Consider the nonparametric regression model $Y_i^{(n)} = g(x_i^{(n)}) + \varepsilon_i^{(n)}, i = 1, \ldots, n,$ where $g$ is an unknown regression function and assumed to be bounded and real valued on $A \subseteq \mathbb{R}^p$, $x_i^{(n)}$'s are known and fixed design points and $\varepsilon_i^{(n)}$'s are assumed to be both dependent and non-identically distributed random variables. This paper investigates the asymptotic properties of the general nonparametric regression estimator $g_n(x) = \sum_{i=1}^{n} W_n(x_i) Y_i^{(n)}$, where the weight function $W_n(x)$ is of the form $W_n(x) = W_n(x; x_1^{(n)}, x_2^{(n)}, \ldots, x_n^{(n)})$. The estimator $g_n(x)$ is shown to be weak, mean square error, and universal consistent under very general conditions on the temporal dependence and heterogeneity of $\varepsilon_i^{(n)}$'s. Asymptotic distribution of the estimator is also considered. © 1990 Academic Press, Inc.

1. INTRODUCTION AND DEFINITIONS

Let $p$ be an integer and $A$ be a compact set in $\mathbb{R}^p$. Consider the nonparametric regression model

$$Y_i^{(n)} = g(x_i^{(n)}) + \varepsilon_i^{(n)}, \quad i = 1, 2, \ldots, n,$$

where $g$ is an unknown regression function and assumed to be bounded and real valued on $A$, $x_i^{(n)}$'s are known and fixed design points from $A$ and the random errors $\varepsilon_i^{(n)}$'s are assumed to be either an $L^d$-mixingale
(1 \leq q \leq 2) or near epoch dependent with respect to certain mixing sequences. Further, the expected value of \( \varepsilon_i^{(n)} \)'s is assumed to be zero.

The estimator we will study in this paper is the general nonparametric regression estimator of the following form

\[
g_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i^{(n)}, \quad x \in A \subset \mathbb{R}^p
\]

where the weight function \( W_{ni}(x) \) is of the form \( W_{ni}(x) = W_{ni}(x; x_1^{(n)}, x_2^{(n)}, \ldots, x_n^{(n)}) \).

The above estimator was proposed by Georgiev [10] and discussed by Georgiev and Greblicki [11] for dimension \( p = 1 \), Georgiev [12] for \( p > 1 \) in the case where \( \varepsilon_i^{(n)} \)'s are independent but not necessarily identically distributed random variables. In his paper, Georgiev [12] considered weak, mean square error, and strong consistency as well as asymptotic normality of \( g_n(x) \). Müller [16] derived results on weak and universal consistency of \( g_n(x) \) in the model with independent, identically distributed disturbances.

The aim of this paper is to extend some of the results in Georgiev [12] and Müller [16] to the estimation of the regression model where the disturbances are dependent and heterogeneous processes.

The notion of universal consistency was first introduced by Stone [19] for the random design model and adapted by Müller [16] for the fixed design model with identically distributed disturbances. For the fixed design model with nonidentically distributed random errors being considered in this paper, we modify the definition of Müller [16] in the following straightforward way: An array of weights \( (W_{ni}(x)), 1 \leq i \leq n, n \in \mathbb{N} \), is called fixed design universally consistent at a fixed point \( x \in A \), if the uniform integrability of \( \{ |g_n(x) - g(x)|^q : i < n \} \) for some \( 1 \leq q \leq 2 \) and the continuity of \( g \) in \( x \) imply that \( E|g_n(x) - g(x)|^q \to 0 \) as \( n \to \infty \).

The remainder of this paper is organized as follows. Section 2 states and proves the weak, mean square error and universal consistency of \( g_n(x) \) under various assumptions on the temporal dependence and relative “magnitude” of \( \varepsilon_i^{(n)} \)'s. In particular, \( \{\varepsilon_i^{(n)}\} \) is assumed to be an \( L^q \)-mixingale for some \( 1 \leq q \leq 2 \). By restricting \( \varepsilon_i^{(n)} \)'s to be a subclass of mixingale, i.e., near epoch dependent with respect to certain mixing sequences, Section 3 then obtains the asymptotic normality of \( g_n(x) \). Several examples are also given in due course of the paper to illustrate the application of various results obtained to kernel and nearest neighbor estimators.

The present section is closed with several definitions which will be used in later part of the paper.

Let \( (\Omega, \mathcal{F}, P) \) denote a probability space and \( \{X_{ni} : i = 1, \ldots, k_n ; n = 1, 2, \ldots\} \) be a triangular array on \( (\Omega, \mathcal{F}, P) \), where \( k_n \to \infty \) as \( n \to \infty \). Let \( \{\mathcal{F}_n : i = 0, 1, \ldots ; n = 1, 2, \ldots\} \) be an array of sub-\( \sigma \) fields of \( \mathcal{F} \) such that...
\{\mathcal{F}_n\} \text{ is nondecreasing in } i \text{ for each } n. \text{ Further, let } \| \cdot \|_q \text{ denote } L^q(P) \text{ norm, i.e., } \|X_n\|_q = (E|X_n|^q)^{1/q}.

**Definition 1.1.** The triangular array \{X_{n,i}, \mathcal{F}_n\} is an \(L^q\)-mixingale if there exist nonnegative constants \{C_i: i = 1, \ldots, k, \ n = 1, 2, \ldots\} and \{\psi_m: m = 0, 1, \ldots\} such that \(\psi_m \downarrow 0\) as \(m \to \infty\) and for all \(i = 1, \ldots, k, \ n \geq 1, \) and \(m \geq 0,\) we have

(a) \(\|E(X_{n,i} | \mathcal{F}_{n,i-m})\|_q \leq C_n \psi_m\) and
(b) \(\|X_{n,i} - E(X_{n,i} | \mathcal{F}_{n,i+m})\|_q \leq C_n \psi_{m+1}\).

