Uniform Bahadur Representation for Local Polynomial Estimates of M-Regression and Its Application to The Additive Model

Efang Kong∗
EURANDOM, The Netherlands
Oliver Linton†
London School of Economics, UK
Yingcun Xia‡
National University of Singapore, Singapore

SUMMARY

We use local polynomial fitting to estimate the nonparametric M-regression function for strongly mixing stationary processes \{(Y_i, X_i)\}. We establish a strong uniform consistency rate for the Bahadur representation of estimators of the regression function and its derivatives. These results are fundamental for statistical inference and for applications that involve plugging in such estimators into other functionals where some control over higher order terms are required. We apply our results to the estimation of an additive M-regression model.

Key words: Additive model; Bahadur representation; Local polynomial fitting; M-regression; Strongly mixing processes; Uniform strong consistency.

1 Introduction

In many contexts one wants to evaluate the properties of some procedure that is a function or functional of some estimators. It is useful to be able to work with some plausible high level assumptions about those estimators rather than to rederive their properties for each different application. In a fully parametric context it is quite natural to assume that parametric estimators

∗Eurandom, Technische Universiteit Eindhoven, The Netherlands. E-mail address: kong@eurandom.tue.nl.
†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. http://econ.lse.ac.uk/staff/olinton/ indexown.html. E-mail address: o.linton@lse.ac.uk.
‡Department of Statistics and Applied Probability, National University of Singapore, Singapore. http://www.stat.nus.edu.sg/~staxyc. E-mail address: staxyc@nus.edu.sg.
are root-$n$ consistent and asymptotically normal. In some cases this suffices; in other cases one needs to be more explicit in terms of the linear expansion of these estimators, but in any case such expansions are quite natural and widely applicable. In a nonparametric context there is less agreement about the use of such expansions and one often sees standard properties of standard estimators derived anew for a different purpose. It is our objective to provide results that can circumvent this. The types of application we have in mind are estimation of semiparametric models where the parameters of interest are explicit or implicit functionals of nonparametric regression functions and their derivatives, see Powell (1994), Andrews (1994), Chen, Linton and Van Keilegom (2003). Another class of applications includes estimation of structured nonparametric models like additive models, Linton and Nielsen (1995), or generalized additive models, Linton, Sperlich, and Van Keilegom (2007).

We motivate our results in a simple i.i.d. setting. Suppose we have a random sample \( \{Y_i, X_i\}_{i=1}^n \) and consider the Nadaraya-Watson estimator of the regression function \( m(x) = E(Y_i|X_i = x) \),

\[
\hat{m}(x) = \frac{\hat{r}(x)}{\hat{f}(x)} = \frac{n^{-1} \sum_{i=1}^n K_h(x-X_i)Y_i}{n^{-1} \sum_{i=1}^n K_h(x-X_i)},
\]

where \( K \) is a kernel, \( h \) is a bandwidth and \( K_h(.) = K(./h)/h \). Standard arguments (Härdle, 1990) show that (under suitable smoothness conditions)

\[
\hat{m}(x) - m(x) = h^2 b(x) + \frac{1}{nf(x)} \sum_{i=1}^n K_h(x-X_i)\varepsilon_i + R_n(x),
\]

(1)

where \( f(x) \) is the covariate density, \( \varepsilon_i \equiv Y_i - m(X_i) \) is the error term and \( b(x) = [m''(x) + 2m'(x)f'(x)/f(x)]/2 \). The remainder term \( R_n(x) \) is of higher order (almost surely) than the two leading terms. Such expansion is sufficient to derive the central limit theorem for \( \hat{m}(x) \) itself, but generally is not if \( \hat{m}(x) \) is to be plugged into some semiparametric procedure. For example, suppose we need to estimate the parameter \( \theta_0 = \int m(x)^2 dx \) by \( \hat{\theta} = \int \hat{m}(x)^2 dx \), where the integral is over some compact set \( \mathcal{D} \); and we would expect to find \( n^{1/2}(\hat{\theta} - \theta_0) \) to be asymptotically normal. The argument goes like this. First, we obtain the expansion

\[
n^{1/2}(\hat{\theta} - \theta_0) = 2n^{1/2} \int m(x)\{\hat{m}(x) - m(x)\}dx + n^{1/2} \int [\hat{m}(x) - m(x)]^2 dx.
\]
If it can be shown that \( \hat{m}(x) - m(x) = o(n^{-1/4}) \) a.s. uniformly in \( x \in D \) (such results are widely available, see for example Masry (1996)), we have

\[
\begin{align*}
n^{1/2}(\hat{\theta} - \theta_0) &= 2n^{1/2} \int m(x)\{\hat{m}(x) - m(x)\}dx + o(1), \quad \text{a.s.}
\end{align*}
\]

Note that the quantity on the right hand side is the term in assumption 2.6 of Chen, Linton, and Van Keilegom (2003) which is assumed to be asymptotically normal. It is the verification of this condition with which we are now concerned. If we substitute in the expansion (1) we obtain

\[
\begin{align*}
n^{1/2}(\hat{\theta} - \theta_0) &= 2n^{1/2} h^2 \int m(x)b(x)dx + 2n^{1/2} \int \frac{m(x)}{f(x)} n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i dx \\
&\quad + 2n^{1/2} \int m(x)R_n(x)dx + o(1), \quad \text{a.s.}
\end{align*}
\]

If \( nh^4 \rightarrow 0 \), then the first term (the smoothing bias term) is \( o(1) \). By a change of variable, the second term (the stochastic term) can be written as a sum of independent random variables with mean zero

\[
\begin{align*}
n^{1/2} \int m(x)f^{-1}(x)n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i dx &= n^{-1/2} \sum_{i=1}^{n} \xi_n(X_i)\varepsilon_i, \\
\xi_n(X_i) &= \int m(X_i + uh)f^{-1}(X_i + uh)K(u)du,
\end{align*}
\]

and this term obeys the Lindeberg central limit theorem under standard conditions. The problem is that (1) only guarantees that \( \int m(x)R_n(x)dx = o(n^{-2/5}) \) a.s. at best. Actually, in this simple case it is possible to derive a more useful Bahadur expansion (Bahadur (1966)) for the kernel estimator

\[
\begin{align*}
\hat{m}(x) - m(x) &= h^2 b_n(x) + \{E\hat{f}(x)\}^{-1}n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i + R^*_n(x), \quad \text{(2)}
\end{align*}
\]

where \( b_n(x) \) is deterministic and satisfies \( b_n(x) \rightarrow b(x) \) uniformly in \( x \in D \), and \( E\hat{f}(x) \rightarrow f(x) \) uniformly in \( x \in D \), while the remainder term now satisfies

\[
\sup_{x \in D} |R^*_n(x)| = O\left(\frac{\log n}{nh}\right), \quad \text{a.s.} \quad \text{(3)}
\]
This property is a consequence of the uniform rate of convergence of \( \hat{f}(x) - E\hat{f}(x) \), \( n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\{m(X_i) - m(x)\} \) - \( EK_h(x - X_i)\{m(X_i) - m(x)\} \), and \( n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i \) that follow from, for example Masry (1996). Clearly, \( R_n^*(x) \) can be made to be \( o(n^{-1/2}) \), a.s. uniformly over \( D \), by appropriate choice of \( h \); from this property we can easily see that the remainder term \( 2n^{1/2} \int m(x)R_n^*(x)dx = o(1) \) a.s. and one can just work with the two leading terms in (2).

The leading terms are slightly more complicated than in the previous expansion but are still sufficiently simple for many purposes; in particular, \( b_n(x) \) is uniformly bounded so that provided \( nh^4 \to 0 \), the smoothing bias term satisfies \( h^2n^{1/2} \int m(x)b_n(x)dx \to 0 \), while the stochastic term is a sum of mean zero independent random variables

\[
n^{1/2} \int \frac{m(x)}{f(x)} n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i dx = n^{-1/2} \sum_{i=1}^{n} \xi_n(X_i)\varepsilon_i
\]

and obeys the Lindeberg central limit theorem under standard conditions, where \( f(x) = E\hat{f}(x) \).

This argument shows the utility of the Bahadur expansion (2). There are many other applications of this result because a host of probabilistic results are available for random variables like \( n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i \) and integrals thereof.

The one-dimensional Nadaraya-Watson estimator for i.i.d. data is particularly easy to analyze and the above arguments are well known. However, the limitations of this estimator are manyfold and there are good theoretical reasons for working instead with the local polynomial class of estimators (Fan and Gijbels, 1996). In addition, for many data one may have concerns about heavy tails or outliers that point in the direction of using robust estimators like the local median or local quantile method, perhaps combined with local polynomial fitting. We examine a general class of (nonlinear) M-regression function (that is, location functionals defined through minimization of a general objective function) and derivative estimators. We treat a general time series setting where the multivariate data are strong mixing. We establish a uniform strong Bahadur expansion like (2) and (3) with remainder term of order \( (\log n/nh^d)^c \) almost surely, where \( c \) depends on several factors including the smoothness of the M-regression function. Under mild
conditions we can obtain $c = 3/4$, almost optimal based on the results in Kiefer (1967) under i.i.d. setting. The leading terms are linear and functionals of them can be analyzed simply. The remainder term can be made to be $o(n^{-1/2})$ a.s. under restrictions on the dimensionality in relation to the amount of smoothness possessed by the M-regression function. We apply our result to the study of marginal integration estimators (Linton and Nielsen, 1995) in additive nonparametric M-regression where we only need the remainder term to be $o(n^{-p/(2p+1)})$ a.s., where $p$ is a smoothness index.

Bahadur expansions (Bahadur, 1966) have been widely studied and applied, with notable refinements in the i.i.d. setting by Kiefer (1967). A recent paper of Wu (2005) extends these results to a general class of dependent processes and provides a review. The closest paper to ours is Hong (2003) who establishes a Bahadur expansion for essentially the same local polynomial M-regression estimator as ours. However, his results are: (a) pointwise, i.e., for a single $x$ only; (b) the covariates are univariate; (c) for i.i.d. data. Clearly, this limits the range of applicability of his results, and specifically, the application to semiparametric or additive models are perforce precluded.

2 The General Setting

Let $\{(Y_i, X_i)\}$ be a jointly stationary process, where $X_i = (x_{i1}, \ldots, x_{id})^\top$ with $d \geq 1$ and $Y_i$ is a scalar. As dependent observations are considered in this paper, we introduce here the mixing coefficient. Let $F_s^t$ be the $\sigma-$ algebra of events generated by random variables $\{(Y_i, X_i), s \leq i \leq t\}$. The stationary process $\{(Y_i, X_i)\}$ is strongly mixing if

$$\sup_{A \in \mathcal{F}^0, B \in \mathcal{F}_k^\infty} |P[AB] - P[A]P[B]| = \gamma[k] \to 0, \text{ as } k \to \infty,$$

and $\gamma[k]$ is called the strong mixing coefficient.

Suppose $\rho(\cdot; \cdot)$ is a loss function. Our first goal is to estimate the multivariate M-regression function

$$m(x_1, \cdots, x_d) = \arg \min_{\theta} E\{\rho(Y_i; \theta) | X_i = (x_1, \cdots, x_d)\},$$

(4)
and its partial derivatives based on observations \( \{(Y_i, X_i)\}_{i=1}^n \). An important example of the M-function is with loss function \( \rho(y; \theta) = (2q-1)(y-\theta) + |y-\theta| \), corresponding to the \( q \)'th quantile of \( Y_i \) given \( X_i = (x_1, \ldots, x_d)^\top \). Another leading example is the \( L_q \) criterion \( \rho(y; \theta) = |y-\theta|^q \) for \( q > 1 \), which includes the least squares criterion \( \rho(y; \theta) = (y-\theta)^2 \) in which case \( m \) is the expectation of \( Y_i \) given \( X_i \).

Assuming that \( m(\underline{x}) \) has derivatives up to order \( p + 1 \) at \( \underline{x} = (x_1, \ldots, x_d)^\top \), we have the following multivariate \( p \)'th order local polynomial approximation of \( m(\underline{x}) \) for any \( \underline{z} \) close to \( \underline{x} 

\[
m(\underline{z}) = \sum_{0 \leq |r| \leq p} \frac{1}{r!} D^m m(\underline{x})(\underline{z} - \underline{x})^r,
\]

where \( r = (r_1, \ldots, r_d) \), \( |r| = \sum_{i=1}^d r_i \), \( r! = r_1! \times \cdots \times r_d! \),

\[
D^m m(\underline{x}) = \frac{\partial^{|r|} m(\underline{x})}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}, \quad \underline{z}^r = x_1^{r_1} \times \cdots \times x_d^{r_d}, \quad \sum_{0 \leq |r| \leq p} = \sum_j \sum_{j=0} \sum_{r_1=0} \cdots \sum_{r_d=0} \sum_{r_1+\cdots+r_d=j}.
\]

Let \( K(u) \) be a nonnegative weight function on \( R^d \), \( h \) be a bandwidth and \( K_h(u) = K(u/h) \).

With observations \( \{(Y_i, X_i)\}_{i=1}^n \), we consider the following quantity

\[
\sum_{i=1}^n K_h(X_i - \underline{x})\rho\left(Y_i; \sum_{0 \leq |r| \leq p} \beta_r(X_i - \underline{x})^r\right).
\]

Minimizing (6) with respect to \( \beta_r, 0 \leq |r| \leq p \) gives an estimate \( \hat{\beta}_r(\underline{x}) \). The M-function \( m(\underline{x}) \) and its derivatives \( D^m m(\underline{x}) \) are then estimated respectively by

\[
\hat{m}(\underline{x}) = \hat{\beta}_0(\underline{x}) \quad \text{and} \quad \hat{D}^m m(\underline{x}) = r!\hat{\beta}_r(\underline{x}), \quad 1 \leq |r| \leq p.
\]

### 3 Main Results

For any \( M > 2 \), \( \lambda_2 \in (0, 1) \), and \( \lambda_1 \in (\lambda_2, (1 + \lambda_2)/2] \), let

\[
d_n = (nh^d / \log n)^{-(\lambda_1 + \lambda_2/2)}(nh^d \log n)^{1/2}, \quad r(n) = (nh^d / \log n)^{(1-\lambda_2)/2},
\]

\[
M_n^{(1)} = M(\log n)^{-\lambda_1}, \quad M_n^{(2)} = M^{1/4}(\log n)^{-\lambda_2}, \quad T_n = \{r(n)/h\}^d
\]

and \( L_n \) be the smallest integer such that \( \log n(M/2)^{L_n+1} > nM_n^{(2)}/d_n \). We use ||\( \cdot \)|| to denote the Euclidean norm and \( C \) is a generic constant, which may have different values at each appearance.

