Ann. Inst. Statist. Math.
38 (1986), Part A, 539-549

ON THE CONSISTENCY AND FINITE-SAMPLE PROPERTIES OF NONPARAMETRIC KERNEL TIME SERIES REGRESSION, AUTOREGRESSION AND DENSITY ESTIMATORS*

P. M. ROBINSON

(Received Nov. 12, 1984; revised May 20, 1985)

Summary

Kernel estimators of conditional expectations and joint probability densities are studied in the context of a vector-valued stationary time series. Weak consistency is established under minimal moment conditions and under a hierarchy of weak dependence and bandwidth conditions. Prompted by these conditions, some finite-sample theory explores the effect of serial dependence on variability of estimators, and its implications for choice of bandwidth.

1. Introduction

Let $\{X_t;\ t=1,2,\cdots\}$ be a strictly stationary real-valued vector stochastic process and M_a^b be the σ-field of events generated by X_t, $a\leqq t\leqq b$. There is frequently interest in measuring the effect of a stationary q-dimensional M_t^{t+a}-measurable vector Z_t, $a\geqq 0$, on a stationary M_{t+b}^{t+c}-measurable scalar Y_t, $0\leqq b\leqq c$. In particular when $E|Y_1|<\infty$ one wishes to estimate the regression function $m(z)=E(Y_1|Z_1=z)$ for values $z\in R^q$. In parametric time series analysis it is assumed that $m(z)$ is a given function of z and finitely many unknown parameters; moreover that some independence properties hold between the $U_t=Y_t-m(Z_t)$ and Z_s for $s\leqq t$ and all t, or between the full sequences $\{U_t\}$ and $\{Z_t\}$. Examples are time series regression models and distributed lag models with stochastic regressors, and autoregressions. Following [20], [21], we study a nonparametric estimator of $m(z)$ which does not rely on such assumptions and is particularly of value at the exploratory stage, especially when non-Gaussianity or non-linearity is suspected, where

* This research was supported by the ESRC.

Key words and phrases: Time series, density estimators, nonparametric regression and autoregression, mixing conditions, convergence in probability and mean square, finite-sample properties.

despite some recent advances in the theory of nonlinear time series models, considerable obstacles hinder a parametric approach.

We observe X_t for $t=1,\cdots,T$. Introduce a real, integrable, non-null, bounded function on R^q, denoted $k(z)$, and a q-dimensional positive definite matrix h_T of functions of T, which $\to 0$ as $T\to\infty$. We use the abbreviation $K_{tT}=k(h_T^{-1}Z_t)$. Because our results will concern pointwise convergence we choose Z_t so the point of interest is $z=0$, w.l.o.g. We estimate $m=m(0)$ by $m_T=g_T/f_T$, where

$$f_T=(T'H_T)^{-1}\sum_t{}' K_{tT}\ , \qquad g_T=(T''H_T)^{-1}\sum_t{}'' Y_tK_{tT}\ ,$$

and H_T is the determinant of h_T, $\sum'=\sum_1^{T'}$, $\sum''=\sum_1^{T''}$, $T'=T-a$, $T''=T-\max(a, c)$. Thus m_T is a kernel regression estimator of the Nadaraya-Watson type, apart from end-effects. We assume w.l.o.g.

<u>A1.1</u> $\int_{R^q} k(z)dz=1.$

Throughout we assume also

<u>A1.2</u> Z_1 admits a density, denoted $f(z)$.

Then f_T is the usual kernel estimator of $f=f(0)$.

In this paper we establish pointwise convergence in mean square of f_T to f, and pointwise convergence in probability of m_T of m, when the origin is a continuity point of $f(z)$ and $m(z)$ respectively, under minimal moment assumptions on Y_t and under various mixing conditions on X_t: strong mixing, ϕ-mixing and $*$-mixing. Such conditions can variously be checked for a number of processes of interest in time series analysis, including linear and nonlinear moving averages, linear autoregressions and autoregressive moving averages, Markov processes, and many nonlinear functionals of Gaussian processes, under suitable conditions on the autocorrelations of the Gaussian process; see eg, [13], [16].

