The Annals of Statistics
1996, Vol. 24, No. 3, 975–991

FIXED-DESIGN REGRESSION FOR LINEAR TIME SERIES

BY LANH TRAN,[1] GEORGE ROUSSAS, SIDNEY YAKOWITZ[2]
AND B. TRUONG VAN

*Indiana University, University of California, Davis,
University of Arizona and Université de Pau*

This investigation is concerned with recovering a regression function $g(x_i)$ on the basis of noisy observations taken at uniformly spaced design points x_i. It is presumed that the corresponding observations are corrupted by additive dependent noise, and that the noise is, in fact, induced by a general linear process in which the summand law can be discrete, as well as continuously distributed. Discreteness induces a complication because such noise is not known to be strong mixing, the postulate by which regression estimates are often shown to be asymptotically normal. In fact, as cited, there are processes of this character which have been proven not to be strong mixing. The main analytic result of this study is that, in general circumstances which include the non-strong mixing example, the smoothers we propose are asymptotically normal. Some motivation is offered, and a simple illustrative example calculation concludes this investigation. The innovative elements of this work, mainly, consist of encompassing models with discrete noise, important in practical applications, and in dispensing with mixing assumptions. The ensuing mathematical difficulties are overcome by sharpening standard arguments.

1. Introduction. This paper is concerned with a fixed-design regression problem in which the design points $x_{n0}, \dots, x_{nn}$ and the responses $Y_{n0}, \dots, Y_{nn}$ are related as follows:

$$Y_{ni} = g(x_{ni}) + \varepsilon_{ni}, \qquad 0 \le i \le n. \tag{1.1}$$

Here ε_{ni}, $0 \le i \le n$, are the random disturbances and g is a bounded real-valued function defined on a compact subset A of the real line R. Assume that for each n, $\{\varepsilon_{ni},\ 0 \le i \le n\}$ have the same distribution as $\xi_0, \dots, \xi_n$, where ξ_t, $t = 0, \pm 1, \pm 2, \dots,$ is a general weakly stationary linear process,

$$\xi_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}. \tag{1.2}$$

Here $\{Z_t\}$ is a martingale difference sequence relative to an increasing sequence of σ-fields $\mathscr{F}_t$ such that $E_{t-1}(Z_t^2) = \sigma^2$ almost surely (a.s.) for some

---

Received August 1992; revised July 1995.
[1]Research partially supported by NSF Grant DMS-94-03718.
[2]Research partially supported by NSF Grant INT-92-01430 and NIH Grant RO1 AI37535-02.
AMS 1991 *subject classifications*. Primary 62G05; secondary 60J25, 62J02, 62M05, 62M09, 62H12.
Key words and phrases. Fixed design, martingale difference, linear time series.

$\sigma > 0$, where E_t denotes conditional expectation with respect to $\mathscr{F}_t$ (and similarly for variances). It is assumed that

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{1.3}$$

(and thus $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$). For background material on martingale differences, the reader is referred to the books of Hall and Heyde (1980) and Stout (1974). In the stronger case that the Z_t's are independent, identically distributed (i.i.d.) random variables (r.v.'s), the latter condition is equivalent [see, e.g., Rosenblatt (1985), pages 39 and 46, Priestley (1981), page 143, or Bartlett (1955), page 157] to the assumption that the time series is well defined, is stationary and possesses a spectral density. It thus encompasses stationary autoregressive moving-average (ARMA) models and many other time series of interest in the literature. The texts by Brockwell and Davis (1991), Priestley (1981) and Box and Jenkins (1970) offer fine background material regarding such linear time series.

The problem we face here is that of estimating the function g on the basis of sample pairs (x_{ni}, Y_{ni}), $0 \le i \le n$. The estimate is to be a general linear smoother of the form

$$g_n(x) = \sum_{i=1}^{n} w_{ni}(x) Y_{ni}, \tag{1.4}$$

where the weight functions $w_{ni}(x) = w_{ni}(x; \tilde{x}_n)$, $i = 1, \dots, n$, depend on $x \in A$, on the fixed design points $\tilde{x}_n = (x_{n0}, \dots, x_{nn})$ and on the number of observations n.

In the case that the ξ_t's satisfy the strong mixing (or α-mixing) condition, a suitably normalized version of $g_n(x)$ has been shown to be asymptotically normal by Roussas, Tran and Ioannides (1992). [See also Boente and Fraiman (1990).] Conditions for general linear processes to be strong mixing may be found, for example, in Athreya and Pantula (1986), Gorodetskii (1977), Pham and Tran (1985) and Withers (1981a, b). However, all works known to us require that the Z_t variables be continuously distributed or at least [Athreya and Pantula (1986), Theorem 2] satisfy the property that a finite segment of the distribution of the sum (1.2) has a range on which it is absolutely continuous. Yet a great number of processes ξ_t and even Y_t are naturally quantized or alternately quantized by rounding of the observations. Any tabular time records, such as those in Part 5 of Box and Jenkins (1970), have already been rounded or truncated. Some of these, such as closing stock prices, are, by their nature, discrete. Asymptotic normality of nonparametric estimates (1.4) for processes with disturbances as the ones discussed here has not hitherto been established because such noises are not known to be strong mixing. Arguments in Athreya and Pantula (1986) would seem to imply that whenever the Z_t process is discrete, strong mixing is in doubt. The central advance of the present work is that nevertheless, even in nonmixing cases, under conditions to be stated, smoothers (1.4), suitably normalized, are now confirmed to be asymptotically normal.

