Quantile estimation of a general single-index model

Efang Kong*
Technische Universiteit Eindhoven

Yingcun Xia†
National University of Singapore, Singapore

Abstract Single-index model is one of the most popular semiparametric model in Econometrics. In this paper, we define a quantile regressive single-index model, which includes the single-index structure for conditional mean and for conditional variance.

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

1 Introduction

Regression quantiles, along with the dual methods of regression rank scores, can be considered one of the major statistical breakthroughs of the past decades. Its advantages over the other estimation methods have been well investigated. Regression quantile methods provide a much more complete statistical analysis of the stochastic relationships among variables; in addition, they are more robust against possible outliers or extremely values, and can be computed via traditional linear programming methods. Although median regression ideas go back to the 18th century and the work of Laplace, regression quantile methods were first introduced by Koenker and Bassett (1978). The linear regression quantile is very useful, but like linear regression it is not flexible to capture complicated relations. For quantile regression, this disadvantage is even worse. As an example, consider the popular AR(1)-ARCH(1) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sigma_t z_t,\ z_t \sim \text{ IID}$$
$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2,\ \beta_0 > 0,\ \beta_1 \geq 0,$$

*Eurandom, Technische Universiteit Eindhoven, The Netherlands. E-mail address: `kong@eurandom.tue.nl`.
†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`.

which cannot be fitted well by the linear quantile model.

In this paper, we focus on an important special case when the loss function is specified as

$$\rho_\tau(v) = \tau I(v > 0)v + (\tau - 1)I(v \le 0)v, \tag{1}$$

where $0 < \tau < 1$ and $I(.)$ is the identity function, leading to the τth quantile regression, see Koenker and Bassett (1978).

Under nonparametric setting, we can state the problem as follows. Suppose Y is the response variable and $X \in R^d$ are the covariates. For loss function $\rho_\tau(.)$, we are interested in a function x, $m_\tau(x)$, such that

$$m_\tau(x) = \arg\min E\{\rho_\tau[Y - m_\tau(X)]\Big|X = x\} \quad \text{with respect to } m \in L_1 \tag{2}$$

Function $m_\tau(x)$ is called the $\tau-$th quantile nonparametric regression function of Y on X. The application of nonparametric quantile estimation has been intensively investigated in the literature. See for example Koenker (2005) and Kong et al (2008). As in nonparametric estimation of the conditional mean function, there is the "curse of dimensionality" in estimation the general multiple function $m_\tau(x)$. The dimension reduction approach can thus be applied here, which is equivalent to approximate

$$m_\tau(\theta^\top x) = \arg\min E\{\rho_\tau(Y - m(\theta^\top X))|X = x\} \quad \text{with respect to} \;\; \theta \in \Theta \text{ and } m \in L_1 \tag{3}$$

where $\Theta = \{\theta : |\theta| = 1\}$. More ideally, we come to a single-index quantile model

$$Y = m(\theta_0^\top X) + \varepsilon, \quad E(\varphi(\varepsilon)|X) = 0, \; a.s. \tag{4}$$

A typical model is the general single-index model,

$$Y = g(\theta_0^\top X, \varepsilon)$$

where ε is independent of X. Under such model specification, it is easy to see that

$$m_\tau(x) = g_\tau(\theta^\top x) \equiv \min_v\{v : P(g(\theta_0^\top x, \varepsilon) \le v) \ge \tau\}.$$

For the conditional heteroscadiscity model, where $g(\theta_0^\top X, \varepsilon) = g(\theta_0^\top X)\varepsilon$, we even have

$$m_\tau(x) = g(\theta_0^\top X)Q_\tau(\varepsilon)$$

where $Q_\tau(\varepsilon)$ is the $\tau-$th quantile of ε. An interesting special case for this setting is the ARCH(p) model, where $X = (y_{t-1}^2, ..., y_{t-p}^2)^\top$ and $Y = y_t$ in a time series setting.

Our main focus is the estimation of θ_0. Suppose $\{X_i, Y_i\}_{i=1}^n$ are I.I.D. observations from underlying model (4). We propose to estimate the index parameter θ_0 by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \min_{a_j,b_j} \sum_{i=1}^n \sum_{j=1}^n K(\theta^\top X_{ij}/h)\rho(Y_i - a_j - b_j\theta^\top X_{ij}), \quad X_{ij} = X_i - X_j \tag{5}$$

where $K(.)$ is a kernel function and h is a bandwidth. The minimization in (5) can be realized through iteration. First for any initial estimate $\vartheta \in \Theta$, denote by $[\hat{a}_\vartheta(x), \hat{b}_\vartheta(x)]$, the minimizer of

$$\sum_{i=1}^n K(\vartheta^\top X_{ix}/h)\rho(Y_i - a - b\theta^\top X_{ix}) \quad \text{with respect to } a \text{ and } b, \tag{6}$$

where $X_{ix} = X_i - x$. The estimate of θ_0 is then updated by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n \sum_{j=1}^n K(\vartheta^\top X_{ij}/h)\rho\{Y_i - \hat{a}_\vartheta(X_j) - \hat{b}_\vartheta(X_j)\theta^\top X_{ij}\} \tag{7}$$

Repeat (6) and (7) until convergence. The true value θ_0 is thus estimated by the standardized final estimate $\hat{\theta} := \hat{\theta}/|\hat{\theta}|$.

2 Numerical studies

Again, the calculation of the above minimization problem can be decomposed into two minimization problems.

- Fixing $\theta = \vartheta$ and $w_{ij}^\vartheta = K_h(\vartheta^\top X_{ij})$, the estimation of a_j and d_j are

$$\sum_{i=1}^n \rho\{Y_i - a_j - d_j\vartheta^\top X_{ij}\}w_{ij}^\vartheta.$$

- Fixing a_j and d_j, the minimization respect to θ can be done as follows. Again, let

$$Y_{ij}^\vartheta = Y_i(w_{ij}^\vartheta)^{1/2} - a_j(w_{ij}^\vartheta)^{1/2}, \quad X_{ij}^\vartheta = d_j X_{ij}(w_{ij}^\vartheta)^{1/2}.$$

  Then the problem becomes

$$\min_{\vartheta} \sum_{i,j=1}^n \rho\{Y_{ij}^\vartheta - \theta^\top X_{ij}^\vartheta\}$$

  Suppose the solution to the above problem is θ. Standardize it to $\theta := \theta/||\theta||$.

Set $\vartheta = \theta$ and repeat the two steps until convergence. Note that both steps are simple linear quantile regression problems and that several efficient algorithms are available, see Koenker (2005).

Example 2.1 (Single-index median regression) In this example we consider the following model

$$y = \exp\{-5(\theta_0^\top X)^2\} + \varepsilon, \tag{8}$$

where $X \sim \Sigma_0^{1/2} X_0$ with $X_0 \sim N(0, I_5)$ and $\Sigma_0 = (0.5^{|i-j|})_{0 \le i,j \le 5}$. For the noise term, we consider several distributions with both heavy tail and thin tails as well. For simplicity, we consider the median regression only. As a comparison, we also run the MAVE where a least square type estimation is used. With different sample sizes $n = 100, 200$, we carried out 100 replications. The calculation results are listed in Table 1.

Table 1: Estimation errors (and standard errors) for model (8) based on quadratic loss function and 50% quantiles

		Distribution of ε			
size	method	$0.05t(1)$	$0.1(N(0,1)^4 - 3)$	$\sqrt{5}t(5)/20$	N(0,1)/4
100	MAVE	0.3641(0.3526)	0.3530(0.3102)	0.0401(0.0182)	0.0581(0.0263)
	qMAVE	0.0902(0.1074)	0.1512(0.1957)	0.0833(0.0785)	0.1146(0.0651)
200	MAVE	0.3381(0.3389)	0.2859(0.2887)	0.0232(0.0091)	0.0373(0.0147)
	qMAVE	0.0681(0.1415)	0.0581(0.0698)	0.0402(0.0173)	0.0652(0.0272)

the MAVE method with quadratic loss function has very bad performance when the noise has heavy tail (e.g. $t(1)$) or is highly asymmetric (e.g. $N(0,1)^4$). With the absolute value loss function, the performance is much better. Even in the situation when the noise has thin tail and symmetric, qMAVE still performance reasonably well.

3 Assumptions and asymptotic properties

We adopt model (4) throughout and make the additional assumption that $\{(X_i, Y_i)\}_{i=1}^\infty$ are I.I.D. observations. The extension to the case of weakly dependent time series should be straightforward but complicates matters without adding anything conceptually. Furthermore, the following conditions are assumed in the proofs of Theorem 6.1.

(A1) For each $v \in \mathcal{R}$, $\rho(v)$ is absolutely continuous, i.e., there is a function $\varphi(.)$ such that $\rho(v) = \rho(0) + \int_0^v \varphi(t)dt$. The probability density function of ε_i is bounded and continuously differentiable. $E\{\varphi(\varepsilon_i)|X_i\} = 0$ almost surely and $E|\varphi(\varepsilon_i)|^{\nu_1} \le M_0 < \infty$ for some $\nu_1 > 2$.

(A2) $\varphi(.)$ satisfies the Lipschitz condition in (a_j, a_{j+1}), $j = 0, \cdots, m$, where $a_1 < \cdots < a_m$ are finite number of jump discontinuity points of $\varphi(.)$, $a_0 \equiv -\infty$ and $a_{m+1} \equiv +\infty$.

(A3) $K(.)$ has a compact support, say $[-1, 1]^{\otimes d}$ and $|u^j K(u) - v^j K(v)| \leq C\|u - v\|$ for all j with $0 \leq |j| \leq 3$.

(A4) The $m(.)$ defined in (4) is bounded with continuous and bounded partial derivatives up to the third order.

Note that (A1) and (A2) are satisfied in quantile regression. Based on (A1) and (A2), Hong (2003) proved that there is a constant $C > 0$, such that for all small t and all x,

$$E\Big[\{\varphi(Y - t - a) - \varphi(Y - a)\}^2 | X = x\Big] \leq C|t| \tag{9}$$

holds for all (a, x) in a neighborhood of $\{m(x^\top \theta_0), x\}$. Define

$$G(t; x) = E\{\rho\{Y - m(x^\top \theta_0) + t\} | X = x\}, \quad G_i(t, x) = (\partial^i / \partial t^i) G(t; x),\ i = 1, 2, 3. \tag{10}$$

Then it holds that

$$g(x) = G_2(0; x) \geq C > 0$$

and $G_3(t, x)$ is continuous and uniformly bounded for all $x \in \mathcal{D}$ and t near 0. For quantile regression, $g(x) = f_\varepsilon(0|x)$, where $f_\varepsilon(.|x)$ is the conditional probability density function of ε given $X = x$.