The following assumptions are used in our proofs of the results. Let \( \varepsilon_i = Y_i - m(X_i) \).
For each \( y \in \mathcal{R} \), \( \rho(y; \theta) \) is absolutely continuous in \( \theta \), i.e., there is a function \( \varphi(y; \theta) \equiv \varphi(y - \theta) \) such that for any \( \theta \in \mathcal{R} \), 
\[ \rho(y; \theta) = \rho(y; 0) + \int_{\theta}^{y} \varphi(y; t) dt. \]
The probability density function of \( \varepsilon_i \) is bounded, 
\[ E\{\varphi(\varepsilon_i)|X_i\} = 0 \text{ almost surely and } E|\varphi(\varepsilon_i)|^{\nu_1} < \infty \text{ for some } \nu_1 > 2. \]

\( \phi(.) \) satisfies the Lipschitz condition in \((a_j, a_{j+1})\), \( j = 0, \cdots, m \), where \( a_1 < \cdots < a_m \) are
the finite number of jump discontinuity points of \( \phi(.) \), \( a_0 \equiv -\infty \) and \( a_{m+1} \equiv +\infty \).

\( K(.) \) has a compact support, say \([-1, 1]^d \) and 
\[ |H_j(u) - H_j(v)| \leq C\|u - v\| \text{ for all } j \text{ with } 0 \leq |j| \leq 2p + 1, \text{ where } H_j(u) = u^j K(u). \]

The probability density function of \( X \), \( f(.) \) is bounded and with bounded first order derivatives. The joint probability density of \((X_0, X_i)\) satisfies 
\[ f(u, v; l) \leq C < \infty \text{ for all } l \geq 1. \]

For \( r \) with \(|r| = p + 1 \), \( D^2 m(x) \) is bounded with bounded first order derivative.

The bandwidth \( h \to 0 \) satisfies that
\[ nh^d/\log n \to \infty, \; nh^{d+(p+1)/\lambda_2}/\log n < \infty, \; n^{-\left\{r(n)\right\}^{\nu_2/2}d_n \log n/M_n^{(2)}} \to \infty, \]
for some \( 2 < \nu_2 \leq \nu_1 \) and the processes \( \{(Y_i, X_i)\} \) are strongly mixing with mixing coefficient \( \gamma[k] \) satisfying
\[ \sum_{k=1}^{\infty} k^a \{\gamma[k]\}^{1-2/\nu_2} < \infty \text{ for some } a > (p + d + 1)(1 - 2/\nu_2)/d. \]

Moreover, the bandwidth \( h \) and \( \gamma[k] \) should jointly satisfy the following condition
\[ \sum_{n=1}^{\infty} n^{3/2} T_n \left\{ \frac{M_n^{(1)}}{d_n} \right\}^{1/2} \frac{\gamma[r(n)]^{(2\nu_2/2)/M}L_{n/\nu_2}}{r(n)\{2^{\nu_2/2}/M\}L_{n/\nu_2}} \{4M^{2N}\}^{L_n} < \infty, \; \forall M > 0. \]

The conditional density \( f_{X|Y} \) of \( X \) given \( Y \) exists and is bounded. The conditional density 
\[ f_{(X_1, X_{i+1})|Y_1, Y_{i+1}} \text{ of } (X_1, X_{i+1}) \text{ given } (Y_1, Y_{i+1}) \text{ exists and is bounded, for all } l \geq 1. \]

Remark 1. (A1) is imposed for model specification and (A2) is necessary for the remainders in Bahadur representations to achieve optimal rates. To our best knowledge, in all known robust
and likelihood type regressions, $\varphi(\cdot, \cdot)$ satisfies (A2). In this case, it was proved in Hong (2003) that, if the conditional density $f(y|x)$ of $Y$ given $X$ is continuously differentiable with respect to $y$, then there is a constant $C > 0$, such that for all small $t$ and $x$,

$$E\left[\left\{ \varphi(Y; t + a) - \varphi(Y; a) \right\}^2 | X = u \right] \leq C|t|$$

holds for all $(a, u)$ in a neighborhood of $(m(x), x)$. Let

$$G(t, u) = E\{ \varphi(Y; t) | X = u \}, \quad G_i(t, u) = (\partial^i / \partial t^i)G(t, u), \; i = 1, 2.$$\n
Then

$$g(x) = G_1(m(x), x) \geq C > 0, \quad G_2(t, x) \text{ bounded for all } x \in \mathcal{D} \text{ and } t \text{ near } m(x).$$

Assumptions (A3)-(A7) are standard for nonparametric smoothing in multivariate time series analysis, see Masry (1996). Note that condition (11) is more stringent than (4.7b) in Masry (1996), due to the fact that the form of $\rho(\cdot)$ considered here is more general than the simple squared loss.

Let $N_i = \binom{i+d-1}{d-1}$ be the number of distinct $d$–tuples $r$ with $|r| = i$. Arrange these $d$–tuples as a sequence in a lexicographical order (with the highest priority given to the last position so that $(0, \cdots, 0, i)$ is the first element in the sequence and $(i, 0, \cdots, 0)$ the last element). Let $\tau_i$ denote this 1-to-1 map, i.e. $\tau_i(1) = (0, \cdots, 0, i), \cdots, \tau_i(N_i) = (i, 0, \cdots, 0)$. For each $i = 1, \cdots, p$, define a $N_i \times 1$ vector $\mu_i(x)$ with its $k$th element given by $x^{\tau_i(k)}$ and write

$$\mu(x) = (1, \mu_1(x)^T, \cdots, \mu_p(x)^T)^T,$$

which is a column vector of length $N = \sum_{i=0}^{p} N_i \times 1$. Similarly define vectors $\beta_p(x)$ and $\beta$ through the same lexicographical arrangement of $D\mathcal{D}m(x)$ and $\beta_r$ in (6) for $0 \leq |r| \leq p$. Thus (6) can be rewritten as

$$\sum_{i=1}^{n} K_h(X_i - x)\rho(Y_i; \mu(X_i - x)^T \beta).$$

Suppose the minimizer of (15) is denoted as $\hat{\beta}_n(x)$. Let $\hat{\beta}_p(x) = W_p \hat{\beta}_n(x)$, where $W_p$ is the diagonal matrix with diagonal entries the lexicographical arrangement of $r^i$, $0 \leq |r| \leq p$. 
Let \( \nu_k = \int K(u)u^2du \). For \( g(.) \) given in (14), define
\[
\nu_{nl}(x) = \int K(u)u^2g(x + hu)\Phi(x + hu)du.
\]
For \( 0 \leq j, k \leq p \), let \( S_{j,k} \) and \( S_{n,j,k}(x) \) be two \( N_j \times N_k \) matrices with their \((l,m)\) elements respectively given by
\[
[S_{j,k}]_{l,m} = \nu_{lj}(x) + \tau_{k}(m)(x), \quad [S_{n,j,k}(x)]_{l,m} = \nu_{l,n}^{n_j}(x) + \tau_{k}(m)(x).
\]
(16)
Now define the \( N \times N \) matrices \( S_p \) and \( S_{n,p}(x) \) by
\[
S_p = \begin{bmatrix}
S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\
S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\
\vdots & \vdots & \ddots & \vdots \\
S_{p,0} & S_{p,1} & \cdots & S_{p,p}
\end{bmatrix}, \quad S_{n,p}(x) = \begin{bmatrix}
S_{n,0,0}(x) & S_{n,0,1}(x) & \cdots & S_{n,0,p}(x) \\
S_{n,1,0}(x) & S_{n,1,1}(x) & \cdots & S_{n,1,p}(x) \\
\vdots & \vdots & \ddots & \vdots \\
S_{n,p,0}(x) & S_{n,p,1}(x) & \cdots & S_{n,p,p}(x)
\end{bmatrix}.
\]
According to Lemma 6.8, \( S_{n,p}(x) \) converges to \( g(x)f(x)S_p \) uniformly in \( x \in D \) almost surely. Hence for \( |S_p| \neq 0 \), we can define
\[
\beta_n^*(x) = -\frac{1}{nh^d} W_p S_{n,p}(x) H^{-1} \sum_{i=1}^{n} K_h(x_i - x) \varphi(Y_i, \mu(x_i - x)^\top \beta_p(x)) \mu(x_i - x), \quad (17)
\]
where \( \varphi(,; \cdot) \) is as defined in (A1) and \( H \) is the diagonal matrix with diagonal entries the lexicographical arrangement of \( h^{[r]} \), \( 0 \leq |r| \leq p \).

The following asymptotic expression for the mean of \( \beta_n^*(x) \) is an extension of Proposition 2.2 in Hong (2003) to the multivariate case.

**Proposition 3.1** Denote the typical element of \( \beta_n^*(x) \) by \( \beta_n^*_{\tau}(x) \), \( 0 \leq |\tau| \leq p \). If \( f(x) > 0 \), then under (A1)-(A5),
\[
E\beta_n^*_{\tau}(x) = \begin{cases}
-h^{p+1}e_{N(\tau)}W_p S_{n,p-1}^{-1}B_1 \mathbf{m}_{p+1}(x) + o(h^{p+1}), & \text{for } p - |\tau| \text{ odd}, \\
-h^{p+2}e_{N(\tau)}W_p S_{n,p}^{-1}\{f g\}^{-1}(x) \mathbf{m}_{p+1}(x) \{N_p S_p^{-1}B_1 + B_2 \mathbf{m}_{p+2}(x)\} + o(h^{p+2}), & \text{for } p - |\tau| \text{ even},
\end{cases}
\]
where \( N(\tau) = \tau^{-1}_{[\tau]}(x) + \sum_{k=0}^{\lfloor |\tau| \rfloor - 1} N_k, e_i \) is a \( N \times 1 \) vector having 1 as the \( i \)-th entry with all other entries 0, and
\[
B_1 = \begin{bmatrix}
S_{0,p+1} \\
S_{1,p+1} \\
\vdots \\
S_{p,p+1}
\end{bmatrix}, \quad B_2 = \begin{bmatrix}
S_{0,p+2} \\
S_{1,p+2} \\
\vdots \\
S_{p,p+2}
\end{bmatrix}.
\]
Our Bahadur representation for local polynomial estimates is as follows.

**Theorem 3.2** Let (A1)-(A7) hold with \( \lambda_2 = (p+1)/(2(p+s+1)) \) for some \( s \geq 0 \) and \( D \) be any compact subset of \( \mathbb{R}^d \). Then

\[
\sup_{x \in D} |H\{\hat{\beta}_p(x) - \beta_p(x)\} - \beta_n^*(x)\| = O\left(\frac{\log n}{nh^{\lambda(s)}}\right) \quad \text{almost surely,}
\]

where \( |.| \) is taken to be the sup norm and

\[
\lambda(s) = \min \left\{ \frac{p+1}{p+s+1}, \frac{3p+3+2s}{4p+4s+4} \right\}.
\]

**Remark 2.** From above Theorem, we can see that the dependence among the observations doesn’t have effect on the rate of uniform convergence, given that the degree of the dependence, as indicated by the mixing coefficient \( \gamma[k] \), is not very strong, i.e. (10) and (11) are satisfied. This is in accordance with the results in Masry (1996), where for local polynomial estimator with squared loss, the uniform convergence rate is proved to be \((nh^d/\log n)^{-1/2}\), the same as in the independent case.

**Remark 3.** It is of practical interest to provide an explicit rate of decay for the strong mixing coefficient \( \gamma[k] \) of the form \( \gamma[k] = O(1/k^c) \) for some \( c > 0 \) (to be determined) under which Theorem 3.2 holds. It is easily seen that, among all the conditions imposed on \( \gamma[k] \), the summability condition (11) is the most restrictive. We assume that

\[
h = h_n \sim (\log n/n)^{\bar{a}} \quad \text{for some} \quad \frac{1}{2(p+s+1)+d} \leq \bar{a} < \frac{\frac{1}{d} \left(1 - \lambda_1 \nu_2 - 4\lambda_1 + 2(1 + \lambda_2)\right)}{2(1 - \lambda_2)(1 - \bar{a})\nu_2 - 8\bar{a}d + 4(1 - \bar{a})(1 - \lambda_2 - 2\lambda_1)},
\]

so that (9) is satisfied. Algebraic calculations show that the summability condition (11) is satisfied provided that

\[
c > \nu_2 \left(1 - \bar{a}d\right)\left(1 - \lambda_2\right)(4N + 1) + 8N\lambda_1) + 10 + (4 + 8N)\bar{a}d \quad \left(1 - \lambda_1\nu_2 - 4\lambda_1 + 2(1 + \lambda_2)\right) - 1 = c(d,p,\nu_2,\bar{a},\lambda_1,\lambda_2). \quad (18)
\]

Note that we would need the following condition

\[
\nu_2 > 2 + \frac{4\{\bar{a}d + (1 - \bar{a}d)\lambda_1\}}{(1 - \bar{a}d)(1 - \lambda_2)}
\]

to secure positive denominator for (18). It is easy to see that \( c(d,p,\nu_2,\bar{a},\lambda_1,\lambda_2) \) is decreasing in \( \nu_2(\leq \nu_1) \) and therefore there is a tradeoff between the order \( \nu_1 \) of the moment \( E|\varphi(\varepsilon_i)|^{\nu_1} < \infty \).
in (A1) and the decay rate of the strong mixing coefficient $\gamma[k]$: the existence of higher order moments allow for weaker condition on $\gamma[k]$.

The following proposition follows from the above theorem with $s = 0$ and uniform convergence of sum of weakly dependent observations.