Consistency results that may be compared to ours are in [8], [20], [21], but these papers assume X_t is Markovian and our more general weak dependence conditions reveal a possible trade-off between strength of serial dependence and desirable bandwidth size, supporting the view that selection of bandwidth should ideally be influenced by the presence of serial dependence, as well as by marginal features of the process and the size of T. Indeed in the case of a Gaussian process we examine also for *finite* T the relationship between its auto-covariances and those of K_{tT}, demonstrating that positive autocorrelation in the process tends to inflate the variance of a density estimator, and that negative autocorrelation may have the same effect. For a class of

Gaussian processes, and again for finite T, we show that variance can be reduced by increasing H_T. Rates of stochastic convergence, like those implicit in, eg, [6], [14], [15], [19], can be established under stronger conditions than ours, but we choose to impose (so far as convergence of m_T is concerned) moment conditions that are minimal, see A4.2, A5.2, and determine how weak a condition on mixing and bandwidth will then ensure consistency. For other relevant work see, eg, [1]–[5], [7], [9], [10], [18], [23], [24]. Our earlier paper [19] gave central limit results for similar estimators to those considered in the present paper, under the strong mixing condition and for scalar X_t.

2. Mixing conditions

Define functions α, ϕ, ψ on the integers such that for all $s \geqq 1$, $t \geqq 1$ and $A \in M_1^s$, $B \in M_{s+t}^\infty$, $|P(AB)-P(A)P(B)|$ is bounded by $\alpha(t)$, $\phi(t)P(A)$ and $\psi(t)P(A)P(B)$. We say X_t is strong mixing if $\alpha(t) \downarrow 0$, ϕ-mixing if $\phi(t) \downarrow 0$, $*$-mixing if $\psi(t) \downarrow 0$ (as $t \uparrow \infty$). Let U and V be respectively M_1^s- and M_{s+t}^∞-measurable random variables. We use the inequalities (see, eg, [17])

$$|\mathrm{Cov}\,(U, V)| \leqq 4(\mathrm{ess\,sup}\,|U|)(\mathrm{ess\,sup}\,|V|)\alpha(t)\ , \tag{2.1}$$

$$|\mathrm{Cov}\,(U, V)| \leqq 2(\mathrm{ess\,sup}\,|U|)E|V|\phi(t)\ , \tag{2.2}$$

$$|\mathrm{Cov}\,(U, V)| \leqq E|U|E|V|\psi(t)\ . \tag{2.3}$$

3. Consistency of f_T

Let $\|\cdot\|$ denote Euclidean norm. We introduce assumption

A3.1 Either

(i) $k(u)$ is bounded with compact support; or

(ii) $|k(u)| \leqq C\exp(-D\|u\|^\rho)$ for some C, D, $0<C$, $D<\infty$, where $\rho>0$ is such that $\underline{\lim}_{T\to\infty} \|h_T\|^\rho \log H_T > -\infty$; or

(iii) k is bounded and integrable, $\|u\|^q|k(u)| \to 0$ as $\|u\| \to \infty$, and for some $C<\infty$, $\|h_T\|^q \leqq CH_T$.

Under A3.1 it follows from Lemma 8.1 of [19] that for any $\delta>0$

$$\sup_{\|u\|>\delta/\|h_T\|} |k(u)|/H_T \to 0\ , \qquad \text{as } T\to\infty\ . \tag{3.1}$$

A3.2 $TH_T \to \infty$ as $T\to\infty$.

A3.3 $f(z)$ is continuous at $z=0$.

A3.4 For all $t>a$, (Z_1, Z_{1+t}) admits a density, which is bounded uniformly in t in a neighbourhood of the origin in R^{2q}.

A3.5 $(TH_T^2)^{-1}\sum_1^T \alpha(t)\to 0$ as $T\to\infty$.

A3.6 For any $\varepsilon>0$, $(TH_T^2)^{-1}\sum_{\varepsilon T}^T \alpha(t)\to 0$ as $T\to\infty$.