4 Initial estimate of θ_0

We use the average derivative estimation (ADE, Hardle and Stocker, 1989; Chaudhuri et al., 1997) method to obtain an initial estimate of θ_0, observing the fact that

$$E[\partial m(\theta_0^\top X)/\partial X] = \theta_0 E[\partial m(\theta_0^\top X)/\partial(\theta_0^\top X)].$$

First for any $x \in R^d$ and a kernel function $H(.) : R^d \to R^+$ which satisfies $(A3)$, denote by $[\hat{a}(x), \hat{b}(x)]$, the minimizer of the following quantity

$$\sum_{i=1}^{n} H(X_{ix}/h_0)\rho[Y_i - a - b^\top X_{ix}]$$

with respect to a and b. An initial estimate of θ_0 is thus defined as

$$\vartheta = \sum_{j=1}^{n} \hat{b}(X_j) \Big/ \Big| \sum_{j=1}^{n} \hat{b}(X_j) \Big|. \tag{11}$$

The consistency of ϑ is guaranteed by the uniform Bahadur representation of $\{\hat{a}(x), \hat{b}(x)\}$, i.e. with probability 1, for any compact set $\mathcal{D} \in R^d$ such that $f(x) > 0$,

$$\left[\begin{array}{c} \hat{a}(x) - m(\theta_0^\top x) \\ h_0\{\hat{b}(x) - m'(\theta_0^\top x)\theta_0\} \end{array} \right] = \beta_n(x) + O\{(nh_0^d/\log n)^{-3/4}\} \tag{12}$$

uniformly in $x \in \mathcal{D}$, where

$$\beta_n(x) = \frac{1}{nh_0^d} S_n^{-1}(x) \sum_{i=1}^n H(X_{ix}/h_0)\varphi\{Y_i - m(\theta_0^\top x) - m'(\theta_0^\top x)X_{ix}^\top\theta_0\} \left[\begin{array}{c} 1 \\ X_{ix}/h_0 \end{array} \right]$$

and $S_n(x)$ is the $(d+1)\times(d+1)$ matrix with its (j,k) entry given by

$$\nu_{n;j,k}(x) = \int K(u)g(x+h_0u)f(x+h_0u)u_{k-1}u_{j-1}du.$$

where $f(.)$ is the density function of X and $u = (u_1, \cdots, u_d) \in R^d$. If $nh_0^{d+4}/\log n < \infty$, then according to Proposition 3.1 and Corollary 3.3 in Kong et al (2007), we have with probability one,

$$\hat{b}(x) = m'(\theta_0^\top x)\theta_0 + \frac{1}{nh_0^{d+1}} S_n^{-1}(x) \sum_{i=1}^n H(X_{ix}/h_0)\varphi(\varepsilon_i) \left[\begin{array}{c} 1 \\ X_{ix}/h_0 \end{array} \right] + O\Big\{h_0^{-1}\Big(\frac{\log n}{nh_0^d}\Big)^{3/4}\Big\}, \tag{13}$$

uniformly in $x \in \mathcal{D}$. This in turn implies that with probability one,

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^n \hat{b}(X_j) &= m'(\theta_0^\top x)\theta_0 + \frac{1}{n^2h_0^{d+1}} \sum_{i,j=1}^n S_n^{-1}(X_j)H(X_{ij}/h_0)\varphi(\varepsilon_i) \left[\begin{array}{c} 1 \\ X_{ix}/h_0 \end{array} \right] \\ &\quad + O\Big\{h_0^{-1}\Big(\frac{\log n}{nh_0^d}\Big)^{3/4}\Big\}. \end{aligned}$$

Through arguments as in Masry (1996), we know that

$$\frac{1}{nh_0^d}\sum_{i=1}^n H(X_{ij}/h_0)\varphi(\varepsilon_i)\frac{X_{ij}}{h_0} = O\{(nh_0^d/\log n)^{-1/2}\},$$
$$\frac{1}{n^2h_0^{d+1}2}\sum_{i,j=1}^n S_n^{-1}(X_j)H(X_{ij}/h_0)\varphi(\varepsilon_i)\frac{X_{ij}}{h_0} = O\{h_0^{-1}(nh_0^d/\log n)^{-1/2}\}.$$

Therefore, we have established the convergence rate of the initial estimator ϑ in (11)

$$\delta_\vartheta \equiv \theta_0 - \vartheta = O\{h_0^{-1}(nh_0^d/\log n)^{-1/2}\}, \ a.s. \tag{14}$$

Next, we only need to consider parametric space $\Theta_n \equiv \{\vartheta : |\delta_\vartheta| < Ch_0^{-1}(nh_0^d/\log n)^{-1/2}\}$.

5 Asymptotics of $\hat{a}_\vartheta(x)$ and $\hat{b}_\vartheta(x)$

Let $a_j \equiv m(X_j^\top \theta_0)$ and $b_j \equiv m'(X_j^\top \theta_0)$. For any $\vartheta \in \Theta_n$, define

$$m_\vartheta(v) = \arg\min_a E\{\rho(Y-a)|X^\top\vartheta = v\}.$$

Let $f_\vartheta(x)$ (res. F_ϑ(x)) be the value of the probability density (res. distribution) function of $\vartheta^\top X$ at $\vartheta^\top x$ and define

$$G_\vartheta(t,x) = E\{\rho(Y - m_\vartheta(x) + t)|\vartheta^\top X = \vartheta^\top x\},\ G_\vartheta^i(t,x) = (\partial^i/\partial t^i)G_\vartheta(t,x),\ i = 1,2.$$

Suppose $G_\vartheta^2(t,x)$ is continuous and uniformly bounded in the neighborhood of $\{m_\vartheta(x), x\}$ and $g_\vartheta(x) = G_\vartheta^2(m_\vartheta(x),x) > 0$. With abuse of notation, let $m_\vartheta(X_j)$ and $m'_\vartheta(X_j)$ stand for $m_\vartheta(X_j^\top\vartheta)$ and $m'_\vartheta(X_j^\top\vartheta)$ respectively. Denote by $[\hat{a}_j, \hat{b}_j] \equiv [\hat{a}_\vartheta(X_j), \hat{b}_\vartheta(X_j)]$ the solution to (6) with $x = X_j$. Based on the uniform Bahadur representation of the local polynomial estimates of M-regression function (e.g., Kong et al, 2007), we have

$$\begin{bmatrix} \hat{a}_j - m_\vartheta(X_j) \\ h\{\hat{b}_j - m'_\vartheta(X_j)\}\end{bmatrix} = \frac{1}{nh} S_{nj}^{-1}\sum_{i=1}^n K_{ij}^\vartheta \varphi(Y_{ij}^*)\begin{bmatrix} 1 \\ X_{ij}^\top\vartheta/h\end{bmatrix} + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\}$$

uniformly in $X_j \in \mathcal{D}$, where $S_{nj} \equiv S_n(X_j) = \{gf\}_\vartheta(X_j)I[1 + O\{h + (nh/\log n)^{-1/2}\}]$, $K_{ij}^\vartheta = K(X_{ij}^\top\vartheta/h)$ and $Y_{ij}^* = Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X_{ij}^\top\vartheta$. Hereinafter, $\{gf\}_\vartheta(.) = g_\vartheta(.)f_\vartheta(.)$. Moreover, if $nh^5/\log n < \infty$,

$$\hat{a}_j - m_\vartheta(X_j) = \frac{1}{nh}\{gf\}_\vartheta^{-1}(X_j)\sum_{i=1}^n K_{ij}^\vartheta\varphi(Y_{ij}^*) + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\}, \tag{15}$$

$$h\{\hat{b}_j - m'_\vartheta(X_j)\} = \frac{1}{nh}\{gf\}_\vartheta^{-1}(X_j)\sum_{i=1}^n K_{ij}^\vartheta\varphi(Y_{ij}^*)X_{ij}^\top\vartheta/h + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\},$$

uniformly in $X_j \in \mathcal{D}$. In the Appendix, we further show that

$$m_\vartheta(X_j) - a_j = b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\} + o(|\delta_\vartheta|), \tag{16}$$

$$m'_\vartheta(X_j) - b_j = b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + o(|\delta_\vartheta|), \tag{17}$$

$$EK_{ij}^\vartheta\varphi(Y_{ij}^*) = \frac{1}{2}m''(X_j^\top\theta_0)(fg)_\vartheta(X_j)h^3 + O(h^4) + o(h\delta_\vartheta), \tag{18}$$

$$\begin{aligned} EK_{ij}^\vartheta\varphi(Y_{ij}^*)X_{ij}^\top\vartheta &= h^4\Big\{\frac{1}{2}m''(X_j^\top\theta_0)(f\mu)'_\vartheta(X_j) \\ &\quad -\frac{1}{6}m^{(3)}(X_j^\top\theta_0)(f\mu)_\vartheta(X_j)\Big\} + O(h^4\delta_\vartheta), \end{aligned} \tag{19}$$

where $(\nu/\mu)_\vartheta(X_j) \equiv \nu_\vartheta(X_j^\top\vartheta)/\mu_\vartheta(X_j^\top\vartheta)$ and

$$\mu_\vartheta(v) = E[g(X)|X^\top\vartheta = v], \quad \nu_\vartheta(v) = E[g(X)X|X^\top\vartheta = v]. \tag{20}$$

Therefore, equation (15) could be rewritten as

$$\begin{aligned}
\hat{a}_j - a_j &= \frac{1}{2}m''(X_j^\top\theta_0)(\mu/g)_\vartheta(X_j)h^2 + b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\} \\
&\quad + (nh)^{-1}\{gf\}_\vartheta^{-1}(X_j)\sum_{i=1}^n \varphi_1(X_i, Y_i) + O\Big\{\Big(\frac{\log n}{nh}\Big)^{3/4}\Big\} + O(h^4 + h\delta_\vartheta) \qquad (21) \\
\hat{b}_j - b_j &= h^2\Big[\frac{1}{2}m''(X_j^\top\theta_0)\{(f\mu)'/(fg)\}_\vartheta(X_j) - \frac{1}{6}m^{(3)}(X_j^\top\theta_0)\{(f\mu)/(fg)\}_\vartheta(X_j)\Big] \\
&\quad + b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + (nh^2)^{-1}\{gf\}_\vartheta^{-1}(X_j)\sum_{i=1}^n \varphi_2(X_i, Y_i) \\
&\quad + O\Big\{h^4 + h\delta_\vartheta + \Big(\frac{\log n}{nh}\Big)^{3/4}/h\Big\}
\end{aligned}$$

uniformly in j, where $\varphi_1(X_i, Y_i)$ and $\varphi_2(X_i, Y_i)$ are zero-mean I.I.D. random variables

$$\begin{aligned}
\varphi_1(X_i, Y_i) &= K_{ij}^\vartheta\varphi(Y_{ij}^*) - E[K_{ij}^\vartheta\varphi(Y_{ij}^*)]. \qquad (22) \\
\varphi_2(X_i, Y_i) &= K_{ij}^\vartheta\varphi(Y_{ij}^*)X_{ij}^\top\vartheta/h - E[K_{ij}^\vartheta\varphi(Y_{ij}^*)X_{ij}^\top\vartheta/h].
\end{aligned}$$

(21) is on the almost sure convergence of $\hat{a}_j$ and $\hat{b}_j$. As for the asymptotic bias and variance, Welsh (1996) showed that for any x with $f(.) > 0$ in a neighborhood of x,

$$E\hat{b}(x) = m_\vartheta(\vartheta^\top x) + O(h), \quad \mathrm{Var}\hat{b}(x) = O(n^{-1}h^{-3}), \qquad (23)$$

and the $O(.)$ are uniformly in x in any compact set on which $f(.)$ is strictly positive.