**Corollary 3.3** Suppose that conditions in Theorem 3.2 hold with $s = 0$. Then with probability 1, we have,

$$
\sup_{x \in D} |H\{\hat{\beta}_p(x) - \beta_p(x)\} - E\beta^*_n(x) - \frac{1}{nh^d} W_p S^{-1}_n(x) H^{-1} \sum_{i=1}^n K_h(x_i - x) \varphi(\varepsilon_i) \mu(X_i - x)| = O\left(\left\{\frac{\log n}{nh^d}\right\}^{3/4}\right).
$$

**Remark 4.** The rate $(nh^d/\log n)^{-3/4}$ obtained here is not optimal for all such M-regressions, as the rate for the N-W estimator given in (3) is faster. The explanation is that our results are developed for a wider variety of loss functions. This does not rule out the possibility that the rate could be higher for one particular loss function, e.g., the squared loss corresponding to the N-W estimator. It has been proved that the optimal rate of Bahadur representation of sample quantiles is $(\log n/n)^{3/4}$ (Kiefer, 1967), so we expect that the rate given above is indeed optimal for a similar class of problems.

### 4 M-Estimation of the Additive model

The convergence rate of the estimated $m(x_1, \ldots, x_d)$ strongly depends on the dimension of $d$. The rate decreases dramatically as $d$ increases (Stone, 1982). This phenomenon is the so-called “curse of dimensionality”. One approach to reduce the curse is by imposing model structure. A popular model structure is the additive model assuming that

$$
m(x_1, \ldots, x_d) = c + m_1(x_1) + \ldots + m_d(x_d), \tag{19}
$$

where $c$ is an unknown constant and $m_k(\cdot)$, $k = 1, \ldots, d$ are unknown functions which have been normalized such that $Em_k(x_k) = 0$ for $k = 1, \ldots, d$. In this case, the optimal rate of convergence is the same as one dimensional nonparametric regression (Stone, 1986). We consider this case
where \( m(x) \) is the M-regression function defined above. Previous work on additive quantile regression, for example, includes Linton (2001) and Horowitz and Lee (2005) for the i.i.d. case. We are interested in applications to the volatility model
\[
Y_i = \sigma_i \varepsilon_i \quad \text{and} \quad \ln \sigma_i^2 = m(X_i),
\]
where \( X_i = (Y_{i-1}, \ldots, Y_{i-d})^\top \). We suppose that \( \varepsilon_i \) satisfies \( E[\varphi(\ln \varepsilon_i^2; 0)|X_i] = 0 \), whence \( m \) is defined as the conditional M-regression of \( \ln Y_i^2 \) on \( X_i \). Peng and Yao (2003) have applied LAD estimation to parametric ARCH and GARCH models and have shown the superior robustness property of this procedure over Gaussian QMLE with regard to heavy tailed innovations. The heavy tails issue also arises in nonparametric models, which is why our procedures may be useful.

We use the marginal integration method (Linton and Nielsen, 1995) to estimate the additive model, which is known to achieve the optimal rate under some conditions. This involves estimating first the unrestricted M-regression function and then integrating it over some directions. Partition \( X_i = (x_1, \ldots, x_d) \) as \( X_i = (x_{1i}, X_{2i}) \), where \( x_{1i} \) is the one dimensional direction of interest and \( X_{2i} \) is a \( d - 1 \) dimensional nuisance direction and let \( x = (x_1, x_2) \). Define the functional
\[
\phi_1(x_1) = \int m(x_1, x_2) f_2(x_2) dx_2, \quad (20)
\]
where \( f_2(x_2) \) is the joint density of \( X_{2i} \). Under the additive structure (19), \( \phi_1 \) is \( m_1 \) up to a constant. Replace \( m \) in (20) with \( \hat{\beta}_0(x_1, x_2) := \hat{\beta}_0(x) \) defined in (7) and \( \phi_1(x_1) \) can thus be estimated by the sample version of (20):
\[
\tilde{\phi}_1(x_1) = n^{-1} \sum_{i=1}^n \hat{\beta}_0(x_1, X_{2i}).
\]
The application of Corollary 3.3 here may seem somewhat straightforward, however, we need to be cautious about the choice of the bandwidth. As noted by Linton and Härdle (1996) and Hengartner and Sperlich (2005), different bandwidths should be employed for the direction of interest \( X_1 \) and the \( d - 1 \) dimensional nuisance direction \( X_{2} \), say \( h_1 \) and \( h \) respectively. The following corollary is about the asymptotic properties of \( \tilde{\phi}_1(x_1) \).
Corollary 4.1 Suppose the support of $X$ is $\chi = [0, 1]^{\otimes d}$ with strictly positive density function. Let the conditions in Proposition 3.3 hold with $T_n = \{r(n)/\min(h_1, h)\}^d$ and the $h^d$ in all the notations defined in (8) or (9) replaced by $h_1 h^{d-1}$. Especially, (9) is strengthened as

$$nh_1 h^{3(d-1)} / \log^3 n \to \infty, \quad nh_1 h^{d-1} \max(h_1, h)^2(p+1) / \log n < \infty, \quad \frac{1}{n} \{r(n)\}^{\nu_2/2} \log n / M_n^{(2)} \to \infty.$$  

Then we have

$$(nh_1)^{1/2} \{ \hat{\phi}_1(x_1) - \phi_1(x_1) + \{ \max(h_1, h) \}^{p+1} e_1 W_p S_p^{-1} B_1 E m_{p+1}(x_1, X_2) \} \xrightarrow{L} N(0, \tilde{\sigma}^2(x_1)) \tag{22}$$

where '$L$-\textit{t}' stands for convergence in distribution,

$$\tilde{\sigma}^2(x_1) = \left\{ \int_{[0,1]^{\otimes d-1}} \{ f \sigma^2 \}^{-1}(x_1, X_2) f_2(x_2) \sigma^2(x_1, X_2) dX_2 \right\} e_1 S_p^{-1} K_2 K_2^\top S_p^{-1} e_1,$$

$$\sigma^2(x) = E[\varphi^2(\varepsilon) | X = x]$$ and $K_2 = \int_{[0,1]^{\otimes d}} K(y) \mu(y) dy$. In particular for quantile estimation, i.e. $\rho(y; \theta) = (2q - 1)(y - \theta) + |y - \theta|$, we have

$$\tilde{\sigma}^2(x_1) = q(1-q) \left\{ \int_{[0,1]^{\otimes d-1}} f^{-1}(x_1, X_2) f_2^{-2}(0 | x_1, X_2) f_2^2(x_2) dX_2 \right\} e_1 S_p^{-1} K_2 K_2^\top S_p^{-1} e_1.$$

Remark 5. For the conditions in the above corollary to hold, we would need $3d < 2p + 5$, i.e. the order of local polynomial approximation increases as the dimension of the predictor variable $X$ increases. See also the discussion in Hengartner and Sperlich (2005). Note that if we need (22) to admit the following form

$$(nh_1)^{1/2} \{ \hat{\phi}_1(x_1) - \phi_1(x_1) \} \xrightarrow{L} N(e_1 W_p S_p^{-1} B_1 E m_{p+1}(x_1, X_2), \tilde{\sigma}^2(x_1)), \tag{23}$$

then the fastest convergence rate is achieved only when $h_1 \propto n^{-1/(2p+3)}$ and $h = O(h_1)$.

Remark 6. It is trivial to extend this result to the generalized additive case where $G(m(x_1, \ldots, x_d)) = c + m_1(x_1) + \ldots + m_d(x_d)$ for some known smooth function $G$ in which case the marginal integration estimator is the sample average of $G(\hat{m}(x_1, X_2))$. It is also easy to obtain uniform strong Bahadur expansions for $\hat{\phi}_1(x_1)$ themselves like those assumed in Linton, Sperlich, and Van Keilegom (2007).
5 Proof of Theorem, Proposition and Corollaries

Proof of Proposition 3.1. Write \( \beta_n^*(\bar{z}) = -W_p S_{n,p}^{-1}(\bar{z}) \sum_{i=1}^{n} Z_{ni}(\bar{z})/n, \) where

\[
Z_{ni}(\bar{z}) = H^{-1} h^{-d} K_h(X_i - \bar{z}) \varphi(Y_i, \mu(X_i - \bar{z})^\top \beta_p(\bar{z})) \mu(X_i - \bar{z}).
\]

We first focus on \( EZ_{ni}(\bar{z}). \) Based on (13) and (14), we have

\[
E\{ \varphi(Y_i, \mu(X_i - \bar{z})^\top \beta_p(\bar{z})) | X_i \} = G(\mu(X_i - \bar{z})^\top \beta_p(\bar{z}), X_i)
= -g(X_i) \{ m(X_i) - \mu(X_i - \bar{z})^\top \beta_p(\bar{z}) \}
+ G_2(\xi_i(x), X_i) \{ m(X_i) - \mu(X_i - \bar{z})^\top \beta_p(\bar{z}) \}^2/2
\]

for some \( \xi_i(x) \) between \( \mu(X_i - \bar{z})^\top \beta_p(\bar{z}) \) and \( m(X_i) \). Apparently, if \( X_i = \bar{z} + h v \), then

\[
m(X_i) - \mu(X_i - \bar{z})^\top \beta_p(\bar{z}) = h^{p+1} \sum_{|k|=p+1} \frac{D^k m(x)}{k!} v^k + h^{p+2} \sum_{|k|=p+2} \frac{D^k m(x)}{k!} v^k + o(h^{p+2}).
\]

Therefore,

\[
EZ_{ni}(\bar{z}) = h^{p+1} \int K(y) f g(\bar{z} + h v) \mu(v) \sum_{|k|=p+1} \frac{D^k m(x)}{k!} v^k dv
+ h^{p+2} \int K(y) f g(\bar{z} + h v) \mu(v) \sum_{|k|=p+2} \frac{D^k m(x)}{k!} v^k dv + o(h^{p+2})
\equiv T_1 + T_2.
\]

Now arrange the \( N_{p+1} \) elements of the derivatives \( D^k m(\bar{z})/k! \) for \( |k| = p + 1 \) as a column vector \( m_{p+1}(\bar{z}) \) using the lexicographical order introduced earlier and define \( m_{p+2}(\bar{z}) \) in the similar way. Let the \( N \times N_{p+1} \) matrix \( B_{n1} \) and the \( N \times N_{p+2} \) matrix \( B_{n2} \) be defined as

\[
B_{n1}(\bar{z}) = \begin{bmatrix}
S_{n,0,p+1}(\bar{z}) \\
S_{n,1,p+1}(\bar{z}) \\
\vdots \\
S_{n,p,p+1}(\bar{z})
\end{bmatrix}, \quad B_{n2}(\bar{z}) = \begin{bmatrix}
S_{n,0,p+2}(\bar{z}) \\
S_{n,1,p+2}(\bar{z}) \\
\vdots \\
S_{n,p,p+2}(\bar{z})
\end{bmatrix},
\]

where \( S_{n,i,p+1}(\bar{z}) \) and \( S_{n,i,p+2}(\bar{z}) \) is as given by (16). Therefore, \( T_1 = h^{p+1} B_{n1}(\bar{z}) m_{p+1}(\bar{z}), \)
\( T_2 = h^{p+2} B_{n2}(\bar{z}) m_{p+2}(\bar{z}) \), and

\[
E\beta_n^*(\bar{z}) = -W_p h^{p+1} S_{n,p}^{-1}(\bar{z}) B_{n1}(\bar{z}) m_{p+1}(\bar{z}) - W_p h^{p+2} S_{n,p}^{-1}(\bar{z}) B_{n2}(\bar{z}) m_{p+2}(\bar{z}) + o(h^{p+2}).
\]
Let $e_i$, $i = 1, \cdots, d$ be the $d \times 1$ vector having 1 in the $i$th entry and all other entries 0. For $0 \leq j \leq p$, $0 \leq k \leq p + 1$, let $N_{j,k}(x)$ be the $N_j \times N_k$ matrix with its $(l, m)$ element given by

$$\left[ N_{j,k}(x) \right]_{l,m} = \sum_{i=1}^{d} D_{2i} \{ f g \}(x) \int K(u) u^{r_j(l)+r_k(m)+\nu} du,$$

(23)

and use these $N_{j,k}(x)$ to construct a $N \times N$ matrix $N_p(x)$ and a $N \times N_{p+1}$ matrix $\tilde{M}(x)$ via

$$N_p(x) = \begin{bmatrix} N_{0,0}(x) & N_{0,1}(x) & \cdots & N_{0,p}(x) \\ N_{1,0}(x) & N_{1,1}(x) & \cdots & N_{1,p}(x) \\ \vdots & \vdots & \ddots & \vdots \\ N_{p,0}(x) & N_{p,1}(x) & \cdots & N_{p,p}(x) \end{bmatrix}, \quad \tilde{M}(x) = \begin{bmatrix} N_{0,p+1}(x) \\ N_{1,p+1}(x) \\ \vdots \\ N_{p,p+1}(x) \end{bmatrix}.$$  

Then $S_{n,p}(x) = \{ f g \}(x)S_p + hN_p(x) + O(h^2)$, $B_{n1}(x) = \{ f g \}(x)B_1 + h\tilde{M}(x) + O(h^2)$ and $B_{n2}(x) = \{ f g \}(x)B_2 + O(h)$. As $S_{n,p}(x) = \{ f g \}^{-1}(x)S_p^{-1} - h\{ f g \}^{-2}(x)S_p^{-1}N_p(x)S_p^{-1} + O(h^2)$, we have

$$-E\beta^*_n(x) = W_p h^{p+1} \left[ \{ f g \}^{-1}(x)S_p^{-1} - h\{ f g \}^{-2}(x)S_p^{-1}N_p(x)S_p^{-1} \right] \left[ \{ f g \}(x)B_1 + h\tilde{M}(x) \right] \mathbf{m}_{p+1}(x) + W_p h^{p+2} \{ f g \}^{-1}(x)S_p^{-1} \{ f g \}(x)B_2 \mathbf{m}_{p+2}(x) + o(h^{p+2})$$

$$= h^{p+1} W_p S_p^{-1} B_1 \mathbf{m}_{p+1}(x) + h^{p+2} W_p S_p^{-1} \left[ \{ f g \}^{-1}(x) \mathbf{m}_{p+1}(x) \{ \tilde{M}(x) - N_p(x)S_p^{-1}B_1 \} + B_2 \mathbf{m}_{p+2}(x) \right] + o(h^{p+2}).$$

We claim that for elements $E\beta^*_n(x)$ of $E\beta^*_n(x)$ with $p - |r|$ even, the $h^{p+1}$ term will vanish. This means for any given $r$ with $|r| \leq p$ and $r_2$ with $|r_2| = p + 1$,

$$\sum_{0 \leq |r| \leq p} \left\{ S_p^{-1} \right\}_{N(r_1), N(r_2)} \nu_{r+r_2} = 0. \quad (24)$$

To prove this, first note that for any $r_1$ with $0 \leq |r_1| \leq p$ and $r_2$ with $|r_2| = p + 1$,

$$\sum_{0 \leq |r| \leq p} \left\{ S_p^{-1} \right\}_{N(r_1), N(r_2)} \nu_{r+r_2} = \int u^{r_2} K_{\Sigma, p}(u) du, \quad (25)$$

where $K_{\Sigma, p}(u) = \{|M_{\Sigma, p}(u)|/|S_p|\}K(u)$ and $M_{\Sigma, p}(u)$ is the same as $S_p$, but with the $N(r)$ column replaced by $\mu(u)$. Let $c_{ij}$ denote the cofactor of $\{S_p\}_{i,j}$ and expand the determinant of $M_{\Sigma, p}(u)$ along the $N(r)$ column. We see that

$$\int u^{r_2} K_{\Sigma, p}(u) du = |S_p|^{-1} \int \sum_{0 \leq |r| \leq p} c_{N(r_1), N(r_2)} u^{r_2+r} K(u) du.$$

15
(25) thus follows, because $c_{N(\xi),N(\xi)}/|S_p| = \{S_p^{-1}\}_{N(\xi),N(\xi)}$ from the symmetry of $S_p$ and a standard result concerning cofactors. As a generalization of Lemma 4 in Fan et al (1995) to multivariate case, we can further show that for any $\xi_1$ with $0 \leq |\xi_1| \leq p$ and $p - |\xi_1|$ even,

$$\int u^{r_2}K_{\xi,p}(u)du = 0,$$

for any $|r_2| = p + 1$,

which together with (25) leads to (24).