A3.7 $(TH_T)^{-1}\sum_1^T \phi(t)\to 0$ as $T\to\infty$.

A3.8 For any $\varepsilon>0$, $(TH_T)^{-1}\sum_{\varepsilon T}^T \phi(t)\to 0$ as $T\to\infty$.

A3.9 X_t is $*$-mixing.

THEOREM 3.1. *Let* A1.1, A1.2, A3.1, A3.2 *and* A3.3 *hold. Then under either* (i) A3.5; *or* (ii) A3.6 *and* A3.4; *or* (iii) A3.7; *or* (iv) A3.8 *and* A3.4; *or* (v) A3.9,

$$E(f_T-f)^2\to 0 \quad \textit{as } T\to\infty\,. \tag{3.2}$$

PROOF. The LHS of (3.2) is $P+Q$ where $P=\operatorname{Var}(f_T)$ and $Q=(Ef_T-f)^2$. An extension of a theorem of Bochner (see [5], [20]), plus A1.1 and (3.1), implies $Q\to 0$, whereas

$$P=(T'H_T^2)^{-1}\left\{\operatorname{Var}(K_{1T})+\sum_{t=1}^{T'-1}\operatorname{Cov}(K_{1T}, K_{t+1,T})\right\} \tag{3.3}$$

$$\leqq (T'H_T^2)^{-1}\left\{(a+1)\operatorname{Var}(K_{1T})+\sum_{t=a+1}^{T'-1}|\operatorname{Cov}(K_{1T}, K_{t+1,T})|\right\}\,. \tag{3.4}$$

By straightforward application of Lemma 8.2 of [19], $\operatorname{Var}(K_{1T})\leqq CH_T$, where henceforth C is a generic constant. By (2.1),

$$|\operatorname{Cov}(K_{1T}, K_{t+a+1,T})|\leqq 4\sup_u |k(u)|^2\alpha(t)\,, \qquad t\geqq 1\,.$$

Thus, writing $G_T=(TH_T)^{-1}$ and $\alpha(a,b)=(TH_T^2)^{-1}\sum_a^b \alpha(t)$,

$$P\leqq C\{G_T+\alpha(1,T)\}\,, \tag{3.5}$$

to prove the Theorem under (i). When A3.4 is assumed,

$$|\operatorname{Cov}(K_{1T}, K_{t+a+1,T})|\leqq CH_T^2\,, \tag{3.6}$$

using Lemmas 8.2 and 8.3 of [19]. Thus for any N, $1\leqq N\leqq T$,

$$P\leqq C\{G_T+N/T+\alpha(N,T)\} \tag{3.7}$$

and the choice $N\sim\varepsilon T$ with $T\to\infty$ and then $\varepsilon\to 0$ completes the proof under (ii). Under ϕ-mixing, (2.2) implies

$$|\operatorname{Cov}(K_{1T}, K_{t+a+1,T})|\leqq 2(\sup_z |k(z)|)E|K_{1T}|\phi(t)\leqq CH_T\phi(t)\,, \qquad t\geqq 1\,,$$

using also Lemma 8.2 of [19]. Thus, with $\phi(a, b)=G_T \sum_a^b \phi(t)$,

$$P \leqq C\{G_T+\phi(1, T)\} , \tag{3.8}$$

resolving case (iii); with A3.4 also,

$$P \leqq C\{G_T+N/T+\phi(N, T)\} , \tag{3.9}$$

so the proof under (iv) is completed as in (ii). Finally apply (2.3),

$$|\mathrm{Cov}\,(K_{1T}, K_{t+a+1,T})| \leqq (E|K_{1T}|)^2 \psi(t) \leqq CH_T^2 \psi(t) , \qquad t \geqq 1 .$$

Thus,

$$P \leqq C\left\{G_T+T^{-1} \sum_1^T \psi(t)\right\} \rightarrow 0 \tag{3.10}$$

under A3.9, to deal with (v).

