Asymptotic properties of nonparametric M-estimation for mixing functional data

Jia Chen*, Lixin Zhang

Department of Mathematics, Zhejiang University, Yuquan Campus, Hangzhou 310027, China

Abstract

We investigate the asymptotic behavior of a nonparametric M-estimator of a regression function for stationary dependent processes, where the explanatory variables take values in some abstract functional space. Under some regularity conditions, we give the weak and strong consistency of the estimator as well as its asymptotic normality. We also give two examples of functional processes that satisfy the mixing conditions assumed in this paper. Furthermore, a simulated example is presented to examine the finite sample performance of the proposed estimator.

1. Introduction

The nonparametric regression techniques in modelling economic and financial time series data have gained a lot of attention during the last two decades. Both estimation and specification testing have been systematically examined for real-valued i.i.d. or weakly dependent stationary processes. See, for example, Robinson (1989), Härdle et al. (1997), Fan and Yao (2003) and the references therein.

In this paper, we consider the case of mixing functional data. Let \( \{X_i, Y_i\}_{i=1}^{\infty} \) be a stationary process, where \( Y_i \) are real-valued random variables and \( X_i \) take values in some abstract functional space \( H \) (e.g. Banach or Hilbert space) with norm \( \| \cdot \| \) and the deduced distance \( d(u, v) = \| u - v \|, u, v \in H \). In recent years, there is increasing interest in studying the limit theorems for functional data. For example, Araujo and Gine (1980) and Vashashnia et al. (1987) extended certain probability limit theorems to Banach-space-valued random variables. Ferraty and Vieu (2002) studied nonparametric regression estimation, and Ramsay and Silverman (2002) gave an introduction to the basic methods of functional data analysis and provided diverse case studies in several areas including criminology, economics and neurophysiology.

In this paper, we consider the following regression functional:

\[
m(x) = E[Y_1 | X_1 = x], \quad x \in H,
\]

© 2008 Elsevier B.V. All rights reserved.
assuming that $E|Y_1| < \infty$. Masry (2005) established the asymptotic normality of the Nadaraya–Watson (NW) kernel estimator $m_n^\circ(x)$ of $m(x)$, which is defined by
\begin{equation}
    m_n^\circ(x) = \frac{1}{n} \sum_{i=1}^n Y_i K\left(\frac{d(X_i, x)}{h_n}\right)
\end{equation}
where $K(\cdot)$ is a kernel function and $h_n$ is the bandwidth, which tends to zero as $n$ tends to infinity. The NW estimator $m_n^\circ(x)$ can be seen as a local least squares estimator, since it minimizes
\begin{equation}
    \sum_{i=1}^n K\left(\frac{d(X_i, x)}{h_n}\right)(Y_i - \theta)^2
\end{equation}
with respect to $\theta$.

Although NW estimators are central in nonparametric regression, it is known that they are not robust due to the fact that NW estimators can be considered as local least squares estimators and the least squares estimators are not robust. For instance, they are sensitive to outliers and do not perform well when the errors are heavy-tailed. However, outliers or aberrant observations are observed very often in econometrics and finance as well as in many other applied fields. A treatment of outliers is an important step in highlighting features of a data set. So in order to attenuate the lack of robustness of NW estimators, M-type regression estimators are natural candidates for achieving desirable robustness properties. In this paper, we investigate the asymptotic properties of the nonparametric M-estimator $m_n(x)$, which is implicitly defined as the solution to the following problem: find $\theta \in \mathbb{R}$ to minimize
\begin{equation}
    \sum_{i=1}^n K\left(\frac{d(X_i, x)}{h_n}\right) \rho(Y_i - \theta)
\end{equation}
or to satisfy the equation
\begin{equation}
    \sum_{i=1}^n K\left(\frac{d(X_i, x)}{h_n}\right) \psi(Y_i - \theta) = 0,
\end{equation}
where $\rho(\cdot)$ is a given outlier-resistant loss function defined in $\mathbb{R}$ and $\psi(\cdot)$ is its derivative. If the relevant function $\psi$ is $\psi(y) = y$, then the estimator $m_n(x)$ becomes the NW estimator $m_n^\circ(x)$. If $\psi(y) = \text{sign}(y)$, then the resulting estimator of (1.4) is a nonparametric least absolute distance estimator of $m(x)$. Many examples of choices of $\psi$ can be found in the book of Serfling (1980).

There is much literature on M-estimation for real-valued data. For example, Huber (1964) studied M-estimation of a location parameter as well as asymptotic normality of a nonparametric regression M-estimator for i.i.d. observations. Hardle (1989) considered the asymptotic maximal deviation of the nonparametric M-smoother $m_n(x)$ towards the regression curve $m(x)$, i.e. $\sup_{0 \leq x \leq 1} |m_n(x) - m(x)|$. Boente and Fraiman (1989) proposed nonparametric M-estimators for regression and autoregression of mixing processes. Beran (1991) studied M-estimation of a local parameter for Gaussian long-memory processes and Beran et al. (2003) investigated the behavior of a nonparametric kernel M-estimator for a fixed design model with long-memory errors.

To the best of our knowledge, however, nonparametric M-estimation has not been developed for functional data. The purpose of this paper is to investigate the nonparametric M-estimator $m_n(x)$, which is defined as the solution of Eq. (1.4), and to obtain its weak and strong consistency as well as its asymptotic normality under $\alpha$-mixing dependence assumptions. From the asymptotic normality, the asymptotic mean and variance of the error $m_n(x) - m(x)$ can be got.

The rest of the paper is organized as follows. In Section 2, we present the assumptions and main results. The derivations of the main results are given in Section 3. Section 4 includes some examples of functional processes that satisfy the mixing conditions in this paper as well as a simulated example. The simulation results show that the proposed M-estimator works well in practice. Some useful lemmas are given in the Appendix.

2. Assumptions and main results

A stationary process $(X_t, Y_t)_{t=1}^\infty$ is called $\alpha$-mixing if
\begin{align*}
    \alpha(n) &\to 0 \quad as \ n \to \infty,
\end{align*}
where
\begin{align*}
    \alpha(i) = \sup_{A \in \mathcal{F}, \mathcal{B} \subseteq \mathcal{F}_1} |P(AB) - P(A)P(B)|,
\end{align*}
where $\mathcal{F}_1$ is the $\sigma$-algebra generated by $(X_t, Y_t)_{t=1}^\infty$. $\alpha$-Mixing dependence is satisfied by many time series models (see Pham and Tran, 1985 for more details). Hence, it is assumed by many authors in nonparametric inference for real-valued time series,
say Roussas (1990), Masry and Fan (1997) and Cai (2002). For limit theorems for χ-mixing sequences, we refer to Lin and Lu (1996) and the references therein.

Define \( D_i = d(x, X_i) \) and \( m_i = Y_i - m(X_i) \). We first give the assumptions that are necessary in deriving the asymptotic properties of the nonparametric M-estimator \( m_n(x) \).

A1. \( K(\cdot) \) is a nonnegative and bounded kernel function with compact support, say \([0, 1]\).