**Definition 1.2.** Let \(\{Y_{n,i}: i = 0, \pm 1, \ldots, n = 1, 2, \ldots\}\) be a doubly infinite double array of random vectors defined on \((\Omega, \mathcal{F}, P)\). Let \(G_{n,j} = \sigma(Y_{nj}, \ldots, Y_{nk}),\) for all \(j < k, j, k \in \mathbb{Z}, n \in \mathbb{N}\) The process \(\{X_{n,i}\}\) is near epoch dependent with respect to \(\{Y_{n,i}\}\) if and only if \(X_{n,i} \in L^2(P)\) for all \(n, i \in \mathbb{N}\) and there exist constants \{\mu_m \geq 0: m = 0, 1, \ldots\} with \(\mu_m \downarrow 0\) and \(\{d_m > 0: n, i = 1, 2, \ldots\}\) such that

\[\|X_{n,i} - E(X_{n,i} | G_{n,i+m})\|_2 \leq \mu_m d_m\]

**Definition 1.3.** The uniform \((\phi-\) and strong \((\alpha-\) mixing coefficients for \(Y_{n,i}\) are, respectively,

\[\phi_m = \sup_n \sup_i \sup_{G \in G_{n,i}^{- \infty}, H \in G_{n,i+m}} |P(H | G) - P(H)|\]

\[\alpha_m = \sup_n \sup_i \sup_{G \in G_{n,i}^{- \infty}, H \in G_{n,i+m}} |P(G \cap H) - P(G) P(H)|.\]

**Remark 1.1.** Definition 1.1 is given by Andrews [3] when \(q = 1,\) by Mcleish [15] when \(q = 2. L^2\)-mixingale is also called mixingale. Definitions 1.2 and 1.3 are given by Wooldridge and White [21].

2. Consistency Results

The following three basic assumptions will be used throughout this section.

**Assumptions.** (A1) (a) The random errors \(\varepsilon_i^{(n)}\)'s form an \(L^q\)-mixingale with \(1 \leq q \leq 2\) and \(E\varepsilon_i^{(n)} = 0,\) for \(i \leq n;\)

(b) The function \(g\) is bounded on a compact set \(A \subset \mathbb{R}^p.\)

(A2) The weight functions \(W_{n,i}(x)\) satisfy:

(a) \(\sum_{i=1}^{n} W_{n,i}(x) \to 1\) as \(n \to \infty;\)
NONPARAMETRIC MULTIPLE REGRESSION

(b) \( \sum_{i=1}^{n} |W_{ni}(x)| \leq B \) for all \( n \);

(c) \( \sum_{i=1}^{n} |W_{ni}(x)| I_{\|x_{i}^{(n)} - x\| > a} \rightarrow 0 \) as \( n \rightarrow \infty \) for all \( a > 0 \).

\( (A3) \) \( \sup_{i \leq n} |W_{ni}(x)| \rightarrow 0 \) as \( n \rightarrow \infty \).

**Theorem 2.1.** Assume \( (A1) \) and \( (A2) \) hold, then \( \hat{E}_{n}(x) \rightarrow g(x) \) as \( n \rightarrow \infty \) at every continuity point \( x \in A \) of the function \( g \).

**Proof.** See Georgiev [12].

**Remark 2.1.** The non-independence of \( \varepsilon_{i}^{(n)} \)'s or \( Y_{i}^{(n)} \)'s does not affect, in any way, the asymptotic unbiasedness of \( \hat{g}_{n}(x) \) as is well known in the literature.

The following lemma extends Theorem 1 of Pruitt [17] (Lemma 2 of Müller [16]) to the model with \( \varepsilon_{i}^{(n)} \)'s being a martingale difference sequence. In particular, they are not necessarily identically distributed as assumed by Pruitt. It will be used to prove the universal consistency of \( \hat{g}_{n}(x) \), i.e., Theorem 2.3.

**Lemma 2.2.** Let \( \{\varepsilon_{i}^{(n)}, F_{ni} : i \leq n\} \) be a martingale difference sequence. If \( \{|\varepsilon_{i}^{(n)}|^{q} : i \leq n\} \) is uniformly integrable for some \( 1 \leq q \leq 2 \), and the weight functions \( W_{ni}(x) \) satisfy \( (A2)(b) \) and \( (A3) \), then,

(i) \( E|\sum_{i=1}^{n} W_{ni}(x) \varepsilon_{i}^{(n)}|^{q} \rightarrow 0 \) as \( n \rightarrow \infty \);

(ii) \( \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{i}^{(n)} \rightarrow 0 \) in probability as \( n \rightarrow \infty \).

**Proof.** (i) Given any number \( \varepsilon > 0 \), since \( \{|\varepsilon_{i}^{(n)}|^{q} \} \) is uniformly integrable, there exists a constant \( C < \infty \) such that

\[
\sup_{i,n} \|\varepsilon_{i}^{(n)}I(|\varepsilon_{i}^{(n)}| > C)\|_{q} < \varepsilon/4B.
\]

Further, \( (A2)(b) \) and \( (A3) \) imply \( n_{x} = (\sum_{i=1}^{n} W_{ni}^{2}(x))^{-1} \rightarrow \infty \) as \( n \rightarrow \infty \). Therefore, \( \exists N > 0 \), such that \( n > N \) implies

\[
\frac{1}{n_{x}} < \varepsilon^{2}/4C^{2}.
\]

Let \( Y_{ni} = \varepsilon_{i}^{(n)}I(|\varepsilon_{i}^{(n)}| < C) \) and \( Z_{ni} = \varepsilon_{i}^{(n)}I(|\varepsilon_{i}^{(n)}| > C) \). Then