Unless $\alpha(1)=0$, (i) requires at least $TH_T^2 \rightarrow \infty$; on the other hand, A3.2 suffices for (ii) when $\alpha(t)=O(t^{-\theta})$, $\theta \geqq 2$. To distinguish between (iii) and (iv), when $\phi(t)=O(t^{-1})$ (iv) is satisfied with A3.2 but (iii) requires that $TH_T/\log T \rightarrow \infty$. We have not established that all our conditions are necessary for consistency, and the general, nonlinear nature of k and our unwillingness to make more precise distributional assumptions on X_t seriously hinder the derivation of necessary conditions.

4. Some finite-sample theory

Our conditions A3.5–A3.8, as well as a number of Monte Carlo simulations carried out by the author, tend to support the intuition that given two processes, one serially independent and the other dependent, but with identical marginal features, a larger bandwidth might often be chosen in the latter case in order to achieve comparable variances in finite samples. (The usual bandwidth conditions for the CLT of nonparametric estimators, and asymptotic variances and MSE, are by contrast the same for dependent processes as for independent ones, see, e.g., [14], [15], [19], [20].) As usual, asymptotic theory provides no precise guide to the choice of bandwidth for a given data set and, though automatic procedures such as cross-validation can be devised, these are computationally expensive and depend on a somewhat arbitrary choice of objective function, and as in spectral analysis may not adequately substitute for a mixture of judgement, experience and trial and error. We must add that a larger bandwidth typically carries the penalty of increased bias, and it is also possible that the covariances in P above could conceivably make a net *negative* contribution, meaning that serial dependence might in some cases actually *reduce* the variance.

The influence of serial dependence and bandwidth, and their interaction, can be examined analytically for finite T in some special cases. For example, let $Z_t = X_t + \mu$, where X_t is scalar and Gaussian with lag-j autocorrelation ρ_j and (w.l.o.g.) $EX_t = 0$, $\operatorname{Var} X_t = 1$. Varying μ corresponds to varying the point at which the density of X_t is to be estimated. Let $k(z) = I(|z| \leqq 1)/2$, satisfying A3.1. By Price's theorem ([11], p. 87)

$$(4.1) \qquad \frac{d}{d\rho_j} \operatorname{Cov}(K_{tT}, K_{t+j,T}) = E\left(\frac{\partial K_{tT}}{\partial X_t} \frac{\partial K_{t+j,T}}{\partial X_{t+j}}\right) .$$

where the partial derivatives are generalized functions in this case. After straightforward derivation (4.1) is seen to equal

$$(4.2) \qquad \frac{1}{4\pi(1-\rho_j^2)^{1/2}} \exp\left\{-\frac{(\mu^2+H_T^2)}{1+\rho_j}\right\} \left\{\cosh\left(\frac{2H_T\mu}{1+\rho_j}\right) - \exp\left(\frac{-2\rho_j H_T^2}{1-\rho_j^2}\right)\right\} ,$$

which is positive on $\{0 \leqq \rho_j < 1,\ \mu \in R^1,\ (\rho_j, \mu) \neq (0, 0),\ H_T > 0\}$. Thus if, for example, $\rho_j \geqq 0$ for all j and $\rho_j > 0$ for some j, the contribution of the covariance terms to P is positive, and increasing in the ρ_j. Notice that (4.2) need not be positive for all $\rho_j < 0$, in which case the autocovariances of K are not monotonic in ρ_j, and negatively correlated X_t's can give rise to positively correlated K's. Indeed at $\mu = 0$ (4.2) is negative for all $\rho_j < 0$, so the K's are positively correlated whether ρ_j is positive or negative. Thus, although we acknowledged above that the covariances could make a net negative contribution to P, this outcome may be rather unlikely, particularly bearing in mind that processes with some negative ρ_j will often have a number of positive ρ_j also. These observations seem also to suggest that our mixing conditions, which ignore the sign of $\operatorname{Cov}(K_{tT}, K_{t+j,T})$, may produce sharper results than appeared at first sight.

