

In the unbounded case, $Z_i(\mathbf{x}) = Z'_{i1} + Z''_{i2}$, $i = 1, \dots, n$, where

$$Z'_{ij} = b_n^{d/2}\big[c_j\big[K_i(x_j - X_i) - \mathscr{E}K_i(x_j - X_i)\big] + d_j\big\{\big[Y'_iK_i(x_j - X_i)\big] - \mathscr{E}\big[Y'_iK_i(x_j - X_i)\big]\big\}\big]$$

and

$$Z''_{ij} = d_j b_n^{d/2}\big\{\big[Y''_iK_i(x_j - X_i)\big] - \mathscr{E}\big[Y''_iK_i(x_j - X_i)\big]\big\}, \qquad j = 1, 2.$$

Thus

$$n^{-1/2}S_n(\mathbf{x}) = n^{-1/2}T'_n(\mathbf{x}) + n^{-1/2}T''_n(x_1) + n^{-1/2}T''_n(x_2).$$

As in Section 4, by taking the limits as $n \to \infty$ first and as $L \to \infty$ next, we have, by (4.1) [where Assumption A.1(v) is not used],

$$\operatorname{var}\big[n^{-1/2}T''_n(x_1)\big] + \operatorname{var}\big[n^{-1/2}T''_n(x_2)\big] \to 0. \tag{5.5}$$

On the other hand, as in (4.3), for each $L > 0$ and as $n \to \infty$,

$$n^{-1/2}T'_n(x_1) \to_d N\big(0, \tau_{\mathbf{cd}L}^2(\mathbf{x})\big), \tag{5.6}$$

where

$$\tau_{\mathbf{cd}L}^2(\mathbf{x}) = \big\{\big[c_1^2 f(x_1) + 2c_1d_1w_L(x_1) + d_1^2v_L(x_1)\big] + \big[c_2^2 f(x_2) + 2c_2d_2w_L(x_2) + d_2^2v_L(x_2)\big]\big\}\theta_d\int K^2(u)\,du.$$

Also, as in (4.4),

$$\tau_{\mathbf{cd}L}^2(\mathbf{x}) \to \tau_{\mathbf{cd}}^2(\mathbf{x}) \quad \text{as } L \to \infty. \tag{5.7}$$

Its justification is the same as that given in the proof of (4.4). On the basis of (5.5)–(5.7), the arguments employed in Section 4 after relation (4.5) complete the justification of (5.2). The proof of Theorem 2.3 is thus concluded. □

APPENDIX

A.1. Proof of relation (3.13). Under assumptions made in the paper and by means of (3.5), we have

$$\mathscr{E}K_i = \int K(u)f(x - b_iu)\,du \to f(x), \qquad x \in C(f). \tag{A.1.1}$$

This is so by the d-dimensional version of Theorem 1.A in Parzen (1962), which is a modified version of a result in Stein (1970). Likewise,

$$b_i^d \mathscr{E} K_i^2 = \int K^2(u) f(x - b_i u)\, du \to f(x) \int K^2(u)\, du \tag{A.1.2}$$

and also

$$\mathscr{E}(Y_i K_i) = \int K(u) w(x - b_i u)\, du \to w(x), \qquad x \in C(w), \tag{A.1.3}$$

$$b_i^d \mathscr{E}(Y_i K_i)^2 = \int K^2(u) v(x - b_i u)\, du \to v(x) \int K^2(u)\, du, \qquad x \in C(v) \tag{A.1.4}$$

and

$$b_i^d \mathscr{E}(Y_i K_i^2) = \int K^2(u) w(x - b_i u)\, du \to w(x) \int K^2(u)\, du, \qquad x \in C(w). \tag{A.1.5}$$

Next, from (3.9), (3.11) and (3.12), we have

$$n^{-1} \mathscr{E}(S_n'')^2 = n^{-1} \sum_{m=1}^{r} \operatorname{var}(y'_{mn}) + 2n^{-1} \sum_{1 \le i < j \le r} \operatorname{cov}(y'_{in}, y'_{jn}) \tag{A.1.6}$$

and

$$\begin{aligned} \operatorname{var}(y'_{mn}) &= \operatorname{var}\left(\sum_{i=l_m}^{l_m+q-1} Z_i \right) \\ &= \sum_{i=l_m}^{l_m+q-1} \operatorname{var}(Z_i) + 2 \sum_{l_m \le i < j \le l_m+q+1} \operatorname{cov}(Z_i, Z_j). \end{aligned} \tag{A.1.7}$$