\[
\left\| \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{i}^{(n)} \right\|_{q} = \left\| \sum_{i=1}^{n} W_{ni}(x)(\varepsilon_{i}^{(n)} - E(\varepsilon_{i}^{(n)} | F_{ni-1})) \right\|_{q} \leq \left\| \sum_{i=1}^{n} W_{ni}(x)(Y_{ni} - E(Y_{ni} | F_{ni-1})) \right\|_{q} + \left\| \sum_{i=1}^{n} W_{ni}(x)(Z_{ni} - E(Z_{ni} | F_{ni-1})) \right\|_{q}.
\]
\[ \begin{aligned}
& \leq \left\| \sum_{i=1}^{n} W_{ni}(x)(Y_{ni} - E(Y_{ni} | \mathcal{F}_{n-1})) \right\|_{2} \\
& + 2 \sum_{i=1}^{n} \left\| W_{ni}(x) Z_{ni} \right\|_{q} \\
& = \left[ \sum_{i=1}^{n} W_{ni}^{2}(x) E(Y_{ni} - E(Y_{ni} | \mathcal{F}_{n-1}))^{2} \right]^{1/2} \\
& + 2 \sum_{i=1}^{n} |W_{ni}(x)| \left\| Z_{ni} \right\|_{q} \\
& \leq C \left[ \sum_{i=1}^{n} W_{ni}^{2}(x) \right]^{1/2} + 2 \sup_{i,n} \left\| Z_{ni} \right\|_{q} \sum_{i=1}^{n} |W_{ni}(x)| \\
& \leq \frac{C}{\sqrt{n}} + \frac{\varepsilon}{2} \\
& \leq \varepsilon
\end{aligned} \]

for \( n > N \). Therefore, \( E \left| \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{i}^{(n)} \right|^{q} \to 0 \) as \( n \to \infty \).

(ii) Follows from (i) and Markov inequality.

Note the above proof is a modification of Andrews \([3]\).

**Theorem 2.3.** Suppose assumptions (A1), (A2), and (A3) hold. If \( \sup_{i,n} C_{ni} < \infty \), then \( \{ W_{ni}(x) \}_{1 \leq i \leq n} \) is fixed design universally consistent in \( x \in A \). (A2)(a) and (A3) are necessary conditions. If \( W_{ni}(x) > 0, i = 1, \ldots, n \), (A2)(b) and (A2)(c) are necessary, too.

**Proof.** Since \( |g_{n}(x) - g(x)| \leq |g_{n}(x) - Eg_{n}(x)| + |Eg_{n}(x) - g(x)| \), by \( C_{r} \)-inequality with \( r = q \), we obtain

\[
E \left| g_{n}(x) - g(x) \right|^{q} \leq C_{q}(E \left| g_{n}(x) - Eg_{n}(x) \right|^{q} + |Eg_{n}(x) - g(x)|^{q}),
\]

where \( C_{q} = 2^{q-1} \). Then, it suffices to show that \( E \left| g_{n}(x) - Eg_{n}(x) \right|^{q} \to 0 \) as \( n \to \infty \) in view of Theorem 2.1.

Now, for any positive number \( M \) to be chosen,

\[
g_{n}(x) - Eg_{n}(x) = \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{i}^{(n)}
\]
\[
\begin{align*}
&= \sum_{i=1}^{n} W_{ni}(x)(\varepsilon_{i}^{(n)} - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+M})) \\
&\quad + \sum_{i=1}^{n} W_{ni}(x)(E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+M}) - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+M-1})) + \cdots \\
&\quad + \sum_{i=1}^{n} W_{ni}(x)(E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i-M+1}) - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i-M})) \\
&\quad + \sum_{i=1}^{n} \mathcal{W}_{ni}(x) E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i-M}) \\
&= \sum_{i=1}^{n} W_{ni}(x)(\varepsilon_{i}^{(n)} - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+M})) \\
&\quad + \sum_{m=-M+1}^{M+1} \sum_{i=1}^{n} W_{ni}(x)(E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+m}) - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+m-1}))) \\
&\quad + \sum_{i=1}^{n} \mathcal{W}_{ni}(x) E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i-M}) \\
&= \sum_{i=1}^{n} W_{ni}(x)(\varepsilon_{i}^{(n)} - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+M})) \\
&\quad + \sum_{m=-M+1}^{M+1} \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{im}^{(n)} + \sum_{i=1}^{n} \mathcal{W}_{ni}(x) E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i-M}),
\end{align*}
\]

where \( \varepsilon_{im}^{(n)} = E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+m}) - E(\varepsilon_{i}^{(n)} | \mathcal{F}_{n,i+m-1}) \) for \( i = 1, 2, \ldots, m = \ldots, 0, 1, \ldots \). Then \( \{\varepsilon_{im}^{(n)}, \mathcal{F}_{n,i+m}; i \leq n\} \) is a martingale difference sequence for each \( m \) and \( n \). Moreover, the uniform integrability of \( \{|\varepsilon_{i}^{(n)}|^q\} \) implies that of \( \{|\varepsilon_{im}^{(n)}|^q\} \) for fixed \( m \). Hence, Lemma 2.2 implies

\[
E \left| \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{im}^{(n)} \right|^q \rightarrow 0 \quad as \quad n \rightarrow \infty \quad for \quad m = \ldots, 0, 1, \ldots
\]

Now, for any fixed \( \varepsilon > 0 \), there exists a sufficiently large number \( M > 0 \) such that

\[
\psi_{M} \leq \frac{(\varepsilon/2)^{1/q}}{3B[\sup_{i,n} C_{n}]}.
\]

For this fixed \( M \), \( \exists N > 0 \), such that \( n > N \) implies

\[
E \left| \sum_{i=1}^{n} W_{ni}(x) \varepsilon_{im}^{(n)} \right|^q \leq \frac{\varepsilon}{2(2M)^q 3^{q-1}}.
\]
From Eq. (3), by taking the qth moment on both sides, we obtain