The basic fixed-design stochastic sequence model (1.1) and (1.2) is not new. In the case of independent noise, it goes back to Priestley and Chao (1972), Gasser and Müller (1979) and others. In the case of dependent samples, Chu and Marron (1991) have considered a similar model, theirs being only slightly less general in that they postulated the disturbances to be ARMA; they were concerned with bandwidth estimation, not convergence of the regression estimate. Härdle and Tuan (1986) proved asymptotic normality of M-smoothers. Truong (1991) obtained optimal rates of convergence for kernel estimators based on local averages. Härdle and Tuan (1986) and Truong (1991) studied the case when the noise process is a general linear process. Burman (1991) investigated the problem of estimating a regression function with nonrandom design points and dependent errors. In the random-design case, Truong and Stone (1992) and Tran (1993) obtained optimal rates of convergence of local average estimators under α-mixing conditions. Roussas and Tran (1992) have established asymptotic normality of recursive regression estimators under related conditions.

The last mentioned work offers motivation for this type of model in an economic context of an enterprise trying to infer the profit $g(x)$ at level x of some controllable variable, such as price, fat content of hamburgers and so on. Imagine that the level of this variable is changed on a periodic basis. The profit, being influenced by other factors such as weather, unemployment and so forth, one can anticipate that it has temporal dependence. In this context, the fixed-design model here makes even more sense to us, because it presumes the decision-maker is varying his or her control, x, in a systematic, rather than random, manner. In connection with applicability of this line of inquiry, there are a number of works bringing nonparametric notions to bear on generalized linear processes by investigators affiliated with economics departments [e.g., Andrews (1984), Robinson (1983, 1987) and White and Domowitz (1984)], although these studies do not tie the model to an application. Andrews (1984) is intrigued by AR processes which are not strongly mixing, and White and Domowitz (1984) are concerned with conditions implying strong mixing and asymptotic normality. Pham (1986) investigates mixing properties in a setting of bilinear and generalized random coefficient AR models. Hesse (1990) shows that, under conditions general enough to include the same example of an AR sequence which is not strong mixing as in our Section 4, the pth quantile admits a Bahadur-type representation. In brief, this AR example has attracted enough attention that demonstration of asymptotic normality in regression under such disturbances may compound the interest already expressed. Finally, Stoyanov and Robinson (1991) concern themselves with semiparametric and nonparametric inference in the framework of continuous-time stochastic processes.

Another conceivable setting in which our estimation model might be appropriate is that of a ship taking sonar or plumb readings of depth, as it travels in a fixed direction, or a traveling aircraft taking geodesic readings by radar. Then the ideal readings might be disturbed by position-dependent error influenced by currents, winds, temperature, moisture in the air and so forth. Digitizing of the observations would put us into the framework of (1.1) and (1.2), if the disturbances could be viewed as arising from linear time series.

The main result is contained in Theorem 2.1. The assumptions involved are relatively simple and not substantially more complicated than those used even in the i.i.d. case [see, e.g., Georgiev (1988)].

This paper is organized as follows: the basic notation, assumptions and central result are stated in Section 2. The proof is deferred to Section 3, and the concluding section offers an example of a function [taken from Chu and Marron (1991)] contaminated by a linear noise process which is not strong mixing. Computational evidence suggests that nevertheless the linear smoother provides an adequate approximation consistent with Theorem 2.1 of this study. The technique used here is reminiscent of Gordin's [see Gordin (1969) and Hall and Heyde (1980), Theorem 5.1]. However, our method of proof is different; we show that the statistics involved are martingale arrays instead of approximating them by martingales. The letter C will be used to denote constants whose values may vary and are unimportant. All limits are taken as $n \to \infty$ unless indicated otherwise.

2. Assumptions and statements of main results. For simplicity, we will consider the estimate $g_n(x)$ with $x \in A$, where A is a compact set which we will take to be $[0, 1]$ without loss of generality. The weight functions are of the form

$$w_{ni}(x) = ((x_{ni} - x_{n,i-1})/h_n)K((x - x_{ni})/h_n), \tag{2.1}$$

where x_{ni}, $i = 0, 1, \dots, n+1$, are given design points, K is a kernel function and $\{h_n\}$ is a sequence of bandwidths tending to 0. We will write h_n as h for brevity. Denote

$$s_n^2(x) = \mathrm{Var}[g_n(x)].$$

Omitting the argument x for simplicity, we have

$$\begin{aligned} s_n^2 &= \mathrm{Var}\left(\sum_{i=1}^n w_{ni}Y_{ni}\right) = \mathrm{Var}\left(\sum_{i=1}^n w_{ni}\varepsilon_{ni}\right) = \mathrm{Var}\left(\sum_{i=1}^n w_{ni}\xi_i\right) \\ &= \mathrm{Var}(\xi_1)\sum_{i=1}^n w_{ni}^2 + 2\sum_{1\le i<j\le n} w_{ni}w_{nj}\,\mathrm{Cov}\{\xi_i, \xi_j\}, \end{aligned} \tag{2.2}$$

where

$$\mathrm{Cov}\{\xi_i, \xi_j\} = \mathrm{Cov}\{\xi_i, \xi_{i-(i-j)}\} = \sigma^2 \sum_{u=-\infty}^{\infty} \psi_u \psi_{u-i+j}.$$

Let $m = m(n) \uparrow \infty$ be a nondecreasing integer-valued function and define

$$\xi_t^m = \sum_{j=-m}^{m} \psi_j Z_{t-j}, \tag{2.4}$$

$$(s_n^m)^2 = \operatorname{Var}\left(\sum_{i=1}^{n} w_{ni} \xi_i^m \right). \tag{2.5}$$

ASSUMPTION 1. Suppose:

(i) $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$;
(ii) $E_{t-1} Z_t^2 = \sigma^2$ for every t and $\sup_{-\infty < t < \infty} E_{t-1}(|Z_t|^{2+\delta}) = M$ a.s. for some $\delta > 0$ and $M < \infty$.