Since $b_n \downarrow$ [by Assumption A.4(i)], relations (A.1.1)–(A.1.5) lead to

$$\operatorname{var}(Z_i) \le C \quad \text{for all } i. \tag{A.1.8}$$

Hence

$$n^{-1} \sum_{m=1}^{r} \sum_{i=l_m}^{l_m+q-1} \operatorname{Var}(Z_i) \le C\left(\frac{qr}{n}\right) \to 0,$$

by means of the first convergence in Assumption A.5(ii).

Next, it will be shown that the second term on the right-hand side of (A.1.7) also converges to 0, when summed over m. In so doing, we borrow a technique used by Masry (1986). More precisely, divide the set of pairs (i, j) with i, j in $\{1, \dots, n\}$ and $i < j$ as follows:

$$S_1 = \{(i, j) | i, j \in \{1, \dots, n\}, 1 \le j - i \le c_n\},$$

$$S_2 = \{(i, j) | i, j \in \{1, \dots, n\}, c_n + 1 \le j - i \le n - 1\},$$

where $\{c_n\}$ is a suitably chosen sequence of positive numbers tending to ∞ [see

Assumption A.5(i)]. Then

$$\sum_{1\le i<j\le n} \operatorname{cov}(K_i, K_j) = \sum_{S_1} \operatorname{cov}(K_i, K_j) + \sum_{S_2} \operatorname{cov}(K_i, K_j) \stackrel{\text{df}}{=} J_{1n} + J_{2n}.$$

By Assumption A.2(iii),

$$|J_{1n}| \le \sum_{S_1} \left|\operatorname{cov}(K_i, K_j)\right| \le C \sum_{S_1} 1 = C\left(\sum_{j=2}^{n}\sum_{i=1}^{j-2} 1 + \sum_{j=c_{n+1}}^{n}\sum_{i=j-c_n}^{n} 1\right) < C_n,$$

so that $n^{-1}b_n^d|J_{1n}| \le Cc_n b_n^d$. If c_n is chosen as in the first part of Assumption A.5(i), then

$$n^{-1}b_n^d J_{1n} \to 0. \tag{A.1.9}$$

Next, work with J_{2n} and employ Davydov inequality [see, e.g., Deo (1973), or Theorem 7.3 in Roussas and Ioannides (1987)] to obtain, by means of Assumption A.1(ii),

$$\left|\operatorname{cov}(K_i, K_j)\right| \le 10\alpha^{\delta/(2+\delta)}(j-i)\left(\mathscr{E}K_i^{2+\delta}\right)^{1/(2+\delta)}\left(\mathscr{E}K_j^{2+\delta}\right)^{1/(2+\delta)}$$

for some $\delta > 0$ [which is taken to be the same as that in Assumption A.1(iii)]. Set

$$\gamma = \frac{2}{2+\delta}, \quad \text{so that } 0 < \gamma < 1,\ \delta = \frac{2(1-\gamma)}{\gamma}, \qquad 2+\delta = \frac{2}{\gamma},\ \frac{\delta}{2+\delta} = 1-\gamma. \tag{A.1.10}$$

Then

$$\left|\operatorname{cov}(K_i, K_j)\right| \le 10\alpha^{1-\gamma}(j-i)\left(\mathscr{E}K_i^{2/\gamma}\right)^{\gamma/2}\left(\mathscr{E}K_i^{2/\gamma}\right)^{\gamma/2}.$$