\[ E \left| g_n(x) - E g_n(x) \right|^q \]
\[ \leq 3^{q-1} \left[ \left( \sum_{i=1}^{n} W_{ni}(x) (e_i^{(n)} - E(e_i^{(n)} | \mathcal{F}_{n,i+M})) \right)^q \right. \]
\[ + \left. E \left( \sum_{m=-M+1}^{M+1} \sum_{i=1}^{n} W_{ni}(x) e_{im}^{(n)} \right)^q \right] \]
\[ + E \left( \sum_{i=1}^{n} W_{ni}(x) E(e_i^{(n)} | \mathcal{F}_{n,i-M}) \right)^q \]
\[ \leq 3^{q-1} \left[ \left( \sum_{i=1}^{n} |W_{ni}(x)| \|e_i^{(n)} - E(e_i^{(n)} | \mathcal{F}_{n,i+M})\|_q \right)^q \right. \]
\[ + \left. \left( \sum_{m=-M+1}^{M+1} \left( E \left( \sum_{i=1}^{n} W_{ni}(x) e_{im}^{(n)} \right)^{1/q} \right)^q \right)^q \right] \]
\[ + \left( \sum_{i=1}^{n} |W_{ni}(x)| \|E(e_i^{(n)} | \mathcal{F}_{n,i-M})\|_q \right)^q \]
\[ \leq 3^{q-1} \left[ \left( \sum_{i=1}^{n} |W_{ni}(x)| C_{ni} \right)^q \psi_{M+1}^q \right. \]
\[ + \left. \left( \sum_{m=-M+1}^{M+1} \frac{(\varepsilon/2)^{1/q}}{(2M)^{1-1/q}} \right)^q \right] \]
\[ + \left( \sum_{i=1}^{n} |W_{ni}(x)| C_{ni} \right)^q \psi_M^q \]
\[ \leq \left( \sup_{i,n} C_{ni} \right) B \psi_{M+1}^q + \psi_M^q \leq 3 \sup_{i,n} C_{ni} B \psi_{M}^q \]
\[ \leq q \psi_{M}^q + \varepsilon/2 \]
\[ \leq \varepsilon \]

for \( n > N \), where the first inequality is obtained by using \( C_r \)-inequality and the second by applying Hölder's inequality for sums.

For the necessity part, see Müller [16].

Note the idea of splitting \( g_n(x) - E g_n(x) \) into three terms in Eq. (3) is borrowed from Andrews [3].

Remark 2.2. Theorem 2.3 extends Theorem 1 in Müller [16] to the model with dependent and heterogeneous observations.

(a) The mixingale assumption imposed on \( \{e_i^{(n)}\} \) in Theorem 2.3 allows them to have considerable temporal dependence as well as heterogeneity. For example, \( \alpha(\cdot) \), \( \rho(\cdot) \), and \( \phi(\cdot) \) mixing conditions, commonly adopted in nonparametric time series literature (See, for example,
Roussas [18]. Although the observations in Roussas and elsewhere in the literature are assumed to be stationary, the mixing conditions themselves do not impose stationarity on the data.), are special cases of mixingale.

(b) The uniform integrability of \( \{ |e_i^{(n)}| \} \) collapses to the finiteness of the \( q \)th absolute moment of \( \varepsilon \) being assumed by Müller when \( e_i^{(n)} \)'s are identically distributed with \( \varepsilon \).

(c) Assumptions imposed on the weight functions \( W_{ni}(x) \) in Theorem 2.3 are exactly those used by Müller [16].

The weak consistency of \( g_n(x) \) is obtained as a corollary of Theorem 2.3 by letting \( q = 1 \) which extends Theorem 2 in Georgiev [12] to the model with dependent observations.

**Corollary 2.4.** Let the assumptions of Theorem 2.3 hold with \( q = 1 \), in addition \( \{ e_i^{(n)} \} \) is uniformly integrable. If \( \sup_{i,n} C_{ni} < \infty \), then, \( g_n(x) \rightarrow g(x) \) in probability as \( n \rightarrow \infty \) at every continuity point \( x \in A \) of the function \( g \).

**Proof.** Follows from the result of Theorem 2.3 for \( q = 1 \) and Markov inequality.

**Remark 2.3.** Remark 2.2(a) and (c) also apply to the comparison of Corollary 2.4 with Theorem 2 of Georgiev. In addition, the uniform integrability of \( \{ e_i^{(n)} \} \) is weaker than the uniform boundedness of \( \{ e_i^{(n)} \} \) which is assumed by Georgiev in his Theorem 2, since the latter assumption implies the former (see Billingsley [4]).

**Example 2.1.** Consider the multidimensional version of the Priestley and Chao estimate introduced by Ahmad and Lin [2], i.e., the estimate (2) with weights

\[
\tilde{W}_{ni}(x) = K \left( \frac{x - x_i}{a_n} \right) \frac{A(A_i)}{a_n^p},
\]

where \( A_1, \ldots, A_n \) is partition of \( A = [0, 1]^p \) into \( n \) regions such that the volume \( A(A_i) \) is of order \( n^{-1} \), \( K(u) \) is a known \( p \)-dimensional bounded density, \( a_n \) is a sequence of reals converging to zero as \( n \rightarrow \infty \), and \( x_i \in A_i \). Georgiev concluded in his paper that his Theorem 2 extends and improves the weak consistency results of the above estimator given by Ahmad and Lin (see Georgiev [12]) in the sense that Theorem 2 in Georgiev allows \( e_i^{(n)} \)'s to be heterogeneous which is not the case in Ahmad and Lin and the best result for the sequence \( a_n \) allowed by Theorem 2 in Georgiev is \( na_n^p \rightarrow \infty \) while in Ahmad and Lin is \( na_n^{2p} \rightarrow \infty \) as \( n \rightarrow \infty \).
In view of Remarks 2.2 and 2.3, Corollary 2.4 given in this paper extends the weak consistency result of the above estimator improved by Georgiev to the regression model with both dependent and heterogeneous observations. Further, Theorem 2.3 implies that the array of kernel weights \( \hat{W}_n(x) \) is fixed design universally consistent for \( x \in A \) which extends the result of Corollary 2 in Müller [16]. Same conclusions can be made for the multivariate weights

\[
\hat{W}_n(x) = \frac{1}{a_n} \int_{A_i} K \left( \frac{x - u}{a_n} \right) du
\]  


Note if we let \( q = 2 \) in Theorem 2.3, the universal consistency of the weights \( W_n(x) \) implies mean square error consistency of \( g_n(x) \). However, for mean square error consistency, the uniform integrability assumption in Theorem 2.3 can be relaxed by a slightly different proof which is given below.