6 Asymptotics of $\hat{\theta}$

For previously obtained ϑ, $\hat{a}_j$, $\hat{b}_j$, $j = 1, \cdots, n$, suppose $\hat{\theta}$ minimizes $\tilde{\Phi}_n(\theta)$, where

$$\tilde{\Phi}_n(\theta) = \Phi_n(\theta) + \frac{n^2h}{2}(\theta - \vartheta)^\top\vartheta\vartheta^\top(\theta - \vartheta), \quad \Phi_n(\theta) = \sum_{i=1}^n\sum_{j=1}^n K_{ij}^\vartheta\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}).$$

Let $[\hat{a}_j,\ \hat{b}_j] \equiv [\hat{a}_\vartheta(X_j),\ \hat{b}_\vartheta(X_j)]$, $K_{ij}^\vartheta = K(X_{ij}^\top\vartheta/h)$ and $Y_{ij} \equiv Y_i - \hat{a}_j - \hat{b}_jX_{ij}^\top\theta_0$. Then with abuse of notations, $\hat{\theta}$ also minimizes

$$\begin{aligned}
\tilde{\Phi}_n(\theta) &= \Phi_n(\theta) + n^2h\{\frac{1}{2}(\theta - \theta_0)^\top\vartheta\vartheta^\top(\theta - \theta_0) + (\theta_0 - \vartheta)^\top\vartheta\vartheta^\top(\theta - \theta_0)\} \\
\Phi_n(\theta) &= \sum_{i=1}^n\sum_{j=1}^n K_{ij}^\vartheta\{\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_{ij})\}. \qquad (24)
\end{aligned}$$

As $|\vartheta - \theta_0| \le a_{n\vartheta}$, $\vartheta\vartheta^\top = \theta_0\theta_0^\top + O(a_{n\vartheta})$. Hence for any θ with $|\theta - \theta_0| \le a_{n\vartheta}$,

$$\tilde{\Phi}_n(\theta) = \Phi_n(\theta) + n^2h\{\frac{1}{2}(\theta - \theta_0)^\top\theta_0\theta_0^\top(\theta - \theta_0) + (\theta_0 - \vartheta)^\top\theta_0\theta_0^\top(\theta - \theta_0)\} + o(n^2ha_{n\vartheta}^2)$$

Write

$$\Phi_n(\theta) = E[\Phi_n(\theta)] + \delta_\theta^\top\{R_{n1}(\theta) - ER_{n1}(\theta)\} + R_{n2}(\theta) - ER_{n2}(\theta),$$

where

$$R_{n1} = \sum_{i,j} K_{ij}^\vartheta \varphi(Y_{ij})\hat{b}_j X_{ij}, \ R_{n2}(\theta) = \sum_{i,j} K_{ij}^\vartheta \Big[\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_{ij}) - \delta_\theta^\top \varphi(Y_{ij})\hat{b}_j X_{ij}\Big].$$

Applying the results on $E(\Phi_n(\theta))$ in Lemma 6.10, we have

$$\Phi_n(\theta) = \delta_\theta^\top R_{n1} + \frac{1}{2}\delta_\theta^\top G_{n2}\delta_\theta\{1 + o(1)\} + R_{n2}(\theta) - ER_{n2}(\theta), \tag{25}$$

where

$$\begin{aligned} G_{n2} &= \sum_{i,j} E[K_{ij}^\vartheta g(X_i)\hat{b}_j^2 X_{ij}X_{ij}^\top] = n^2 h S_2\{1 + o(1)\}, \\ S_2 &= \int \{m'(X^\top\theta_0)\}^2 \omega_{\theta_0}(X) f_{\theta_0}(X) dX, \end{aligned}$$

and $\omega_\vartheta(x) = E\{g(X)(X - x)(X - x)^\top | X^\top\vartheta = x^\top\vartheta\}$. Naturally,

$$\Phi_n(\theta) = \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) + \frac{1}{2}\delta_\theta^\top(G_{n2} + \theta_0\theta_0^\top)\delta_\theta\{1 + o(1)\} + R_{n2}(\theta) - ER_{n2}(\theta).$$

Our main result is as follows

Theorem 6.1 *Suppose (A1)-(A4) hold. With $\nu_\vartheta(.)$ and $\mu_\vartheta(.)$ as defined in (20), we have*

$$\begin{aligned} \hat{\theta} - \theta_0 &= (S_2 + \theta_0\theta_0^\top)^{-1}\frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) + (S_2 + \theta_0\theta_0^\top)^{-1}\theta_0\theta_0^\top\delta_\vartheta \\ &\quad + (S_2 + \theta_0\theta_0^\top)^{-1}\frac{1}{n}\sum_j b_j^2\{(\nu/\mu)_\vartheta(X_j) - X_j\} \times \{\nu_\vartheta(X_j) - X_j\mu_\vartheta(X_j)\}^\top\delta_\vartheta \\ &\quad + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2}) \qquad (26) \\ &= (S_2 + \theta_0\theta_0^\top)^{-1}\frac{1}{n}\sum_i \varphi(\varepsilon_i) b_i \{\varpi f\}_{\theta_0}(X_i) + (S_2 + \theta_0\theta_0^\top)^{-1}(\Omega_1 + \theta_0\theta_0^\top)\delta_\vartheta \\ &\quad + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2}) \qquad (27) \end{aligned}$$

almost surely, where $\varpi_\theta(x) = E(X|X^\top\theta = x^\top\theta) - x$, $\alpha_n = o(1)$ uniformly in ϑ and

$$\begin{aligned} \Omega_1 &\equiv \lim_{n\to\infty} n^{-1}\sum_j b_j^2 \mu_{\theta_0}(X_j)\{(\nu/\mu)_{\theta_0}(X_j) - X_j\}\{(\nu/\mu)_{\theta_0}(X_j) - X_j\}^\top \\ &= \int \{m'(X^\top\theta_0)\}^2 \mu_{\theta_0}(X)\{(\nu/\mu)_{\theta_0}(X) - X\}\{(\nu/\mu)_{\theta_0}(X) - X\}^\top dF_{\theta_0}(X) \end{aligned}$$

Remark 6.2 By multivariate Cauchy-Schwarz inequality, we have

$$E\{g(X)(X-x)|X^\top\vartheta = x^\top\vartheta\}E\{g(X)(X-x)^\top|X^\top\vartheta = x^\top\vartheta\}^\top$$
$$-E\{g(X)|X^\top\vartheta = x^\top\vartheta\}E\{g(X)(X-x)(X-x)^\top|X^\top\vartheta = x^\top\vartheta\} \le 0.$$

This means that for any X,

$$\omega_\vartheta(X) - \mu_{\theta_0}(X)\{(\nu/\mu)_{\theta_0}(X) - X\}\{(\nu/\mu)_{\theta_0}(X) - X\}^\top \text{ is non-negative definite,}$$

whence $S_2 - \Omega_1 \ge 0$. In Lemma 6.15, we prove that if $\delta_\vartheta \ne 0$,

$$0 < |(S_2 + \theta_0\theta_0^\top)^{-1}(\Omega_1 + \theta_0\theta_0^\top)\delta_\vartheta|/|\delta_\vartheta| < 1. \tag{28}$$

This implies that the impact on $\hat\theta - \theta_0$ of the deviance between ϑ and θ_0 decreases geometrically.

Remark 6.3 We prove Theorem 6.1 under the assumption that $\{(X_i, Y_i)\}_{i=1}^\infty$ are I.I.D. observations. It is possible, however to extend this result for time series observations provided that the time dependency (usually measured by mixing coefficient) are weak enough. For example, the stationary $*-$ mixing processes, which satisfies

$$|P(AB) - P(B)P(A)| < \phi(k)P(B)P(A) \quad \text{and } \phi(k) \to 0, \qquad \text{as } k \to \infty,$$

for all $A \in \mathcal{F}_{-\infty}^a$, $B \in \mathcal{F}_{a+k}^\infty$ and $\mathcal{F}_a^b$ is the $\sigma-$algebra generated by $\{(X_i, Y_i)\}_{i=a}^b$.

The rationality behind the above conjecture is that most of the Lemmas which are used in the proof can be replaced by their counterparts in the time series setting, namely Lemma 6.5 by Theorem 1.4 in Bosq (1998), and Lemma 6.7 by Theorem 2 in Sen (1972), as $*-$ mixing implies all the other types of mixing conditions (Ibragimov et al, 1971). The only issue is results is yet unavailable on law of iterated logarithm for degenerated U-statistics of dependent observations; that is whether Lemma 6.6 is still true for $*-$ mixing processes. Heuristically it is, an evidence is that the corresponding normality is proved in Fan and Li (1999), i.e.

$$n^{-1}\sum_{i\ne j} g(X_i, X_j) \to N(0, \sigma^2) \quad \text{for some constant } \sigma > 0.$$

Proof of Theorem 6.1. Let $a_{n\vartheta} = \max\{(n\log\log n)^{-1/2}, |\delta_\vartheta|\}$. It suffices to prove that

$$\hat\theta - \theta_0 = \{n^2h(S_2 + \theta_0\theta_0^\top)\}^{-1}(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) \qquad a.e. \tag{29}$$

$$(n^2h)^{-1}R_{n1} = \tfrac{1}{n}\sum_i \varphi(\varepsilon_i)b_i\{\varpi f\}_{\theta_0}(X_i) + \Omega_1\delta_\vartheta + \alpha_n|\vartheta - \theta_0| + o(n^{-1/2}) \qquad a.e. \tag{30}$$

As the first step to prove (29), we show in Lemma 6.12 and Lemma 6.14 that for each fixed θ,

$$(n^2ha_{n\vartheta}^2)^{-1}[R_{n2}(\theta) - ER_{n2}(\theta)] = o(1) \text{ almost surely.} \quad (31)$$

This together with (25) and the fact that $G_{n2} = n^2hS_2\{1 + o(1)\}$ imply that for any fixed θ,

$$(n^2ha_{n\vartheta}^2)^{-1}[\tilde{\Phi}_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta] \to 0 \text{ almost surely.}$$

As both $\tilde{\Phi}_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta)$ and $\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta$ are convex in θ, it follows from Lemma 6.4 that for any compact set $\Theta_{n\theta} \subset \Theta_n$(convex open set),

$$\sup_{\theta\in\Theta_{n\theta}} (n^2ha_{n\vartheta}^2)^{-1}|\tilde{\Phi}_n(\theta) - \delta_\theta^\top(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta| \to 0 \text{ almost surely.} \quad (32)$$