Working as above,

$$\mathscr{E}K_i^{2/\gamma} = b_i^{-(1+\delta)d} q_i(x), \qquad q_i(x) \stackrel{\text{df}}{=} \int K^{2/\gamma}(u) f(x - b_i u)\, du \tag{A.1.11}$$

with

$$q_i(x) \to f(x)\int K^{2/\gamma}(u)\, du, \qquad x \in C(f). \tag{A.1.12}$$

Therefore, by also employing the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} n^{-1}b_n^d|J_{2n}| &\le 10 n^{-1} b_n^d \sum_{S_2} \alpha^{1-\gamma}(j-i)\left[b_i^{-(1+\delta)d}q_i(x)\right]^{\gamma/2}\left[b_j^{-(1+\delta)d}q_j(x)\right]^{\gamma/2} \\ &\le 10 b_n^{-(1-\gamma)d}\sum_{l=c_n}^{\infty}\alpha^{1-\gamma}(l)\left[\sum_{i=1}^{n}\frac{1}{n}\left(\frac{b_n}{b_i}\right)^{(2-\gamma)d} q_i^{\gamma}(x)\right]. \end{aligned} \tag{A.1.13}$$

The second term on the right-hand side above converges to a finite quantity by

Assumption A.4(ii) and (A.1.12). The first term on the same side converges to 0 by means of the second part of Assumption A.5(i). Then combining relations (A.1.9) and (A.1.13), we obtain the following lemma.

LEMMA A.1.1. *Under Assumptions* A.1(i)–(ii), A.2(i), (iii), A3(i)–(ii), A.4(i)–(ii) *and* A.5(i),

$$n^{-1}b_n^d \sum_{1\le i<j\le n} \left|\operatorname{cov}(K_i, K_j)\right| \to 0, \qquad x \in C(f)$$

and

$$\sum_{1\le i<j\le n} \operatorname{cov}(K_i, Y_j K_j) = \sum_{S_1} \operatorname{cov}(K_i, Y_j K_j) + \sum_{S_2} \operatorname{cov}(K_i, Y_j K_j)$$
$$\stackrel{\text{df}}{=} J_{1n} + J_{2n}.$$

On the other hand,

$$\operatorname{cov}(K_i, Y_j K_j) = \int\!\!\int\!\!\int wK(u)K(v)\big[f_{(X_i, X_j)|Y_j}(x - b_i u, x - b_j v|w)$$
$$-f_{X_j|Y_j}(x - b_j v|w) f_{X_i}(x - b_i u)\big] f_{Y_j}(w)\, du\, dv\, dw.$$

Then, by Assumption A.2(iv) and the first part of Assumption A.5(i),

$$n^{-1}b_n^d J_{1n} \to 0. \tag{A.1.14}$$

On the set S_2, working as above, we have

$$\left|\operatorname{cov}(K_i, Y_j K_j)\right| \le 10\alpha^{1-\gamma}(j-i)\left(\mathscr{E}K_i^{2/\gamma}\right)^{\gamma/2}\left(\mathscr{E}|Y_j K_j|^{2/\gamma}\right)^{\gamma/2},$$

$$\mathscr{E}|Y_j K_j|^{2/\gamma} = b_j^{-(1+\delta)d} h_j(x), \qquad h_j(x) \stackrel{\text{df}}{=} \int K^{2/\gamma}(u) h(x - b_j u)\, du,$$

where h is as in Assumption A.2(ii). [Recall that $2/\gamma = 2 + \delta$ from (A.1.10).] Also,

$$h_j(x) \to h(x), \qquad x \in C(h). \tag{A.1.15}$$

Next, working as in (A.1.12) and utilizing (A.1.11), we have by the Cauchy–Schwarz inequality

$$n^{-1}b_n^d|J_{2n}| \le 10 b_n^{-(1-\gamma)d} \sum_{l=c_n}^{\infty} \alpha^{1-\gamma}(l)$$
$$\times \left\{\left[\sum_{i=1}^{n} \frac{1}{n}\left(\frac{b_n}{b_i}\right)^{(2-\gamma)d} q_i^{\gamma}(x)\right]\left[\sum_{j=1}^{n} \frac{1}{n}\left(\frac{b_n}{b_j}\right)^{(2-\gamma)d} h_j^{\gamma}(x)\right]\right\}^{1/2}.$$

Therefore, by the second part of Assumption A.5(i), relations (A.1.12), (A.1.15), Theorem 1.A in Parzen (1962) and the Toeplitz lemma, we have

$$n^{-1}b_n^d J_{2n} \to 0. \tag{A.1.16}$$

Then combining relations (A.1.14) and (A.1.16), we obtain the following lemma.