A2. \( \rho(\cdot) \) is a convex function. Let \( \psi(\cdot) \) be the derivative of \( \rho(\cdot) \), then it is monotone and continuous almost everywhere.

A3. (i) \( F(u, x) := P(D_1 \leq u) = \rho(u f_1(x) \) as \( u \to 0 \), where \( \rho(0) = 0 \) and \( \rho(u) \) is absolutely continuous in a neighborhood of 0 and \( f_1(x) \) is some positive constant.

(ii) \( \sup_{x \in \mathcal{X}} P(D_1 \leq u) \leq \rho(u f_2(x)) \) as \( u \to 0 \), where \( \rho(u) \) \( \to 0 \) as \( u \to 0 \) and \( \rho(h_n) \cdot \rho^2(h_n) \) is bounded, \( f_2(\cdot) \) is a nonnegative function defined in \( \mathcal{X} \);

(iii) For \( i = 1, 2, h_i(h_n) := (h_n/\rho(h_n)) \int_0^1 K_s(v)\rho(s) \, dv \to K_i \) as \( n \to \infty \), where \( K_1 \) and \( K_2 \) are two positive constants.

A4. Uniformly for \( s \) in a neighborhood of \( s = m(x) \).

(i) There exists a function \( \lambda_1(\cdot) \) defined in \( \mathcal{X} \), such that as \( |u| \to 0 \),

\[
E[\psi(\epsilon_1 + u)|X_1 = z] = \lambda_1(z)u + o(u),
\]

where \( \lambda_1(z) \) is continuous in a neighborhood of \( x \) and \( \lambda_1(x) \neq 0 \);

(ii) \( E[\psi(Y_1 - s)|r^2 < \infty \) for some \( r > 2 \) and if we let \( g_1(z, s) = E[\psi(Y_1 - s)|X_1 = z] \) and \( g_2(z, s) = E[\psi(Y_1 - s)|X_1 = z] \) then \( g_1(z, s) \) and \( g_2(z, s) \) are continuous in a neighborhood of \( x \) with respect to \( z \). Furthermore, we have, for \( z \) in a neighborhood of \( x \),

\[
\max(E[\psi(\epsilon_1 + u) - \psi(\epsilon_1 - s)|X_1 = z], E[\psi(\epsilon_1 + u) - \psi(\epsilon_1)|X_1 = z]) \leq 2\lambda_2(u),
\]

where \( \lambda_2(u) \) is continuous at \( u = 0 \) with \( \lambda_2(0) = 0 \);

(iii) \( g_j(x_1, x_2; s) = E[\psi(Y_1 - s)|X_1 = x_1, X_2 = x_2, i \neq j, \) is uniformly bounded in a neighborhood of \( (x_1, x_2) = (x, x) \).

A5. \( |m(z) - m(x)| \leq C d^\beta(z, x) \) for \( z \) in a neighborhood of \( x \), where \( \beta \) is some positive constant.

A6. \( \sum_{i=1}^\infty (d(i))^{1-2/\beta} < \infty, \delta > 1 - 2/\beta \).

For simplicity, we will write \( h \) instead of \( h_n \) in the rest of the paper.

Remark 1. A1 is standard in nonparametric kernel estimation except that \( K(\cdot) \) is a one-sided kernel on \([0, 1]\). This is because \( K(\cdot) \) takes the form \( K(d(x, x_i)/h) \) in our paper and \( d(x, x_i) \geq 0 \). Therefore, we assume that \( K(\cdot) \) has a compact support \([0, 1]\) as Masry (2005). In A2, we impose some mild restrictions on \( \psi(\cdot) \). In Härdle (1984, 1989), \( \psi(\cdot) \) is assumed to be monotone, antisymmetric, bounded and twice continuously differentiable. With contrast to their assumption, A2 is much milder and covers some well-known special cases such as the least squares estimate (LSE), the least absolute distance estimate (LADE) and the mixed LSE and LADE. Assumptions A3, A5 and A6 are the same as Conditions 1, 3' and 4 in Masry (2005). Gasser et al. (1998) also used the condition \( F(u, x) = \rho(u f_1(x) \) as \( u \to 0 \), and refer to \( f_1(\cdot) \) as the functional probability density of \( X \). For an insight of A3, we take \( \mathcal{H} = \mathcal{H}(\mathcal{X}) \) as an example and assume that \( f_1(\cdot) \) is the probability density function of \( X_1 \). Then, \( F(u, x) = C_1 u^2 f_1(x) \) as \( u \to 0 \), where \( C_1 \) is the volume of the unit ball in \( \mathcal{H}(\mathcal{X}) \). Let \( f_2(x) = \sup_{i \neq j} f_j(x, x), \) where \( f_j(x, x) \) is the joint density of \( X_1 \) and \( X_2 \) at \( (x, x) \). Then,

\[
\sup_{i \neq j} P(D_1 \leq u, D_2 \leq u) \leq (C_1 u^2)^2 f_2(x) \) as \( u \to 0 \). So in this case, \( \rho(u) = C_1 u^2 \) and \( \rho(u) = O(\rho^2(u)) \) as \( u \to 0 \). Furthermore, by an elementary calculation, we know that \( l_i(h) = d(i)^{1/2} K(\cdot)^d - 1 \) \( dv \) for \( i = 1, 2 \), which is obviously finite. Masry (2005) also gave an alternative condition on \( F(u, x) \) (i.e. \( 0 < C_3 h^2 f_1(x) \leq F(h, x) \leq C_4 \rho(h f_1(x)) \) when calculating the variance of the proposed NW estimator. A4 imposes some uniform continuity assumptions on certain conditional moments of \( \psi(\epsilon_1 + \cdot) \) and A4 (i) and (ii) are similar to (M3) and (M4) in Bai et al. (1992). Assumption A6 concerns with the decaying rate of the mixing coefficient, which is related to the moment order \( r \) in A4.

Our main results are stated as follows.

**Theorem 1.** Suppose that \( m(x) \) is an isolated root of \( E[\psi(Y_1 - 0)|X_1 = x] = 0 \) and A1–A6 hold. Moreover, the bandwidth \( h \) satisfies \( h \to 0 \) and \( n \rho(h) \to \infty \). Then, we have

\[
m_n(x) \xrightarrow{p} m(x), \quad n \to \infty.
\]

(2.1)

**Theorem 2.** Assume that the assumptions of Theorem 1 hold. Besides, there exists a sequence of positive constants \( \{A_n\} \), such that \( A_n \to \infty \) and the quantities \( A_n, h \) and \( \alpha(\cdot) \) satisfy

\[
\sum_{n=1}^{\infty} A_n^{-1} < \infty, \quad \sum_{n=1}^{\infty} \left( \frac{A_n}{\alpha(h)} \right)^{1/2} 
\frac{n^{1/2} \alpha([n^{1/2}])}{A_n^2 \log^2 n} < \infty, \quad \frac{n \rho^2(h)}{A_n^2 \log^2 n} \to \infty.
\]

(2.2)
Then,
\[ m_n(x) \xrightarrow{a.s.} m(x), \quad n \to \infty. \]  

(2.3)

Remark 2. To gain an insight as to what kind of \( h \) and \( A_n \) satisfy the assumptions above, let us suppose that \( A_n = n^{1/2} \varphi(\ln n)^{-2} \), \( \varphi(l) = O(l^{-\delta}) \), where \( \delta > \max \left( \frac{1}{2}, \frac{\beta}{2} + 1 \right) \) and \( h \) is chosen such that \( \varphi(h) = O(n^{-1/2} \ln^{1/2} \ln n) \). Then, Corollary 2.1 and A6 hold.