Let $\eta_n = \{n^2h(S_2 + \theta_0\theta_0^\top)\}^{-1}(R_{n1} + \theta_0\theta_0^\top\delta_\vartheta)$. Now we are ready to prove the equivalent of (29), i.e. : with probability 1, for any $\delta > 0$, $|\hat{\theta} - \theta_0 - \eta_n|/a_{n\vartheta} \le \delta$ for large n.
First note that as $\theta_0 + \eta_n$ is bounded with probability 1, Θ_n can be chosen to contain B_n^δ(a closed ball with center $\theta_0 + \eta_n$ and radius $a_{n\vartheta}\delta$). Replace $\Theta_{n\theta}$ in (32) by B_n^δ, we have

$$\Delta_n \equiv \sup_{\theta\in B_n^\delta} (n^2ha_{n\vartheta}^2)^{-1}|\tilde{\Phi}_n(\theta) - \delta_\theta^\top(R_{n1} - \theta_0\theta_0^\top\delta_\vartheta) - \frac{1}{2}n^2h\delta_\theta^\top(S_2 + \theta_0\theta_0^\top)\delta_\theta| = o(1), \text{ a. e.} \quad (33)$$

Now consider the behavior of $\tilde{\Phi}_n(\theta)$ outside B_n^δ. Suppose $\theta = \theta_0 + \eta_n + a_{n\vartheta}\beta\nu$, for some $\beta > \delta$ and ν a unit vector. Define θ^* as the boundary point of B_n^δ that lies on the line segment from $\theta_0 + \eta_n$ to θ, i.e. $\theta^* = \theta_0 + \eta_n + a_{n\vartheta}\delta\nu$. Convexity of $\Phi_n(\theta)$ and the definition of Δ_n imply

$$\begin{aligned}
\frac{\delta}{\beta}\tilde{\Phi}_n(\theta) + (1 - \frac{\delta}{\beta})\tilde{\Phi}_n(\theta_0 + \eta_n) &\ge \tilde{\Phi}_n(\theta^*) \\
&\ge \frac{1}{2}n^2h\delta^2a_{n\vartheta}^2\nu^\top(S_2 + \theta_0\theta_0^\top)\nu \\
&\quad -\frac{1}{2}(n^2h)^{-1}R_{n1}^\top(S_2 + \theta_0\theta_0^\top)^{-1}R_{n1} - n^2ha_{n\vartheta}^2\Delta_n \\
&\ge \frac{1}{2}n^2h\delta^2a_{n\vartheta}^2\nu^\top(S_2 + \theta_0\theta_0^\top)\nu + \tilde{\Phi}_n(\theta_0 + \eta_n) - 2n^2ha_{n\vartheta}^2\Delta_n.
\end{aligned}$$

It follows that

$$\inf_{|\theta-\theta_0-\eta_n|>\delta a_{n\vartheta}} \tilde{\Phi}_n(\theta) \ge \tilde{\Phi}_n(\theta_0 + \eta_n) + \frac{\beta}{\delta}n^2ha_{n\vartheta}^2[\frac{1}{2}\delta^2\nu^\top(S_2 + \theta_0\theta_0^\top)\nu - 2\Delta_n].$$

As $S_2 + \theta_0\theta_0^\top$ are positive definite, then according to (33), with probability 1, $\delta^2\nu^\top S_2\nu > 4\Delta_n$ for large enough n. This implies that for any $\delta > 0$ and for large enough n, the minimum of $\tilde{\Phi}_n(\theta)$ must occur within B_n^δ. This implies (29).

To derive (30), recall that

$$R_{n1}(\theta) = \sum_{i,j} K_{ij}^{\vartheta}\varphi(\varepsilon_i)b_jX_{ij} + \sum_{i,j} K_{ij}^{\vartheta}\varphi(\varepsilon_i)(\hat{b}_j - b_j)X_{ij} + \sum_{i,j} K_{ij}^{\vartheta}\hat{b}_jX_{ij}\{\varphi(Y_{ij}) - \varphi(\varepsilon_i)\} \quad (34)$$

For the first term above, by Lemma 7.8 in Xia and Tong (2006), i.e.

$$\sup_{\vartheta\in\Theta}\Big|\frac{1}{n^2h}\sum_{i,j}\Big\{K(\vartheta^\top X_{ij}/h)\phi_j(\vartheta) - E_j[K(\vartheta^\top X_{ij}/h)\phi_j(\vartheta)]\Big\}\xi_i(\vartheta)\Big| = O\Big(\frac{\log n}{nh}\Big)$$

we have

$$(n^2h)^{-1}\sum_{i,j} K_{ij}^{\vartheta}\varphi(\varepsilon_i)b_jX_{ij} = n^{-1}\sum_i \varphi(\varepsilon_i)E[K_{ij}^{\vartheta}b_jX_{ij}/h] + O\{(nh/\log n)^{-1}\}$$

$$\begin{aligned} E_j[K_{ij}^{\vartheta}b_jX_{ij}/h] &= b_i\{\varpi f\}_{\vartheta}(X_i) + \delta_{\vartheta}m''(X_i\theta_0)\{\Sigma f\}_{\theta_0}(X_i) + o(|\delta_{\vartheta}|^2 + h^3) \\ &\quad + h^2[m''(X_i^\top\theta_0)\{\varpi f\}'_{\theta_0}(X_i) + m^{(3)}(X_i^\top\theta_0)\{\varpi f\}_{\theta_0}(X_i)] \end{aligned}$$

where $\varpi_{\vartheta} = E(X - x|X^{\vartheta} = x^\top\vartheta)$, $\Sigma_{\vartheta} = E((X-x)(X-x)^\top|X^{\vartheta} = x^\top\vartheta)$. Therefore,

$$\frac{1}{n^2h}\sum_{i,j} K_{ij}^{\vartheta}\varphi(\varepsilon_i)b_jX_{ij} = \frac{1}{n}\sum_i \varphi(\varepsilon_i)b_i\{\varpi f\}_{\theta_0}(X_i) + o\{n^{-1/2} + \delta_{\vartheta}\} \quad (35)$$

For the second and third term in (34), we will show in Lemma 6.11 that

$$\frac{1}{nh}\sum_i K_{ij}^{\vartheta}\varphi(\varepsilon_i)(\hat{b}_j - b_j)X_{ij} = o(n^{-1/2}) + O\{\delta_{\vartheta}(nh/\log n)^{-1/2})\},$$

$$\frac{1}{nh}\sum_{i,j} K_{ij}^{\vartheta}\hat{b}_jX_{ij}\{\varphi(Y_{ij}) - \varphi(\varepsilon_i)\} = \delta_{\vartheta}^\top\sum_j b_j^2\{(\nu/\mu)_{\vartheta}(X_j) - X_j\}\{\nu_{\vartheta}(X_j) - X_j\mu_{\vartheta}(X_j)\}^\top + o(n\delta_{\vartheta})$$

uniformly in ϑ. This together with (34), (35) and (48) leads to (30). ■

Appendix

Proof of (16) and (17). Using the property of conditional expectation

$$\begin{aligned} E\{\rho(Y-a)|X^\top\vartheta = x^\top\vartheta\} &= E[E\{\rho(Y-a)\}|X\}|X^\top\vartheta = x^\top\vartheta] \\ &= E[G(m(\theta_0^\top X) - a)|X^\top\vartheta = x^\top\vartheta] \\ &= E[G\{m(\theta_0^\top X) - a; X\}|X^\top\vartheta = x^\top\vartheta]. \end{aligned}$$

Using the differentiability of $G(t;X)$, i.e.

$$G\{m(\theta_0^\top X) - a; X\} = G(0;X) + g(X)(m(\theta_0^\top X) - a)^2/2 + O\{(m(\theta_0^\top X) - a)^3\}$$

For each a near $m(\theta_0^\top X)$ (whence $m(\theta_0^\top x)$),

$$E[G\{m(\theta_0^\top X)-a;X\}|X^\top\vartheta=x^\top\vartheta]-E[G(0;X)|X^\top\vartheta=x^\top\vartheta]$$
$$\to E[g(X)(m(\theta_0^\top X)-a)^2|X^\top\vartheta=x^\top\vartheta]/2.$$

As $\rho(.)$ is convex, this convergence is uniform over all a near $m(\theta_0^\top X)$, which implies that the minimizer of $E[G\{m(\theta_0^\top X)-a;X\}|X^\top\vartheta=x^\top\vartheta]$ is also (asymptotically) the minimizer of $E[g(X)(m(\theta_0^\top X)-a)^2|X^\top\vartheta=x^\top\vartheta]$. We have

$$m(\theta_0^\top X)=m(\theta_0^\top x)+m'(\theta_0^\top x)\theta_0^\top(X-x)+C\{\theta_0^\top(X-x)\}^2,$$

$$E[g(X)(m(\theta_0^\top X)-a)|X^\top\vartheta=x^\top\vartheta]=\{m(\theta_0^\top x)-a\}\mu_\vartheta(x^\top\vartheta)+m'(\theta_0^\top x)\delta_\vartheta^\top\{\nu_\vartheta(x^\top\vartheta)-x\mu_\vartheta(x^\top\vartheta)\}$$
$$+O(|\delta_\vartheta|^2).$$

It is easily understood that the first statement in (16) is true.