LEMMA A.1.2. *Under Assumptions* A.1(i)–(iii), A.2(i)–(ii), (iv), A.3(i)–(ii), A.4(i)–(ii) *and* A.5(i),

$$n^{-1}b_n^d \sum_{1 \le i < j \le n} \left|\operatorname{cov}(K_i, Y_j K_j)\right| \to 0, \qquad x \in C(f) \cap C(h).$$

Finally, we occupy ourselves with $\operatorname{cov}(Y_i K_i, Y_j K_j)$. First,

$$\left|\operatorname{cov}(Y_i K_i, Y_j K_j)\right| \le \mathscr{E}|Y_i K_i Y_j K_j| + \mathscr{E}|Y_i K_i|\mathscr{E}|Y_j K_j|, \tag{A.1.17}$$

and by means of Assumptions A.1(iii) and A.3(i),

$$\mathscr{E}|Y_i K_i| \le C \quad \text{and likewise} \quad \mathscr{E}|Y_j K_j| \le C \qquad \text{for all } i \text{ and } j,\ 1 \le i < j. \tag{A.1.18}$$

Also,

$$\mathscr{E}|Y_i K_i Y_j K_j| \le C\mathscr{E}|Y_i Y_j| \le C\mathscr{E}|Y_1|^2 \le C \quad [\text{by Assumption A.1(iii)}].$$

That is,

$$\mathscr{E}|Y_i K_i Y_j K_j| \le C \quad \text{for all } i, j,\ 1 \le i < j. \tag{A.1.19}$$

Therefore, by means of (A.1.17)–(A.1.19), we have on S_1:

$$n^{-1}b_n^d|J_{n1}| \to 0. \tag{A.1.20}$$

On S_2,

$$\left|\operatorname{cov}(Y_i K_i, Y_j K_j)\right| \le 10\alpha^{1-\gamma}(j-i)\left(\mathscr{E}|Y_i K_i|^{2/\gamma}\right)^{\gamma/2}\left(\mathscr{E}|Y_j K_j|^{2/\gamma}\right)^{\gamma/2}. \tag{A.1.21}$$

But

$$\mathscr{E}|Y_i K_i|^{2/\gamma} = b_i^{-(1+\delta)d} h_i(x), \tag{A.1.22}$$

where $h_i(x) = \int K^{2/\gamma}(u)h(x - b_i u)\,du$ and h is as in Assumption A.2(ii) and

$$h_i(x) \to h(x), \qquad x \in C(h). \tag{A.1.23}$$

On the basis of (A.1.21)–(A.1.23), we proceed as in (A.1.13) to obtain

$$n^{-1}b_n^d|J_{2n}| \to 0. \tag{A.1.24}$$

Finally, combining (A.1.20) and (A.1.24), we have the following lemma.

LEMMA A.1.3. *Under Assumptions* A.1(i)–(iii), A.2(i)–(ii), (v), A.3(i)–(ii), A.4(i)–(ii) *and* A.5(i),

$$n^{-1}b_n^d \sum_{1 \le i < j \le n} \left|\operatorname{cov}(Y_i K_i, Y_j K_j)\right| \to 0, \qquad x \in C(w) \cap C(h).$$

Next,

$$\operatorname{cov}(Z_i, Z_j) = b_n^d c_1^2 \operatorname{cov}(K_i, K_j) + b_n^d c_1 d_1 \operatorname{cov}(K_i, Y_j K_j) + b_n^d c_1 d_1 \operatorname{cov}(Y_i K_i, K_j) + b_n^d d_1^2 \operatorname{cov}(Y_i K_i, Y_j K_j),$$

and by Lemmas A.1.1–A.1.3,

$$n^{-1} b_n^d \sum_{m=1}^{r} \sum_{l_m \le i < j \le l_m + q - 1} |\operatorname{cov}(K_i, K_j)| \le n^{-1} b_n^d \sum_{1 \le i < j \le n} |\operatorname{cov}(K_i, K_j)| \to 0,$$