Theorem 3. Assume that the conditions of Theorem 1 hold. Besides, there exists a sequence of positive integers \( \{v_n\} \), such that \( v_n \to \infty \), \( v_n = O((n \varphi(h))^{1/2}) \) and \( (n \varphi(h))^{1/2} v_n \to 0 \) as \( n \to \infty \). Then, for all \( x \) that satisfy \( \lambda_1(x) \neq 0 \), we have
\[ (n \varphi(h))^{1/2} (m_n(x) - m(x) - \mu_n) \xrightarrow{d} N(0, \sigma^2(x)), \]

(2.4)

where \( \mu_n = \frac{B_n}{\lambda_1(x)f_1(x)K_1} \), \( \sigma^2(x) = \frac{\sigma^2(m(x))}{(\lambda_1(x)f_1(x)K_1)^2} \),
\[ B_n = \frac{1}{\varphi(h)} E(A_1 \psi(Y_1 - m(x))) = O(h^\beta), \quad A_1 = \int K \left( \frac{d(x, X_i)}{h} \right) dF(u, x), \quad \sigma^2(s) = g_1(x, s)f_1(x)K_2 \]
and \( K_1, K_2 \) are defined in A3 and \( g_1(x, s) \) is defined in A4.

Remark 3. In Masry (2005), a smoothness condition on the regression function \( m(x) \) is also imposed, that is \( |m(u) - m(v)| \leq Cd^\beta (u, v) \) for all \( u, v \in \mathcal{H} \) for some \( \beta > 0 \), and the asymptotic normality of the NY type estimator \( m_n^*(x) \) is given as
\[ (n \varphi(h))^{1/2} (m_n^*(x) - m(x) - B_n^*) \xrightarrow{d} N(0, \sigma^*(x)), \]
where the bias term \( B_n^* = O(h^\beta) \), which is of the same order as \( \mu_n \) in (2.4).

3. Proofs of the main results

In order to avoid unnecessary repetitions, we suppose that all limits are taken as \( n \) tends to infinity except special notations and the limits in the proofs of Theorems 1 and 2 hold uniformly in a neighborhood of \( s = m(x) \).

Proof of Theorem 1. Define \( A_1 = K(d(x, X_i)/h) \), \( H_n(x, s) = \sum_{i=1}^{n} A_1 \psi(Y_i - s) \) and \( H(x, s) = E[\psi(Y_1 - s)|X_1 = x]f_1(x)K_1 \), where \( \lambda_1(x) \) is defined in A4. We first prove
\[ H_n(x, s) \xrightarrow{P} H(x, s) \]
and then prove (2.1) by the monotony and continuity of the function \( \psi(\cdot) \). We prove (3.1) by verifying the following two statements:
\[ H_n(x, s) - EH_n(x, s) \xrightarrow{P} 0 \]
and
\[ EH_n(x, s) \to H(x, s). \]

The validity of (3.3) is obvious by an elementary calculation. In fact, by A1, A3 and A4,
\[ EH_n(x, s) = \frac{1}{\varphi(h)} EA_1 \psi(Y_1 - s) \]
\[ = \frac{1}{\varphi(h)} E(A_1 E[\psi(Y_1 - s)|X_1]) \]
\[ = E[\psi(Y_1 - s)|X_1 = x][1 + o(1)] \frac{1}{\varphi(h)} EA_1 \]
\[ = E[\psi(Y_1 - s)|X_1 = x][1 + o(1)] \frac{1}{\varphi(h)} \int K \left( \frac{u}{h} \right) F(du, x) \]
\[ = E[\psi(Y_1 - s)|X_1 = x][1 + o(1)] \frac{h}{\varphi(h)} f_1(x) \int_0^1 K(v) \varphi'(hv) dv \]
\[ \to E[\psi(Y_1 - s)|X_1 = x]f_1(x)K_1 = H(x, s).\]
To prove (3.2), we first show
\[ n\varphi(h)\text{Var}[H_n(x, s)] \rightarrow \sigma^2(s), \] (3.5)
where \( \sigma^2(s) = g_1(x, s)f_1(x)K_2 \). Let \( Z_n,i(s) = A_1i\psi(Y_i - s) \), then
\[ H_n(x, s) = \frac{1}{n\varphi(h)} \sum_{i=1}^{n} Z_n,i(s) \]
and
\[ (n\varphi(h))^{1/2}(H_n(x, s) - EH_n(x, s)) = \frac{1}{\sqrt{n\varphi(h)}} \sum_{i=1}^{n}(Z_n,i(s) - EZ_n,i(s)). \] (3.6)
Notice that
\[ n\varphi(h)\text{Var}[H_n(x, s)] = \frac{1}{\varphi(h)}\text{Var}(Z_{n,1}) + \frac{1}{n\varphi(h)} \sum_{ij=1 \atop i \neq j}^{n} \text{Cov}(Z_{n,i}, Z_{n,j}). \] (3.7)
By (3.4) and the definition of \( Z_{n,1}(s) \), we have
\[ EZ_{n,1}(s) = O(\varphi(h)). \] (3.8)
By A5, we know that when \( d(X_1, x) \leq h, m(X_1) = m(x) + O(h^\beta) \). Hence,
\[
\text{Var}(Z_{n,1}) = E[A_1^2\psi(Y_1 - s)]^2 + O(\varphi^2(h)) \\
= E[A_1^2E[\psi^2(Y_1 - s)|X_1]] + O(\varphi^2(h)) \\
= g_1(x, s)(1 + o(1))h_1(x) \int_0^1 \kappa^2(v)\varphi'(hv)dv + O(\varphi^2(h)).
\]
This, in conjunction with A3(iii) and \( \varphi(h) \rightarrow 0 \), gives
\[ \frac{1}{\varphi(h)}\text{Var}(Z_{n,1}) \rightarrow \sigma^2(s). \] (3.9)
Next, we will show
\[ \sum_{ij=1 \atop i \neq j}^{n} |\text{Cov}(Z_{n,i}, Z_{n,j})| = o(n\varphi(h)). \] (3.10)
Let \( 0 < a_n < n \) be a sequence of positive integers that tends to infinity and its value will be specified later. Then,
\[ \sum_{ij=1 \atop i \neq j}^{n} |\text{Cov}(Z_{n,i}, Z_{n,j})| \leq \sum_{0<|i-j| \leq a_n} |\text{Cov}(Z_{n,i}, Z_{n,j})| + \sum_{|i-j|>a_n} |\text{Cov}(Z_{n,i}, Z_{n,j})| \\
=: J_1 + J_2. \] (3.11)
In view of (3.8), A1, A3(ii) and A4(iii), we have
\[
J_1 \leq \sum_{0<|i-j| \leq a_n} [|EZ_{n,i}Z_{n,j}| + (EZ_{n,1})^2] \\
= \sum_{0<|i-j| \leq a_n} [|Eg_{1d_j}(X_i, X_j; s)A_1A_j| + O(\varphi^2(h))] \\
\leq \sum_{0<|i-j| \leq a_n} |C|P(D_i \leq h, D_j \leq h) + O(\varphi^2(h))] \\
\leq Cn a_n[f_2(x)\varphi(h) + O(\varphi^2(h))],
\] (3.12)
where \( C \) denotes a positive constant that may take different values in different places.
As for the covariances in $J_2$, we employ A4 and Corollary A.2 (Davydov’s lemma) in Hall and Heyde (1980) and get
\[ |\text{Cov}(Z_{n,i}, Z_{n,j})| \leq 8|E|\psi(Y_1 - s)\Delta_1(t)^{2/r}|z((i - j))|^{1 - 2/r} \]
\[ \leq C|P(D_1 \leq h)|^{2/r}|z((i - j))|^{1 - 2/r} \]
\[ \leq C|f_1(x)\phi(h)|^{2/r}|z((i - j))|^{1 - 2/r}. \]