To prove (17), consider for $t\to 0$,

$$\begin{aligned}E[g(X)(m(\theta_0^\top X)-a)|X^\top\vartheta=x^\top\vartheta+t]&=\{a-m(\theta_0^\top x)\}\mu_\vartheta(x^\top\vartheta+t)+m'(\theta_0^\top x)E[g(X)\{t\\&\quad+\delta_\vartheta^\top(X-x)\}|X^\top\vartheta=x^\top\vartheta+t]+O(|\delta_\vartheta|^2)\\&=\{a-m(\theta_0^\top x)\}\mu_\vartheta(x^\top\vartheta+t)+tm'(\theta_0^\top x)\mu_\vartheta(x^\top\vartheta+t)\\&\quad+m'(\theta_0^\top x)\delta_\vartheta^\top E[g(X)(X-x)|X^\top\vartheta=x^\top\vartheta+t]\\&\quad+O(t^2|\delta_\vartheta|^2).\end{aligned}$$

Therefore,

$$m_\vartheta(\vartheta^\top x+t)=m(\theta_0^\top x)+tm'(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\frac{\nu_\vartheta(x^\top\vartheta+t)-x\mu_\vartheta(x^\top\vartheta+t)}{\mu_\vartheta(x^\top\vartheta+t)}+O(|\delta_\vartheta|^2),$$

$$m_\vartheta(\vartheta^\top x)=m(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\frac{\nu_\vartheta(x^\top\vartheta)-x\mu_\vartheta(x^\top\vartheta)}{\mu_\vartheta(x^\top\vartheta)}+O(|\delta_\vartheta|^2).$$

Suppose the first order derivative of $\mu_\vartheta(.)$ and $\nu_\vartheta(.)$ are both Lipschitz continuous. We have

$$\begin{aligned}&m_\vartheta(\vartheta^\top x+t)-m_\vartheta(\vartheta^\top x)\\=\;&tm'(\theta_0^\top x)+m'(\theta_0^\top x)\delta_\vartheta^\top\frac{\nu_\vartheta(x^\top\vartheta+t)-\nu_\vartheta(x^\top\vartheta)}{\mu_\vartheta(x^\top\vartheta)}-m'(\theta_0^\top x)\delta_\vartheta^\top x\frac{\mu_\vartheta(x^\top\vartheta+t)-\mu_\vartheta(x^\top\vartheta)}{\mu_\vartheta(x^\top\vartheta)}\\&+O(t^2)+m'(\theta_0^\top x)\delta_\vartheta^\top\{\nu_\vartheta(x^\top\vartheta+t)-x\mu_\vartheta(x^\top\vartheta+t)\}\frac{\mu_\vartheta(x^\top\vartheta)-\mu_\vartheta(x^\top\vartheta+t)}{\mu_\vartheta(x^\top\vartheta)\mu_\vartheta(x^\top\vartheta+t)}\\=\;&tm'(\theta_0^\top x)+tm'(\theta_0^\top x)\delta_\vartheta^\top(\nu'/\mu)_\vartheta(x^\top\vartheta)-tm'(\theta_0^\top x)\delta_\vartheta^\top x(\mu'/\mu)_\vartheta(x^\top\vartheta)\\&-tm'(\theta_0^\top x)\delta_\vartheta^\top\{\nu_\vartheta(x^\top\vartheta+t)-x\mu_\vartheta(x^\top\vartheta+t)\}(\mu'/\mu^2)_\vartheta(x^\top\vartheta)+O(t^2)\\=\;&tm'(\theta_0^\top x)+tm'(\theta_0^\top x)\delta_\vartheta^\top\{(\nu'/\mu)-x(\mu'/\mu)-(\nu-x\mu)(\mu'/\mu^2)\}+O(t^2)\end{aligned}$$

and (17) thus follows. ■

Proof of (18) and (19) Note that by (16), (17) and the continuity of $G(t; X)$ in t defined in (10), we have

$$\begin{aligned} m_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X_{ij}^\top\vartheta &= m(X_i^\top\theta_0) - m(X_j^\top\theta_0) - b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\} \\ &\quad -\{b_j + b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j)\}X_{ij}^\top\vartheta + o(|\delta_\vartheta|) \\ &= b_jX_{ij}^\top\delta_\vartheta + \frac{1}{2}m''(X_j^\top\theta_0)(\theta_0^\top X_{ij})^2 - \frac{1}{6}m^{(3)}(X_j^\top\theta_0)(\theta_0^\top X_{ij})^3 \\ &\quad -b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j)X_{ij}^\top\vartheta - b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\} \\ &\quad +o(|\delta_\vartheta|) + O\{(\theta_0^\top X_{ij})^4\} \end{aligned}$$

Therefore,

$$\begin{aligned} &E[\varphi\{Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X_{ij}^\top\vartheta\}|X_i] \\ &= b_j\delta_\vartheta^\top g(X_i)X_{ij} - b_j\delta_\vartheta^\top\{(\nu/\mu)_\vartheta(X_j) - X_j\}g(X_i) - b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j)X_{ij}^\top\vartheta g(X_i) \\ &\quad +\frac{1}{2}m''(X_j^\top\theta_0)g(X_i)(\theta_0^\top X_{ij})^2 - \frac{1}{6}m^{(3)}(X_j^\top\theta_0)g(X_i)(\theta_0^\top X_{ij})^3 + o(|\delta_\vartheta|) + O(h^4) \end{aligned} \quad (36)$$

and thus

$$E_i[K_{ij}^\vartheta\varphi\{Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X_{ij}^\top\vartheta\}] = \frac{1}{2}m''(X_j^\top\theta_0)(gf)_\vartheta(X_j)h^3 + o(h|\delta_\vartheta|) + O(h^4).$$

This is (18). As for (19), i.e.

$$\begin{aligned} &E_i[K_{ij}^\vartheta X_{ij}^\top\vartheta\varphi\{Y_i - m_\vartheta(X_j) - m'_\vartheta(X_j)X_{ij}^\top\vartheta\}] \\ &= \frac{1}{2}m''(X_j^\top\theta_0)(f\mu)'_\vartheta(X_j)h^4 - \frac{1}{6}m^{(3)}(X_j^\top\theta_0)(f\mu)_\vartheta(X_j)h^4 + O(h^4\delta_\vartheta + h^6), \end{aligned}$$

it can be proved similarly based on (36) and the following facts

$$\begin{aligned} &E[g(X_i)X_{ij}|X_i^\top\vartheta = X_j^\top\vartheta + hu] \\ &= \nu_\vartheta(X_j^\top\vartheta + hu) - X_j\mu_\vartheta(X_j^\top\vartheta + hu) \\ &= \nu_\vartheta(X_j^\top\vartheta) + hu\nu'_\vartheta(X_j^\top\vartheta) - X_j\mu_\vartheta(X_j^\top\vartheta) - huX_j\mu'_\vartheta(X_j^\top\vartheta) + O(h^2) \end{aligned}$$

$$E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu] = \mu_\vartheta(X_j^\top\vartheta) + hu\mu'_\vartheta(X_j^\top\vartheta) + O(h^2)$$

$$\begin{aligned} \int K(u)E[g(X_i)X_{ij}|X_i^\top\vartheta = X_j^\top\vartheta + hu]hudu &= h^2\{(f\nu')_\vartheta(X_j^\top\vartheta) - X_j(f\mu')_\vartheta(X_j^\top\vartheta)\} \\ &\quad +h^2\{(f'\nu)_\vartheta(X_j^\top\vartheta) - X_j(f'\mu)_\vartheta(X_j^\top\vartheta)\} + O(h^3) \end{aligned}$$

$$\int K(u)E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu]hudu = h^2(\mu'f + \mu f')_\vartheta(X_j^\top\vartheta) + O(h^3)$$

$$\int K(u)E[g(X_i)|X_i^\top\vartheta = X_j^\top\vartheta + hu]h^2u^2du = h^2(\mu f)_\vartheta(X_j^\top\vartheta) + O(h^3)$$

Lemma 6.4 *Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a sequence of random convex functions defined on a convex, open subset Θ of R^d. Suppose $\lambda(\theta)$ is a real valued function on Θ such that $\lambda_n(\theta)$ tends to $\lambda(\theta)$ for each θ almost surely, Then for each compact set K of Θ, with probability* 1,

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \to 0.$$

Proof The condition can be restated as follows: for any fixed $\theta \in \Theta$, there exists some $\Omega_\theta \subseteq \Omega$, such that $P(\Omega_\theta) = 1$ and

$$\lambda_n(\omega, \theta) - \lambda(\theta) \to 0, \text{ for any } \omega \in \Omega_\theta.$$

The conclusion can be restated that for each compact set K of Θ, there exists some $\Omega_0 \subseteq \Omega$, such that

$$P(\Omega_0) = 1 \quad \text{and } \sup_{\theta \in K} |\lambda_n(\omega, \theta) - \lambda(\theta)| \to 0, \text{ for any } \omega \in \Omega_0.$$

For such uniformity of the convergence, it is enough to consider the case where K is a cube with edges parallel to the coordinate directions $e_l, \cdots, e_d$. Every compact subset of Θ can be covered by finitely many such cubes.

Let $\Im_0 \equiv K$ and $K^{+\delta_0}$ be the larger cube constructed by adding an extra layer of cubes with sides δ_0 to K. Suppose $\delta_0 > 0$ is small enough such that $K^{+\delta_0} \subset \Theta$. Define $\mho_0$ for the finite set of all vertices of all the cubes that make up $K^{+\delta^0}$.

Now for $k = 1, 2, \cdots$, let $\epsilon_k = k^{-1}$. As convexity implies continuity, there is a $0 < \delta^k < \delta^{k-1}$ such that $\lambda(.)$ varies by less than $\epsilon_k/(d+1)$ over each cube of side $3\delta^k$ that intersects K. Partition each cube in $\Im_{k-1}$ into a union of cubes with side at most δ^k and denote by $\Im_k$ the resulted union of cubes. Then expand K to a larger cube $K^{+\delta^k}$ by adding an extra layer of these δ^k-cubes around each face. As $\delta^k < \delta^{k-1}$, $K^{+\delta^k} \subset K^{+\delta^{k-1}}$ is still within Θ. Define

$$\begin{aligned}
\mho_k &= \{ \text{ vertices of all the } \delta^k - \text{cubes that make up } K^{+\delta^k}\} \bigcup \mho_{k-1} \\
&\equiv \{ \text{ vertices of all the } \delta^k - \text{cubes that make up } K^{+\delta^k}\} \bigcup \{\mho_{k-1} \bigcap K^c\}
\end{aligned}$$

and

$$\Omega_k = \bigcap_{\theta \in \mho_k} \Omega_\theta.$$

As $\mho_k$ is finite, we have $P(\Omega_k) = 1$ and

$$\text{for any } \omega \in \Omega_k, \quad M_n^k(\omega) = \sup_{\theta \in \mho_k} |\lambda_n(\omega, \theta) - \lambda(\theta)| \to 0. \tag{37}$$

We first establish the connection between $M_n^k(\omega)$ and the upper bound for $\lambda_n(\omega,\theta)-\lambda(\theta)$, over $\theta\in K$, for any given $\omega\in\Omega_k$.

For any fixed $k=1,2,\cdots$, each θ in K lies within a δ^k-cube with vertices $\{\theta_i\}\in\mho_k$; it can be written as a convex combination $\sum_i \alpha_i\theta_i$ of those vertices, i.e.

$$\theta=\sum_{\theta_i\in\mho_k}\alpha_i\theta_i,\qquad \sum_{\theta_i\in\mho_k}\alpha_i=1.$$

Then for any given $\omega\in\Omega_k$, convexity of $\lambda_n(\omega,\theta)$ in θ gives

$$\begin{aligned}\lambda_n(\omega,\theta) &\le \sum_{\theta_i\in\mho_k}\alpha_i\lambda_n(\omega,\theta_i)\\ &= \sum_{\theta_i\in\mho_k}\alpha_i\{\lambda_n(\omega,\theta_i)-\lambda(\theta_i)\}+\sum_{\theta_i\in\mho_k}\alpha_i\{\lambda(\theta_i)-\lambda(\theta)\}+\lambda(\theta)\\ &\le M_n^k(\omega)+\max_{\theta_i\in\mho_k}|\lambda(\theta_i)-\lambda(\theta)|+\lambda(\theta).\end{aligned}$$

Therefore,

$$\lambda_n(\omega,\theta)-\lambda(\theta)\le M_n^k(\omega)+\epsilon_k. \tag{38}$$

Next we establish the companion lower bound. For any fixed $k=1,\cdots$, each θ in K lies within a δ^k-cube with a vertex θ_0 in $K\bigcap\mho_k$:

$$\theta=\theta_0+\sum_{i=1}^d\delta_ie_i,\ \text{with }|\delta_i|\le\delta^k,\ i=1,\cdots,d.$$

Without loss of generality, suppose $\delta_i\ge 0$ for each $i=1,\cdots,d$. Define

$$\theta_{ik}=\theta_0-\delta_i'e_i,\qquad \text{where }\delta_i'\equiv\min\{c\ge\delta_k:\theta_0-ce_i\in\mho_k\},\ i=1,\cdots,d$$

Note that as $\theta_0\in K\bigcap\mho_k$, δ_i' must exist and $\delta_i'<2\delta^k$, for all $i=1,\cdots,d$.