ASSUMPTION 2. Suppose $C_1/n \le |x_{ni} - x_{n,i-1}| \le C_2/n$ for all $1 \le i \le n$, $1 \le n < \infty$.

ASSUMPTION 3. The kernel function K is nonnegative, bounded and continuous almost everywhere on R and has a majorant; that is, $K(x) \le H(x)$, all $x \in R$, where H is symmetric, bounded, nonincreasing on $[0, \infty)$ with $\int H(y)\,dy < \infty$, where the integral is over R.

ASSUMPTION 4. Suppose:

(i) $nh \to \infty$;
(ii) $\liminf nhs_n^2 \ge C$ for some positive constant C.

We now proceed with the formulation of the main result of the paper.

THEOREM 2.1. *For each $x \in A$, a compact subset of R, under Assumptions* 1–4,

$$\frac{g_n(x) - Eg_n(x)}{s_n(x)} \to_d N(0,1).$$

3. Some auxiliary results. We will implicitly assume that Assumptions 1 to 4 are satisfied unless otherwise stated.

LEMMA 3.1. *Suppose $h \to 0$ and $nh \to \infty$ and Assumptions* 2 *and* 3 *are satisfied. Then*

$$\sum_{i=1}^{n} [(x_{ni} - x_{n,i-1})/h] K((x - x_{ni})/h) \to \int K(y)\,dy, \tag{3.1}$$

$$\sum_{i=1}^{n} [(x_{ni} - x_{n,i-1})/h] K^2((x - x_{ni})/h) \to \int K^2(y)\,dy > 0. \tag{3.2}$$

This is the lemma on pages 5 and 6 in Georgiev and Greblicki (1986).

LEMMA 3.2. *There exist two positive constants C_1^* and C_2^* such that*

$$C_1^* \le nh \sum_{i=1}^{n} w_{ni}^2 \le C_2^*$$

for all n.

PROOF. By (3.1),

$$\sum_{i=1}^{n} w_{ni} \to \int K(y)\,dy. \tag{3.3}$$

Also,

$$w_{ni} \le C_2 (nh)^{-1} K((x - x_{ni})/h) \le C(nh)^{-1},$$

so that

$$\max_{1 \le i \le n} w_{ni} \le C(nh)^{-1}. \tag{3.4}$$

Then

$$\sum_{n=1}^{\infty} w_{ni}^2 \le \max_{1 \le i \le n} w_{ni} \sum_{i=1}^{n} w_{ni} \le C \max_{1 \le i \le n} w_{ni} \le C(nh)^{-1}. \tag{3.5}$$

Also,

$$\begin{aligned} \sum_{i=1}^{n} w_{ni}^2 &= \sum_{i=1}^{n} \left[\frac{x_{ni} - x_{n,i-1}}{h} K\left(\frac{x - x_{ni}}{h} \right) \right]^2 \\ &\ge \frac{C}{nh} \sum_{i=1}^{n} \frac{x_{ni} - x_{n,i-1}}{h} K^2 \left(\frac{x - x_{ni}}{h} \right) \ge \frac{C}{nh}, \end{aligned} \tag{3.6}$$

where the last inequality is obtained by using Lemma 3.1. Relations (3.6) and (3.5) complete the proof. □

LEMMA 3.3. *It holds that*

$$\lim nh\left(s_n^2 - (s_n^m)^2\right) = 0.$$

PROOF. Define $C(k)$ and $C_m(k)$ as follows: $C(k) = \operatorname{Cov}\{\xi_i, \xi_{i+k}\}$ and $C_m(k) = \operatorname{Cov}\{\xi_i^m, \xi_{i+k}^m\}$. Then

$$s_n^2 = \sum_{|k| \le n-1} C(k) \sum_{i=1}^{n-|k|} w_{ni} w_{n,i+k}. \tag{3.7}$$

Equation (3.7) still holds if s_n^2 is replaced by $(s_n^m)^2$ as defined in (2.5) and $C(k)$ is replaced by $C_m(K)$. By (3.3) and (3.4),

$$\begin{aligned} s_n^2 - (s_n^m)^2 &\le \left(\max_{1 \le j \le n} w_{nj} \right) \left(\sum_{i=1}^{n} w_{ni} \right) \sum_{|k| \le n-1} |C(k) - C_m(k)| \\ &\le C(nh)^{-1} \sum_{|k| \le n-1} |\Delta_m(k)|, \end{aligned}$$

where $\Delta_m(k) = C(k) - C_m(k)$. The proof will be completed by showing that $\sum_{|k| \le n-1} |\Delta_m(k)| \to 0$. Clearly,

$$C(k) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|k|},$$

$$C_m(k) = \begin{cases} \sigma^2 \sum_{j=-m}^{m-|k|} \psi_j \psi_{j+|k|}, & \text{for } |k| \le 2m, \\ 0, & \text{for } |k| > 2m. \end{cases}$$