To carry things a stage further, we deduce from (3.3), (4.2) that

$$(4.3) \qquad P - (TH_T^2)^{-1} \operatorname{Var}(K_{1T}) \doteqdot \{(2\pi H_T^2)^{-1} \exp(-H_T^2 - \mu^2) \cdot [\cosh(2H_T\mu) - 1]\} \sum_{j=1}^{T-1} (1 - j/T)\rho_j$$

neglecting a term of $O\left(\sum_1^T \rho_j^2\right)$, which, on a nondegenerate set of $\{\rho_j\}$ values, is smaller in magnitude than the right-hand side of (4.3), and as $\sup_j |\rho_j| \to 0$ is a negligible fraction of the right-hand side of (4.3). The factor in braces in (4.3) is a positive, decreasing function of H_T, whereas the summation is increasing in the ρ_j. This approximate analysis suggests that the increase in variance owing to (predominantly positive) autocorrelation can indeed be reduced by increasing H_T; increased bias, thence in some cases increased MSE, may result, though bias is often of smaller order than variance, to an extent depending on smoothness of f and choice of k.

5. Consistency of m_T

We now assume

A5.1 $m(z)$ is continuous at $z=0$.

A5.2 $E|Y_1-m(Z_1)|<\infty$; $E(|Y_1-m(Z_1)||Z_1=z)$ is continuous at $z=0$.

However, (unlike [21]) we no longer assume continuity of $f(z)$ but

A5.3 $f(z)$ is bounded and bounded away from zero in a neighbourhood of the origin.

Like [21], we intensify A3.2 to

A5.4 $TH_T\uparrow\infty$ as $T\to\infty$.

We likewise introduce A5.5–A5.8 by changing $\to$ to $\downarrow$ in A3.5–A3.8; A5.5–A5.8 are negligibly stronger than A3.5–A3.8 and are readily checked in cases such as $\alpha(t)$ or $\phi(t)$ is $O(t^{-\theta})$ and $H_T^{-1}=O(T^{-\nu})$, for suitable values of θ and ν. Our final requirement is satisfied for all reasonable choices of k, and, like A3.1, allows for k that are not everywhere nonnegative.

A5.9 For some $\varepsilon>0$, $k(z)>0$ for $\|z\|<\varepsilon$.

THEOREM 5.1. *Let* A1.1, A1.2, A3.1, A5.1, A5.2, A5.3, A5.4 *and* A5.9 *hold. Then under* (i)′ A5.5; *or* (ii)′ A5.6 *and* A3.4; *or* (iii)′ A5.7; *or* (iv)′ A5.8 *and* A3.4; *or* (v)′ A3.9

$$p\lim_{T\to\infty} m_T=m\ . \tag{5.1}$$

PROOF. For some $\delta>0$,

$$\begin{aligned} Ef_T &\geqq H_T^{-1}\left\{\int_{\|z\|\leqq\delta} k(h_T^{-1}z)f(z)dz-\left|\int_{\|z\|>\delta} k(h_T^{-1}z)f(z)dz\right|\right\} \\ &\geqq \inf_{\|z\|\leqq\delta} f(z)\int_{\|z\|\leqq\delta/\|h_T\|} k(z)dz-\sup_{\|z\|>\delta/\|h_T\|}|k(z)|/H_T\geqq\lambda-o(1)>0 \end{aligned} \tag{5.2}$$

for large enough T, by A1.1, A5.1 and A5.3. For such T, $m_T-m\sim\sum''(Y_t-m)K_{tT}/\sum' K_{tT}$, so for any η such that $0<\eta<\lambda$,

$$P(|m_T-m|>\eta)\leqq P(|(T''H_T)^{-1}\textstyle\sum''(Y_t-m)K_{tT}|>\eta^2)+P(f_T<\eta)\ .$$

Using (5.2), $P(f_T<\eta)\leqq P(|f_T-Ef_T|>\lambda-\eta)\to 0$ by Tchebycheff's inequality and $P \to 0$, established in the proof of Theorem 3.1, on noting that (i)′–(v)′ imply (i)–(v) respectively. Now write

$$(T''H_T)^{-1} \sum'' (Y_t - m) K_{tT} = (T''H_T)^{-1} \sum'' U_t K_{tT} + (T''H_T)^{-1} \sum'' \{m(Z_t) - m\} K_{tT} .$$