Write θ_0 as a convex combination of θ and these θ_{ik}:

$$\theta_0=\frac{\prod_{j=1}^d\delta_j'}{\prod_{j=1}^d\delta_j'+\sum_{j=1}^d\delta_j\prod_{l\ne j}\delta_l'}\theta+\sum_{i=1}^d\frac{\delta_i\prod_{j\ne i}\delta_j'}{\prod_{j=1}^d\delta_j'+\sum_{j=1}^d\delta_j\prod_{l\ne j}\delta_l'}\theta_{ik}.$$

Denote these convex weights by β and $\{\beta_i\}$. As $\delta_j\le\delta^k\le\delta_j'$, we have $\beta\ge 1/(d+1)$ and

$$\begin{aligned}\beta\lambda_n(\omega,\theta) &\ge \lambda_n(\omega,\theta_0)-\sum_i\beta_i\lambda_n(\omega,\theta_{ik})\quad(\text{ convexity of }\lambda_n(\omega,\theta)\text{ in }\theta)\\ &\ge \lambda(\theta_0)-\sum_i\beta_i\lambda(\theta_{ik})-2M_n^k(\omega)\quad(\text{ from (37)})\\ &\ge \lambda(\theta)-\epsilon_k/(d+1)-\sum_i\beta_i[\lambda(\theta)+\epsilon_k/(d+1)]-2M_n^k(\omega)\\ &= \beta\lambda(\theta)-2\epsilon_k/(d+1)-2M_n^k(\omega)\end{aligned}$$

where the third inequality is due to the definition of δ^k and the fact that there exists a cube of side $3\delta^k$ which contains both θ_{ik} and θ_0. As $\beta \geq 1/(d+1)$,

$$\lambda_n(\omega,\theta) - \lambda(\theta) \geq -2\epsilon_k - 2(d+1)M_n^k(\omega).$$

This together with (38) implies that for any $k = 1, 2, \cdots$, there exists some $\Omega_k(\supseteq \Omega_{k+1})$ such that $P(\Omega_k) = 1$ and

$$\forall \omega \in \Omega_k,\ \sup_{\theta \in K} |\lambda_n(\omega,\theta) - \lambda(\theta)| \leq (d+1)M_n^k(\omega) + 2k^{-1}.$$

Let $\Omega_0 \equiv \bigcap_{k=1}^{\infty} \Omega_k$. As Ω_k is a decreasing sequence and $P(\Omega_k) = 1$, we have $P(\Omega_0) = 1$ and for any $\omega \in \Omega_0$,

$$\sup_{\theta \in K} |\lambda_n(\omega,\theta) - \lambda(\theta)| \leq (d+1)M_n^k(\omega) + 2k^{-1},\ \text{for all } k \geq 1. \tag{39}$$

Note that as $n \to \infty$, $M_n^k(\omega) \to 0$ for each fixed k, as in (37). Take limit of both sides of (39)

$$\lim_{n\to\infty} \sup_{\theta \in K} |\lambda_n(\omega,\theta) - \lambda(\theta)| \leq \lim_{n\to\infty} M_n^k(\omega) + k^{-1} = k^{-1},\ \text{for all } k \geq 1.$$

This is equivalent to that with probability 1,

$$\lim_{n\to\infty} \sup_{\theta \in K} |\lambda_n(\omega,\theta) - \lambda(\theta)| \to 0.$$

We now list a number of facts in the literature that will be used in our proofs later.

Lemma 6.5 *[Bernstein's inequality] Let $X_1, \cdots, X_n$ be independent zero-mean real valued random variables and there exists $c > 0$ such that the following Cramer's condition are satisfied*

$$E|X_i|^k \leq c^{k-2} k! EX_i^2 < +\infty,\ i = 1, \cdots, n;\ k = 3, 4, \cdots$$

Let $S_n = \sum_{i=1}^n X_i$, then

$$P(|S_n| \geq t) \leq 2\exp\Big(-\frac{t^2}{4\sum_{i=1}^n EX_i^2 + 2ct}\Big),\ t > 0.$$

Lemma 6.6 [Theorem 1.1. Giné et al] Let $X,\ Y,\ X_i,\ i = 1, \cdots,$ be i.i.d. random variables taking values in S and let $g : S \times S \to R$ be a measurable function of two variables. Then,

$$\limsup_n \frac{1}{n \log\log n} \sum_{i \neq j} g(X_i, X_j) < \infty \quad a.s.$$

if an only if the following three conditions hold:

(a) $g(.)$ is integrable and $Eg(X, y) = 0$ for almost all $y \in S$.

There exists some $C < \infty$ such that

(b) For all $u \geq 10$ $E\{g^2(X, Y) \wedge u\} \leq C \log\log u$ and

(c) $\sup\{Eg(X,Y)f_1(X)f_2(Y); Ef_1^2(X) \leq 1,\ Ef_2^2(X) \leq 1,\ \|f_1\|_\infty < \infty,\ \|f_2\|_\infty < \infty\} \leq C$

Lemma 6.7 [Korolyuk et al, 1989] Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables. With a symmetric kernel $\Phi : X^m \to R$, we consider the U-statistic

$$U_n = \binom{n}{m} \sum_{l \leq i_1 < \cdots < i_m \leq n} \Phi(X_{i_1}, \cdots, X_{i_m})$$

Let $\theta = E\Phi(X_1, \cdots, X_m) < \infty$ and for $c = 0, 1, \cdots, m$, define

$$\Phi_c(x_1, \cdots, x_c) = E(\Phi(X_1, \cdots, X_m)|X_1 = x_1, \cdots, X_c = x_c), \quad , \Phi_0 = \theta,\ \Phi_m = \Phi$$
$$g_c(x_1, \cdots, x_c) = \sum_{d=0}^{c} (-1)^{c-d} \sum \sum_{l \leq j_1 < \cdots < j_d \leq c} \Phi_d(x_{j1}, \cdots, x_{jd}), \quad \sigma_1^2 = Eg_1^2(X_1)$$

Suppose $\sigma_1^2 > 0$ and for all $c = 1, \cdots, m$, $Eg_c^{2c/(2c-1)} < \infty$. The with probability 1,

$$\limsup_{n \to \infty} \frac{n^{1/2}(U_n - \theta)}{(2m^2\sigma_1^2 \log\log n)^{1/2}} = 1 \quad \blacksquare$$

Lemma 6.8 [Berbee's Lemma] Let (X, Y) be a $R^d \times R^{d'}-$valued random vector. Then there exists a $R^{d'}-$valued random vector Y^* which has the same distribution as Y and

$$Y^* \text{ is independent of } X;\ P(Y^* \neq Y) = \beta(\sigma(X), \sigma(Y)) \tag{40}$$

where $\sigma(X)$ and $\sigma(Y))$ are the $\sigma-$algebra generated by X and Y respectively, and

$$\beta[\sigma(X), \sigma(Y)] = E \sup_{A \in \sigma(Y)} |P(A) - P(A|\sigma(X))|$$

Lemma 6.9 $\beta[\sigma(X_1, Y_1), \sigma(\hat{a}_j, \hat{b}_j)] = O\{(nh/\log^3 n)^{-1/4}\}$

Proof By the definition,

$$\beta[\sigma(X_1, Y_1), \sigma(\hat{a}_j, \hat{b}_j)] = E \sup_{A \in \sigma(\hat{a}_j, \hat{b}_j)} |P(A) - P(A|\sigma(X_1, Y_1))|$$

Rewrite (21) as

$$\hat{a}_j = E\hat{a}_j + \frac{1}{nh}\sum_{l\neq i}^{n} K_{lj}\varphi_1(X_l, Y_l) + \frac{1}{nh}K_{ij}\varphi_1(X_i, Y_i) + O\{(nh/\log n)^{-3/4}\}$$

$$\hat{b}_j = E\hat{b}_j + \frac{1}{nh^2}\sum_{i=2}^{n} K_{lj}\varphi_2(X_l, Y_l) + \frac{1}{nh^2}K_{ij}\varphi_2(X_i, Y_i) + O\{(nh/\log n)^{-3/4}/h\} \quad (41)$$

where $\{\varphi_1(X_l, Y_l)\}_{l=1}^n$, $\{\varphi_2(X_l, Y_l)\}_{l=1}^n$, are two sequences of bounded and identically distributed zero-mean random variables. Let $\tau_n = (nh/\log n)^{-3/4}$ and

$$S_1 = \frac{1}{nh}\sum_{l\neq i}^{n} K_{lj}\varphi_1(X_l, Y_l) = \sum_{,l\leq i-m} + \sum_{l\neq i, i-m<l<i+m} + \sum_{l\geq i+m} = S_{11} + S_{12} + S_{13},$$

$$S_2 = \frac{1}{nh^2}\sum_{l\neq i}^{n} K_{lj}\varphi_2(X_l, Y_l) = \sum_{,l\leq i-m} + \sum_{l\neq i, i-m<l<i+m} + \sum_{l\geq i+m} = S_{21} + S_{22} + S_{23}.$$

Note that S_1/σ_1 and S_2/σ_2 are asymptotically normal with $\sigma_1 = O((nh)^{1/2})$ and $\sigma_1 = O((nh^3)^{1/2})$. Using the fact that both $\varphi_1(.)$ and $\varphi_2(.)$ are bounded and

$$\{S_1 \leq t_1 - E\hat{a}_j\} \subseteq \{S_{11} + S_{13} \leq t_1 - E\hat{a}_j + Cm(nh)^{-1}\}$$

$$\{S_2 \leq t_2 - E\hat{b}_j\} \subseteq \{S_{21} + S_{23} \leq t_2 - E\hat{b}_j + Cm(nh^2)^{-1}\}$$

we have

$$\begin{aligned}
& P\{\hat{a}_j \leq t_1, \hat{b}_j \leq t_2 | Y_i, X_i\} \\
\leq\ & P\{S_1 \leq t_1 - E\hat{a}_j + C\tau_n, S_2 \leq t_1 - E\hat{b}_j + C\tau_n/h | Y_i, X_i\} \\
\leq\ & P\Big[S_{11} + S_{13} \leq t_1 - E\hat{a}_j + C\tau_n + Cm(nh)^{-1}, S_{21} + S_{23} \leq t_2 - E\hat{b}_j C\tau_n/h + Cm(nh^2)^{-1} | Y_i, X_i\Big] \\
\leq\ & P\Big[(S_{11} + S_{13})/\sigma_1 \leq (t_1 - E\hat{a}_j)/\sigma_1, (S_{21} + S_{23})/\sigma_2 \leq (t_2 - E\hat{b}_j)/\sigma_2\Big] \\
& + \phi(m) + C(nh)^{1/2}\tau_n + Cmn^{-1/2}h^{-1} \\
=\ & P[\hat{a}_j \leq t_1, \hat{b}_j \leq t_2] + C\phi(m) + C(nh)^{1/2}\tau_n + Cmn^{-1/2}h^{-1}
\end{aligned}$$