**Theorem 2.5.** Suppose assumptions (A1) with \( q = 2 \), (A2) and (A3) hold. In addition \( \epsilon_i^{(n)} \)'s are \( L^2 \)-bounded with \( C_n = \sup_{i,n} \| \epsilon_i^{(n)} \|_2 = C \) for all \( i,n \) and the \( L^2 \)-mixingale numbers \( \{ \psi_m \} \) satisfy \( \sum_{m=1}^\infty \psi_m < \infty \). Then, \( E[g_n(x) - g(x)]^2 \to 0 \) as \( n \to \infty \) at every continuity point \( x \in A \) of the function \( g \).

**Proof.** Note \( E[g_n(x) - g(x)]^2 = \text{Var}[g_n(x)] + [Eg_n(x) - g(x)]^2 \). It suffices to show \( \text{Var}[g_n(x)] \to 0 \) as \( n \to \infty \). Observe

\[
\text{Var}[g_n(x)] = E \left[ \sum_{i=1}^n W_{ni}(x) \epsilon_i^{(n)} \right]^2
\]

\[
= \sum_{i=1}^n W_{ni}^2(x) E[\epsilon_i^{(n)}]^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n W_{ni}(x) W_{nj}(x) E(\epsilon_i^{(n)} \epsilon_j^{(n)})
\]

\[
\leq C^2/n_x + 2 \sum_{j=1}^{n-1} \sum_{i=j+1}^n W_{ni}(x) W_{nj}(x) E(\epsilon_i^{(n)} \epsilon_j^{(n)})
\]  

and

\[
|E(\epsilon_i^{(n)} \epsilon_j^{(n)})| \leq 2C\psi_{(i-j)/2}.
\]

It follows that
\[
\text{Var}[g_n(x)] \leq \frac{C^2}{n_x} + 4C \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} |W_{ni}(x) W_{nj}(x)| \psi_{[(i-j)/2]}
\]
\[
\leq \frac{C^2}{n_x} + 4C \sup_{i \leq n} |W_{ni}(x)| \sum_{j=1}^{n-1} |W_{nj}(x)| \psi_{[(i-j)/2]}
\]
\[
\leq \frac{C^2}{n_x} + 4CB \sup_{i \leq n} |W_{ni}(x)| \left(\sum_{i=1}^{[n/2]} \psi_i\right)
\]
\[
\to 0
\] (9)

as \(n \to \infty\), since (A2)(b) and (A3) imply \(n_x \to \infty\) as \(n \to \infty\).

Remark 2.4. (a) It is obvious from inequality (9) that Theorem 2.5 still holds true if we replace assumption \(\sum_{m=1}^{\infty} \psi_m < \infty\) by (10), i.e.,
\[
\sup_{i \leq n} |W_{ni}(x)| \left[\sum_{i=1}^{n} \psi_i\right] \to 0 \quad \text{as} \quad n \to \infty.
\] (10)

(b) When \(\varepsilon_i^{(n)}\)'s are independent r.v.'s, condition (10) holds automatically, since \(\psi_i = 0\) for \(i \geq 1\). Hence, Theorem 2.5 is a natural generalization of Theorem 3 in Georgiev (1988) to the model with dependent observations.

Example 2.2. Consider the \(k\)-NN estimator given in Mack [14] with weights
\[
\hat{W}_{ni}(x) = K\left(\frac{x - x_i}{R_n}\right) / \sum_{i=1}^{n} K\left(\frac{x - x_i}{R_n}\right), \quad i = 1, 2, ..., n
\] (11)

where \(K(\cdot)\) is a bounded, nonnegative weight function satisfying \(K(u) = 0\), for \(\|u\| \geq 1\). \(R_n\) is the Euclidean distance between \(x\) and its \(k\)th nearest neighbor, and \(k = k_n \to \infty, k_n/n \to 0\) as \(n \to \infty\). The fixed design points \(x_1, ..., x_n\) are the same as those in Example 2.1.

It can be shown that \(R_n \to 0\), and \(nR_n^p \to \infty\), as \(n \to \infty\). Therefore, \(R_n\) in \(k\)-NN estimator plays the same role as \(h_n\) in Nadaraya–Watson estimator discussed in Georgiev [13]. Assumptions (A2) and (A3) can be verified for \(\hat{W}_{ni}(x)\) by using Lemma 2 of Georgiev along the same line as that of the proof of Theorem 1 in Georgiev [13].

We can conclude from the results obtained so far in this section that the \(k\)-NN estimator is weak and mean square error consistent upon the satisfaction of the other conditions in Corollary 2.4 and Theorem 2.5, respectively. Further, the array of nearest neighbor weights \((\hat{W}_{ni}(x))\) is fixed design universally consistent, provided the remaining conditions in Theorem 2.3 are satisfied.
The final result of this section gives the asymptotic variance of $g_n(x)$.

**Theorem 2.6.** Suppose assumptions (A1) with $q = 2$, (A2)(b), and (A3) hold. In addition, $\{\varepsilon_i^{(n)}\}$ is $L^2$-bounded with $C_{ni} = \sup_{i, n} \| \varepsilon_i^{(n)} \|_2 = C$ for all $i, n$ and the $L^2$-mixingale numbers $\{\psi_m\}$ satisfy $\sum_{m=1}^{\infty} \psi_m < \infty$. If the following conditions are satisfied,

(A4) (a) $\sup_{n} n_x \sup_{i \leq n} |W_{ni}(x)| < \infty$;
(b) $\sqrt{n_x} \sup_{i \leq n} |W_{ni}(x)| = O(n^{-\alpha/2})$ for some $1 > \alpha > 0$;
(c) $\sum_{i=1}^{n} W_{ni}^2(x) E |\varepsilon_i^{(n)}|^2 = \sigma_0^2/n_x + o(1/n_x)$ for some $\sigma_0^2 > 0$;

then, $n_x \text{Var}[g_n(x)] \to \sigma_0^2$ as $n \to \infty$.