$$n^{-1} b_n^d \sum_{m=1}^{r} \sum_{l_m \le i < j \le l_m + q - 1} |\operatorname{cov}(K_i, Y_j K_j)| \le n^{-1} b_n^d \sum_{1 \le i < j \le n} |\operatorname{cov}(K_i, Y_j K_j)| \to 0,$$

$$n^{-1} b_n^d \sum_{m=1}^{r} \sum_{l_m \le i < j \le l_m + q - 1} |\operatorname{cov}(Y_i K_i, Y_j K_j)| \le n^{-1} b_n^d \sum_{1 \le i < j \le n} |\operatorname{cov}(Y_i K_i, Y_j K_j)| \to 0,$$

so that

$$n^{-1} \sum_{m=1}^{r} \sum_{l_m \le i < j \le l_m + q - 1} |\operatorname{cov}(Z_i, Z_j)| \le n^{-1} \sum_{1 \le i < j \le n} |\operatorname{cov}(Z_i, Z_j)| \to 0. \tag{A.1.25}$$

Combining relations (A.1.7)–(A.1.18) and (A.1.25), we then have the following lemma.

LEMMA A.1.4. *Under Assumptions* A.1(i)–(iii), A.2(i)–(v), A.3(i)–(ii), A.4(i)–(ii), A.5(i) *and the first convergence in* A.5(ii),

$$n^{-1} \sum_{m=1}^{r} \operatorname{var}(y'_{mn}) \to 0, \qquad x \in C(f) \cap C(w) \cap C(v) \cap C(h).$$

Now, by (3.11),

$$\sum_{1 \le i < j \le r} |\operatorname{cov}(y'_{in}, y'_{jn})| \le \sum_{1 \le i < j \le r} \sum_{k=l_i}^{l_i+q-1} \sum_{l=l_j}^{l_j+q-1} |\operatorname{cov}(Z_k, Z_l)|.$$

At this point, recall that any two Z_k, Z_l have indices which differ by at least p.

Therefore, the right-hand side above is bounded by

$$C\sum_{k=1}^{n-p}\sum_{l=k+p}^{n}|\operatorname{cov}(Z_k, Z_l)|.$$

In other words,

$$n^{-1}\sum_{1\le i<j\le r}\left|\operatorname{cov}(y'_{in}, y'_{jn})\right| \le Cn^{-1}\sum_{1\le i<j\le n}\left|\operatorname{cov}(Z_i, Z_j)\right|,$$

and this last expression tends to 0 by (A.1.25). This result and Lemma A.1.4 imply, by way of relation (A.1.6), that

$$n^{-1}\mathscr{E}(S''_n)^2 \to 0. \tag{A.1.26}$$

Finally, from (3.11), (3.12) and (A.1.8),

$$\begin{aligned} n^{-1}\mathscr{E}(S'''_n)^2 &= n^{-1}\sum_{i=r(p+q)+1}^{n}\operatorname{var}(Z_i) \\ &\le C\frac{n-r(p+q)}{n} = C\left(1-\frac{pr}{n}-\frac{qr}{n}\right) \to 0, \end{aligned} \tag{A.1.27}$$

since $qr/n \to 0$ by Assumption A.5(ii), and $pr/n \to 1$, as is easily seen. Relations (A.1.26) and (A.1.27) complete the proof of (3.13). □

A.2. Proof of relation (3.15). From (3.9) and (3.11), it is clear that

$$\begin{aligned} \operatorname{var}(y_{mn}) &= b_n^d\sum_{i=k_m}^{k_m+p-1}\operatorname{var}(c_1K_i + d_1Y_iK_i) \\ &\quad + 2b_n^d\sum_{k_m\le i<j\le k_m+p-1}\operatorname{cov}(c_1K_i + d_1Y_iK_i, c_1K_j + d_1Y_jK_j). \end{aligned} \tag{A.2.1}$$

Therefore, the desired result will follow from (A.2.1), in conjunction with Lemma A.1.4 and relation (A.1.27), by showing that

$$n^{-1}b_n^d\sum_{i=1}^{n}\operatorname{var}(c_1K_i + d_1Y_iK_i) \to \tau_{cd}^2(x) \tag{A.2.2}$$

and

$$n^{-1}b_n^d\sum_{1\le i<j\le n}\left|\operatorname{cov}(c_1K_i + d_1Y_iK_i, c_1K_j + d_1Y_jK_j)\right| \to 0. \tag{A.2.3}$$