The second term $\to 0$ in L_1 by application of Lemma 8.7 of [19]. We handle the first term by arguments based on truncated variables, $U_t' = U_t I(|U_t| < n)$, $U_t'' = U_t - U_t'$, for appropriate n. Let

$$R' = (T''H_T)^{-1} \sum'' \{U_t' - E(U_t' | Z_t)\} K_{tT} ,$$

$$R'' = (T''H_T)^{-1} \sum'' \{U_t'' - E(U_t'' | Z_t)\} K_{tT} ,$$

then the theorem follows if $R' \to 0$ in L_2 and $R'' \to 0$ in L_1. First consider case (i)′. We obtain (cf. (3.4), (3.5)), since U_t' is M_{t+b}^{t+c}-measurable and $E(U_t' | Z_t)$ is M_t^{t+a}-measurable,

$$ER'^2 \leqq C\{(TH_T^2)^{-1} \operatorname{Var}(U_1' K_{1T}) + n^2 \alpha(1, T)\} . \tag{5.3}$$

Now $\operatorname{Var}(U_1' K_{1T}) \leqq CnE|U_1 K_{1T}|$ and by Lemma 8.2 of [19]

$$E(U_1 K_{1T}) \leqq CH_T \{\sup_{\|z\| \leqq \delta} E(|U_1| | Z_1 = z) \sup_{\|z\| \leqq \delta} f(z) + E|U_1|\} . \tag{5.4}$$

Thus taking δ small and applying A5.2 and A5.3, $\operatorname{Var}(U_1' K_{1T}) \leqq CnH_T$. For any $\eta > 0$ choose $n = \min[\eta G_T^{-1}, \{\eta / \alpha(1, T)\}^{1/2}]$, so

$$ER'^2 \leqq C\eta . \tag{5.5}$$

Now

$$E|R''| \leqq 2(T''H_T)^{-1} \sum'' E|U_t'' K_{tT}| \leqq 2H_T^{-1} E\{E(|U_1''| | Z_1) |K_{1T}|\}$$
$$\leqq C\{\sup_{\|z\| \leqq \delta} E(|U_1''| | Z_1 = z) \sup_{\|z\| \leqq \delta} f(z) + E|U_1''|\}$$

for some $\delta > 0$, using Lemma 8.2 of [19]. We have

$$E(|U_1''| | Z_1 = z) \leqq E\{|U_1| I(|U_1| > \eta G_T^{-1}) | Z_1 = z\} + E\{|U_1| I(|U_1| > \{\eta / \alpha(1, T)\}^{1/2} | Z_1 = z\} .$$

For small δ, both terms on the right-hand side are continuous on the set $\|z\| \leqq \delta$ ([21], p. 1388 and A5.2) and both $\downarrow 0$ as $T \to \infty$ because of A5.4 and A4.5. Thus by [22], Theorem 7.13 $\limsup_{T \to \infty} \sup_{\|z\| \leqq \delta} E(|U_1''| | Z_1 = z) = 0$, and because of A5.2 and A5.5 imply $E|U_1''| \to 0$ as $T \to \infty$, it follows that $R'' \to 0$ in L_1. Then (5.5) implies $ER'^2 \to 0$ because η is arbitrary. In case (ii)′ we obtain first, cf. (3.6), $ER'^2 \leqq Cn\{G_T + nNT^{-1} + n\alpha(N, T)\}$. Choose $N = \varepsilon T$ for $\varepsilon > 0$, then $\eta > 0$ and $n = \min[\eta G_T^{-1}, (\eta/\varepsilon)^{1/2}, \{\eta / \alpha(\varepsilon T, T)\}^{1/2}]$ to give (5.5). In this case we have

$$E(|U_1''| | Z_1 = z) \leqq E(|U_1| I(|U_1| > \eta G_T^{-1}) | Z_1 = z)$$

$$+E(|U_1|I(|U_1|>(\eta/\varepsilon)^{1/2})|Z_1=z)$$
$$+E(|U_1|I(|U_1|>\{\eta/\alpha(\varepsilon T, T)\}^{1/2}|Z_1=z) \ .$$