Thus

$$\Delta_m(k) = \begin{cases} \sigma^2 \left(\sum_{j<-m} \psi_j \psi_{j+|k|} + \sum_{j>m} \psi_j \psi_{j-|k|} \right), & \text{for } |k| \le 2m, \\ C(k), & \text{for } |k| > 2m. \end{cases}$$

Hence

$$\sum_{|k| \le n-1} |\Delta_m(k)| \le \sigma^2 \left(\sum_{|j|>m} |\psi_j| \right) \sum_{j=-\infty}^{\infty} |\psi_j| + \sum_{|k|>2m} |C(k)|. \tag{3.8}$$

The right-hand side of (3.8) tends to 0 since $\sum_{k=-\infty}^{\infty} |C(k)| < \infty$, by Assumption 1. □

LEMMA 3.4. *It holds that* $(s_n / s_n^m) \to 1$.

PROOF. By Lemma 3.3,

$$\frac{nh\left(s_n^2 - (s_n^m)^2\right)}{nhs_n^2} \to 0$$

since nhs_n^2 is bounded by Assumption 4(ii). Thus $(s_n^2 - (s_n^m)^2)/s_n^2 \to 0$, from which the lemma follows. □

LEMMA 3.5. *It holds that*

$$(s_n^m)^{-1} \max_{1 \le i \le n} w_{ni} \to 0.$$

PROOF. By Lemma 3.4, to establish the proof, it is sufficient to show that $s_n^{-2}(\max_{1 \le i \le n} w_{ni})^2 \to 0$. This follows since $(s_n^2)^{-1} \max_{1 \le i \le n} w_{ni}$ stays bounded, as $n \to \infty$, by Assumption 4(ii), and $\max_{1 \le i \le n} w_{ni} \to 0$ by (3.4). □

For $n > 2m$, define c_{ni}^m by

$$c_{ni}^m = \begin{cases} \sum_{j=1}^{m+i} \alpha_{nj}^m \psi_{j-i}, & \text{for } 1 - m \le i \le m, \\ \sum_{j=-m}^{m} \alpha_{n,i+j}^m \psi_j, & \text{for } m + 1 \le i \le n - m, \\ \sum_{j=0}^{n-i+m} \alpha_{n,n-j}^m \psi_{n-i-j}, & \text{for } n - m + 1 \le i \le n + m, \end{cases}$$

where $\alpha_{ni}^m = (s_n^m)^{-1} w_{ni}$.

LEMMA 3.6. *It holds that*

$$\max_{1-m \le i \le n+m} |c_{ni}^m| \to 0.$$

PROOF. Using the definition of c_{ni}^m,

$$\max_{1-m \le i \le n+m} |c_{ni}^m| \le (s_n^m)^{-1} \Big(\max_{1 \le i \le n} w_{ni} \Big) \sum_{i=-\infty}^{\infty} |\psi_i| \to 0,$$

by Assumption 1(i) and Lemma 3.5. □

Define

$$\alpha_{ni} = (s_n)^{-1} w_{ni}, \qquad T_n^m = \sum_{i=1}^{n} \alpha_{ni}^m \xi_i^m, \qquad T_n = \sum_{i=1}^{n} \alpha_{ni} \xi_i. \tag{3.9}$$

Thus

$$\operatorname{Var}(T_n^m) = 1. \tag{3.10}$$

Also, define A_{nk} by

$$A_{nk} = \sum_{i=1-m}^{k-m} c_{ni}^m Z_i, \qquad 1 \le k \le n + 2m. \tag{3.11}$$

LEMMA 3.7. *For all $n > 2m$,*

$$A_{n,n+2m} = T_n^m.$$

PROOF. From the definition of ξ_t^m in (2.4),

$$T_n^m = \sum_{k=1}^{n} \sum_{j=-m}^{m} \alpha_{nk}^m \psi_j Z_{k-j}.$$

Set $i = k - j$. Then $1 - m \le i \le n + m$ since $1 \le k \le n$ and $-m \le j \le m$. Also, $1 - i \le j \le n - i$, whereas $-m \le j \le m$. Thus $\max\{1 - i, -m\} \le j \le \min\{m, n - i\}$. Consequently,

$$T_n^m = \sum_{i=1-m}^{n+m} g(i, m, n) Z_i,$$

where

$$g(i, m, n) \equiv \sum_{j=\max\{1-i,\,-m\}}^{\min\{m,\,n-i\}} \alpha_{n,\,i+j}^{m} \psi_j.$$

If $1 - m \le i \le m$, then $\max\{1 - i, -m\} = 1 - i$ and $\min\{m, n - i\} = m$ since $n > 2m$. In this case,

$$g(i, n, m) = \sum_{j=1-i}^{m} \alpha_{n,\,i+j}^{m} \psi_j = c_{ni}^{m}.$$

If $m + 1 \le i \le n - m$, then $\max\{1 - i, -m\} = -m$ and $\min\{m, n - i\} = m$ because $n > 2m$. Therefore,

$$g(i, n, m) = \sum_{j=-m}^{m} \alpha_{n,\,i+j}^{m} \psi_j = c_{ni}^{m}.$$

If $n - m + 1 \le i \le n + m$, then $\max\{1 - i, -m\} = -m$ and $\min\{m, n - i\} = n - i$ and then

$$g(i, n, m) = \sum_{j=-m}^{n-i} \alpha_{n,\,i+j}^{n-i} \psi_j = c_{ni}^{m},$$

again by using the fact that $n > 2m$. □

The following lemma follows from Theorem 2.5 in Helland (1982).