From (A.1.1), Assumption A.4(i)–(ii) and the Toeplitz lemma,

$$n^{-1}b_n^d\sum_{i=1}^{n}(\mathscr{E}K_i)^2 = \sum_{i=1}^{n}\frac{1}{n}\left(\frac{b_n}{b_i}\right)^d b_i^d(\mathscr{E}K_i)^2 \to 0, \tag{A.2.4}$$

and likewise, by means of (A.1.2) and for $x \in C(f)$,

$$n^{-1}b_n^d \sum_{i=1}^{n} (\mathscr{E}K_i)^2 = \sum_{i=1}^{n} \frac{1}{n}\left(\frac{b_n}{b_i}\right)^2 \int K^2(u) f(x - b_i u)\, du \to \theta_d f(x) \int K^2(u)\, du. \tag{A.2.5}$$

Hence (A.2.4) and (A.2.5) imply

$$n^{-1}b_n^d \sum_{i=1}^{n} \operatorname{var}(K_i) \to \theta_d f(x) \int K^2(u)\, du, \qquad x \in C(f). \tag{A.2.6}$$

Utilizing relations (A.1.3) and (A.1.4) in a similar fashion,

$$n^{-1}b_n^d \sum_{i=1}^{n} \operatorname{var}(Y_i K_i) \to \theta_d v(x) \int K^2(u)\, du, \qquad x \in C(w) \cap C(v). \tag{A.2.7}$$

Finally, relations (A.1.1) and (A.1.5), together with Assumption A.4(i)–(ii) and the Toeplitz lemma, yield as above

$$n^{-1}b_n^d \sum_{i=1}^{n} \operatorname{cov}(K_i, Y_i K_i) \to \theta_d w(x) \int K^2(u)\, du, \qquad x \in C(f) \cap C(w). \tag{A.2.8}$$

Relations (A.2.6)–(A.2.8) taken together imply then that, for $x \in C(f) \cap C(w) \cap C(v)$,

$$n^{-1}b_n^d \sum_{i=1}^{n} \operatorname{var}(c_1 K_i + d_i Y_i K_i) \to \left[c_1^2 f(x) + 2c_1 d_1 w(x) + d_1^2 v(x)\right]\theta_d \int K^2(u)\, du.$$

Since the right-hand side above is $\tau_{cd}^2(x)$, by (3.2), relation (A.2.2) has been established.

As for relation (A.2.3), it is an immediate consequence of Lemmas A.1.1–A.1.3.

A.3. Proof of relation (4.1). Clearly,

$$\operatorname{var}\left(n^{-1/2}T_n''\right) = d_1 b_n^d n^{-1} \operatorname{var} \sum_{i=1}^{n} (Y_i'' K_i) + 2d_1^2 b_n^d n^{-1} \sum_{1 \le i < j \le n} \operatorname{cov}(Y_i'' K_i, Y_j'' K_j). \tag{A.3.1}$$

Consider first the covariance terms and let S_1 and S_2 be as in Section A.1.

The individual covariances are bounded as follows on S_1:

$$\left|\operatorname{cov}(Y_i''K_i, Y_j''K_j)\right| \le \int\int\int\int |vz| K(u)K(w) \times \left| f_{X_i,Y_i,X_j,Y_j}(x - b_i u, v, x - b_j w, z) - f_{X_i,Y_i}(x - b_i u, v)\, f_{X_j,Y_j}(x - b_j w, z)\right| du\, dv\, dw\, dz.$$

From this point on, the proof is the same as that in Lemma A.1.3, regarding the set S_1. As for the set S_2,

$$\left|\operatorname{cov}(Y_i''K_i, Y_j''K_j)\right| \le 10\alpha^{1-\gamma}(j-i)\left(\mathscr{E}|Y_i K_i|^{2/\gamma}\right)^{\gamma/2}\left(\mathscr{E}|Y_j K_j|^{2/\gamma}\right)^{\gamma/2},$$

and the proof is then the same as that of the corresponding part in Lemma A.1.3. Thus

$$\text{(A.3.2)} \qquad b_n^d n^{-1} \sum_{1 \le i < j \le n} \operatorname{cov}(Y_i''K_i, Y_j''K_j) \to 0 \quad \text{as } n \to \infty.$$