Hence,
\[
J_2 \leq C \sum_{|i - j| > a_n} |f_1(x)\phi(h)|^{2/r}|z((i - j))|^{1 - 2/r}
\leq Cn|f_1(x)\phi(h)|^{2/r} \sum_{l = a_n + 1}^{\infty} |z(l)|^{1 - 2/r}
\leq Cn|f_1(x)\phi(h)|^{2/r} \sum_{l = a_n + 1}^{\infty} \beta^l |z(l)|^{1 - 2/r}. \tag{3.13}
\]

Now we choose $a_n = [(\phi(h))^{-(1 - 2/r)/\delta}]$, then by (3.12) and (3.13), we have
\[ J_1 \leq Cn|\phi(h)|^{2 - (1 - 2/r)/\delta} \]
and
\[ J_2 \leq Cn|\phi(h)| \sum_{l = a_n + 1}^{\infty} \beta^l |z(l)|^{1 - 2/r}. \]

By A6, we have
\[ J_1 = o(n|\phi(h)|) \tag{3.14} \]
and
\[ J_2 = o(n|\phi(h)|). \tag{3.15} \]

Combining (3.11), (3.14) and (3.15), we get (3.10). By (3.7), (3.9) and (3.10), we know that (3.5) is true. Therefore,
\[ H_n(x, s) - EH_n(x, s) = O_p \left( \frac{1}{\sqrt{n|\phi(h)|}} \right) = o_p(1). \]

This implies that (3.2) holds. Thus we have proved (3.1).

The remaining proof of Theorem 1 is rather standard in the context of M-estimation and we omit it here. For details, see Huber (1964). □

**Proof of Theorem 2.** We first prove
\[ H_n(x, s) - EH_n(x, s) \overset{d}{\rightarrow} 0, \quad n \rightarrow \infty, \tag{3.16} \]
then the rest of the proof of $m_n(x) \overset{d}{\rightarrow} m(x)$ is similar to that of $m_n(x) \overset{p}{\rightarrow} m(x)$. To prove (3.16), we first employ a truncation method. Let $|A_n| \uparrow \infty$ be the sequence that satisfies (2.2) in Theorem 2 and denote
\[ Z_{n,i}^A = \psi^A(Y_i - s)A_i, \quad H_n^A(x, s) = \frac{1}{n|\phi(h)|} \sum_{i = 1}^{n} Z_{n,i}^A, \]
where $\psi^A(y) = \psi(y)I(|\psi(y)| \leq A_n)$. Then,
\[ |H_n(x, s) - EH_n(x, s)| \leq |H_n^A(x, s) - EH_n^A(x, s)| + |EH_n^A(x, s) - EH_n(x, s)| + |EH_n^A(x, s) - EH_n(x, s)|. \tag{3.17} \]

Since $\sum_{n=1}^{\infty} A_n^{-r} < \infty$, we have, for any $\varepsilon > 0$,
\[ \sum_{n=1}^{\infty} P(|\psi(Y_n - s)| > \varepsilon A_n) \leq \sum_{n=1}^{\infty} e^{-\varepsilon} E|\psi(Y_n - s)|A_n^{-r} < \infty. \]
By Theorem 4.2.1 in Galambos (1978), we know that
\[ P \left( \max_{1 \leq i \leq n} 1|\psi(Y_i - s)| > A_n \right) \text{ i.o.} = 0. \]

Therefore, for \( n \) sufficiently large,
\[ H_n(x, s) = H^A_n(x, s) \quad \text{a.s.} \tag{3.18} \]

Besides, by A3(iii), A4(ii), \( r > 2 \) and \( A_n \to \infty \), we have
\[ |E H_n^A(x, s) - E H_n(x, s)| \leq \frac{1}{\varphi(h)} [E|\psi(Y_1 - s)| K(|\psi(Y_1 - s)| > A_n) A_1] \]
\[ \leq A_n^{1-r} \frac{1}{\varphi(h)} [E|\psi(Y_1 - s)| K(|\psi(Y_1 - s)| > A_n) A_1] \]
\[ \leq A_n^{1-r} \frac{1}{\varphi(h)} [E|\psi(Y_1 - s)| A_1] \]
\[ = g_2(x, s) (1 + o(1)) A_n^{1-r} \frac{1}{h} E A_1 \]
\[ = g_2(x, s) (1 + o(1)) \frac{1}{h} \int_0^1 K(v) \phi'(hv) \, dv \to 0. \] \( \tag{3.19} \)

As \( |Z_{n, i}^A/\varphi(h)| \leq A_n/\varphi(h) \), we now apply Lemma 1 in the Appendix and get for each \( q = 1, \ldots, [n/2], \)
\[ P(|H^A_n - E H^A_n| > \sigma) \leq 4 \exp \left( -\frac{\sigma^2}{8v^2(q)} \right) + 22 \left( 1 + \frac{4A_n}{\varphi(h)} \right)^{1/2} q^2 \left( \left[ \frac{n}{2q} \right] \right), \] \( \tag{3.20} \)

where \( v^2(q) = 2\sigma^2(q)/p^2 + A_n/2p(h) \), \( p = [n/2q] \) and
\[ \sigma^2(q) = \max_{0 \leq j \leq 2q-1} \text{Var} \left( \frac{Z_{n, i}^A}{\varphi(h)} + \cdots + \frac{Z_{n, j+1}^A}{\varphi(h)} \right), \]

which, by (3.5), is bounded above by \( Cp/\varphi(h) \). If we choose \( q = [n^{1/2}] \), then (3.20) is bounded above by
\[ 4 \exp(-Cn^{1/2} \varphi(h)/A_n) + C \left( \frac{A_n}{\varphi(h)} \right)^{1/2} n^{1/2} \varphi([n^{1/2}]) \]