LEMMA 3.8. *Let* $\{\zeta_{nk};\ k = 1, 2, \dots,\ n = 1, 2, \dots\}$ *be an array of random variables defined on a probability space* $(\Omega, \mathscr{F}, P)$ *and let* $\{\mathscr{F}_{nk};\ k = 0, 1, \dots, n = 1, 2, \dots\}$ *be an array of σ-fields such that* $\mathscr{F}_{n,\,k-1} \subset \mathscr{F}_{nk} \subset \mathscr{F}$ *for each n and* $k \ge 1$. *Suppose* ζ_{nk} *is a martingale difference array adapted to* $\{\mathscr{F}_{nk}\}$; *that is,* ζ_{nk} *is* $\mathscr{F}_{nk}$*-measurable and* $E_{k-1}\{\zeta_{nk}\} = 0$ *a.s. for* $k \ge 1$ *and* $n \ge 1$. *For an increasing sequence* $\{k_n\}$ *of integers, suppose that*

$$\sum_{k=1}^{k_n} E_{k-1}\left\{|\zeta_{nk}|^{2+\delta}\right\} \to_P 0, \tag{3.12}$$

$$\sum_{k=1}^{k_n} \mathrm{Var}_{k-1}\{\zeta_{nk}\} \to_P 1. \tag{3.13}$$

Then

$$\zeta_{n,1} + \cdots + \zeta_{n,k_n} \to_d N(0, 1).$$

LEMMA 3.9. *Let* A_{nk} *be given by* (3.11). *Then*

$$A_{n,\,n+2m} \to_d N(0, 1).$$

PROOF. First,

$$\sum_{k=1-m}^{n+m} E_{k-1}\{|c_{nk}^m Z_k|^{2+\delta}\}$$
$$\le \Big(\max_{1-m\le k\le n+2m} E_{k-1}|Z_k|^{2+\delta}\Big)\max_{1-m\le k\le n+2m}|c_{nk}^m|^\delta \sum_{k=1-m}^{n+m}(c_{nk}^m)^2.$$

The right-hand side tends to 0 by Assumption 1(i), Lemma 3.6 and by noticing that $\sum_{k=1-m}^{n+2m}(c_{nk}^m)^2 = \sigma^{-2}$ by (3.10). Thus (3.12) of Lemma 3.8 is satisfied with $\zeta_{nk} = c_{nk}^m Z_k$ and $k_n = n + 2m$.

Second, $\sum_{k=1}^{n+m} \mathrm{Var}_{k-1}\{c_{nk}^m Z_k\} = 1$ by (3.10) and Lemma 3.7 and, consequently, (3.13) is satisfied. It is easy to see that $\sum_{k-1-m}^{0} \mathrm{Var}_{k-1}\{c_{nk}^m Z_k\} \to 0$. The lemma follows by a simple argument using Lemma 3.8.

LEMMA 3.10. *It holds that*

$$E(T_n - T_n^m)^2 \to 0.$$

PROOF. Define q_{1n} and q_{2n} by

$$q_{1n} = 2\bigg(\frac{1}{s_n} - \frac{1}{s_n^m}\bigg)^2 E\bigg(\sum_{i=1}^n w_{ni}\xi_i\bigg)^2,$$
$$q_{2n} = \frac{2}{(s_n^m)^2} E\bigg(\sum_{i=1}^n w_{ni}(\xi_i - \xi_i^m)\bigg)^2.$$

Then, by the C_r-inequality,

$$E(T_n - T_n^m)^2 \le q_{1n} + q_{2n}.$$

By Lemma 3.2, Assumption 1 and (3.4),

$$E\bigg(\sum_{i=1}^n w_{ni}\xi_i\bigg)^2 \le C\bigg(\sum_{i=1}^n w_{ni}^2\bigg) + 2\sigma^2\Big(\max_{1\le j\le n} w_{nj}\Big)\bigg(\sum_{-\infty}^{\infty}|\psi_u|\bigg)^2 \sum_{i=1}^n w_{ni} \le C(nh)^{-1}.$$

However, by Assumption 4,

$$q_{1n} \le \frac{C}{nh}\bigg(\frac{1}{s_n} - \frac{1}{s_n^m}\bigg)^2 \le C\big(1 - (s_n/s_n^m)\big)^2 (nhs_n^2)^{-1},$$

which tends to 0 by Lemma 3.4. After an easy but tedious computation and by using Assumption 1,

$$q_{2n} \le C(s_n^m)^{-2}\bigg(\sum_{i=1}^n w_{ni}^2 + \max_{1\le j\le n} w_{nj}\sum_{i=1}^n w_{ni}\bigg)o(1), \tag{3.14}$$

which tends to 0 by Lemma 3.2 and relations (3.4) and (3.3). The presence of the $o(1)$ term in (3.14) is obtained by using Assumption 1. □