As for the variance terms, we have

$$\operatorname{var}(Y_i''K_i) \le b_i^{-d} \int K^2(u) \bar{v}_L(x - b_i u)\, du,$$

where $\bar{v}_L(x) = \int_{(|t| > L)} t^2 f_{X_1,Y_1}(x, t)\, dt$, so that

$$\text{(A.3.3)} \qquad \bar{v}_L(x) \to 0 \quad \text{as } L \to \infty.$$

Therefore, $\operatorname{var}(Y_i''K_i) \le b_i^{-d} x_{iL}$, where

$$x_{iL} = \int K^2(u) \bar{v}_L(x - b_i u)\, du \to \bar{v}_L \int K^2(u)\, du \quad \text{as } i \to \infty \text{ for fixed } L.$$

It follows that

$$\text{(A.3.4)} \qquad b_n^d n^{-1} \sum_{i=1}^{n} \operatorname{var}(Y_i''K_i) \le \sum_{i=1}^{n} \frac{1}{n}\left(\frac{b_i}{b_n}\right)^d x_{iL}.$$

In (A.3.4), for each fixed L, take the limit as $n \to \infty$ to obtain

$$\text{(A.3.5)} \qquad \limsup_n b_n^d n^{-1} \sum_{i=1}^{n} \operatorname{var}(Y_i''K_i) \le \theta_d \bar{v}_L(x) \int K^2(u)\, du.$$

Next, take the limit in (A.3.5) as $L \to \infty$ and use relation (A.3.3) to get

$$\text{(A.3.6)} \qquad b_n^d n^{-1} \sum_{i=1}^{n} \operatorname{var}(Y_i''K_i) \to 0.$$

Relations (A.3.2) and (A.3.6) establish (4.1) by way of (A.3.1). □

A.4. Proof of relation (4.4). From the definition of the quantities $\tau_{cd}^2(x)$ and $\tau_{cdL}^2(x)$ through relations (3.2) and (4.3), respectively, all we have to show is that, as $L \to \infty$, $w_L(x) \to w(x)$, $v_L(x) \to v(x)$. However, this is true because of the dominated convergence theorem and the assumption that $h(x) = \int |y|^{2+\delta} f_{X_1, Y_1}(x, y)\, dy < \infty$ $(\delta > 0)$ [see Assumption A.2(ii)]. □

A.5. Proof of relation (5.4). For simplicity, write y_{mn}, y_{mn1} and y_{mn2} instead of $y_{mn}(\mathbf{x})$, $y_{mn}(x_1)$ and $y_{mn}(x_2)$, respectively. Then

$$n^{-1} \sum_{m=1}^{r} \operatorname{var}(y_{mn}) = n^{-1} \sum_{m=1}^{r} \operatorname{var}(y_{mn1}) + n^{-1} \sum_{m=1}^{r} \operatorname{var}(y_{mn2}) + 2n^{-1} \sum_{m=1}^{r} \operatorname{cov}(y_{mn1}, y_{mn2}),$$

and the sum of the first two terms on the right-hand side above converges to $\tau_{\mathbf{cd}}^2(\mathbf{x})$ by (3.15). As for the third term on the same side, we have by a simple computation,

$$n^{-1} \sum_{m=1}^{r} \operatorname{cov}(y_{mn1}, y_{mn2}) = n^{-1} \sum_{m=1}^{r} \sum_{i=k_m}^{k_m+p-1} \operatorname{cov}(Z_{i1}, Z_{i2}) + n^{-1} \sum_{m=1}^{r} \sum_{i \neq j} \operatorname{cov}(Z_{i1}, Z_{j2}),$$

and the second term on the right-hand side above converges to 0 as was seen in (A.1.25). Regarding the first term on the same side, it is easily seen that it also converges to 0, because of the following result.