**Proof.** Note

\[
n_x \text{Var}[g_n(x)] = n_x \left[ \sum_{i=1}^{n} W_{ni}^2(x) E |\varepsilon_i^{(n)}|^2 \right]
\]
\[
+ 2n_x \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} W_{ni}(x) W_{nj}(x) E(\varepsilon_i^{(n)}\varepsilon_j^{(n)})
\]
\[
= \sigma_0^2 + o(1) + 2n_x \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} W_{ni}(x) W_{nj}(x) E(\varepsilon_i^{(n)}\varepsilon_j^{(n)})
\]
\[
= \sigma_0^2 + o(1) + 2I,
\]
where the definition of $I$ is obvious from Eq. (12). Then, it is sufficient to show $I \to 0$ as $n \to \infty$.

For $N_n < n - 2$ to be chosen,

\[
I = n_x \sum_{j=1}^{N_n} \sum_{i=j+1}^{n} W_{ni}(x) W_{nj}(x) E(\varepsilon_i^{(n)}\varepsilon_j^{(n)})
\]
\[
+ n_x \sum_{j=1}^{n-1} \sum_{i=j+1}^{N_n} W_{ni}(x) W_{nj}(x) E(\varepsilon_i^{(n)}\varepsilon_j^{(n)})
\]
\[
= I_1 + I_2.
\]

It follows from Eq. (8) that

\[
|I_2| \leq n_x \sum_{j=N_n+1}^{n-1} \sum_{i=j+1}^{n} |W_{ni}(x)| |W_{nj}(x)| 2C \psi_{(i-j/2)}^{(n/2)}
\]
\[
\leq \left[ n_x \sup_{i \leq n} |W_{ni}(x)| \right] \sum_{j=N_n+1}^{n-1} |W_{nj}(x)| 2C \left( \sum_{i=1}^{n/2} \psi_i \right)
\]
which tends to zero as $n \to \infty$, provided that $N_n \to \infty$ as $n \to \infty$, by (A4)(a), (A2)(b), and the assumption that $\sum_{m=1}^{\infty} \psi_m < \infty$. 

Similarly,

\[ |I_i| \leq \sum_{j=1}^{N_n} \sum_{i=j+1}^{j+N_n} |W_{ni}(x)| |W_{nj}(x)| E |(e_i^{(n)}e_j^{(n)})| \]

\[ + n_x \sum_{j=1}^{N_n} \sum_{i=j+N_n}^{n} |W_{ni}(x)| |W_{nj}(x)| E |(e_i^{(n)}e_j^{(n)})| \]

\[ \leq C^2 n_x \left[ \sup_{i \leq n} |W_{ni}(x)| \right]^2 N_n^2 \]

\[ + 2 C n_x \sup_{i \leq n} |W_{ni}(x)| \left[ \sum_{j=1}^{N_n} |W_{nj}(x)| \right] \left[ \sum_{i=\lceil N_n/2 \rceil}^{\lceil n/2 \rceil} \psi_i \right] \]

\[ \leq C^2 \left[ \sqrt{n_x} \sup_{i \leq n} |W_{ni}(x)| \right]^2 N_n^2 \]

\[ + 2 C B \left[ n_x \sup_{i \leq n} |W_{ni}(x)| \right] \left[ \sum_{i=\lceil N_n/2 \rceil}^{\lceil n/2 \rceil} \psi_i \right] \]  

(14)

which tends to zero as \( n \to \infty \), if we take \( N_n = o(n^{1/2}) \), by (A4)(a), (b), and the assumption that \( \sum_{m=1}^{\infty} \psi_m < \infty \).

**Remark 2.5.** If \( e_i^{(n)} \) are identically distributed with variance \( \sigma_0^2 \), then,

\[ \sum_{i=1}^{n} W_{ni}(x) E |e_i^{(n)}|^2 = \sigma_0^2/n_x. \]

Therefore, (A4)(c) is satisfied and \( n_x \) can be regarded as the effective sample size at \( x \). (A4)(a) requires that no observation should be a very large multiple of \( 1/n_x \).

**Example 2.3.** Suppose \( e_i^{(n)} \) are identically distributed with variance \( \sigma_0^2 \) and the weights given by

(i) \( W_{ni}(x) \) with \( p = 1 \) defined in Eq. (6). Then it can be shown

\[ n_x = na_n/\int_0^1 K^2(x) \, dx + o(na_n) \]

provided \( na_n \to \infty \) as \( n \to \infty \). Theorem 2.6 shows that, for the particular weights \( W_{ni}(x) \),

\[ na_n \text{Var} [\hat{g}_n(x)] \to \sigma_0^2 \int_0^1 K^2(x) \, dx. \]

This is a well-known result given elsewhere in the literature. See, for example, Gasser and Müller [7].
(ii) \( K(u) = I_{[|u| \leq 1]} \) with \( p = 1 \) in Eq. (11), i.e., \( W_{ni}(x) = I_{[|x - x_i| \leq R_n]} \). Hence,
\[
\sum_{i=1}^{n} W_{ni}^2(x) = \sum_{i=1}^{n} I_{[|x - x_i| \leq R_n]} \frac{1}{k_n^2} = \frac{1}{k_n}.
\]
Theorem 2.6 implies
\[
k_n \text{Var}[g_n(x)] \rightarrow \sigma^2_0.
\]

3. Asymptotic Normality

In order to derive the asymptotic normality of the estimator, \( g_n(x) \), we restrict \( \epsilon_{i}^{(n)} \)'s to be near epoch dependent with respect to certain mixing sequences (see (A5)). The result is given by Georgiev (1988) for independent \( \epsilon_{i}^{(n)} \)'s.

**Theorem 3.1.** Suppose assumptions (A1)(b), (A2)(b), (A3), and (A4) with \( 1 > \alpha > \frac{3}{2} \) hold. In addition, the following conditions are satisfied:

- (A5) (a) for some \( r > 2 \), \( \sup_{i,n} \| \epsilon_{i}^{(n)} \|_r < \infty \);
- (b) \( \{ \epsilon_{i}^{(n)} \} \) is near epoch dependent with respect to \( \{ Y_{ni} \} \) with \( \{ \mu_m \} \) of size \(- (2 - \alpha) / (2 \alpha - 1) \) and constants \( d_{ni} = 1 \);
- (c) \( \{ Y_{ni} \} \) is mixing with either \( \{ \phi_{m} \} \) of size \(- (2 - \alpha) r / (2 \alpha - 1) (r - 1) \) or \( \{ \gamma_{m} \} \) of size \(- 2(2 - \alpha) r / (2 \alpha - 1) (r - 2) \) for \( r > 2 \).