PROOF OF THEOREM 2.1. The proof follows from Lemmas 3.9 and 3.10. □

4. An application.

EXAMPLE 4.1. Let ξ_t be the first-order autoregressive process defined by

$$\xi_t = \tfrac{1}{2}\xi_{t-1} + Z_t, \tag{4.1}$$

where Z_t's are independent symmetric Bernoulli r.v.'s taking values 1 and -1. So that $EZ_t = 0$ and $\sigma^2 = 1$. It is well known that ξ_t is absolutely continuous and does not satisfy the strong mixing condition. This follows from Theorem 5.2 in Withers (1981b); also, see Andrews (1984). Rewrite (4.1) as a general linear process as follows:

$$\xi_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad \psi_j = \begin{cases} 0, & \text{for } j \le -1, \\ 2^{-j}, & \text{for } j \ge 0. \end{cases} \tag{4.2}$$

Take $A = [0,1]$, and, for illustrative purposes, we will simplify things by taking the kernel to be the $U[-1,1]$ p.d.f.; that is,

$$K(x) = \tfrac{1}{2} I_{[-1,1]}(x); \tag{4.3}$$

the weights as in (2.1) with the further restriction that

$$x_{ni} = \frac{i}{n},\ i = 0, 1, \dots, n, \quad \text{and} \quad h = n^{-\theta},\ 0 < \theta < 1. \tag{4.4}$$

Then

$$w_{ni}(x) = \frac{1}{n^{1-\theta}} K\left(\frac{x - (i/n)}{n^{-\theta}}\right) = \frac{1}{2n^{1-\theta}} I_{(|x-(i/n)| \le n^{-\theta})}(x), \tag{4.5}$$

and the resulting estimate is

$$g_n(x) = \frac{1}{2n^{1-\theta}} \sum_{i=1}^{n} I_{(|x-(i/n)| \le n^{-\theta})}(x) Y_{ni}.$$

For this, it will be shown that

$$\frac{g_n(x) - Eg_n(x)}{s_n} \to_d N(0,1), \qquad x \in (0,1).$$

This will be done by showing that Assumption 4(ii) holds, since it is clear that Assumptions 1, 2 and 4(i) hold.

LEMMA 4.1. *Suppose* $0 < \theta < 1$, $0 < x < 1$ *and* K *is the uniform* $[-1,1]$ *kernel, and define*

$$Q_n \equiv \sum_{1 \le i < j \le n} \frac{1}{n^{1-\theta}} K\left(\frac{x - (i/n)}{n^{-\theta}}\right) K\left(\frac{x - (j/n)}{n^{-\theta}}\right) 2^{-(j-1)}.$$

Then

$$\lim Q_n = \tfrac{1}{2}.$$

PROOF. We have

$$Q_n = \frac{1}{n^{1-\theta}} \sum_{i=1}^{n} K\left(\frac{x-(i/n)}{n^{-\theta}}\right) \sum_{j=i+1}^{n} K\left(\frac{x-(j/n)}{n^{-\theta}}\right) 2^{-(j-i)}$$
$$= \frac{1}{n^{1-\theta}} \sum_{S_1} \left(\frac{1}{2}\right) \sum_{S_2} \left(\frac{1}{2}\right) 2^{-(j-i)},$$

where

$$S_1 = \{i\colon nx - n^{1-\theta} \le i \le nx + n^{1-\theta} \text{ and } 1 \le i \le n-1\}$$
$$= \{i\colon \max\{nx - n^{1-\theta}, 1\} \le i \le \min\{nx + n^{1-\theta}, n-1\}\}$$
$$= \{i\colon nx - n^{1-\theta} \le i \le nx + n^{1-\theta}\} \text{ for sufficiently large } n,$$

and, for each $i \in S_1$,

$$S_2 = \{j\colon nx - n^{1-\theta} \le j \le nx + n^{1-\theta} \text{ and } i+1 \le j \le n\}$$
$$= \{j\colon \max\{nx - n^{1-\theta}, i+1\} \le j \le \min\{nx + n^{1-\theta}, n\}\}$$
$$= \{j\colon i+1 \le j \le nx + n^{1-\theta}\} \text{ for sufficiently large } n.$$

Therefore, by simple calculations,

$$Q_n = \frac{1}{4n^{1-\theta}} \sum_{nx - n^{1-\theta} \le i \le nx + n^{1-\theta}} 2^i \sum_{i+1 \le j \le nx + n^{1-\theta}} 2^{-j}$$
$$= \frac{1}{4n^{1-\theta}} \left(2n^{1-\theta} - 1 - \frac{1}{2^{2n^{1-\theta}}}\right) \to \frac{1}{2}. \qquad \square$$

LEMMA 4.2. *$\liminf nhs_n^2 \ge C$ for some positive constant C.*

PROOF. From the expression of the ψ_j's in (4.2),

$$\operatorname{Var}(\xi_1) = \sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} 4^{-j} = \frac{4}{3}, \tag{4.6}$$

and, with $1 \le i < j \le n$,

$$\operatorname{Cov}\{\xi_i, \xi_j\} = \sum_{u=-\infty}^{\infty} \psi_u \psi_{u-i+j} = \sum_{u=0}^{\infty} \psi_u \psi_{u+(j-i)} = \left(\frac{4}{3}\right)\left(\frac{1}{2^{j-i}}\right). \tag{4.7}$$

Furthermore, by (4.2), (4.5), (4.7) and the definition of Q_n in Lemma 4.1,

$$\begin{aligned} &\sum_{1 \le i < j \le n} w_{ni} w_{nj} \operatorname{Cov}\{\xi_i, \xi_j\} \\ &\quad = \sum_{1 \le i < j \le n} w_{ni} w_{nj} \left(\sum_{u=-\infty}^{\infty} \psi_u \psi_{u-i+j}\right) = \frac{4Q_n}{3n^{1-\theta}}. \end{aligned} \tag{4.8}$$