By \( n\varphi(h)/A_n \log^2 n \to \infty \) and \( \sum_{n=1}^\infty (A_n/\varphi(h))^{1/2} n^{1/2} \varphi([n^{1/2}]) \to \infty \), we know
\[ \sum_{n=1}^\infty P(|H^A_n - E H^A_n| > \sigma) < \infty, \]

which, by the Borel–Cantelli lemma, implies \( H^A_n - E H^A_n \to 0 \) a.s. Combining this with (3.17)–(3.19), we get (3.16). \( \square \)

**Remark 4.** Notice that if we choose \( q \gg n^{1/2} \) (i.e. \( q/n^{1/2} \to \infty \)), then by (3.20) we know that in order \( \sum_{n=1}^\infty P(|H^A_n - E H^A_n| > \sigma) < \infty \), we need more rigid conditions on the mixing coefficient \( \alpha(l) \) but weaker conditions on the bandwidth \( h \). For example, if we choose \( q = [n^{2/3}] \), then in order that the last term in (3.20) is summable, it suffices that \( n(\varphi(h))/A_n \log n \to \infty \) and \( \sum_{n=1}^\infty (A_n/\varphi(h))^{1/2} n^{1/2} \varphi([n^{1/2}]) \) \( < \infty \). Hence, there is a tradeoff between the bandwidth \( h \) and the mixing coefficient \( \alpha(l) \). In the proof of Theorem 2, we just choose \( q = [n^{1/2}] \).

**Proof of Theorem 3.** We first prove that for any \( c > 0 \),
\[ \sup_{(n\varphi(h))^{1/2} \mu - m(x) \leq c} \left| \sum_{i=1}^n A_i [\rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))] \right| \]
\[ = o_p(1). \]
As \( \rho(\cdot) \) is convex, we have
\[
\begin{align*}
& |A_i| \rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))| \\
& \leq |A_i| |s - m(x)| |\psi(Y_i - s) - \psi(Y_i - m(x))| \\
& = |A_i||s - m(x)||\psi(\epsilon_i + m(X_i) - s) - \psi(\epsilon_i + m(X_i) - m(x))| \\
& \leq |A_i||s - m(x)||\psi(\epsilon_i + m(X_i) - s) - \psi(\epsilon_i)| + |\psi(\epsilon_i + m(X_i) - m(x)) - \psi(\epsilon_i)||. 
\end{align*}
\]

(3.21)

By A5, we know that when \( (n\varphi(h))^{1/2}|s - m(x)| \leq c \) and \( d(x, Xi) \leq h \), we have
\[
|m(X_i) - s| \leq |m(X_i) - m(x)| + |m(x) - s| \leq Ch^\beta + (n\varphi(h))^{-1/2}c 
\]
and
\[
|m(X_i) - m(x)| \leq Ch^\beta. 
\]

(3.22)

(3.23)

Let \( \chi_{ni} = A_i|\psi(\epsilon_i + m(X_i) - m(x)) - \psi(\epsilon_i)| \), then by Theorem 17.2.2 in Ibragimov and Linnik (1971) and A6, we have
\[
\text{Var} \left( \sum_{i=1}^{n} \chi_{ni} \right) \leq nE\chi_{ni}^2 + Cn \sum_{i=2}^{n} |\text{Cov}(\chi_{ni}, \chi_{ni})| \\
\leq nE\chi_{ni}^2 + Cn \sum_{i=1}^{n} |\varphi(i)|^{1-2r}E|\chi_{ni}|^r |\chi_{ni}|^2 \\
\leq nE\chi_{ni}^2 + CnE|\chi_{ni}|^r |\chi_{ni}|^2. 
\]

(3.24)

Furthermore, by A3, A4 and (3.23), we have
\[
E\chi_{ni}^2 = E[E\chi_{ni}^2|X_1]| \leq C \epsilon[A^2_2|\varphi(\varphi(\varphi(h))^{-1/2})| \leq C \varphi(h)\lambda_2(h^\beta). 
\]

Similarly, we have
\[
E|\chi_{ni}|^r \leq C \varphi(h)\lambda_2(h^\beta). 
\]

Therefore, by (3.24), we know that
\[
\text{Var} \left( \sum_{i=1}^{n} \chi_{ni} \right) = O(n\varphi(h)\lambda_2(h^\beta)) = O(n\varphi(h)). 
\]

(3.25)

If we let \( \chi_{ni}' = |\psi(\epsilon_i + m(X_i) - s) - \psi(\epsilon_i)| \), then by (3.22), we can similarly prove
\[
\text{Var} \left( \sum_{i=1}^{n} \chi_{ni}' \right) = O(n\varphi(h)\lambda_2(h^\beta + (n\varphi(h))^{-1/2})) = O(n\varphi(h)). 
\]

(3.26)

From (3.21), (3.25) and (3.26), we know that the following equation holds uniformly for \( (n\varphi(h))^{1/2}|s - m(x)| \leq c, 
\[
\text{Var} \left( \sum_{i=1}^{n} \left[ A_i|\rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))| \right] \right) = O((s - m(x))^2 n\varphi(h)) = O(1). 
\]

This implies that
\[
\begin{align*}
& \sum_{i=1}^{n} A_i|\rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))| \\
& - E[A_i|\rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))|] \\
& = O(1) 
\end{align*}
\]

(3.27)