Proof of Theorem 3.2. Let $\lambda = \lambda(s)$. By Lemma 6.1 and Lemma 6.9, we know that with probability 1, for some $C_1 > 1$ and all large $M$,

$$\sup_{\xi \in D} \sup_{\alpha \in B_{(1)}^k} \left| \sum_{i=1}^n \Phi_{ni}(\xi;\alpha,\beta) - \frac{nh^d}{2}(H\alpha)^T Snp(\xi)H(\alpha + 2\beta) \right| \leq C_1 M^{3/2}(d_{n1} + d_n) \leq 2C_1 M^{3/2}(nh^d)^{1-\lambda_1}(\log n)^{2\lambda_1},$$

when $n$ is large, (26)

where $d_{n1} = (nh^d)^{1-\lambda_1-\lambda_2}(\log n)^{\lambda_1+2\lambda_2}$. Note that from (17), we can write

$$\sum_{i=1}^n K_{ni}\varphi(Y_i;\mu_{ni}\beta_p(\xi))\mu_{ni}^T \mu_{ni} \alpha = nh^d\beta_n^*(\xi)^TW_p^{-1}Snp(\xi)H\alpha.$$

Replace $B_{n1}^{(1)}$ in (26) with $B_{nk}^{(1)} = \{ \alpha \in \mathbb{R}^N : k \leq M^{-1}(nh^d/\log n)^{\lambda_1}|H\alpha| \leq k + 1 \}$ and $M$ with
\((k+1)M\). We have, by the definition of \(\Phi_{ni}(\underline{x};\alpha,\beta)\), that

\[
\inf_{\underline{x} \in \mathcal{D}} \inf_{\alpha \in B_{nk}^{(1)}} \inf_{\beta \in B_{nk}^{(2)}} \left\{ \sum_{i=1}^{n} \rho(Y_i; \mu_{ni}(\alpha + \beta + \beta_p(\underline{x})))K_{ni} - \sum_{i=1}^{n} \rho(Y_i; \mu_{ni}(\beta + \beta_p(\underline{x})))K_{ni} \right. \\
+ nh^d(W_p^{-1}\beta_n^*(\underline{x}) - H\beta)^\top S_{np}(\underline{x})H\alpha \} \\
\geq \inf_{\underline{x} \in \mathcal{D}} \inf_{\alpha \in B_{nk}^{(1)}} \inf_{\beta \in B_{nk}^{(2)}} \left\{ \frac{nh^d}{2}(H\alpha)^\top S_{np}(\underline{x})H\alpha - 2CM^{3/2}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} \right. \\
\geq \left\{ C_3(kM)^{2/2} - 2C_1(k+1)^{3/2}M^{3/2} \right\}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} \\
\geq (8 - 2^{5/2})C_1C_4^{3/2}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} > 0 \text{ almost surely,} \tag{27}
\]

where the last term is independent of the choice of \(k \geq 1\). The last inequality is derived as follows. As \(S_p > 0\), suppose its minimum eigenvalue is \(\tau_1 > 0\). As \(S_{np}(\underline{x}) \to g(\underline{x})f(\underline{x})S_p\) uniformly in \(\underline{x} \in \mathcal{D}\) by Lemma 6.8 and \(g(\underline{x})f(\underline{x})\) is bounded away from zero by (A5) and (14), there exists some constant \(C_3 > 0\), such that for all \(\underline{x} \in \mathcal{D}\), the minimum eigenvalue of \(S_{np}(\underline{x})\) is greater than \(C_3\). The last inequality thus holds if \(M \geq C_4 = (16C_1/C_3)^2\). Note that

\[
\bigcup_{k=1}^{\infty} B_{nk}^{(1)} = \left\{ \alpha | \in \mathcal{R}^N : \left( \frac{nh^d}{\log n} \right)^{\lambda_1} |H\alpha| \geq M \right\} := B_{nk}^{N}. \tag{28}
\]

Therefore, from (27) and (28), we have

\[
\inf_{\underline{x} \in \mathcal{D}} \inf_{\alpha \in B_{nk}^{(1)}} \inf_{\beta \in B_{nk}^{(2)}} \left\{ \sum_{i=1}^{n} \rho(Y_i; \mu_{ni}(\alpha + \beta + \beta_p(\underline{x})))K_{ni} - \sum_{i=1}^{n} \rho(Y_i; \mu_{ni}(\beta + \beta_p(\underline{x})))K_{ni} \\
+ nh^d(W_p^{-1}\beta_n^*(\underline{x}) - H\beta)^\top S_{np}(\underline{x})H\alpha \right. \} > 0 \text{ almost surely.} \tag{29}
\]

Note that by (30), Lemma 6.10 and Proposition 3.1, we have \(|\beta_n^*(\underline{x})| \leq C_3(nh^d/\log n)^{-\lambda_2}\) uniformly in \(\underline{x} \in \mathcal{D}\) almost surely. Namely, \(\beta_n^*(\underline{x}) \in B_{nk}^{(2)}\) for all \(\underline{x} \in \mathcal{D}\), if \(M > C_3^4\). This implies that if \(M > \max(C_3^4, C_4)\), (29) still holds with \(\beta\) replaced with \(H^{-1}W_p^{-1}\beta_n^*(\underline{x})\). Therefore,

\[
\inf_{\underline{x} \in \mathcal{D}} \inf_{\alpha \in B_{nk}^{(1)}} \left\{ \sum_{i=1}^{n} K_{ni}\rho(Y_i; \mu_{ni}(\alpha + H^{-1}W_p^{-1}\beta_n^*(\underline{x}) + \beta_p(\underline{x}))) \\
- \sum_{i=1}^{n} K_{ni}\rho(Y_i; \mu_{ni}(H^{-1}W_p^{-1}\beta_n^*(\underline{x}) + \beta_p(\underline{x}))) \right\} > 0,
\]

which is equivalent to Theorem 3.2. \(\square\)
Proof of Corollary 3.3. As $1 + \lambda_2 \geq 2\lambda_1$, it’s sufficient to prove that with probability 1,

\[ \beta_n(x) - E\beta_n(x) - \frac{1}{n h^d} W_pS_{n_p}^{-1}(x)H^{-1} \sum_{i=1}^{n} K_h(x - x_i)\varphi(x_i)\mu(x - x_i) = O\left(\frac{\log n}{n h^d}\right)^{(1+\lambda_2)/2}, \quad (30) \]

uniformly in $x \in D$. As $\varphi(x_i) \equiv \varphi(Y_i, m(X_i))$ and $E\varphi(x_i) = 0$, the term on the left hand side of (30) stands for

\[ W_pS_{n_p}^{-1}(x) \frac{1}{n h^d} \sum_{i=1}^{n} \{Z_{ni}(x) - EZ_{ni}(x)\}, \]

where

\[ Z_{ni}(x) = H^{-1}K_h(x - x_i)\mu(x - x_i)\left(\varphi(Y_i, \mu(x - x_i))^{\top}\beta_p(x) - \varphi(x_i)\right). \]

Next, like what we did in Lemma 6.1, we cover $D$ with number $T_n$ cubes $D_k = D_{n,k}$ with side length $l_n = O(T_n^{-1/d})$ and centers $x_k = x_{n,k}$. Write

\[
\sup_{x \in D} \left| \sum_{i=1}^{n} Z_{ni}(x) - EZ_{ni}(x) \right| \leq \max_{1 \leq k \leq T_n} \left| \sum_{i=1}^{n} Z_{ni}(x_k) - EZ_{ni}(x_k) \right| \\
+ \max_{1 \leq k \leq T_n} \sup_{x \in D_k} \left| \sum_{i=1}^{n} Z_{ni}(x) - Z_{ni}(x_k) \right| \\
+ \max_{1 \leq k \leq T_n} \sup_{x \in D_k} \left| \sum_{i=1}^{n} EZ_{ni}(x) - EZ_{ni}(x_k) \right| \\
=Q_1 + Q_2 + Q_3.
\]

As $Z_{ni}(x) - Z_{ni}(x_k) = H^{-1}K_h(x - x_i)\mu(x - x_i)\left(\varphi_{ni}(x, 0) - \varphi_{ni}(x_k, 0)\right)$, through approaches similar to that for $\xi_3$ in the proof of Lemma 6.2, we can show that

\[ Q_2 = O\left(\left(\frac{nh^d}{\log n}\right)^{(1-\lambda_2)/2} \log n\right) \]

almost surely

and so is $Q_3$. To bound $Q_1$, first note that $EZ_{ni}^2(x_k) = O(h^{p+1+d})$ uniformly in $i$ and $k$. As $|Z_{ni}(x)| \leq C$ for some constant $C$ by (A2), we can see that from Lemma 6.5

\[ \sum_{i=1}^{n} EZ_{ni}^2(x_k) + \sum_{i<j}^{} |Cov(Z_{ni}(x_k), Z_{nj}(x_k))| \leq C_2nh^{p+1+d}. \]

Finally by Lemma 6.4 with $B_1 = C_1$, $B_2 \equiv Cnh^{p+1+d}$, $\eta = A_3(nh^d/\log n)^{(1-\lambda_2)/2} \log n$ and $r_n = r(n)$, we have (note that $nB_1/\eta \to \infty$ indeed)

\[ \lambda_n\eta = A_3/(2C_1) \log n, \quad \lambda_n^2B_2 = C_2/(4C_1^2) \log n. \]
Therefore,
\[
P\left( \max_{1 \leq k \leq T_n} \left| \sum_{i=1}^{n} Z_{ni}(x_k) - EZ_{ni}(x_k) \right| \geq A_3(nh^d / \log n)^{(1-\lambda_2)/2} \log n \right) \leq T_n / n^a + CT_n \Psi_n,
\]
where \( a = A_3/(8C_1) - C_2/(4C^2_1) \). By selecting \( A_3 \) large enough, we can ensure that \( T_n / n^a \) is summable over \( n \). As \( T_n \Psi_n \) is summable over \( n \) from (11), we can conclude that
\[
Q_1 = O\left\{ \left( \frac{nh^d}{\log n} \right)^{(1-\lambda_2)/2} \log n \right\} \text{ almost surely.}
\]
This together with Lemma 6.8 completes the proof. \( \Box \)

**Proof of Corollary 4.1.** Through the proof lines for Theorem 3.2 and Corollary 3.3, it’s not difficult to see that Corollary 3.3 still holds under the conditions imposed here. Under the additive structure (19), we thus have
\[
\bar{\phi}_1(x_1) = \phi_1(x_1) + \frac{1}{n} \sum_{i=1}^{n} m_2(X_{2i}) - h^{p+1}e_1W_1S_p^{-1}B_1 \frac{1}{n} \sum_{i=1}^{n} m_{p+1}(x_1, X_{2i}) + \frac{1}{n^2h_1h^{d-1}}e_1 \sum_{j=1}^{n} \varphi(\varepsilon_j) \sum_{i=1}^{n} S_p^{-1}(x_1, X_{2i}) K(X_{1,xj}/h_1, X_{2,ij}/h) \mu(X_{1,xj}/h_1, X_{2,ij}/h)
\]
\[
+ o_p(\{\max(h_1,h)\}^{p+1}) + O_p(\{nh^d/(\log n)^{-3/4}\}),
\]
where \( X_{1,xj} = X_{1ij} - x, X_{2,ij} = X_{2i} - X_{2j} \) and \( e_1 \) is as in Proposition 3.1. Note that by (21), \( (nh^{1/2})(nh^d/(\log n)^{-3/4}) \rightarrow \infty \), the \( O_p(.) \) term can thus be safely ignored.