Also, by means of (4.5),

$$\sum_{i=1}^{n} w_{ni}^2 = \sum_{i=1}^{n} \frac{1}{4n^{2(1-\theta)}} I_{\left(\left|x-\frac{i}{n}\right| \le n^{-\theta}\right)}(x) \tag{4.9}$$

$$= \frac{1}{4n^{2(1-\theta)}} \sum_{S_1} 1 = \frac{2n^{1-\theta}+1}{4n^{2(1-\theta)}}.$$

Therefore, by way of (2.2), (2.3), (4.6), (4.8), (4.9) and Lemma 4.1,

$$s_n^2 = \left(\frac{4}{3}\right) \frac{2n^{1-\theta}+1}{4n^{2(1-\theta)}} + \frac{8Q_n}{3n^{1-\theta}},$$

so that

$$nhs_n^2 = n^{1-\theta} s_n^2 = \frac{2}{3} + \frac{1}{3n^{1-\theta}} + \frac{8Q_n}{3} \to \frac{2}{3} + \frac{4}{3} = 2,$$

which completes the proof of the lemma. □

This section concludes with some experimentation, the objective of which is to see if, with a moderate number of data, under the fixed-design model (1.1) and (1.2) and the postulates of this section, we can recover a regression function g with accuracy suggested by the normal approximation. The regression function, taken from Chu and Marron (1991), is

$$g(x) = 50x^3(1-x)^3, \tag{4.10}$$

the domain A of g is the unit interval and the (non-α-mixing) disturbances are as determined by (4.1) and (4.2), σ^2 being 1. The weights of our linear smoother are as given by (2.1) using the kernel (4.3) and the oft-selected value $\theta = 1/5$.

We need a numerical value for s_n to use Theorem 2.1 as a means to get a confidence interval for $g(x)$. From (2.2) and accounting for the kernel (4.3),

$$s_n^2 = \operatorname{Var}(\xi_1)\left(\sum_{i=1}^{n} w_{ni}\right) + 2 \sum_{1 \le i < j \le n} w_{ni} w_{nj} \operatorname{Cov}\{\xi_i, \xi_j\}$$

$$= \frac{1}{(2[n^{4/5}])^2}\left[\operatorname{Var}(\xi_1)\left(\sum_{i=-[n^{4/5}]+1}^{[n^{4/5}]} 1\right) + 2 \sum_{-[n^{4/5}]+1 \le i < j \le [n^{4/5}]} \operatorname{Cov}\{\xi_i, \xi_j\}\right],$$

where $[n^{4/5}]$ designates the integer part of the indicated number. Use that $\sigma = 1$ and for $j \ge 0$,

$$\operatorname{Cov}\{\xi_0, \xi_j\} = \sum_{t \ge 0} \psi_t \psi_{t+j} = \sum_{t \ge 0} 2^{-t} 2^{-t-j} = \frac{4}{3} \times 2^{-j}. \tag{4.11}$$

Then, at last, conclude from (4.11) that

$$s_n^2 = \frac{4}{(3 \cdot 2[n^{4/5}])} + \frac{2}{(3 \cdot 2[n^{4/5}])} = \frac{1}{[n^{4/5}]}.$$

This estimate is predicated on the assumption that x_{ni} is situated so that it is further than $|n^{4/5}|$ points from either endpoint of A.

Figure 1 shows the results of two independent simulations under the parameters described above for $n = 2000$. The curve in the center is the true regression (4.10), and it is bordered above and below by functions

$$g_{2000} \pm \delta, \quad \text{where } \delta = 2 s_{2000} \approx 9.56 \times 10^{-2}.$$

The factor 2 multiplies the theoretical standard deviation as a crude way to account for bias, which is considerable [because $g_{2000}(x)$ depends on points at

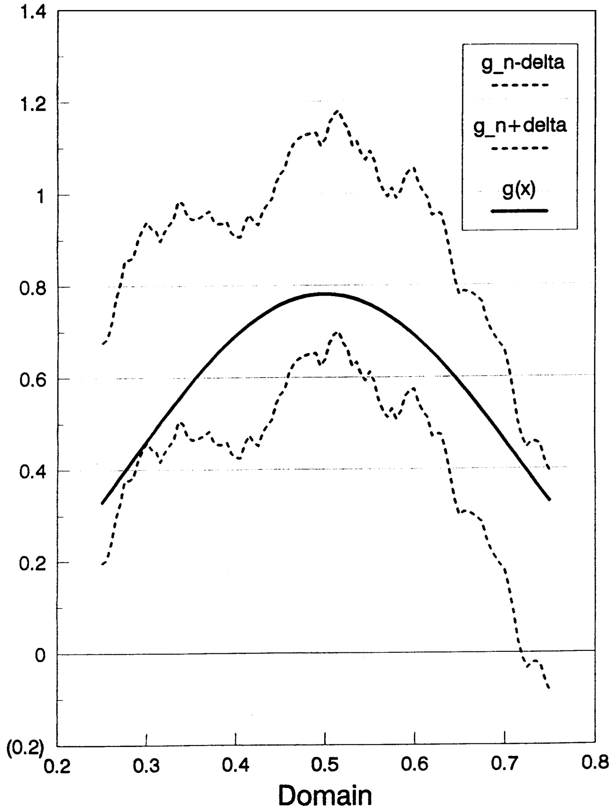

FIG. 1.