Similarly, we have

$$P\{\hat{a}_j \geq t_1, \hat{b}_j \geq t_2 | Y_1, X_1\} \geq P[\hat{a}_j \geq t_1, \hat{b}_j \geq t_2] - \phi(m) - C(nh)^{1/2}\tau_n - Cmn^{-1/2}h^{-1}$$

Therefore,

$$|P\{\hat{a}_j \leq t_1, \hat{b}_j \leq t_2 | Y_1 = y, X_1\} - P\{\hat{a}_j \leq t_1, \hat{b}_j \leq t_2\}| \leq \phi(m) + C(nh)^{1/2}\tau_n + Cmn^{-1/2}h^{-1}.$$

If $m = n^a$ and $\phi(m) = m^{-k}$ for some $a = 1/10$ and $k > 0$, such that $n^{ak} = O\{(nh)^{1/2}\tau_n\}$ with $\tau_n = (nh/\log, n)^{-3/4}$, we have the desired result.

Lemma 6.10 *Under the assumptions (A1)–(A5), we have*

$$E\Phi_n(\theta) = \delta_\theta^\top ER_{n1}(\theta) + \delta_\theta^\top G_{n2}\delta_\theta + o(n^2h|\delta_\theta|^2).$$

Proof Apparently it suffices to show that

$$\begin{aligned} & EK_{ij}^\vartheta\{\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})\} \\ = \;& \delta_\theta^\top E[K_{ij}^\vartheta\varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})\hat{b}_jX_{ij}] + \delta_\theta^\top E[K_{ij}^\vartheta X_{ij}X_{ij}^\top g(X_1)b_j^2]\delta_\theta + o(|\delta_\theta|^2). \end{aligned}$$

By the continuity of $E[\rho(Y_i - \hat{a}_j - t\hat{b}_j)|X_i]$ in t, we have

$$\begin{aligned} & EK_{ij}^\vartheta\{\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})\} \qquad (42) \\ = \;& \delta_\theta^\top E[K_{ij}^\vartheta\varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})\hat{b}_jX_{ij}] + \delta_\theta^\top E[K_{ij}^\vartheta X_{ij}X_{ij}^\top\{\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_jt)\hat{b}_j/\partial t\}|_{t=X_{ij}^\top\theta_0}]\delta_\theta \\ & +\delta_\theta^\top E\Big[K_{ij}^\vartheta X_{ij}X_{ij}^\top\Big\{\{\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_jt)\hat{b}_j/\partial t\}|_{t=X_{ij}^\top\theta_0} - \{\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_jt)\hat{b}_j/\partial t\}|_{t^*}\Big\}\Big]\delta_\theta \end{aligned}$$

where t^* is some value between $\theta^\top X_{ij}$ and $\theta_0^\top X_{ij}$.

Apply Lemma 6.8 and Lemma 6.9 to the second term in (42). There exists $[\tilde{a}_j, \tilde{b}_j]$ which has the same distribution as $[\hat{a}_j, \hat{b}_j]$, is independent of (Y_1, X_1) and $P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]) = O\{(nh/\log^3 n)^{-1/4}\}$. Thus

$$\begin{aligned} & E[\varphi(Y_1 - \hat{a}_j - \hat{b}_j(t+\delta))\hat{b}_j] - E[\varphi(Y_1 - \hat{a}_j - \hat{b}_jt)\hat{b}_j] \\ = \;& E[\varphi\{Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)\}\tilde{b}_j] - E[\varphi(Y_1 - \tilde{a}_j - \tilde{b}_jt)\tilde{b}_j] \\ & +E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_jt)\tilde{b}_j - \varphi(Y_1 - \hat{a}_j - \hat{b}_jt)\hat{b}_j\}I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}] \\ & -E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta))\tilde{b}_j - \varphi(Y_1 - \hat{a}_j - \hat{b}_j(t+\delta))\hat{b}_j\}I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}] \\ = \;& E[\varphi\{Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)\}\tilde{b}_j] - E[\varphi(Y_1 - \tilde{a}_j - \tilde{b}_jt)\tilde{b}_j] \\ & +E[\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j(t+\delta)) - \varphi(Y_1 - \hat{a}_j - \hat{b}_jt)\}\hat{b}_jI\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}] \\ & -E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)) - \varphi(Y_1 - \tilde{a}_j - \tilde{b}_jt)\}\tilde{b}_jI\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}] \\ \equiv \;& \mathrm{T}_1 + \mathrm{T}_2 + \mathrm{T}_3 \qquad (43) \end{aligned}$$

First due to the independency of $[\tilde{a}_j, \tilde{b}_j]$ from (X_1, Y_1) and the continuity of $G_1(.;X)$,

$$\begin{aligned} \mathrm{T}_1 &= E[\{G_1(a_1 - \tilde{a}_j - \tilde{b}_j(t+\delta)|X_1) - G_1(a_1 - \tilde{a}_j - \tilde{b}_jt|X_1)\}\tilde{b}_j] \\ &= \delta E[G_2(a_1 - \tilde{a}_j - \tilde{b}_jt|X_1)\tilde{b}_j^2] + o(\delta). \qquad (44) \end{aligned}$$

Next, we show that $\mathrm{T}_2 = o(\delta)$ and $\mathrm{T}_3 = o(\delta)$.

Define $[t_1, t_2] \equiv [(\hat{a}_j - E\hat{a}_j)/\sigma_1, (\hat{b}_j - E\hat{b}_j)/\sigma_2]$ are asymptotically normal, where

$$\sigma_1 \equiv \{\mathrm{Var}\hat{a}_j\}^{1/2} = O\{(nh)^{-1/2}\},\ \sigma_2 \equiv \{\mathrm{Var}\hat{b}_j\}^{1/2} = O\{(nh^3)^{-1/2}\}.$$

Define $[\tilde{t}_1, \tilde{t}_2]$ similarly from $\tilde{a}_j$ and $\tilde{b}_j$. Note that due to the weak dependency nature of the time series, the conditional probability density function of Y_i given $[t_1, t_2]$ is uniformly bounded. Without loss of generality, assume $\delta > 0$. As for any given values of $\hat{a}_j$ and $\hat{b}_j$ (i.e. t_1 and t_2),

$$|\varphi(Y_i - \hat{a}_j - \hat{b}_j(t+\delta)) - \varphi(Y_i - \hat{a}_j - \hat{b}_j t)| \le CP\{t \le (Y_i - \hat{a}_j)/\hat{b}_j \le t + \delta | t_1, t_2\} = O(\delta).$$

Therefore, $|\mathrm{T}_2| \le C\delta \int \hat{b}_j f(t_1, t_2) g(t_1, t_2) dt_1 dt_2 = o(\delta)$, where

$$g(t_1, t_2) = \int_{[\tilde{t}_1, \tilde{t}_2] \ne [t_1, t_2]} f(\tilde{t}_1, \tilde{t}_2 | t_1, t_2) d\tilde{t}_1 d\tilde{t}_2, \quad \text{with} \int f(t_1, t_2) g(t_1, t_2) dt_1 dt_2 = \beta[\sigma(X_i, Y_i), \sigma(\hat{a}_j, \hat{b}_j)]$$

Similarly we can show that $\mathrm{T}_3 = o(\delta)$. This together with (43) and (44) yields

$$\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_j t)\hat{b}_j / \partial t = E[G_2(a_1 - \tilde{a}_j - \tilde{b}_j t | X_1)\tilde{b}_j^2] + o(1), \tag{45}$$

where $o(1)$ is uniform in i, j and t. Apply this result to the third term in (42), we have

$$\begin{aligned} & \{\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_j t)\hat{b}_j / \partial t\}|_{t = X_{ij}^\top \theta_0} - \{\partial E\varphi(Y_i - \hat{a}_j - \hat{b}_j t)\hat{b}_j / \partial t\}|_{t^*} \\ = \;& E[G_2(\tilde{a}_j + \tilde{b}_j t | X_1)\tilde{b}_j^2] - E[G_2(\tilde{a}_j + \tilde{b}_j t^* | X_1)\tilde{b}_j^2] + o(1) = o(1). \end{aligned}$$

This together with (42) and (45) leads to

$$\begin{aligned} & EK_{ij}^\vartheta \{\rho(Y_i - \hat{a}_j - \hat{b}_j \theta^\top X_{ij}) - \rho(Y_i - \hat{a}_j - \hat{b}_j \theta_0^\top X_{ij})\} \\ = \;& \delta_\theta^\top E[K_{ij}^\vartheta \varphi(Y_i - \hat{a}_j - \hat{b}_j \theta_0^\top X_{ij})\hat{b}_j X_{ij}] + \delta_\theta^\top E[K_{ij}^\vartheta X_{ij} X_{ij}^\top G_2(a_1 - \tilde{a}_j - \tilde{b}_j t | X_1)\tilde{b}_j^2]\delta_\theta + o(|\delta_\theta|^2) \\ = \;& \delta_\theta^\top E[K_{ij}^\vartheta \varphi(Y_i - \hat{a}_j - \hat{b}_j \theta_0^\top X_{ij})\hat{b}_j X_{ij}] + \delta_\theta^\top E[K_{ij}^\vartheta X_{ij} X_{ij}^\top g(X_1) b_j^2]\delta_\theta + o(|\delta_\theta|^2) \end{aligned}$$

where for the last equation follows from the continuity of $G_2(.|X_1)$ and (23). ■

Lemma 6.11 *Define $Z_{ij} = K_{ij}^\vartheta \hat{b}_j X_{ij}\{\varphi(Y_{ij}) - \varphi(\varepsilon_i)\}$. Then*

$$(nh)^{-1} \sum_j E_i Z_{ij} = \delta_\vartheta^\top \sum_j b_j^2 \{(\nu/\mu)_\vartheta(X_j) - X_j\}\{\nu_\vartheta(X_j) - X_j \mu_\vartheta(X_j)\}^\top + o(|\delta_\vartheta| + n^{-1/2}) \tag{46}$$

$$\sum_{i,j} (Z_{ij} - E_i Z_{ij}) = o(n^2 h \delta_\vartheta) \tag{47}$$

$$(nh)^{-1} \sum_i K_{ij} \varphi(\varepsilon_i)(\hat{b}_j - b_j) X_{ij} = o(n^{-1/2}) + O\{\delta_\vartheta (nh/\log n)^{-1/2}\} \tag{48}$$

uniformly in ϑ, if $nh^4 \to \infty$ and $nh^5/\log n < \infty$.