uniformly for \( (n\varphi(h))^{1/2}|s - m(x)| \leq c \). Moreover, by Lemma 1 in Bai et al. (1992), we have
\[
E[\rho(\epsilon_1 + u) - \rho(\epsilon_1)]|X_1| = \frac{1}{2}\lambda_1(z)u^2 + o(u^2) \quad \text{as } |u| \to 0. 
\]
By (3.22) and (3.23), we know that
\[
\sum_{i=1}^{n} E(A_i \mid \rho(Y_i - s) - \rho(Y_i - m(x)) = nE(A_1 \mid \rho(Y_1 - s) - \rho(Y_1 - m(x)))
\]
\[
= \frac{1}{2} nE(A_1 \hat{\lambda}_1(X_1)[(m(X_1) - s)^2 - (m(X_1) - m(x))^2)](1 + o(1))
\]
\[
= \frac{1}{2} (m(x) - s)nE(A_1 \hat{\lambda}_1(X_1)(2m(X_1) - s - m(x))[1 + o(1)]
\]
(3.28)
holds uniformly for \((n\sigma(h))^{1/2} |s - m(x)| \leq c\). On the other hand,
\[
(s - m(x)) \sum_{i=1}^{n} E[A_i \psi(Y_i - m(x))] = (s - m(x))nE(A_1 \hat{\lambda}_1(X_1)(m(X_1) - m(x)))(1 + o(1)).
\]
(3.29)
Combining (3.28) with (3.29), we get
\[
E \sum_{i=1}^{n} (A_i \mid \rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x)))
\]
\[
= \frac{1}{2} (m(x) - s)^2 nE(A_1 \hat{\lambda}_1(X_1))(1 + o(1))
\]
\[
= \frac{1}{2} (m(x) - s)^2 \hat{\lambda}_1(x)f_1(x)K_1 n\sigma(h)(1 + o(1)).
\]
(3.30)
By (3.27) and (3.30), we know that
\[
\left| \sum_{i=1}^{n} A_i \mid \rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x)) - \frac{1}{2} (m(x) - s)^2 \hat{\lambda}_1(x)f_1(x)K_1 n\sigma(h) \right| = o_p(1)
\]
holds uniformly for \((n\sigma(h))^{1/2} |s - m(x)| \leq c\). As \(\sum_{i=1}^{n} A_i \mid \rho(Y_i - s) - \rho(Y_i - m(x)) + \psi(Y_i - m(x))(s - m(x))\) is convex in \(s\) and \(\frac{1}{2} (m(x) - s)^2 \hat{\lambda}_1(x)f_1(x)K_1 n\sigma(h)\) is continuous and convex in \(s\), by Theorem 10.8 of Rockafellar (1970), we have
\[
\sup_{(n\sigma(h))^{1/2} |s - m(x)| \leq c} \left| \sum_{i=1}^{n} A_i \mid \psi(Y_i - s) - \psi(Y_i - m(x)) \right| = o_p(1).
\]
(3.31)
By arguments similar to the proof of (3.31) and by Theorem 2.5.7 of Rockafellar (1970), we have
\[
\sup_{(n\sigma(h))^{1/2} |s - m(x)| \leq c} \left| \frac{1}{\sqrt{n\sigma(h)}} \sum_{i=1}^{n} A_i \mid \psi(Y_i - s) - \psi(Y_i - m(x)) \right| + \sqrt{n\sigma(h)} \hat{\lambda}_1(x)f_1(x)K_1 (s - m(x)) = o_p(1).
\]
(3.32)
Denote \(\theta_n = m_n(x) - m(x)\) and
\[
\tilde{\theta}_n = \frac{1}{n\sigma(h)} \hat{\lambda}_1(x)f_1(x)K_1 \left( \sum_{i=1}^{n} A_i \psi(Y_i - m(x)) \right) = \hat{\lambda}_1(x)f_1(x)K_1 \left( H_n(x, m(x)) \right).
\]
Using the same technique as that used in the proof of Theorem 2.4 in Bai et al. (1992), we obtain
\[
\sqrt{n\sigma(h)}(\theta_n - \tilde{\theta}_n) = o_p(1).
\]
(3.33)
Hence, in order to prove Theorem 3, it remains for us to prove
\[
(n\sigma(h))^{1/2} (H_n(x, m(x)) - EH_n(x, m(x))) \overset{d}{\to} N(0, \sigma^2(m(x))),
\]
(3.34)
and
\[
EH_n(x, m(x)) = B_n = O(h^\beta).
\]
(3.35)
Eq. (3.35) is obvious. In fact

\[
|B_n| \leq \frac{1}{\phi(h)} E(A_1 E(\psi(Y_1 - m(x)))X_1))
\]

\[
= \frac{1}{\phi(h)} E(A_1 \lambda_1(X_1)(m(X_1) - m(x)))
\]

\[
\leq Ch^p \lambda_1(x)f_1(x) \frac{h}{\phi(h)} \int_0^1 K(u\phi'(hu))du
\]

\[
= O(h^p).
\]

Next, we will prove (3.34). Let \(\bar{Z}_{n,i}(m(x)) = (1/\sqrt{h})(Z_{n,i}(m(x)) - EZ_{n,i}(m(x)))\), then

\[
(n\phi(h))^{1/2}(H_n(x, m(x)) - EH_n(x, m(x))) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{Z}_{n,i}(m(x)).
\]

For simplicity, we will write \(\bar{Z}_{n,i} \) instead of \(\bar{Z}_{n,i}(m(x))\). It is easy to see that (3.34) is equivalent to

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{Z}_{n,i} \sim N(0, \sigma^2(m(x))).
\]  

(3.36)

To show this, we adopt the small- and large-block technique. Namely, partition \(\{1, \ldots, n\}\) into \(2k_n + 1\) subsets with small-block of size \(v = v_n\), which is defined in Theorem 3, and large-block of size \(u = u_n\) and \(k_n = [n(u_n + v_n)], \) where \([a]\) denotes the integer part of \(a\). For the choice of \(u_n\), note that by the conditions in Theorem 3, there exists a sequence of positive integers \(\{q_n\}\) such that \(q_n \to \infty, q_nv_n = O((n\phi(h))^{1/2})\) and \(q_n(n/\phi(h))^{1/2}x(v_n) \to 0\). Let \(u_n = ((n\phi(h))^{1/2}/q_n\), then

\[
\frac{v_n}{u_n} \to 0, \quad \frac{u_n}{n} \to 0, \quad \frac{u_n}{(n\phi(h))^{1/2}} \to 0, \quad \frac{n}{u_n} \to \infty.
\]  

(3.37)

Define

\[
\bar{z}_{j} = \sum_{i=(j+1)(u+v)+1}^{j(u+v)+u} \bar{Z}_{n,i}, \quad j = 0, \ldots, k_n - 1,
\]

\[
\bar{\eta}_j = \sum_{i=(j+1)(u+v)+u+1}^{j(u+v)+u+u} \bar{Z}_{n,i}, \quad j = 0, \ldots, k_n - 1
\]

and

\[
\bar{z}_{k_n} = \sum_{i=(k_n)(u+v)+1}^{n} \bar{Z}_{n,i}
\]

then

\[
\sum_{i=1}^n \bar{Z}_{n,i} = \sum_{j=0}^{k_n-1} \bar{z}_{j} + \sum_{j=0}^{k_n-1} \eta_j + \bar{z}_{k_n} =: Q_{n1} + Q_{n2} + Q_{n3}. \]

We will show (3.36) via

\[
\frac{1}{n} \left| E\exp(it\sqrt{n}Q_{n1}) - \prod_{j=0}^{k_n-1} E \exp(it\sqrt{n} \bar{z}_{j}) \right| \to 0.
\]  

(3.40)

\[
\frac{1}{n} \sum_{j=0}^{k_n-1} E\bar{z}_{j}^2 \to \sigma^2(m(x))
\]  

(3.41)

and

\[
\frac{1}{n} \left| E\bar{z}_{j}^2 1(|\bar{z}_j| \geq \varepsilon \sigma(m(x))\sqrt{n}) \right| \to 0
\]  

(3.42)

for any \(\varepsilon > 0\).
Let \( \eta_i \) and define
\[
\eta_i = \sum_{j=0}^{k_n-1} \text{Cov}(\eta_i, \eta_j)
\]
which tends to zero by (3.37).

It is obvious that (3.38)–(3.41) still hold when we substitute \( \eta_i \) by \( \eta_i + \epsilon \).