As before the first term, and for fixed ε, the last term, $\to 0$ uniformly over $\|z\|\leqq\delta$ as $T\to\infty$. As $\varepsilon\to 0$ the second term vanishes and the proof is completed as above. Similar methods are used for ϕ-mixing: in case (iii)$'$ (cf. (3.8)), $ER'^2\leqq Cn\{G_T+\phi(1, T)\}$, so take $n=G_T+\phi(1, T)$; in case (iv)$'$ (cf. (3.9)), $ER'^2\leqq Cn\{G_T+\phi(\varepsilon T, T)+n\varepsilon\}$, so choose $n=\min(\eta(G_T+\phi(\varepsilon T, T))^{-1}, (\eta/\varepsilon)^{1/2})$. Under $*$-mixing we obtain (cf. (3.10))

$$ER'^2\leqq C(TH_T^2)^{-1}\left[\mathrm{Var}\,(U_1'K_{1T})+\{E|U_1'K_{1T}|\}^2\sum_1^T\psi(t)\right]$$
$$\leqq C\left\{G_Tn+T^{-1}\sum_1^T\psi(t)\right\}$$

using $E|U_1'K_{1T}|\leqq E|U_1K_{1T}|$ and (5.4). Choose $n=\eta G_T^{-1}$ and apply A5.2, to complete the proof of the theorem.

6. Time series regression

Prompted by time series regressions that contain no lagged dependent variables, there is some interest in the case $Y_t=m(Z_t)+U_t$, where

A6.1 $\{U_t\}$ is independent of $\{Z_t\}$.

Then A5.2 becomes simply

A6.2 $E|U_1|<\infty$.

By virtue of A6.1 we can impose a different mixing condition on the residuals U_t from those on X_t. Let $\beta(t)$ be the strong mixing coefficient for U_t.

A6.3 $(TH_T)^{-1}\sum_1^T\beta(t)\downarrow 0$ as $T\to\infty$.

A6.4 U_t is strong mixing.

THEOREM 6.1. *Let* A1.1, A1.2, A3.1, A5.1, A5.3, A5.4, A5.9, A6.1 *and* A6.2 *hold. Then under either* (i)$''$ A3.5 *and* A6.3; *or* (ii)$''$ A3.6, A6.4 *and* A3.4; *or* (iii)$''$ A3.7 *and* A6.3; *or* (iv)$''$ A3.8, A6.4 *and* A3.4; *or* (v)$''$ A3.9 *and* A6.3; *or* (vi)$''$ A3.9, A6.4 *and* A3.4, *then* (5.1) *is true.*

PROOF. The first point at which the proof departs from that of Theorem 5.1 is when ER'^2 is first majorized, (5.3). Because $U_t'-E(U_t'|Z_t)=U_t'-E(U_t')$ is independent of Z_u for all t, u, we have, with $\beta=G_T\sum_1^T\beta(t)$,

$$(6.1) \qquad E(R'^2) \leqq C(TH_T^2)^{-1}\left\{\text{Var}(U_1')E(K_{1T}^2)+\sum_1^T E|K_{1T}K_{1+t,T}| \cdot |\text{Cov}(U_1', U_{1+t}')|\right\} \leqq Cn(G_T+n\beta)$$

in the cases in which A3.4 is not imposed. Thus choose $n=\min[\eta G_T^{-1}, (\eta/\beta)^{1/2}]$ and proceed as in the proof of Theorem 5.1; on combining with Theorem 3.1, we have proved (i)″, (iii)″ and (v)″. When A3.4 is assumed, (6.1) implies $ER'^2 \leqq Cn(G_T+n\beta')$, where $\beta'=H_T\beta$. Choose $n=\min[\eta G_T^{-1}, (\eta/\beta')^{1/2}]$ and note that $\beta(t)\downarrow 0$ implies $\beta'\downarrow 0$, to complete the proof.