By central limit theorem for strongly mixing processes (Bosq, 1998, Theorem 1.7), we have
\[
\frac{1}{n} \sum_{i=1}^{n} m_2(X_{2i}) = O_p(n^{-1/2}), \quad \frac{1}{n} \sum_{i=1}^{n} m_{p+1}(x_1, X_{2i}) = E m_{p+1}(x_1, X_{2}) + O_p(n^{-1/2}).
\]
As the expectations of all other terms in (31) are 0, the leading term in the asymptotic bias of \( \bar{\phi}_1(x_1) - \phi_1(x_1) \) is thus given by
\[
- \{\max(h_1,h)\}^{p+1}e_1W_1S_p^{-1}B_1Em_{p+1}(x_1, X_{2}).
\]

Again through standard arguments in Masry (1996), we can see that
\[
\frac{1}{nh^d} \sum_{i=1}^{n} S_p^{-1}(x_1, X_{2i}) K_h(X_{1,xj}, X_{2,ij}) \mu(X_{1,xj}/h_1, X_{2,ij}/h)
\]
\[
= S_p^{-1}(x_1, X_{2j}) f_2(X_{2j}) \int_{[0,1]^{d-1}} \{K \mu\}(X_{1,xj}/h_1, v) dv \left\{ 1 + O \left( \left\{ \frac{\log n}{nh^{d-1}} \right\}^{1/2} \right) \right\}
\]

19
uniformly in \(1 \leq i \leq n\). Therefore, the leading term in the asymptotic variance of \(\hat{\phi}_1(x_1) - \phi_1(x_1)\) is the variance of the following term

\[
(nh_1)^{-1}e_1 \sum_{j=1}^{n} \varphi(\varepsilon_j)S_{np}^{-1}(x_1, X_{2j})f_2(X_{2j}) \int_{[0,1]^{d-1}} \{K\mu\}(X_{1,xj}/h_1, \varepsilon) d_2,
\]

which is asymptotically

\[
(nh_1)^{-1} \left\{ \int_{[0,1]^{d-1}} fg^{-1}(x_1, X_2)f_2^2(X_2)\sigma^2(x_1, X_2)dX_2 \right\} e_1 S_{p}^{-1} K_2 K_2^\top S_{p}^{-1} e_1. \tag{32}
\]

If \(\rho(y; \theta) = (2q - 1)(y - \theta) + |y - \theta|\) and \(\varphi(\theta) = 2qI\{\theta > 0\} + (2q - 2)I\{\theta < 0\}\), we have

\[
g(x) = 2f_\varepsilon(0|x)\text{ and } \sigma^2(x) = E[\varphi^2(\varepsilon)|X = x] = 4q^2 (1 - F_\varepsilon(0)) + 4(1 - q)^2 F_\varepsilon(0) = 4q(1 - q),
\]

which when substituted into (32), yields the asymptotic variance for the quantile regression estimator,

\[
\tilde{\sigma}^2(x_1) = q(1 - q) \left\{ \int_{[0,1]^{d-1}} f^{-1}(x_1, X_2)f_\varepsilon^{-2}(0|x_1, X_2)f_2^2(X_2)dX_2 \right\} e_1 S_{p}^{-1} K_2 K_2^\top S_{p}^{-1} e_1. \tag{33}
\]

### 6 Lemmas

**Lemma 6.1** Under assumptions \((A1) - (A6)\), we have for all large \(M\),

\[
\sup_{x \in D} \sup_{\alpha \in B_n^{(i)}(\alpha, \beta)} \left| \sum_{i=1}^{n} R_{ni}(x; \alpha, \beta) \right| \leq M^{3/2} d_n \text{ almost surely},
\]

where \(B_n^{(i)} = \{\beta \in R^N : |H_n\beta| \leq M_n^{(i)}\}, i = 1, 2\).

**Proof.** Since \(D\) is compact, it can be covered by a finite number \(T_n\) of cubes \(D_k = D_{n,k}\).
with side length \( l_n = O(T_n^{-1/d}) = O\{h(nh^d/\log n)^{-(1-\lambda_2)/2}\} \) and centers \( x_k = x_{n,k} \). Write

\[
\sup_{x \in D} \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^{n} R_{ni}(x; \alpha, \beta) \right| \leq \max_{1 \leq k \leq T_n} \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^{n} \Phi_{ni}(x_k; \alpha, \beta) - E\Phi_{ni}(x_k; \alpha, \beta) \right| \\
+ \max_{1 \leq k \leq T_n} \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^{n} \left\{ \Phi_{ni}(x_k; \alpha, \beta) - \Phi_{ni}(x_k; \alpha, \beta) \right\} \right| \\
+ \max_{1 \leq k \leq T_n} \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^{n} \left\{ E\Phi_{ni}(x_k; \alpha, \beta) - E\Phi_{ni}(x_k; \alpha, \beta) \right\} \right| \\
\equiv Q_1 + Q_2 + Q_3.
\]

In Lemma 6.2, it is shown that \( Q_2 \leq M^{3/2}d_n/3 \) almost surely and thus \( Q_3 \leq M^{3/2}d_n/3 \).

Now all we need to do is to quantify \( Q_1 \). To this end, we partition \( B_n^{(i)}, i = 1, 2 \), into a sequence of disjoint subrectangles \( D_1^{(i)}, \ldots, D_{J_1}^{(i)} \) such that

\[
|D_j^{(i)}| = \sup \left\{ |H_n(\alpha - \beta)| : \alpha, \beta \in D_j^{(i)} \right\} \leq 2M^{-1}M_n^{(i)}/\log n, \ 1 \leq j \leq J_1.
\]

Obviously \( J_1 \leq (M \log n)^N \). Choose a point \( \alpha_j \in D_j^{(1)} \) and \( \beta_{k_1} \in D_{k_1}^{(2)} \). Then

\[
Q_1 \leq \max_{1 \leq k \leq T_n} \sup_{1 \leq j, k_1 \leq J_1} \left| \sum_{i=1}^{n} \left\{ R_{ni}(x_k; \alpha_j, \beta_{k_1}) - R_{ni}(x_k; \alpha, \beta) \right\} \right| \\
+ \max_{1 \leq k \leq T_n} \sup_{1 \leq j, k_1 \leq J_1} \left| \sum_{i=1}^{n} R_{ni}(x_k; \alpha_j, \beta_{k_1}) \right| = H_{n1} + H_{n2}.
\]  

(34)

We first consider \( H_{n1} \). For each \( j_1 = 1, \ldots, J_1 \) and \( i = 1, 2 \), partition each rectangle \( D_j^{(i)} \) further into a sequence of subrectangles \( D_{j_1}^{(i)}, \ldots, D_{j_1}^{(i)} \). Repeat this process recursively as follows. Suppose after the \( l \)th round, we get a sequence of rectangles \( D_j^{(i)} \) with \( 1 \leq j_k \leq J_k, 1 \leq k \leq l \), then in the \((l+1)\)th round, each rectangle \( D_j^{(i)} \) is partitioned into a sequence of subrectangles \( \{D_{j_1}^{(i)}, \ldots, j_{l+1}^{(i)}, 1 \leq j_l \leq J_l\} \) such that

\[
|D_j^{(i)}| = \sup \left\{ |H_n(\alpha - \beta)| : \alpha, \beta \in D_j^{(i)} \right\} \leq 2M_n^{(i)}/(M^l \log n), \ 1 \leq j_{l+1} \leq J_{l+1},
\]

where \( J_{l+1} \leq M^N \). End this process after the \((L_n + 1)\)th round, with \( L_n \) given at the beginning of Section 3. Let \( D_l^{(i)}, i = 1, 2 \), denote the set of all subrectangles of \( D_0^{(i)} \) after the \( l \)th round of
partition and a typical element $D_{j_1,j_2,...,j_l}$ of $D_l$ is denoted as $D_{(j_l)}$. Choose a point $\alpha_{(j_l)} \in D_{(j_l)}$ and $\beta_{(j_l)} \in D_{(j_l)}$ and define

$$V_l = \sum_{(j_l), (k_l)} \sum_{i=1}^{n} \mathbb{P}\left\{ \left| \sum_{r=i}^{n} \{ R_{ni}(\bar{x}_k; \alpha_{j_l}, \beta_{k_l}) - R_{ni}(\bar{x}_k; \alpha_{j_l+1}, \beta_{j_l+1}) \} \right| \geq \frac{M^{3/2}d_n}{2l} \right\}, \quad 1 \leq l \leq L_n,$$

$$Q_l = \sum_{(j_l), (k_l)} \sum_{i=1}^{n} \mathbb{P}\left\{ \sup_{\alpha \in D_{(j_l)}} \sum_{i=1}^{n} \left| \sum_{r=i}^{n} \{ R_{ni}(\bar{x}_k; \alpha_{j_l}, \beta_{k_l}) - R_{ni}(\bar{x}_k; \alpha, \beta) \} \right| \geq \frac{M^{3/2}d_n}{2l} \right\}, \quad 1 \leq l \leq L_n + 1.$$ 

By (A4), it is easy to see that for any $\alpha \in D_{(j_{l+n+1})} \in D_{l+n+1}$ and $\beta \in D_{(k_{l+n+1})} \in D_{l+n+1},$

$$|R_{ni}(\bar{x}_k; \alpha, \beta) - R_{ni}(\bar{x}_k; \alpha_{j_{l+n+1}}, \beta_{k_{l+n+1}})| \leq \frac{CM^{(2)}_n}{M^{l+n+1}\log n},$$

which together with the choice of $L_n$ implies that $Q_{l+n+1} = 0$. As $Q_l \leq V_l + Q_l, \quad 1 \leq l \leq L_n,$

$$P(H_{n1} > \frac{M^{3/2}d_n}{2}) \leq T_n Q_1 \leq T_n \sum_{l=1}^{L_n} V_l. \quad (35)$$

To quantify $V_l$, let

$$W_n = \sum_{i=1}^{n} Z_{ni} , \quad Z_{ni} \equiv R_{ni}(\bar{x}_k; \alpha_{j_l}, \beta_{k_l}) - R_{ni}(\bar{x}_k; \alpha_{j_l+1}, \beta_{j_l+1}). \quad (36)$$

Note that by (A2), we have, uniformly in $\bar{x}, \alpha$ and $\beta$, that

$$|\Phi_{ni}(\bar{x}; \alpha, \beta)| \leq CM^{(1)}_n. \quad (37)$$

Therefore, $|Z_{ni}| \leq CM^{(1)}_n$. With Lemma 6.6, we can apply Lemma 6.4 to $V_l$ with

$$B_1 = C_1M^{(1)}_n, \quad B_2 = nh^d(M^{(1)}_n)^2M^{(2)}_n \{M^l \log n\}^{-2/\nu_2},$$

$$r_n = r_n^l \equiv (2^{\nu_2/2} / M^2)^{2l/\nu_2} r(n), \quad q = n/r_n^l, \quad \eta = M^{3/2}d_n/2l,$$

$$\lambda_n = (2C_1M^{(1)}_n r_n^l)^{-1}, \quad \Psi(n) = Cq^{3/2}/\eta^{1/2} \gamma [r_n^l] r_n^l M^{(1)}_n \}^{1/2}.$$ 

Note that $nM^{(1)}_n / \eta \to \infty, r_n^l \to \infty$ for all $1 \leq l \leq L_n$ from (9) and

$$\lambda \eta = CM^{1/2} \log n M^{2l/\nu_2} / 2^{2l}, \quad \lambda^2 B_2 = C \log n^{1-2/\nu_2} M^{2l/\nu_2} / 2^{2l} = o(\lambda \eta),$$

which hold uniformly for all $1 \leq l \leq L_n$. Therefore,

$$V_l \leq \left( \prod_{j=1}^{l+1} J_j^2 \right) 4 \exp \{-C_1 \log n (M/2^{\nu_2})^{2l/\nu_2} \} + C_2 \tau_n^l,$$
where, as \( J_1 \leq 2(M \log n)^N \) and \( J_l \leq 2M^N \) for \( 2 \leq l \leq L_n \), \( \tau_n^l \) is given by
\[
\tau_n^l = 4^l M^{2N(l+1)}(\log n)^{2N} n^{3/2} \frac{\gamma[r_n^l] \{M_n^{(1)}\}^{1/2}}{r_n^l (d_n)^{1/2}}.
\]

It is tedious but easy to check that for \( M \) large enough,
\[
T_n \sum_{l=1}^{L_n} \left( \prod_{j=1}^{l+1} J_j^2 \right) 4 \exp \{ -C_1 \log n(M/2^{\nu_2})^{2/\nu_2} \}
\]
is summable over \( n \). (38)

As \( \gamma[r_n]/r_n^l \) is increasing in \( l \), we have
\[
T_n \sum_{l=1}^{L_n} \tau_n^l \leq T_n (\log n)^{2N} n^{3/2} \left\{ \frac{M_n^{(1)}}{d_n} \right\}^{1/2} \frac{\gamma[r_n^l]}{r_n^l} \prod_{l=1}^{L_n} 4^l M^{2N(l+1)},
\]
which is again summable over \( n \) according to (11). This along with (35) and (38) implies that \( H_{n1} \leq M^{3/2} d_{n}/2 \) almost surely, by the Borel-Cantelli lemma.

For \( H_{n2} \), first note that
\[
P(H_{n2} > \eta) \leq T_n J_1^2 \sup_{\bar{z} \in D} \sup_{\alpha \in \beta_n^{(1)}} P(\sum_{i=1}^{n} R_{ni}(\bar{z}; \alpha, \beta) > \eta).
\]
(39)

For any given \( \alpha, \beta \), using the facts along with Lemma 6.7, we apply Lemma 6.4 to quantify
\[
P(\sum_{i=1}^{n} R_{ni}(\bar{z}; \alpha, \beta) > \eta),
\]
with \( r_n = r(n) \), \( B_1 = 2C_1 M_n^{(1)} \), \( B_2 = C_2 nh^d(M_n^{(1)})^2 M_n^{(2)} \), \( \lambda_n = \{r(n)M_n^{(1)}\}^{-1}/4C_1 \) and \( \eta = M^{3/2} d_{n} \). Note that \( nB_1/\eta \to \infty \), and
\[
\lambda_n \eta/4 = (nh^d)^{1-\lambda_2}/(\log n)^{(1+\lambda_2)/2}/(16C_1 r(n)) = M^{1/2} \log n/(16C_1),
\]
\[
\lambda_n^2 B_2 = M^{1/4}(nh^d)^{1-\lambda_2}/(\log n)^{\lambda_2}/(16C_1^2 r^2(n)) = M^{1/4} \log n/(16C_1^2),
\]
\[
\Psi(n) \equiv q_n \{nB_1/\eta\}^{1/2} \gamma[r_n] = T_n J_1^2 q(n)^{3/2}/\eta^{1/2} \gamma[r(n)] \{r(n)M_n^{(1)}\}^{1/2},
\]
where \( \Psi(n) \) is summable over \( n \) by condition (11). Therefore,
\[
P(H_{n2} > \eta) \leq 2T_n J_1^2/n^b + \Psi(n), \quad b = \frac{1}{16C_1} (M^{1/2} - M^{1/4} C_2/C_1).
\]
(40)

By selecting \( M \) large enough, we can ensure that (40) is summable. Thus, for \( M \) large enough, \( H_{n2} \leq M^{3/2} d_{n} \) almost surely. By (34), we know for large \( M \), \( Q_1 \leq M^{3/2} d_{n} \) almost surely. \( \square \)

The quantification of \( Q_2 \) is very involved, so we put it as a separate Lemma.
Lemma 6.2 Under the conditions in Lemma 6.1, \( Q_2 \leq M^{3/2}d_n/3 \) almost surely.