LEMMA A.5.1. *Let K satisfy Assumption* A.3(i)–(ii) *and let* $0 < b_n \to 0$. *For* $g\colon \mathbb{R}^d \to \mathbb{R}$ *integrable, set*

$$g_n(x, y) = b_n^{-d} \int K\left(\frac{x - t}{b_n}\right) K\left(\frac{y - t}{b_n}\right) g(t)\, dt, \qquad x, y \in \mathbb{R}^d.$$

Then $g_n(x, y) \to 0$, *provided* $x \neq y$ *and* $x \in C(g)$.

PROOF. This lemma is the d-dimensional version of Lemma 2 in Masry (1986). For the details, the interested reader is referred to Lemma 9.1 in Roussas and Tran (1989).

The justification of relation (5.4) is then completed and therefore the proof of Theorem 2.3 is concluded. □

Acknowledgments. The authors are indebted to a number of referees and an Associate Editor whose thoughtful and constructive comments helped improve an earlier version of this paper.

REFERENCES

AHMAD, I. A. and LIN, P. (1976). Nonparametric sequential estimation of a multiple regression function. *Bull. Math. Statist.* **17** 63–75.

COLLOMB, G. (1984). Propriétés de convergence presque complete du predicteur a noyau. *Z. Wahrsch. Verw. Gebiete* **66** 441–460.

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

DEO, C. M. (1973). A note on empirical processes of strong mixing sequences. *Ann. Probab.* **1** 870–875.

DEVROYE, L. and WAGNER, T. J. (1980). On the L_1 convergence of kernel estimators of regression function with applications in discrimination. *Z. Wahrsch. Verw. Gebiete* **51** 15–25.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

GREBLICKI, W. and PAWLAK, M. (1987). Necessary and sufficient conditions for a recursive kernel regression estimate. *J. Multivariate Anal.* **23** 67–76.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

KRZYŻAK, A. and PAWLAK, M. (1984). Almost everywhere convergence of recursive function estimate and classification. *IEEE Trans. Inform. Theory* **30** 91–93.

LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Toronto.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415.

MASRY, E. (1986). Recursive probability density estimation for weakly dependent processes. *IEEE Trans. Inform. Theory* **32** 254–267.

MASRY, E. (1987). Almost sure convergence of recursive density estimation for weakly dependent processes. *Statist. Probab. Lett.* **5** 249–254.

MASRY, E. and GYÖRFI, L. (1987). Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal.* **22** 79–93.

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

NADARAYA, E. A. (1970). Remarks on nonparametric estimates for density function and regression curves. *Theory Probab. Appl.* **15** 134–137.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

PHAM, D. T. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*, 1st ed. Wiley, New York.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–297.

ROBINSON, P. M. (1986). On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression and density estimators. *Ann. Inst. Statist. Math.* **38** 539–549.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

ROSENBLATT, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II* (P. R. Krishnaiah, ed.) 25–31. Academic, New York.

ROUSSAS, G. G. (1989). Asymptotic normality of the kernel estimate under dependence conditions: Application to hazard rate. *J. Statist. Plann. Inference* **25** 81–104.

ROUSSAS, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Process. Appl.* **36** 107–116.

ROUSSAS, G. G. and IOANNIDES, D. (1987). Moment inequalities for mixing sequences of random variables. *Stochastic Anal. Appl.* **5** 61–120.

ROUSSAS, G. G. and TRAN, L. T. (1989). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. Technical Report 187, Div. Statistics, Univ. California, Davis.

ROUSSAS, G. G. and TRAN, L. T. (1990). Joint asymptotic normality of kernel estimates under dependence conditions, with application to hazard rate. Technical Report 189, Div. Statistics, Univ. California, Davis.

SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at finite number of distinct points. *Ann. Math. Statist.* **43** 84–89.

STEIN, E. M. (1970). *Singular Integrals and Differentiability of Functions*. Princeton Univ. Press.

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

TRAN, L. T. (1990). Recursive density estimation under a weak dependence condition. *Ann. Inst. Statist. Math.* **43** 305–329.

VOLKONSKII, V. A. and ROZANOV, YU. A. (1959). Some limit theorems for random functions. I. *Theory Probab. Appl.* **4** 178–197.

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

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405