Then, \( (g_n(x) - Eg_n(x)) / \sqrt{\text{Var}[g_n(x)]} \rightarrow N(0, 1) \) in distribution as \( n \rightarrow \infty \).

The proof of Theorem 3.1 parallels that of Theorem 3.13 in Wooldridge [20] and is based on a series of lemmas given in Wooldridge, two of which are restated here in order to make the idea of the proof clear.

**Lemma 3.2.** Suppose that \( E(X_{ni}^2) < \infty \), \( E(X_{ni}) = 0 \), \( E(S_{n}^2) \rightarrow 1 \), \( S_n \), \( S''_n \) and \( S'''_n \) are given as in Eq. (15) and (16), respectively, and there exist sequences \( \{ p_n \} \), \( \{ q_n \} \), and \( \{ k_n \} \) such that

- (a) \( E(S''_n^2) \rightarrow 0 \) as \( n \rightarrow \infty \);
- (b) \( \sum_{i=0}^{k_{n} - 2} \sum_{j=1}^{k_{n} - 1} E(\xi_{n,i}^j \xi_{n,j}^1) \rightarrow 0 \);
- (c) \( k_n \rightarrow \infty \);
- (d) \( |E(\exp(iuS_n') - \prod_{i=0}^{k_{n} - 1} E(\exp(iu\xi_{n,i})))| \rightarrow 0 \) for all \( u \in R \);
- (e) For each \( \varepsilon > 0 \),
\[
\sum_{i=0}^{k_{n} - 1} E(I_{[|x| > \varepsilon]} \xi_{n,i}^2) \rightarrow 0;
\]
that is, the Lindeberg condition holds for $\{\xi_{nj}\}$. Then

$$S_n \to N(0, 1).$$

**Lemma 3.3.** Let $\{\xi_{nt}, t = 1, \ldots, k_n; n = 1, \ldots\}$ be a triangular array of r.v.'s on $(\Omega, \mathcal{F}, P)$ with $k_n \to \infty$. Suppose there exist constants $\{\beta_{nt} > 0\}$ such that

1. $\lim_{n \to \infty} \sum_{t=1}^{k_n} \beta_{nt}^2 < \infty$;
2. $\lim_{n \to \infty} \max_{1 \leq t \leq k_n} \beta_{nt}^2 = 0$;
3. $\{\xi_{nt}/\beta_{nt}^2: t = 1, \ldots, k_n; n = 1, 2, \ldots\}$ is uniformly integrable. Then $\{\xi_{nt}\}$ satisfies the Lindeberg condition.

**Proof of Theorem 3.1.** As in the proof of Proposition 2.9 in Wooldridge and White [21], $\{\varphi_t^{(n)}, \mathcal{F}_t^{(n)}\}$ is an $L^2$-mixingale with $\psi_t = 2\varphi^{1/2} + 2\mu^{1/2}$ or $\psi_t = 5\varphi^{1/2} - 2\mu^{1/2}$ of size $-(2 - \alpha)/(2x - 1) < -1$ by (A5) which implies $\sum_{i=1}^{\infty} \psi_i < \infty$. Theorem 2.6 leads to the conclusion that $n_x \text{Var} [g_n(x)] = \sigma_0^2 + o(1)$. Therefore, $(g_n(x) - E_g(x))/\sqrt{\text{Var} [g_n(x)]}$ has the same asymptotic distribution as $(g_n(x) - E_g(x))/(\sigma_0/\sqrt{n_x})$. The rest of the proof is then devoted to the asymptotic normality of $(g_n(x) - E_g(x))/(\sigma_0/\sqrt{n_x}) = S_n$, say.

Without loss of generality, assume $\sigma_0 = 1$. Then

$$S_n = \sqrt{n_x} \sum_{i=1}^{n} W_{ni}(x) \varphi_i^{(n)} = \sum_{j=1}^{n} X_{ni},$$

where $X_{ni} = \sqrt{n_x} W_{ni}(x) \varphi_i^{(n)}$ and it is an $L^2$-mixingale with respect to $\{\mathcal{F}_i\}$ with $\psi_i$ as its mixingale number and $C_{ni} \propto \sqrt{n_x} |W_{ni}(x)|$ (here $\propto$ denotes proportionality).

The proof below is a slight modification of Theorem 3.13 in Wooldridge. We only supply its sketch.

For nonnegative integer functions $p_n, q_n,$ and $k_n,$ let

$$S_n' = \sum_{j=0}^{k_n-1} \xi_{nj}, \quad S_n'' = \sum_{j=0}^{k_n} \eta_{nj},$$

where

$$\xi_{nj} = \sum_{i = j(p_n + q_n) + 1}^{(j + 1)p_n + jq_n} X_{ni}, \quad 0 \leq j \leq k_n - 1$$

$$\eta_{nj} = \sum_{i = (j + 1)p_n + jq_n + 1}^{(j + 1)p_n + jq_n + 1} X_{ni}, \quad 0 \leq j \leq k_n - 1$$

$$\eta_{nk_n} = \sum_{i = k_n(p_n + q_n) + 1}^{n} X_{ni}.$$
Note that $S_n = S'_n + S''_n$, each $\xi_{nj}$ contains $p_n$ terms (the big blocks), $\eta_{nj}$ contains $q_n$ terms for $0 \leq j \leq k_n - 1$, and $\eta_{nk}$ contains at most $p_n + q_n + 1$ terms (the small blocks).

First, we verify the conditions in Lemma 3.3 for $\xi_{nj}$ defined above. Take $\beta_{nj} = \sum_{i=0}^{j+1} C_{ni}$ and note that $\sum_{j=1}^{k_n - 1} \beta_{nj} = \sum_{j=1}^{k_n - 1} C_{nj} \propto 1$ and $\beta_{nj} \lesssim p_n \sup_{i \leq n} C_{ni} = p_n m_n^*$, say, where

$$m_n^2 = n \sup_{i \leq n} |W_n(x)|^2 = O(n^{-2}) \quad \frac{3}{2} < \alpha < 1. \quad (18)$$

Since $X_n^2/C_{nj}^2 = \delta^{(n)}_j$ is uniformly integrable by (A5)(a), the same argument as in Wooldridge implies that $\{\xi_{nj}/\beta_{nj}\}$ is uniformly integrable. Therefore, (e) holds if $p_n m_n^2 \to 0$ as $n \to \infty$.