TABLE 1
Results of a computational experiment

Entry	at 0.25	at 0.5	at 0.75
	$n = 200$		
$g(x)$	0.3296	0.78125	0.3296
m_n	0.3596	0.7282	0.33255
$\hat{\sigma}_n$	0.2392	0.2381	0.2319
$2s_n$	0.288	0.288	0.288
	$n = 500$		
$g(x)$	0.3296	0.78125	0.3296
m_n	0.3185	0.7290	0.3646
$\hat{\sigma}_n$	0.1522	0.1697	0.1894
$2s_n$	0.1665	0.1665	0.1665
	$n = 1000$		
$g(x)$	0.3296	0.78125	0.3296
m_n	0.3301	0.7470	0.3460
$\hat{\sigma}_n$	0.1305	0.1222	0.1182
$2s_n$	0.1262	0.1262	0.1262

a distance of $0.22 \simeq 1/2000^{0.2}$ on either side of x]. In any event, the assumption that the square of the bias is the same order of magnitude as the standard deviation lies at the heart of the principle whereby a bandwidth rate of $\theta = 1/5$ is derived. In generating the data, the disturbance simulator is first "exercised" for 100 observations so that it approximates stationary behavior when its points are put to use.

In further experimentation, for various sample sizes n, 100 replications of this experiment were performed and their means and sample variances, for each fixed n, were computed at domain points $x = 0.25$, 0.5 and 0.75. The results of these calculations are related in Table 1, where they can be compared to their theoretical values. In Table 1, m_n and $\hat{\sigma}_n$ are the sample means and variances, respectively.

Acknowledgments. The authors would like to thank the Editor and the referees for many useful comments.

REFERENCES

ANDREWS, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934.

ATHREYA, K. B. and PANTULA, S. G. (1986). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23** 880–892.

BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes and Special References to Methods and Applications*, 1st ed. Cambridge Univ. Press, London.

BOENTE, G. and FRAIMAN, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from noisy processes. *Ann. Statist.* **18** 891–906.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco.

BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

BURMAN, P. (1991). Regression function estimation from dependent observations. *J. Multivariate Anal.* **36** 263–279.

CHU, C. K. and MARRON, S. (1991). Comparison of two bandwidth selectors with dependent errors. *Ann. Statist.* **19** 1906–1918.

GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression function. In *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **747** 23–68. Springer, Berlin.

GEORGIEV, A. A. (1988). Consistent nonparametric multiple regression. The fixed design case. *J. Multivariate Anal.* **25** 100–110.

GEORGIEV, A. A. and GREBLICKI, W. (1986). Nonparametric function recovering from noisy observations. *J. Statist. Plann. Inference* **13** 1–14.

GORDIN, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** 1174–1176.

GORODETSKII, V. V. (1977). On the strong mixing property for linear sequences. *Theory Probab. Appl.* **22** 411–413.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

HÄRDLE, W. and TUAN, D. P. (1986). Some theory on *M*-smoothing of time series. *J. Time Ser. Anal.* **7** 191–204.

HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–93.

HESSE, C. H. (1990). A Bahadur-type representation for empirical quantiles for a large class of stationary, possibly infinite-variance linear processes. *Ann. Statist.* **18** 188–202.

PHAM, D. T. (1986). The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Process. Appl.* **23** 291–300.

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

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, New York.

PRIESTLEY, M. B. and CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

ROBINSON, P. M. (1987). Time series residuals with application to probability density estimation. *J. Time Ser. Anal.* **8** 329–344.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston.

ROUSSAS, G. G. and TRAN, L. T. (1992). Asymptotic normality of the recursive kernel regression estimates under dependence conditions, and time series. *Ann. Statist.* **20** 98–120.

ROUSSAS, G. G., TRAN, L. T. and IOANNIDES, D. A. (1992). Fixed design regression for time series: asymptotic normality. *J. Multivariate Anal.* **40** 262–291.

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

STOYANOV, J. M. and ROBINSON, P. M. (1991). Semiparametric and nonparametric inference from irregular observations on continuous time stochastic processes. In *Nonparametric Functional Estimation and Related Topics* (G. Roussas, ed.). *NATO ASI Series* **335** 553–557. Kluwer, Dordrecht.

TRAN, L. T. (1993). Nonparametric function estimation for time series by local average estimators. *Ann. Statist.* **21** 1040–1057.

TRUONG, Y. K. (1991). Nonparametric curve estimation with time series errors. *J. Statist. Plann. Inference* **28** 167–183.

TRUONG, Y. K. and STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.

WHITE, H. and DOMOWITZ, I. (1984). Nonlinear regression with dependent observations. *Econometrica* **52** 143–161.

WITHERS, C. S. (1981a). Conditions for linear processes to be strong mixing. *Z. Wahrsch. Verw. Gebiete* **57** 479–480.

WITHERS, C. S. (1981b). Central limit theorems for dependent variables. I. *Z. Wahrsch. Verw. Gebiete* **57** 509–534.

LANH TRAN
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47408
E-MAIL: tran@indiana.edu

GEORGE ROUSSAS
DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616
E-MAIL: ggroussas@ucdavis.edu

SIDNEY YAKOWITZ
DEPARTMENT OF SYSTEMS AND
INDUSTRIAL ENGINEERING
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721

B. TRUONG VAN
DEPARTMENT OF MATHEMATICS
UNIVERSITÉ DE PAU
PAU 64000
FRANCE