An example of a nonlinear time series regression structure of the form $Y_t=m(Z_t)+U_t$ such that $\{Z_t\}$ and $\{U_t\}$ are independent, is in [12], where Z_t consists of lagged values of a scalar observable input variable X_t and $m(Z_t)$ is a bilinear function of the lagged X_t, motivated by a linear distributed lag model whose coefficients are themselves linear in lagged X_t. In [12] X_t is assumed Gaussian, whence our mixing conditions on X_t are conditions on the decay of X_t's autocorrelations.

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

REFERENCES

[1] Ahmad, I. A. (1979). Strong consistency of density estimators by orthogonal series methods for dependent variables with applications, *Ann. Inst. Statist. Math.*, **31**, 279-288.

[2] Ahmad, I. A. (1981). A note on nonparametric density estimation for dependent variables using a delta sequence, *Ann. Inst. Statist. Math.*, **33**, 247-254.

[3] Ahmad, I. A. (1982). Integrated mean square properties of density estimation by orthogonal series methods for dependent variables, *Ann. Inst. Statist. Math.*, **34**, 339-350.

[4] Bosqu, D. (1983). Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires, *Zeit. Wahrscheinlichkeitsth.*, **64**, 541-553.

[5] Cacoullos, T. (1966). Estimation of a multivariate density, *Ann. Inst. Statist. Math.*, **18**, 179-189.

[6] Chanda, K. C. (1983). Density estimation for linear processes, *Ann. Inst. Statist. Math.*, **35**, 439-446.

[7] Collomb, G. (1982). Propriétés de convergence presque complète du predictor à noyau, Preprint.

[8] Georgiev, A. (1984). Nonparametric system identification by kernel method, *IEEE Trans. Automatic Control*, **29**, 356-359.

[9] Greblicki, W., Rutkowska, D. and Rutkowski, L. (1983). An orthogonal series estimate of time-varying regression, *Ann. Inst. Statist. Math.*, **35**, 215-228.

[10] Gyorfi, L. (1981). Strong consistent density estimate from ergodic sample, *J. Multivariate Anal.*, **11**, 81-84.

[11] Hannan, E. J. (1970). *Multiple Time Series*, John Wiley and Sons, New York.

[12] Hinich, M. J. (1979). Estimating the lag structure of a nonlinear time series model, *J. Amer. Statist. Ass.*, **74**, 449-456.

[13] Ibragimov, I. A. (1970). On the spectrum of stationary Gaussian processes satisfying the strong mixing condition. II Sufficient conditions. Mixing rate, *Theory Prob. Appl.*, **15**, 23-36.

[14] Masry, E. (1983). Probability density estimation from sampled data, *IEEE Trans. Inf. Theory*, **29**, 696-709.

[15] Pham, T. D. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation, *Math. Operationsforsch. Statist., ser. Statist.*, **12**, 61-73.

[16] Pham, T. D. and Tran, L. T. (1980). The strong mixing properties of the autoregressive moving average time series model, *Seminaire de Statistique*, Grenoble, 59-76.

[17] Phillip, W. (1969). The central limit problem for mixing sequences of random variables, *Zeit. Wahrscheinlichkeitsth.*, **12**, 155-171.

[18] Prakasa Rao, B. L. S. (1978). Density estimation for Markov processes using delta-sequences, *Ann. Inst. Statist. Math.*, **30**, 321-328.

[19] Robinson, P. M. (1983). Nonparametric estimators for time series, *J. Time Series Anal.*, **4**, 185-207.

[20] Roussas, G. (1967). Nonparametric estimation in Markov processes, *Ann. Inst. Statist. Math.*, **21**, 73-87.

[21] Roussas, G. (1969). Nonparametric estimators of the transition distribution function of a Markov process, *Ann. Math. Statist.*, **40**, 1386-1400.

[22] Rudin, W. (1976). *Principles of Mathematical Analysis*, McGraw-Hill, New York.

[23] Takahata, H. (1977). On recursive density estimators for a class of stationary processes, *Bull. Tokyo Gakugei Univ.*, Ser. IV, **29**, 10-18.

[24] Takahata, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, *Bull. Tokyo Gakugei Univ.*, Ser. IV, **32**, 11-32.