Proof. Let \( X_{ik} = X_i - \bar{x}_k \), \( \mu_{ik} = \mu(X_{ik}) \) and \( K_{ik} = K_h(X_{ik}) \). It is easy to see that we can write \( \Phi_{ni}(\bar{x}; \alpha, \beta) - \Phi_{ni}(x; \alpha, \beta) = \xi_{i1} + \xi_{i2} + \xi_{i3} \), where

\[
\xi_{i1} = (K_{ik} \mu_{ik} - K_{ix} \mu_{ix})^\top \alpha \int_0^1 \left\{ \varphi_{ni}(\bar{x}_k; \mu_{ik}(\beta + \alpha t)) - \varphi_{ni}(\bar{x}_k; 0) \right\} dt,
\]

\[
\xi_{i2} = K_{ix} \mu_{ix} \alpha \int_0^1 \left\{ \varphi_{ni}(\bar{x}_k; \mu_{ik}(\beta + \alpha t)) - \varphi_{ni}(x; \mu_{ix}(\beta + \alpha t)) \right\} dt,
\]

\[
\xi_{i3} = K_{ix} \mu_{ix} \alpha \{ \varphi_{ni}(x; 0) - \varphi_{ni}(\bar{x}_k; 0) \}.
\]

Therefore, \( P(Q_2 > M^{3/2}d_n/3) \leq T_n(P_{n1} + P_{n2} + P_{n3}) \), where

\[
P_{nj} = \max_{1 \leq k \leq T_n} \mathbb{P} \left( \sup_{\alpha \in B_n^{(1)}}, \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \xi_{ij} \right| \geq M^{3/2}d_n/9 \right), \quad j = 1, 2, 3.
\]

If \( \sum_n T_n P_{nj} < \infty \), \( j = 1, 2, 3 \), then by Borel-Cantelli lemma we have \( Q_2 \leq M^{3/2}d_n \) almost surely.

First we study \( P_{n1} \). For any fixed \( \alpha \in B_n^{(1)} \) and \( \beta \in B_n^{(2)} \), let \( I_{ik}^{\alpha, \beta} = 1 \), if there exists some interval \( [t_1, t_2] \subseteq [0, 1] \), such that there are discontinuity points of \( \varphi(Y_t; \theta) \) between \( \mu_{ik}(\beta p(\bar{x}_k) + \beta + \alpha t) \) and \( \mu_{ik}(\beta p(\bar{x}_k)) \) for all \( t \in [t_1, t_2] \); and \( I_{ik}^{\alpha, \beta} = 0 \), otherwise. Write \( \xi_{i1} = \xi_{i1} I_{ik}^{\alpha, \beta} + \xi_{i1}(1 - I_{ik}^{\alpha, \beta}) \). Note that by (A3), \( |(K_{ik} \mu_{ik} - K_{ix} \mu_{ix})^\top \alpha| \leq C_2 M_n^{(1)} l_n/h \). Then by (A2) and the fact that \( |\mu_{ik}(\beta + \alpha t)| \leq C M_n^{(2)} \), we have \( |\xi_{i1} (1 - I_{ik}^{\alpha, \beta})| \leq C M_n^{(2)} M_n^{(1)} l_n/h \) uniformly in \( i, \alpha, \beta \) and \( x \in D_k \). Therefore,

\[
P \left( \sup_{\alpha \in B_n^{(1)}}, \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \xi_{i1} (1 - I_{ik}^{\alpha, \beta}) \right| > \frac{M^{3/2}d_n}{18} \right) \leq P \left( \sum_{i=1}^n I \{ |X_{ik}| \leq 2h \} > \frac{M^{1/4}d_n}{18C} \right), \tag{41}
\]

where we have used the fact that \( \xi_{i1} = \xi_{i1} I \{ |X_{ik}| \leq 2h \} \) since \( l_n = o(h) \). By Lemma 6.5, it follows that \( \text{Var}(\sum_{i=1}^n I \{ |X_{ik}| \leq 2h \}) = O(nh^d) \). We can thus apply Lemma 6.4 to the term on the right hand side of (41) with \( B_1 = 1, \eta = M^{1/4}d/(18C), B_2 = nh^d, r_n = r(n) \). It’s easy to check that \( \lambda_n \eta = CM^{1/4} \log n (nh^d/\log n)^{(1+\lambda_2)/2}, \lambda_n B_2 = o(\lambda_n \eta) \) and \( T_n \Phi_n \) is summable over \( n \) under condition (11). Therefore we have proved that

\[
T_n P \left( \sup_{\alpha \in B_n^{(1)}}, \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \xi_{i1} (1 - I_{ik}^{\alpha, \beta}) \right| > \frac{M^{3/2}d_n}{18} \right) \text{ is summable over } n, \tag{42}
\]
and that \( \sum_n T_n P_{n1} < \infty \), is thus equivalent to

\[
T_n P \left( \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^n \xi_i I_{ik}^{\alpha,\beta} \right| > M^{3/2} d_n / 18 \right) \text{ is summable over } n. \tag{43}
\]

First note that \( I_{ik}^{\alpha,\beta} = I\{\xi_i \in S_{ik}^{\alpha,\beta}\} \), where

\[
S_{ik}^{\alpha,\beta} = \bigcup_{j=1}^m \bigcup_{t \in [0,1]} \left[ a_j - A(X_i, \bar{x}_k) + \mu_{ik}^\top (\beta + \alpha t), a_j - A(X_i, \bar{x}_k) \right]
\]

\[
\subseteq \bigcup_{j=1}^m \left[ a_j - CM_n(2), a_j + CM_n(2) \right] \equiv D_n, \text{ for some } C > 0,
\]

\[
A(x_1, x_2) = (p + 1) \sum_{|\beta|=p+1} \frac{1}{\beta!} (x_1 - x_2)^\beta \int_0^1 D^2 m(x_2 + w(x_1 - x_2)) (1 - w)^p dw,
\]

where the fact that \( A(X_i, \bar{x}_k) = O(h^{p+1}) = O(M_n^{(2)}) \) uniformly in \( i \) with \( |X_{ik}| \leq 2h \) is used in the derivation of \( S_{ik}^{\alpha,\beta} \subseteq D_n \). As \( I_{ik}^{\alpha,\beta} \leq I\{\xi_i \in D_n\} \), we have \( |\xi_1| I_{ik}^{\alpha,\beta} \leq |\xi_1| U_{ni} \), where \( U_{ni} \equiv I(|X_{ik}| \leq 2h) I\{\xi_i \in D_n\} \), which is independent of the choice of \( \alpha \) and \( \beta \). Therefore,

\[
P \left( \sup_{\alpha \in B_n^{(1)}, \beta \in B_n^{(2)}} \left| \sum_{i=1}^n \xi_i I_{ik}^{\alpha,\beta} \right| > M^{3/2} d_n / 18 \right) \leq P \left( \sum_{i=1}^n U_{ni} > M^{1/2} n h d M_n^{(2)} / (18C) \right)
\]

\[
\leq P \left( \sum_{i=1}^n (U_{ni} - EU_{ni}) > \frac{M^{1/2} n h d M_n^{(2)}}{36C} \right), \tag{44}
\]

where the first inequality is because \( |\xi_1| \leq CM_n^{(1)} l_n / h \) and the second one because \( EU_{ni} = O(h d M_n^{(2)}) \) by (A1). As \( EU_{ni}^2 = EU_{ni} \), by Lemma 6.5, we know that \( \text{Var}(\sum_{i=1}^n U_{ni}) = C n h d M_n^{(2)} \).

We can then apply Lemma 6.4 to the last term in (44) with

\[
B_2 = C n h d M_n^{(2)}, \quad B_1 \equiv 1, \quad r_n = r(n), \quad \eta \equiv M^{1/2} n h d M_n^{(2)} / (36C).
\]

Apparently, \( \lambda_n \eta = C \log n (nh^d / \log n)^{(1-\lambda_2)/2} \) and \( \lambda_n^2 B_2 = o(\lambda_n \eta) \). As in this case \( T_n \Psi_n \) is still summable over \( n \) based on (11), (43) thus indeed holds.

For \( P_{n2} \), first note that using approach for \( P_{n1} \), we can show that

\[
\sum_{i=0}^{n-d} \{ \xi_{i2} - \tilde{\xi}_{i2} \} \leq M^{3/2} d_n / 18 \text{ almost surely,}
\]

25
where
\[
\tilde{\xi}_{i2} = K_{ik} \mu_{ik}^\top \alpha \int_0^1 \left\{ \varphi_{ni}(\mathbf{Z}_k; \mu_{ik}^\top (\beta + \alpha t)) - \varphi_{ni}(\mathbf{x}; \mu_{i\alpha}^\top (\beta + \alpha t)) \right\} dt.
\]

Therefore, we would have \( \sum T_n P_{n2} < \infty \), if
\[
T_n P \left( \sup_{\alpha \in B_n^{(1)}} \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \tilde{\xi}_{i2} \right| \geq M^{3/2} d_n / 18 \right) \text{ is summable over } n. \tag{45}
\]

For any fixed \( \alpha \in B_n^{(1)} \), \( \beta \in B_n^{(2)} \) and \( \mathbf{z} \in D_k \), let \( I_{i;k,x}^{\alpha,\beta} = 1 \), if there exists some interval \([t_1, t_2] \subseteq [0, 1]\), such that
\[
Y_i - \mu_{i\beta}^\top (\beta_p(\mathbf{z}) + \beta + \alpha t) \leq a_j \leq Y_i - \mu_{i\beta}^\top (\beta_p(\mathbf{x}) + \beta + \alpha t), \ t \in [t_1, t_2] \tag{46}
\]
with \( a_j \in \{a_1, \ldots, a_m\} \); and \( I_{i;k,x}^{\alpha,\beta} = 0 \), otherwise. Write \( \tilde{\xi}_{i2} = \tilde{\xi}_{i2} I_{i;k,x}^{\alpha,\beta} + \tilde{\xi}_{i2} (1 - I_{i;k,x}^{\alpha,\beta}) \). Note that
\[
K_{ik} \mu_{ik}^\top \alpha = O(M_{i}^{(1)}) \quad \text{and} \quad \varphi_{ni}(\mathbf{Z}_k; \mu_{ik}^\top (\beta + \alpha t)) - \varphi_{ni}(\mathbf{x}; \mu_{i\alpha}^\top (\beta + \alpha t)) = O(M_{i}^{(2)} l_n / h) \text{ if } I_{i;k,x}^{\alpha,\beta} = 0.
\]

Then again as \( \tilde{\xi}_{i2} = \tilde{\xi}_{i2} I\{ |X_{i;k}| \leq 2h \} \), we have similar to (42) that
\[
T_n P \left( \sup_{\alpha \in B_n^{(1)}} \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \tilde{\xi}_{i2}(1 - I_{i;k,x}^{\alpha,\beta}) \right| > M^{3/2} d_n / 18 \right) \text{ is summable over } n.
\]

Therefore, by (45), \( \sum T_n P_{n2} < \infty \), if it can be shown that
\[
T_n P \left( \sup_{\alpha \in B_n^{(1)}} \sup_{\beta \in B_n^{(2)}} \left| \sum_{i=1}^n \tilde{\xi}_{i2} I_{i;k,x}^{\alpha,\beta} \right| \geq M^{3/2} d_n / 36 \right) \text{ is summable over } n. \tag{47}
\]

To this end, define \( \epsilon_i = \epsilon_i + A(\mathbf{X}_i, \mathbf{Z}_k) \). Then \( I_{i;k,x}^{\alpha,\beta} = 1 \), i.e. (46) is equivalent to
\[
A(\mathbf{X}_i, \mathbf{Z}_k) - A(\mathbf{X}_i, \mathbf{x}) + \mu_{i\beta}^\top (\beta + \alpha t) \leq \epsilon_i - a_j \leq \mu_{i\beta}^\top (\beta + \alpha t), \ t \in [t_1, t_2]. \tag{48}
\]

Let \( \delta_n \equiv M_{i}^{(2)} l_n / h \). Then \( |A(\mathbf{X}_i, \mathbf{Z}_k) - A(\mathbf{X}_i, \mathbf{x})| \leq C \delta_n \) and \( |(\mu_{i\beta} - \mu_{i\alpha})^\top \beta| \leq C \delta_n \) and we can say that from (48),
\[
-2C \delta_n + \mu_{i\beta}^\top (\beta + \alpha t) \leq \epsilon_i - a_j \leq \mu_{i\beta}^\top (\beta + \alpha t) + 2C \delta_n, \ t \in [t_1, t_2]. \tag{49}
\]

Without loss of generality, assume \( \mu_{i\beta}^\top \alpha > 0 \). Then (49) implies that
\[
-2C \delta_n + \mu_{i\beta}^\top (\beta + \alpha t_2) \leq \epsilon_i - a_j \leq \mu_{i\beta}^\top (\beta + \alpha t_1) + 2C \delta_n, \tag{50}
\]
which in turn means that if $I^{α,β}_{i;k,x} = 1$, then $|ξ_i| ≤ C(t_2 - t_1)|μ_i^α|^α ≤ Cδ_n$, uniformly in $i, α ∈ B_n^{(1)}, β ∈ B_n^{(2)}$ and $x ∈ D_k$. Therefore, as $ξ_i = 2ξ_i I(∥X_{ik}∥ ≤ 2h)$, we have

$$P\left( \sup_{α ∈ B_n^{(1)}} \sup_{β ∈ B_n^{(2)}} \left| \sum_{i=1}^n \xi_i I^{α,β}_{i;k,x} \right| ≥ \frac{M^{3/2}d_n}{36} \right) \leq P\left( \sup_{α ∈ B_n^{(1)}} \sup_{β ∈ B_n^{(2)}} \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I^{α,β}_{i;k,x} \geq \frac{M^{5/4}nhdM_n^{(1)}}{36C} \right). \quad (51)$$