By (3.37), the proofs of (3.9) and (3.10) and the definition of \( k_n \), we know
\[
I_1 = k_n \left[ vE_{n,1}^2 + \sum_{i=1}^{v} \text{Cov}(\tilde{Z}_{n,i}, \tilde{Z}_{n,i}) \right]
\]
\[
= k_n \left[ v(\sigma^2(m(x)) + o(1)) + o(v) \right]
\]
\[
= k_n v(\sigma^2(m(x)) + o(1)) = o(n).
\]

Note that
\[
|l_2| \leq \sum_{i=0}^{k_n-1} \sum_{j=1}^{(i+1)(u+v)} \sum_{i=0}^{(j+1)(u+v)} |\text{Cov}(\tilde{Z}_{n,i}, \tilde{Z}_{n,j})|.
\]

For \( i_1 \in \{u + v + u + 1, u + v + u + 2, \ldots, (i + 1)(u + v)\} \) and \( i_2 \in \{u + v + u + 1, u + v + u + 2, \ldots, (j + 1)(u + v)\}, i \neq j \), it is obvious that \( |l_1 - l_2| \geq u \). Hence,
\[
|l_2| \leq \sum_{i=0}^{k_n-1} \sum_{i=0}^{(i+1)(u+v)} \sum_{j=0}^{(j+1)(u+v)} \sum_{i=0}^{(j+1)(u+v)+1} |\text{Cov}(\tilde{Z}_{n,i}, \tilde{Z}_{n,j})| = o(n).
\]

As for (3.38) and (3.39), we only prove (3.38), since the proof of (3.39) is similar. By stationarity,
\[
E \sum_{j=0}^{k_n-1} \frac{1}{n} E_{n,j}^2 = \frac{k_n}{n} \sigma^2(m(x)) + o(1).
\]

A simple computation gives
\[
\frac{1}{n} \sum_{j=0}^{k_n-1} E_{n,j}^2 = \frac{k_n u_n}{n} \sigma^2(m(x)) + o(1).
\]

which implies (3.41) by \( k_n u_n/n \to 1 \).

It remains for us to prove (3.42). As \( \psi \) is not necessarily bounded, we employ a truncation method. Denote \( \psi_L(-) = \psi(-) 1(|\psi(-)| \leq L) \) and define \( \tilde{Z}_{n,i} = \psi_L(Y_n - m(x)), \tilde{Z}_{n,i}^L = \left( \tilde{Z}_{n,i} - E \tilde{Z}_{n,i}^L \right) / \sqrt{\Omega_n} \) and \( \tilde{z}_j, j = 0, \ldots, k_n - 1 \), be large-block with \( \tilde{Z}_{n,i} \) replaced by \( \tilde{Z}_{n,i}^L \). By the boundedness of \( \psi_L(-) \), we have
\[
|\tilde{z}_j|^2 \leq C \frac{u_n}{\sqrt{\Omega_n}} = o(\sqrt{n}).
\]

Let \( \sigma^2_L = E[\psi_L^2(Y_1 - s)|X_1 = x]K_{2f_1}(x) \), then by (3.43), we have
\[
\frac{1}{n} \sum_{j=0}^{k_n-1} E_{n,j}^2 \geq \sigma_L^2(\sqrt{n}) \to 0 \quad \text{for any} \quad \epsilon > 0.
\]

It is obvious that (3.38)–(3.41) still hold when we substitute \( \psi \) with \( \psi_L \) in each expression. As a result,
\[
\frac{1}{\sqrt{n}} \sum_{j=0}^{k_n-1} \tilde{z}_j^2 \to \frac{d}{\Phi(0, \sigma^2_L(m(x)))}.
\]

Let \( \tilde{z}_j = \tilde{z}_j - \tilde{z}_j^L \). Reminding (3.47), we only need to show
\[
\frac{1}{n} E \left( \sum_{j=0}^{k_n-1} \tilde{z}_j \right)^2 \to 0.
\]
as first $n \to \infty$ and then $L \to \infty$. In fact, a calculation similar to the proof of (3.38) gives
\[
\frac{1}{n} E \left( \frac{\|x\|^2}{\sum_{j=0}^{k-1} s_j} \right)^2 = \frac{k_n}{n} E \left( \frac{d}{E} \right)^2 + \frac{1}{n} \sum_{ij=0}^{k-1} \text{Cov} \left( \frac{d}{E}, \frac{d}{E} \right)
\]
\[
= \frac{k_n}{n} E \left( \psi_L(Y_1 - m(x)) \right) \|x\| = x f_1(x) K_2(1 + o(1)) + o(1)
\]
\[
\xrightarrow{n \to \infty} f_1(x) K_2(1 + o(1)), \quad \xrightarrow{L \to \infty} 0,
\]
where $\psi_L = \psi - \psi_L$. Thus, the proof of Theorem 3 is completed. □

**Remark 5.** Beran et al. (2003) studied the following fixed design model
\[
y_i = g(t_i) + \varepsilon_i, \quad t_i = \frac{i}{n}, \quad i = 1, 2, \ldots, n
\]
with long-memory errors $\varepsilon_i$ and got the order $O(h^2)$ for the bias of the proposed nonparametric M-estimator $\hat{g}(t)$. In their paper, $K(\cdot)$ was assumed to be a symmetric kernel on $[-1, 1]$, so $\int_1^1 u K(u) du = 0$. After applying the third order Taylor expansion to $g(t)$, $O(h^2)$ was obtained for $E \hat{g}(t) - g(t)$. However, as $\mathcal{H}$ is an abstract functional space in our paper, we do not apply Taylor expansion directly to $m(X_i)$. By imposing the smoothness condition A5 on the regression function $m(\cdot)$, we get
\[
E m_n(x) - m(x) = o_n = O(B_n) = O \left( h^0 h_1(x)f_1(x) \frac{h}{\phi(h)} \int_0^1 K(u) \phi'(hu) du \right) = O(h^0).
\]

4. Examples and numerical implementation

We first give two examples that satisfy the $\omega$-mixing conditions assumed in this paper. The first one is a multivariate process and the second one is a sequence of random curves (or functions).

**Example 1.** Let $\mathcal{H} = \mathcal{B}^d$. For $x = (x_1, \ldots, x_d) \in \mathcal{B}^d$, $\|x\|$ is defined as $\|x\| = \sqrt{x_1^2 + \cdots + x_d^2}$, which is the Euclid norm of $x$. For any $d \times d$ matrix $A$, $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Suppose that $(X_i)$ is a multivariate MA($\infty$) process defined by
\[
X_t = \sum_{i=0}^{\infty} A_i \varepsilon_{t-i},
\]
where $A_i$ are $d \times d$ matrices and $\|A_n\| \to 0$ exponentially fast as $n \to \infty$, $\varepsilon_t$ are i.i.d. $\mathcal{B}^d$-valued random vectors with mean $0$. If the probability density function of $\varepsilon_t$ exists (such as multivariate normal, exponential, and uniform distribution), then by Pham and Tran (1985), $(X_i)$ is $\omega$-mixing with $\omega(n) \to 0$ exponentially fast. It can be easily verified that A5 holds. If we choose $A_n$ such that $\sum_{t=1}^{\infty} A_n^t < \infty$ and $\rho_n(h) / A_n^2 \log^n n \to \infty$, then $\sum_{t=1}^{\infty} (A_n / \rho(h))^{1/2} n^{1/2} (\|h^n\| / \rho(h)) < \infty$. So the conditions in (2.2) are satisfied.