Now we are able to verify (a) to (e) in Lemma 3.2 by using the bounds for them derived in Wooldridge. What we need is to find $p_n$, $q_n$, and $k_n$ with $k_n \to \infty$ such that

$$k_n q_n m_n^2 + (p_n + q_n) m_n^2 + k_n^2 q_n^2 m_n^2 \psi_{[q_n/2]} + q_n m_n^2 + k_n^2 p_n^2 \psi_{[q_n/2]}$$

$$+ p_n m_n^2 + k_n \min(\sqrt{\psi_{[q_n/2]}}, \phi_{[q_n/2]})$$

$$+ k_n |u| p_n m_n \mu_{[q_n/4]} + k_n |u| m_n \to 0.$$ 

If $k_n = \lceil n/(p_n + q_n) \rceil$, then for $\phi$-mixing, the above sum has the same order as

$$q_n p_n^{-1} n^{1-\delta} [\delta^2 m_n^2] + (p_n + q_n) n^{-\delta} [\delta^2 m_n^2]$$

$$+ (q_n/p_n)^2 n^{-\delta} [\delta^2 m_n^2] \psi_{[q_n/2]} + [n^2 m_n^2] n^{2-\delta} \psi_{[q_n/2]}$$

$$+ p_n n^{-\delta} [\delta^2 m_n^2] + n p_n^{-1} \phi_{[q_n/2]}$$

$$+ n^{1-\alpha} [\delta^2 m_n^2] \mu_{[q_n/4]} + n^{1-\alpha/2} p_n^{-1} [\delta^2 m_n^2].$$

Since $n^2 m_n^2 = O(1)$, take $p_n = [n^{1-\delta/2}]$, $q_n = [n^{1-\delta}]$ for some $0 < \delta < 1$ to be chosen. Then it suffices to show

$$n^{2-\delta} \psi_{[q_n/2]} + n^{\delta/2} \phi_{[\mu_n/4]} + n^{1-\delta/2} \mu_{[q_n/4]} \to 0,$$

provided the following is true: $1 - \alpha < \delta/2$ and $\delta < \alpha$.

Now since $\psi_{i} = 2 \phi^{1-1/r}_{[q_n]} + \mu_{[q_n]}$ and by (A5) there exist $\theta > (2-\alpha)/r$ and (2x - 1)/(r - 1) > 1, $\gamma > (2-\alpha)/(2x-1)$ such that $\phi_{i} < i^{-\theta}$, $\mu_{i} < i^{-\gamma}$, so the term of interest is on the order of

$$n^{2-\delta} \phi_{[q_n/4]} + n^{\delta/2} \phi_{[\mu_n/4]} + n^{2-\delta} \mu_{[q_n/4]}$$

$$< n^{2-\delta} q_n^{-\theta(1-r)/r} + n^{2-\delta} q_n^{-\gamma} + n^{\delta/2} q_n^{-\theta}$$

$$= O(n^{2-\delta} q_n^{-\theta(1-r)/r} + n^{2-\delta} q_n^{-\gamma(1-\delta)} + n^{\delta/2} n^{-\theta(1-\delta)}). \quad (19)$$
Equation (19) holds if there exists $2(1-\alpha) < \delta < \alpha$ such that

$$\theta(1-\delta)(r-1)/r > 2-\alpha$$
$$\gamma(1-\delta)^2 > 2-\alpha$$

(20)
$$\theta(1-\delta) > \delta/2.$$  

For the above system of inequalities to hold, it is sufficient to choose $\delta$ such that it satisfies the following condition:

$$2(1-\alpha) < \delta < \min \left\{ \alpha, 1 - \frac{r(2-\alpha)}{\theta(r-1)}, 1 - \frac{2-\alpha}{\gamma}, \frac{\theta}{1/2 + \theta} \right\}.$$  

Such $\delta$ exists by the fact that $\theta > (2-\alpha)r/(2\alpha-1)(r-1)$, $\gamma > (2-\alpha)/(2\alpha-1)$, and $1 > \alpha > \frac{3}{2}$. This completes the proof for $\phi$-mixing by Lemma 3.2.

The same proof holds true for $\alpha$-mixing.

Example 3.1. As an example of the application of Theorem 3.1, consider the estimator $\hat{g}_n(x)$ with weights $\hat{W}_{ni}(x)$ as defined in Eq. (6) for $p = 1$.

Lemma 5 and Eq. (17) in Müller [16] imply

$$n_x \sup_{i < n} |\hat{W}_{ni}(x)| \leq O \left( \frac{1}{na_n} \right) = O(1)$$
$$\sqrt{n_x} \sup_{i < n} |\hat{W}_{ni}(x)| \leq O \left( \frac{1}{\sqrt{na_n}} \right) = O \left( \frac{1}{\sqrt{na_n}} \right).$$

Therefore, if $a_n \propto n^{-(1-\alpha)}$ for some $1 > \alpha > \frac{3}{2}$, then, assumption (A4) is satisfied for identically distributed disturbances. Consequently, as long as (A5) is satisfied, $\hat{g}_n(x)$ has asymptotic normal distribution.

The above example shows assumption (A4)(b) is not very restrictive in the sense that for kernel estimator $\hat{g}_n(x)$, the requirement for the bandwidth to be of order $n^{-(1-\alpha)}$ for some $1 > \alpha > \frac{3}{2}$ does not exclude the choice of optimal bandwidth in terms of minimizing MSE of $\hat{g}_n(x)$ which is $a_n^* \propto n^{-(2k+1)/(2k+1)}$ for $k \geq 2$ (see Gasser and Müller [8] for the derivation of $a_n^*$).

References