We will bound $I^{α,β}_{i;k,x}$ by a random variable that is independent of the choice of $α ∈ B_n^{(1)}$ and $x ∈ D_k$. By the definition of $I^{α,β}_{i;k,x}$ and (50), the necessary condition for $I^{α,β}_{i;k,x} = 1$ is given by

$$ε_i ∈ \bigcup_{j=1}^m [a_j + μ^αβ - 2M_n^{(1)}, a_j + μ^αβ + 2M_n^{(1)}] \equiv D_n^{β_i}, \quad (52)$$

which is indeed independent of the choice of $α$ and $x ∈ D_k$. Therefore,

$$P\left( \sup_{β ∈ B_n^{(2)}} \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I^{α,β}_{i;k,x} \geq \frac{M^{5/4}nhdM_n^{(1)}}{36C} \right) \leq P\left( \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I{ε_i ∈ D_n^{β_i}} \geq \frac{M^{5/4}nhdM_n^{(1)}}{36C} \right). \quad (53)$$

Now we partition $B_n^{(2)}$ into a sequence of subrectangles $S_1, \ldots, S_m$, such that

$$|S_l| = \sup \left\{ |H_\beta(β - β')| : β, β' ∈ S_l \right\} ≤ M_n^{(1)}, \quad 1 ≤ l ≤ m.$$}

Obviously, $m ≤ (M_n^{(2)}/M_n^{(1)})^N = M^{-3N/4}(nh^d/\log n)^{(λ_1 - λ_2)N}$. Choose a point $β_l ∈ S_l$ for each $1 ≤ l ≤ m$, and thus

$$P\left( \sup_{β ∈ B_n^{(2)}} \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I{ε_i ∈ D_n^{β}} \geq \frac{M^{5/4}nhdM_n^{(1)}}{36C} \right) \leq mP\left( \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I{ε_i ∈ D_n^{β_l}} \geq \frac{M^{5/4}nhdM_n^{(1)}}{72C} \right)$$

$$+ mP\left( \sup_{β' ∈ S_l} \sum_{i=1}^n I(∥X_{ik}∥ ≤ 2h) I{ε_i ∈ D_n^{β_l}} - I{ε_i ∈ D_n^{β'}} \geq \frac{M^{5/4}nhdM_n^{(1)}}{72C} \right) \equiv m(T_1 + T_2). \quad (54)$$
We deal with $T_1$ first. Let

$$U_{ni}^j \equiv I\{|X_{ik}| \leq 2h\}I\{\epsilon_i \in D_{ni}^\beta\}. \quad (55)$$

Then by the definition of $D_{ni}^\beta$ given in (52), $EU_{ni}^j = O(h^d M_n^{(1)}) < M^{5/4} n^d M_n^{(1)}/(144C)$ for large $M$ and we have

$$T_1 \leq P\left(\sum_{i=1}^n (U_{ni}^j - EU_{ni}^j) \geq \frac{M^{5/4} n h^d M_n^{(1)}}{144C}\right).$$

We can thus apply Lemma 6.4 to the quantity on the right hand side with $B_1 \equiv 1$, $B_2$ given by (66), $r_n = r(n)$ and $\eta \propto M^{5/4} n h^d M_n^{(1)}$, and $\lambda_n = 1/(2r_n)$. It follows that

$$\lambda_n \eta = CM^{5/4} \log n (nh^d/\log n)^{(1+\lambda_2)/2-\lambda_1}, \quad \lambda_n^2 B_2 = C \log n (nh^d/\log n)^{-2(\lambda_1-\lambda_2)/\nu_2}.$$ 

As $(1 + \lambda_2)/2 \geq \lambda_1$ and $\lambda_2 < \lambda_1$, we have $T_1 = O(n^{-b})$ for any $b > 0$.

For $T_2$, note that as $|\mu_{ik}^\top (\beta - \beta_l)| \leq CM_n^{(1)}$ for any $\beta \in S_l$, $1 \leq l \leq m$, we have

$$|I\{\epsilon_i \in D_{ni}^\beta\} - I\{\epsilon_i \in D_{ni}^\beta\}| = I\{\epsilon_i \in D_{ni}^\beta \setminus D_{ni}^\beta\}
\leq I\{\epsilon_i \in \bigcup_{j=1}^m [a_j + \mu_{ik}^\top \beta_l - CM_n^{(1)}, a_j + \mu_{ik}^\top \beta_l + CM_n^{(1)}]\} \equiv U_{ni},$$

for some $C > 0$, which is independent of the choice of $\beta \in S_l$. Therefore,

$$T_2 \leq P\left(\sum_{i=1}^n I\{|X_{ik}| \leq 2h\}U_{ni} \geq \frac{M^{5/4} n h^d M_n^{(1)}}{72C}\right),$$

which can be dealt with similarly as with $T_1$ and thus $T_2 = O(n^{-b})$ for any $b > 0$. Thus from (51), (53) and (54), we can claim that (47) is true and thus $T_n P_{n2}$ is summable over $n$.

The quantification of $P_{n3}$ is much simpler, as there is no $\beta$ involved in $\xi_{i3}$. For any given $x \in D_k$, let $I_{i;k,x} = 1$, if there is a discontinuity point of $\varphi(Y_i; \theta)$ between $\mu_{ik}^\top \beta_p(x_k)$ and $\mu_{ik}^\top \beta_p(x)$; and $I_{i;k,x} = 0$ otherwise. Write $\xi_{i3} = \xi_{i3} I_{i;k,x} + \xi_{i3}(1 - I_{i;k,x})$. Again by (A2) and the fact that $|K_{ix} \mu_{ix}^\top \alpha| = O(M_n^{(1)})$ and $|\mu_{ik}^\top \beta_p(x_k) - \mu_{ix}^\top \beta_p(x)| = |A(x_i, x_k) - A(x_i, x)| = O(M_n^{(2)} l_n/h)$, we have similar to (42) that

$$T_n P\left(\sup_{\alpha \in B_{n3}^{(1)}} \sup_{x \in D_k} \left|\sum_{i=1}^n \xi_{i3}(1 - I_{i;k,x})\right| > M^{3/2} d_n/18\right) \text{ is summable over } n.$$
It's easy to see that $I_{i;k,x} \leq I\{\varepsilon_i + A(X_i, x_k) \in S_{i;k,x}\}$, where

$$S_{i;k,x} = \bigcup_{j=1}^{m} \bigcup_{t \in [0,1]} \left[ a_j - |A(X_i, x_k) - A(X_i, \bar{x})|, a_j + |A(X_i, x_k) - A(X_i, \bar{x})| \right]$$

$$\subseteq \bigcup_{j=1}^{m} [a_j - CM_n^{(2)} l_n/h, a_j + CM_n^{(2)} l_n/h] \equiv D_n, \text{ for some } C > 0.$$ 

Therefore, $|\xi_{i3}|I_{i;k,x} = |\xi_{i3}|I\{||X_{i;k}| \leq 2h\}I_{i;k,x} \leq U_{ni}$, where

$$U_{ni} \equiv M_n^{(1)}I\{||X_{i;k}| \leq 2h\}I \{\varepsilon_i + A(X_i, x_k) \in D_n\},$$

which is independent of the choice of $\alpha \in B_n^{(1)}$ and $x \in D_k$. Thus

$$T_nP\left( \sup_{\alpha \in B_n^{(1)}, x \in D_k} \left| \sum_{i=1}^{n} \xi_{i3}I_{i;k,x} \right| > M^{3/2}d_n/18 \right) \leq T_nP\left( \sum_{i=1}^{n} |U_{ni} - EU_{ni}| > M^{3/2}d_n/36 \right), \quad (56)$$

where we have used the fact that $EU_{ni} = O(h^d M_n^{(1)} M_n^{(2)} l_n/h) = O(d_n/n)$. We will have

$$\sum T_nP_{n3} < \infty \text{ if the right hand side in (56) is summable over } n, \text{ i.e.}$$

$$T_nP\left( \sum_{i=1}^{n} |U_{ni} - EU_{ni}| > M^{3/2}d_n/36 \right) \text{ is summable over } n. \quad (57)$$

It's easy to check that Lemma 6.5 again holds with $\psi_\alpha(X_i, Y_i)$ standing for $U_{ni}$. Applying Lemma 6.4 to (57) with $B_1 \equiv M_n^{(1)}$, $B_2 \equiv Cnh^d (M_n^{(1)})^2 M_n^{(2)} l_n/h$, $\eta \equiv M^{3/2}d_n/36$ and $r_n = r(n)$, we have (note that $nB_1/\eta \to \infty$ indeed)

$$\lambda_n\eta/4 = CM^{1/2} \log n, \quad \lambda_n^2 B_2 = C r_n^{-2/\nu_2} \log n = o(\lambda_n\eta).$$

Thus, $T_n\Psi_n$ again is summable over $n$ and (57) indeed holds. \hfill \Box

The next Lemma is due to Davydov (Hall and Heyde (1980), Collary A2).

**Lemma 6.3** Suppose that $X$ and $Y$ are random variables which are $\mathcal{G}$− and $\mathcal{H}$− measurable, respectively, and that $E|X|^p < \infty$, $E|Y|^q < \infty$, where $p$, $q > 1$, $p^{-1} + q^{-1} < 1$. Then

$$|E X Y - EXE Y| \leq 8\|X\|_p\|Y\|_q\{a(\mathcal{G}, \mathcal{H})\}^{1-p^{-1}-q^{-1}}.$$

The next lemma is some excerpts from the proof of Theorem 2 in Masry (1996).
Lemma 6.4 Suppose \( \{Z_i\}_{i=1}^{\infty} \) is a zero-mean strictly stationary processes with strongly mixing coefficient \( \gamma[k] \), and that \( |Z_i| \leq B_1, \sum_{i=1}^{n} E Z_i^2 + \sum_{i<j} |\text{Cov}(Z_i, Z_j)| \leq B_2 \). Then for any \( \eta > 0 \) and integer series \( r_n \to \infty \), if \( nB_1/\eta \to \infty \) and \( q_n \equiv [n/r_n] \to \infty \), we have

\[
P(|\sum_{i=1}^{n} Z_i| \geq \eta) \leq 4 \exp\{-\frac{\lambda_n \eta}{4} + \lambda_n^2 B_2\} + \Psi(n),
\]

where \( \Psi(n) = q_n \{nB_1/\eta\}^{1/2} \gamma[r_n], \lambda_n = 1/\{2r_nB_1\} \).

Proof. We partition the set \( \{1, \cdots, n\} \) into \( 2q \equiv 2q_n \) consecutive blocks of size \( r \equiv r_n \) with \( n = 2qr + v \) and \( 0 \leq v < r \). Write

\[
V_n(j) = \sum_{i=(j-1)r+1}^{jr} Z_i, \quad j = 1, \cdots, 2q
\]

and

\[
W_n' = \sum_{j=1}^{q} V_n(2j-1), \quad W_n'' = \sum_{j=1}^{q} V_n(2j), \quad W_n''' = \sum_{i=2qr+1}^{n} Z_i.
\]

Then \( W_n = \sum_{i=1}^{n} Z_i = W_n' + W_n'' + W_n''' \). The contribution of \( W_n''' \) is negligible as it consists of at most \( r \) terms compared of \( qr \) terms in \( W_n' \) or \( W_n'' \). Then by the stationarity of the processes, for any \( \eta > 0 \),

\[
P(W_n > \eta) \leq P(W_n' > \eta/2) + P(W_n'' > \eta/2) = 2P(W_n' > \eta/2). \tag{58}
\]

To bound \( P(W_n' > \eta/2) \), using recursively Bradley’s Lemma, we can approximate the random variables \( V_n(1), V_n(3), \cdots, V_n(2q-1) \) by independent random variables \( V_n^*(1), V_n^*(3), \cdots, V_n^*(2q-1) \), which satisfy that for \( 1 \leq j \leq q \), \( V_n^*(2j-1) \) has the same distribution as \( V_n(2j-1) \) and

\[
P\left(|V_n^*(2j-1) - V_n(2j-1)| > u\right) \leq 18(\|V_n(2j-1)\|_\infty/u)^{1/2} \sup_{A,B} |P(AB) - P(A)P(B)|, \tag{59}
\]

where \( u \) is any positive value such that \( 0 < u \leq \|V_n(2j-1)\|_\infty < \infty \) and the supremum is taken over all sets of \( A \) and \( B \) in the \( \sigma \)-algebras of events generated by \( \{V_n(1), V_n(3), \cdots, V_n(2j-3)\} \) and \( V_n(2j-1) \) respectively. By the definition of \( V_n(j) \), we can see that \( \sup |P(AB) - P(A)P(B)| = \gamma[r_n] \). Write

\[
P(W_n' > \frac{\eta}{2}) \leq P\left(\left|\sum_{j=1}^{q} V_n^*(2j-1)\right| > \frac{\eta}{4}\right) + P\left(\left|\sum_{j=1}^{q} V_n(2j-1) - V_n^*(2j-1)\right| > \frac{\eta}{4}\right)
\]

\[
\equiv I_1 + I_2. \tag{60}
\]
We bound $I_1$ as follows. Let $\lambda = 1/(2B_1r)$. Since $|Z_i| \leq B_1$, $\lambda |V_n(j)| \leq 1/2$, then using the fact that $e^x \leq 1 + x + x^2/2$ holds for $|x| \leq 1/2$, we have

$$E\left( e^{\pm \lambda V_n^*(2j-1)} \right) \leq 1 + \lambda^2 E\{V_n(j)^2\} \leq e^{\lambda^2 E\{V_n^*(2j-1)^2\}}. \quad (61)$$

By Markov inequality, (61) and the independence of the $\{V_n^*(2j-1)\}_{j=1}^q$, we have

$$I_1 \leq e^{-\lambda \eta/4} \left[ E \exp \left( \lambda \sum_{j=1}^q V_n^*(2j-1) \right) + E \exp \left( -\lambda \sum_{j=1}^q V_n^*(2j-1) \right) \right]$$

$$\leq 2 \exp \left( -\lambda \eta/4 + \lambda^2 \sum_{j=1}^q E\{V_n^*(2j-1)^2\} \right)$$

$$\leq 2 \exp \left( -\lambda \eta/4 + C_2 \lambda^2 B_2 \right). \quad (62)$$

We now bound the term $I_2$ in (60). Notice that

$$I_2 \leq \sum_{j=1}^q P \left( \left| V_n(2j-1) - V_n^*(2j-1) \right| > \frac{\eta}{4q} \right).$$