**Example 2.** Assume that $\{V_t\}$ is a real-valued autoregressive moving average (ARMA) process, then it admits a Markovian representation
\[
V_t = H Z_t, \quad Z_t = F Z_{t-1} + G e_t,
\]
where $Z_t$ are random variables, $H, F, G$ are appropriate constants and $e_t$ are i.i.d. random variables with density function $g(\cdot)$. If $|F| < 1$, $E|e_t|^\delta < \infty$ and $\int g(x) - g(x - \theta) \, dx = O(\theta^\gamma)$ for some $\delta > 0$ and $\gamma > 0$, then by Pham and Tran (1985), $(V_t)$ is an $\omega$-mixing process with $\omega(n) \to 0$ at an exponential rate. Let $X_t = h(V_t)$, where $h(\cdot)$ is some real-valued function, then $\{X_t(t)\}$ is a sequence of $\omega$-mixing random curves with exponentially decaying mixing rate. Hence, for suitably chosen $h$ and $v_n$, the $\omega$-mixing conditions in Theorem 3 hold.

Although there are cases of time series that are mixing, the mixing condition is difficult to check in general. So in the context of asymptotic theory of statistics, it is usually a common practice to assume the mixing property as a precondition.

**Numerical implementation:** Next, we will give a numerical example to show the efficiency of the M-estimation discussed in this paper. We choose $\mathcal{H} = C[0, 1]$, which is the space composed of all of the continuous functions defined in $[0, 1]$. We consider
The average MADE of Table 1 bandwidth to minimize the mean absolute deviation error (MADE), which is defined in (4.4).

Consider the following model:

\[ Y_i = m(X_i) + e_i, \quad m(x) = 4 \int_0^1 (x'(t))^2 \, dt, \] (4.3)

which was also used in Laksaci et al. (2007). In view of the smoothness of the random curves \( X_i(t) \), we choose the distance in \( \mathcal{H} \) as \( d(x,y) = \left( \int_0^1 (x'(t) - y'(t))^2 \, dt \right)^{1/2} \) for \( x = x(t) \in \mathcal{H} \) and \( y = y(t) \in \mathcal{H} \). For more discussion on the choices of distance in functional nonparametric estimation, see Ferraty and Vieu (2006).

In this example, we apply the quadratic kernel function \( K(u) = \frac{3}{2}(1 - u^2)1(u \in [0,1]) \) in order to compare the performance of the proposed M-estimator \( m_n \) with the NW estimator \( m_n^\star \), we consider the following different distributions of the errors \( e_i \):

(i) standard normal distribution: \( e_i \sim \mathcal{N}(0,1) \);
(ii) contaminated errors: \( e_i \sim 0.4\mathcal{N}(0,2^2) + 0.9\mathcal{N}(0,1) \);
(iii) Cauchy distribution: \( e_i \sim \mathcal{C}(0, 1) \).

Furthermore, we choose \( \psi(y) = \max(-1.75, \min(y, 1.75)) \), which is Huber’s function with \( c = 1.75 \).

A difficult problem in simulation is the choice of a proper bandwidth. Here, we apply an iterative procedure to obtain \( m_n \). For a predetermined sequence of \( h \)’s from a wide range, say from 0.05 to 1 with an increment 0.02, we choose the bandwidth to minimize the mean absolute deviation error (MADE), which is defined in (4.4).

As it may be not easy to find the solution to (1.4), we apply an iterative procedure to obtain \( m_n \). For any original value \( \theta_0 \), calculate \( \theta_i \) iteratively by

\[ \theta_i = \theta_{i-1} + \left( \sum_{i=1}^n \psi'(Y_i - \theta_{i-1}) A_i \right)^{-1} \sum_{i=1}^n \psi(Y_i - \theta_{i-1}) A_i. \]

This procedure terminates at \( \theta_0 \) when \( |\theta_{i0} - \theta_{i0-1}| < 1 \times 10^{-4} \) or \( i = 500 \), then let \( m_n = \theta_{i0} \). We compare the MADE of \( m_n \) with that of the NW estimator \( m_n^\star \). The MADEs of \( m_n \) and \( m_n^\star \) are defined as

\[ \frac{1}{n} \sum_{i=1}^n [m_n(X_i) - m(X_i)] \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n [m_n^\star(X_i) - m(X_i)]. \] (4.4)

respectively. The simulation data consist of 500 replications of sample sizes \( n = 300 \) and 800. The results are reported in Table 1.

From Table 1, we know that the proposed M-estimator \( m_n \) is more robust than the NW estimator \( m_n^\star \). This is more apparent when the errors \( e_i \) are heavy-tailed (distribution (iii)). So when dealing with outliers or aberrant observations, \( m_n \) performs better than the NW estimator \( m_n^\star \).

**Acknowledgments**

We show our gratitude to the referees for their constructive comments, which help greatly to improve the presentation of this paper. We would also like to thank Dr. Li, D.G. for his help in the completion of this paper. This work was supported by the National Natural Science Foundation of China (Grant no. 10771192).

**Appendix**

The first lemma follows by Theorem 1.3 of Bosq (1998) and the second lemma is due to Volkonskii and Rozanov (1959).
Lemma 1. Suppose that \( \{X_t\} \) is a stationary \( \alpha \)-mixing process with mean 0 and \( S_n = \sum_{t=1}^{n} X_t \). If \( P(\{|X_t| \leq b\} = 1, t = 1, \ldots, n \) for each \( q = 1, \ldots, \lfloor n/2 \rfloor \) and \( \epsilon > 0 \),

\[
P(S_n > n\epsilon) \leq 4 \exp \left( -\frac{\epsilon^2 q}{8\sigma^2(q)} \right) + 22 \left( 1 + \frac{4b}{\epsilon} \right)^{1/2} q \alpha \left( \frac{n}{2q} \right),
\]

where \( \sigma^2(q) = 2\sigma^2(q)/p^2 + b_\nu/2, p = n/(2q) \), and

\[
\sigma^2(q) = \max_{0 \leq j \leq 4q-1} E(\{|j|p + 1 - |j|p\}X_1 + X_2 + \cdots + X_{l(j+1)p-|j|p} + (|j|p + |j|p)X_{l(j+1)p-|j|p+1})^2.
\]

Lemma 2. Let \( V_1, \ldots, V_L \) be \( \alpha \)-mixing random variables that can take values in the complex space and are measurable with respect to the \( \sigma \)-algebras \( \mathcal{F}_{i_1}, \ldots, \mathcal{F}_{i_L} \), respectively, with \( 1 \leq i_1 < i_2 < \cdots < i_L \leq n \), \( i_{j+1} - i_j \geq w \geq 1 \) for \( 1 \leq l \leq L - 1 \). Assume that \( |V_j| \leq 1 \) for \( j = 1, \ldots, L \), then

\[
\left| E \left( \prod_{j=1}^{L} V_j \right) - E \left( \prod_{j=1}^{L} E(V_j) \right) \right| \leq 16(L - 1)\alpha(w),
\]

where \( \alpha(w) \) is the mixing coefficient.

References