If $\|V_n(2j-1)\|_\infty \geq \eta/(4q)$, substitute $\eta/(4q)$ for $u$ in (59),

$$I_2 \leq 18q \{(\|V_n(2j-1)\|/\eta)/(4q)\}^{1/2} \gamma[r_n] \leq C q^{3/2} \eta^{1/2} \gamma[r_n](r_n B_1)^{1/2}, \quad (63)$$

If $\|V_n(2j-1)\|_\infty < \eta/(4q)$, let $u \equiv \|V_n(2j-1)\|_\infty$ in (59) and we have

$$I_2 \leq C q \gamma[r_n],$$

which is of smaller order than (63), if $nB_1/\eta \to \infty$. Thus by (58), (60), (62) and (63),

$$P(W_n > \eta) \leq 4 \exp \{-\lambda_n \eta/4 + C_2 B_2 \lambda_n^2\} + C \Psi_n,$$

where the constant $C$ is independent of $n$. \qed

**Lemma 6.5** For any $x \in \mathbb{R}^d$, let $\psi_x(X_i, Y_i) = I(|X_i| \leq h)\psi_x(X_i, Y_i)$, a measurable function of $(X_i, Y_i)$ with $|\psi_x(X_i, Y_i)| \leq B$ and $V = E\psi_x^2(X_i, Y_i)$. Suppose the mixing coefficient $\gamma[k]$ satisfies (10). Then

$$\text{Cov} \left( \sum_{i=1}^n |\psi_x(X_i, Y_i)| \right) = n V \left[ 1 + o \left( \left( B^2 h^{p+d+1}/V \right)^{1-2/\nu_2} \right) \right].$$

31
\textbf{Proof.} Denote } \psi_{ix}(X_i, Y_i) \text{ by } \psi_{ix}. \text{ First note that }

\begin{align*}
V = E\psi_{ix}^2 &= h^d \int_{|u| \leq 1} E(\psi_{ix}^2 | X_j = x + hu) f(x + hu) du, \\
\sum_{i < j} |\text{Cov}(\psi_{ix}, \psi_{jx})| &= \sum_{l=1}^{n-d} (n-l-d+1) |\text{Cov}(\psi_{0x}, \psi_{lx})| \leq n \sum_{l=1}^{n-d} |\text{Cov}(\psi_{0x}, \psi_{lx})| \\
&= n^{d-1} + n \sum_{l=0}^{n} + n \sum_{l=\pi_n}^{n-d} = nJ_{21} + nJ_{22} + nJ_{23},
\end{align*}

where } \pi_n = h^{(\nu+1)/a}. \text{ For } J_{21}, \text{ there might be an overlap between the components of } X_0 \text{ and } X_J, \text{ for example, when } X_i = (X_{i-d}, \cdots, X_{i-1}), \text{ where } \{X_i\} \text{ is a univariate time series. Without loss of generality, let } u', u'' \text{ and } u''' \text{ of dimensions } l, d - l \text{ and } l \text{ respectively, be the } d + l \text{ distinct random variables in } (X_0/h, X_J/h). \text{ Write } u_1 = (u'^T, u''^T)^T \text{ and } u_2 = (u'^T, u'''^T)^T. \text{ Then by Cauchy inequality, we have }

\begin{align*}
&|\text{Cov}(\psi_{0x}, \psi_{lx})| \leq h^d V \int_{|u| < 1} \left| f(x + hu_1, x + hu_2, l) - f(x + hu_1, x + hu_2, l + d - 1) du'du''du''',
\end{align*}

where by (A4) and (A5), the integral is bounded. Therefore,

\begin{align*}
nJ_{21} \leq CnV \sum_{l=1}^{n} h^l = o(nV).
\end{align*}

For } J_{22}, \text{ there is no overlap between the components of } X_0 \text{ and } X_J. \text{ Let } X_0x = hu \text{ and } X_Jx = hv \text{ and we have }

\begin{align*}
|\text{Cov}(\psi_{0x}, \psi_{lx})| &\leq h^{2d} \int_{|u| < 1} E\left(\psi_{0x} \psi_{lx} | X_i = x + hu \right) du dv \\
&\times \left[ f(x + hu, x + hv; l + d - 1) - f(x + hu)\right] \\
&= Ch^d V,
\end{align*}

where the last equality follows from (A4), (A5) and (64). Therefore, as } \pi_n h^d \to 0,

\begin{align*}
nJ_{22} = O\{n\pi_n h^d V\} = o(nV).
\end{align*}
For $J_{23}$, using Davydov’s lemma (Lemma 6.3) we have

$$|\text{Cov}(\psi_{0x}, \psi_{lx})| \leq 8\gamma [l - d + 1]^{1-2/\nu_2} \{E|\psi_{lx}|^{\nu_2}\}^{2/\nu_2}, \quad \text{as } \nu_2 > 2. \quad (65)$$

As $|\psi_{lx}| \leq B$, $E|\Phi_{ni}|^{\nu_2} \leq B^{\nu_2-2}V$,

$$J_{23} \leq CB^{(\nu-2)/\nu_2}V^{2/\nu_2}/\pi_n^a \sum_{l=\pi_n+1}^{\infty} l^a \{\gamma[l - d + 1]\}^{1-2/\nu_2},$$

where the summation term is $o(1)$ as $\pi_n \to \infty$. Thus $J_{23} = o\left\{V \left(B^2 h^{d+1}/V\right)^{1-2/\nu_2}\right\}$, which completes the proof.

\[\square\]

**Lemma 6.6** Suppose (A2)- (A6) hold. Then for $U_{ni}^l, l = 1, \cdots, m$ defined in (55) and $Z_{ni}, l = 1, \cdots, L_n$ defined in (36), we have

$$\sum_{i=1}^{n} E(U_{ni}^l)^2 + \sum_{i<j} \text{Cov}(U_{ni}^l, U_{nj}^l) \leq Cn h^d M_n^{(1)} \{M_n^{(2)}/M_n^{(1)}\}^{1-2/\nu_2}, \quad (66)$$

$$\sum_{i=1}^{n} E Z_{ni}^2 + \sum_{i<j} \text{Cov}(Z_{ni}, Z_{nj}) = n h^d (M_n^{(1)})^2 M_n^{(2)} \{M^l \log n\}^{-2/\nu_2}, \quad (67)$$

uniformly in $x_k, 1 \leq k \leq T_n$.

**Proof.** We only prove (67), which is more involved than (66). To simplify the notations, denote $\alpha_{ji}, \beta_{ki}, \alpha_{ji}$ and $\beta_{ji}$ by $\alpha_1, \beta_1, \alpha_2$ and $\beta_2$, respectively. Clearly,

$$\int_{u} u^T H(\alpha_{i2}+\beta_2) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt = \int_{u} u^T H(\alpha_1+\beta_1) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt,$$

and

$$Z_{ni} = \int_{u} u^T H(\alpha_1+\beta_1) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt - \int_{u} u^T H(\alpha_{i2}+\beta_2) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt$$

$$= \int_{u} u^T H(\alpha_1+\beta_1) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt - \int_{u} u^T H(\alpha_{i2}+\beta_2) \{\varphi_{ni}(x_k; t) - \varphi_{ni}(x_k; 0)\} dt$$

$$- \int_{u} u^T H(\alpha_{i1}+\beta_1) \{\varphi_{ni}(x_k; t + u^T H(\beta_2 - \beta_1)) - \varphi_{ni}(x_k; 0)\} dt \equiv \Delta_1 + \Delta_2.$$
Therefore, $E\{Z_{ni}\}^2 = h^d \int K^2(u) f(x_k + hu) E\{(\Delta_1 + \Delta_2)^2 | X_i = x_k + hu \} du$. The conclusion is thus obvious observing that by Cauchy inequality and (12),

\[
E(\Delta_1^2 | X_i = x_k + hu) \leq |u^\top H\alpha_1 u^\top H(\beta_2 - \beta_1)u^\top H\alpha_1| \leq 2(M_n^{(1)})^2 M_n^{(2)}/(M^4 \log n),
\]

\[
E(\Delta_2^2 | X_i = x_k + hu) \leq \{u^\top H(\alpha_2 - \alpha_1)\}^2 (|u^\top H\alpha_1| + |u^\top H\alpha_1| + 2|u^\top H\beta_1|) \leq 4(M_n^{(1)})^2 M_n^{(2)}/(M^4 \log n),
\]

where we used the facts that $|\alpha_1 - \alpha_2| \leq 2M_n^{(1)}/(M^4 \log n)$ and $|\beta_1 - \beta_2| \leq 2M_n^{(2)}/(M^4 \log n)$. Therefore, $E\{Z_{ni}\}^2 = Ch^d (M_n^{(1)})^2 M_n^{(2)}/(M^4 \log n)$. As $|Z_{ni}| \leq CM_n^{(1)}$ and $h^{p + 1}/M_n^{(2)} < \infty$, the rest of the proof can be completed following the proof of Lemma 6.5.

Lemma 6.7 Suppose (A2)-(A6) hold.

\[
\sum_{i=1}^{n} E\Phi_{ni}^2 + \sum_{i<j} |\text{Cov}(\Phi_{ni}, \Phi_{nj})| \leq Cnh^d (M_n^{(1)})^2 M_n^{(2)},
\]

(68)

uniformly in $x \in \mathcal{D}, \alpha \in B_n^{(1)}, \beta \in B_n^{(2)}$.

**Proof.** By Cauchy inequality and (12), we have

\[
E\Phi_{ni}^2 = h^d \int K^2(u) \mathbb{E}\left[ \int_{\mu(u)^\top H(\alpha + \beta)}^{\mu(u)^\top H(\alpha + \beta)} (\varphi_{ni}(x; t) - \varphi_{ni}(x; 0)) dt \right]^2 |X_i = x + hu| f(x + hu) du
\]

\[
\leq h^d \int f(x + hu)K^2(u)\mu(u)^\top H\alpha \int_{\mu(u)^\top H\beta}^{\mu(u)^\top H(\alpha + \beta)} \mathbb{E}\left[ (\varphi_{ni}(x; t) - \varphi_{ni}(x; 0))^2 \right] |X_i = x + hu| dt du
\]

\[
\leq h^d \int K^2(u)\mu(u)^\top H\alpha \int_{\mu(u)^\top H\beta}^{\mu(u)^\top H(\alpha + \beta)} C|t| dt f(x + hu) du = O\left\{ h^d (M_n^{(1)})^2 M_n^{(2)} \right\},
\]

(69)

uniformly in $x \in \mathcal{D}$, $\alpha \in B_n^{(1)}$ and $x \in B_n^{(2)}$. (68) thus follows from (69) and Lemma 6.5.

Lemma 6.8 Let (A3) - (A6) hold. Then

\[
\sup_{x \in \mathcal{D}} \left| S_{np}(x) - g(x)f(x)S_p \right| = O(h + (nh^d/\log n)^{-1/2}) \text{ almost surely.}
\]

**Proof.** The result is almost the same as Theorem 2 in Masry (1996). Especially if (11) holds, then the requirement (3.8a) there on the mixing coefficient $\gamma[k]$ is met.
Lemma 6.9 Denote \( d_{n1} = (nh^d)^{1-\lambda_1-2\lambda_2} (\log n)^{\lambda_1+2\lambda_2} \) and let \( \lambda_1 \) and \( B_{ni}^{(i)} \), \( i = 1, 2 \), be as in Lemma 6.1. Suppose that (A1) – (A5) and (9) hold. Then there is a constant \( C > 0 \) such that for each \( M > 0 \) and all large \( n \),

\[
\sup_{x \in D} \sup_{\alpha \in B_{n1}^{(1)}, \beta \in B_{n1}^{(2)}} \left| \sum_{i=1}^{n} E \Phi_{ni}(x; \alpha, \beta) - \frac{nh^d}{2} (H\alpha)^T S_{np}(x) H(\alpha + 2\beta) \right| \leq CM^{3/2} d_{n1}.
\]

Proof. Recall that \( G(t, u) = E(\phi(Y; t)|X = u) \),

\[
E \Phi_{ni}(x; \alpha, \beta) = h^d \int K(y) f(x + hu) dy \times \int_{\mu(u)^T}^{\mu(u)^T + \beta} G(t + \mu(u)^T H\beta_p(x), x + hu) - G(\mu(u)^T H\beta_p(x), x + hu) \right\} dt.
\]

By (A3) and (A5), we have

\[
G(t + \mu(u)^T H\beta_p(x), x + hu) - G(\mu(u)^T H\beta_p(x), x + hu) = tG_1(\mu(u)^T H\beta_p(x), x + hu) + \frac{t^2}{2} G_2(\xi_n(t, u), x + hu),
\]

where \( \xi_n(t, u) \) falls between \( \mu(u)^T H\beta_p(x) \) and \( t + \mu(u)^T H\beta_p(x) \), and the term \( O(h^{p+1}) \) is uniform in \( x \in D \). Therefore, the inner integral in (70) is given by

\[
\frac{1}{2} g(x + hu)(H\alpha)^T \mu(u)^T H(\alpha + 2\beta) + O\left\{ M^{3/2} \left( \frac{\log n}{nh^d} \right)^{\lambda_1+2\lambda_2} \right\}
\]

uniformly in \( x \in D \), where we have used the fact that \( nh^{d+(p+1)/\lambda_2}/\log n < \infty \). By the definition of \( S_{np}(x) \), the proof is thus completed. \( \square \)

Lemma 6.10 Under conditions in Theorem 3.2, we have

\[
\sup_{x \in D} \left| \frac{1}{nh^d} W_p S_{n1}(x) H^{-1} \sum_{i=1}^{n} K_h(X_i - x) \phi(\xi_i) \mu(X_i - x) \right| = O\left\{ \left( \frac{\log n}{nh^d} \right)^{1/2} \right\} \text{ almost surely.}
\]

Proof. Note that, under conditions Theorem 3.2, the conditions imposed by Masry (1996) in Theorem 5 also hold. Specifically, (4.5) there follows from (9) and (4.7b) there can be derived from (11). Therefore, following the proof lines there, we can show that

\[
\sup_{x \in D} \left| \frac{1}{nh^d} H^{-1} \sum_{i=1}^{n} K_h(X_i - x) \phi(\xi_i) \mu(X_i - x) \right| = O\left\{ \left( \frac{\log n}{nh^d} \right)^{1/2} \right\},
\]

35
which together with Lemma 6.8, yields the desired results. □

REFERENCES