Proof To prove (46), we apply Lemma 6.8 and Lemma 6.9. Suppose $[\tilde{a}_j, \tilde{b}_j]$ has the same distribution as $[\hat{a}_j, \hat{b}_j]$ and is independent of (X_1, Y_1). Therefore, $P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}) = O(n^{-1/2})$. Let $\mathcal{X} = \sigma(X_1, ..., X_n)$.

$$
\begin{aligned}
& E[\{\varphi(Y_1 - \hat{a}_j - t\hat{b}_j) - \varphi(\varepsilon_1)\}\hat{b}_j|\mathcal{X}] \\
= \; & E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j t) - \varphi(\varepsilon_1)\}\tilde{b}_j|\mathcal{X}] - E[\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j t) - \varphi(\varepsilon_1)\}\tilde{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}] \\
& + E[\{\varphi(Y_1 - \hat{a}_j - \hat{b}_j t) - \varphi(\varepsilon_1)\}\hat{b}_j I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}|\mathcal{X}] \\
= \; & \mathrm{T}_1 - \mathrm{T}_2 + \mathrm{T}_3,
\end{aligned} \tag{49}
$$

and $E_i Z_{ij} = E[K_{1j} X_{1j}(\mathrm{T}_1 - \mathrm{T}_2 + \mathrm{T}_3)]$. First by the independency of $[\tilde{a}_j, \tilde{b}_j]$ and (X_1, Y_1) and the continuity of $G_1(.|X)$, we have

$$
\begin{aligned}
\mathrm{T}_1 &= E[\{G_1(a_1 - \tilde{a}_j - \tilde{b}_j t; X_1) - G_1(0; X_1)\}\tilde{b}_j] \\
&= g(X_1) E\{\tilde{b}_j(a_1 - \tilde{a}_j - \tilde{b}_j t)\} + O\{E(a_1 - \tilde{a}_j - \tilde{b}_j t)^2\}.
\end{aligned} \tag{50}
$$

Using the expansions of $(\tilde{a}_j, \tilde{b}_j)$ as given in (21), we have

$$
\begin{aligned}
& a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^\top \theta_0 = a_1 - a_j + a_j - \tilde{a}_j - \tilde{b}_j X_{1j}^\top \theta_0 \\
= \; & \frac{1}{2} m''(X_j^\top \theta_0)\{(X_{1j}^\top \theta_0)^2\} - \frac{1}{2} m''(X_j^\top \theta_0) h^2 + O\{(X_{1j}^\top \theta_0)^3\} \\
& - b_j \delta_\vartheta^\top \{(\nu/\mu)_\vartheta(X_j) - X_j\} - b_j \delta_\vartheta^\top \{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) X_{1j}^\top \theta_0 \\
& - h^2 \Big[\frac{1}{2} m''(X_j^\top \theta_0)\{(f\mu)'/(fg)\}_\vartheta(X_j) - \frac{1}{6} m^{(3)}(X_j^\top \theta_0)(f\mu)_\vartheta(X_j)\Big] X_{1j}^\top \theta_0 \\
& + \{gf\}_\vartheta^{-1}(X_j) \frac{1}{nh} \sum_{i=1}^n \varphi_1(\tilde{X}_i, \tilde{Y}_i) - \Big[\{gf\}_\vartheta^{-1}(X_j) \frac{1}{nh^2} \sum_{i=1}^n \varphi_2(\tilde{X}_i, \tilde{Y}_i)\}\Big] X_{1j}^\top \theta_0 \\
& + O\{(nh/\log n)^{-3/4}(1 + \delta_\vartheta/h) + h^3\}
\end{aligned} \tag{51}
$$

where $\varphi_1(\tilde{X}_i, \tilde{Y}_i)$, $\varphi_2(\tilde{X}_i, \tilde{Y}_i)$ are IID zero-mean random variables and are independent of (X_1, Y_1). Therefore, $E(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^\top \theta_0)^2 = o(|\delta_\vartheta| + n^{-1/2})$ uniformly in ϑ and

$$
\begin{aligned}
E[K_{1j}^\vartheta X_{1j} \mathrm{T}_1] &= E[K_{1j}^\vartheta g(X_1)\tilde{b}_1(a_1 - \tilde{a}_j - \tilde{b}_j X_{1j}^\top \theta_0)] + o(h|\delta_\vartheta| + n^{-1/2}h) \\
&= h\delta_\vartheta^\top b_j^2 \{(\nu/\mu)_\vartheta(X_j) - X_j\}\{\nu_\vartheta(X_j) - X_j \mu_\vartheta(X_j)\} + O(h^2|\delta_\vartheta|) + o(hn^{-1/2})
\end{aligned} \tag{52}
$$

uniformly in ϑ, where we have used (23).

Now based on the expansion of $\tilde{b}_j - b_j$, we have

$$
\begin{aligned}
E[K_{1j}^\vartheta X_{1j} \mathrm{T}_2] &= b_j E[K_{1j}^\vartheta X_{1j}\{\varphi(Y_1 - \tilde{a}_j - \tilde{b}_j X_{ij}^\top \theta_0) - \varphi(\varepsilon_1)\} I\{[\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]\}] + o(n^{-1/2}h) \\
&= o(n^{-1/2}h) + o(h\delta_\vartheta)
\end{aligned} \tag{53}
$$

uniformly in ϑ, where the last step is due to the fact that $P([\tilde{a}_j, \tilde{b}_j] \neq [\hat{a}_j, \hat{b}_j]) = O\{(nh/\log^3 n)^{-1/4}\}$. Similarly we have $E[K^{\vartheta}_{1j} X_{1j} \mathrm{T}_2] = o(n^{-1/2}h) + o(h\delta_\vartheta)$. This together with (52) and (53) yields (46).

To prove (47), first note that

$$\begin{aligned}\varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij}) - \varphi(\varepsilon_i) &= [\varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij}) - \varphi(Y_i - a_j - b_j\theta_0^\top X_{ij})] \\ &\quad +[\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(\varepsilon_i)].\end{aligned}$$

Therefore, based on Lemma 6.14, it suffices to show that

$$\sum_{i,j}(Z_{ij} - EZ_{ij}) = O(nh\delta_\vartheta), \quad Z_{ij} = K^{\vartheta}_{ij}\hat{b}_j X_{ij}\{\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(\varepsilon_i)\}$$

Due to Borel-Cantelli Lemma and the fact that for any $\epsilon > 0$,

$$P\{|\sum_{i,j}(Z_{ij} - EZ_{ij})| \geq \epsilon n^2 h\delta_\vartheta\} \leq nP\{|\sum_i (Z_{ij} - EZ_{ij})| \geq \epsilon nh\delta_\vartheta\},$$

the problem is further reduced to prove that for any $\epsilon > 0$, the quantity

$$P\{|\sum_i (Z_{ij} - EZ_{ij})| \geq \epsilon nh\delta_\vartheta\}$$

is summable over n.

Let $\tilde{Z}_{ij} = K^{\vartheta}_{ij} b_j X_{ij}\{\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(\varepsilon_i)\}$. As $\tilde{Z}_{ij}$ is bounded and $E\tilde{Z}^2_{ij} = h(h^2 + \delta^2_\vartheta)$, applying Bernstein's inequality, we have

$$P\{|\sum_i (Z_{ij} - EZ_{ij})| \geq \epsilon nh\delta_\vartheta\} \leq C\exp\Big\{ - \frac{\epsilon^2 n^2 h^2 \delta^2_\vartheta}{nh^3 + nh\delta^2_\vartheta + \epsilon nh\delta_\vartheta}\Big\} = o(n^{-2}).$$

Now it remains to show that

$$(\hat{b}_j - b_j)\sum_i K^{\vartheta}_{ij} X_{ij}\{\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(\varepsilon_i)\} = o(nh\delta_\vartheta) \tag{54}$$

By expansion of $\hat{b}_j - b_j$ in (21),

$$\begin{aligned}\hat{b}_j - b_j &= h^2\Big[\frac{1}{2}m''(X_j^\top\theta_0)\{(f\mu)'/(fg)\}_\vartheta(X_j) - \frac{1}{6}m^{(3)}(X_j^\top\theta_0)\{(f\mu)/(fg)\}_\vartheta(X_j)\Big] \\ &\quad + b_j\delta_\vartheta^\top\{(\mu\nu' - \mu'\nu)/\mu^2\}_\vartheta(X_j) + \frac{1}{nh^2}\sum_{i=1}^n \varphi_2(X_i, Y_i) + O\{(nh/\log n)^{-3/4}/h\}\end{aligned}$$

where $E\varphi_2(X_i,Y_i)=0$. As $\varphi(.)$ is bounded, we need not worry about the deterministic(bias) term in $\hat b_j - b_j$. For the stochastic part, write

$$\begin{aligned}
& (nh^2)^{-1}\sum_{i,l} K^{\vartheta}_{ij}K^{\vartheta}_{lj}X_{ij}\varphi_2(X_l,Y_l)[\varphi(Y_i-a_j-b_j\theta_0^\top X_{ij})-\varphi(\varepsilon_i)] \\
= \; & (nh^2)^{-1}\sum_{i} K^{\vartheta}_{ij}X_{ij}\varphi_2(X_i,Y_i)[\varphi(Y_i-a_j-b_j\theta_0^\top X_{ij})-\varphi(\varepsilon_i)] \\
& +(nh^2)^{-1}\sum_{i\neq l} K^{\vartheta}_{ij}X_{ij}\varphi_2(X_l,Y_l)[\varphi(Y_i-a_j-b_j\theta_0^\top X_{ij})-\varphi(\varepsilon_i)] \qquad (55)
\end{aligned}$$

Again as both $\varphi(.)$ and $\varphi_2(.)$ are bounded, handling of the first term is trivial. Now define

$$\varphi_1(X_i,Y_i)=K^{\vartheta}_{ij}X_{ij}\varphi(Y_i-a_j-b_j\theta_0^\top X_{ij})-\varphi(\varepsilon_i),$$

whence $c\equiv E\varphi_1(X_i,Y_i)=O(h^3+h\delta^2_\vartheta)$ and the second term in (55) is $(nh^2)^{-1}(\mathrm{T}_1+c\mathrm{T}_2)$, where

$$\mathrm{T}_1=\sum_{i<l}[\varphi_2(X_l,Y_l)\{\varphi_1(X_i,Y_i)-c\}+\varphi_2(X_i,Y_i)\{\varphi_1(X_l,Y_l)-c\}],\ \mathrm{T}_2=\sum_{i<l}\{\varphi_2(X_l,Y_l)+\varphi_2(X_i,Y_i)\}$$

By the law of the iterated logarithm of U-statistics in Giné et al (Lemma 6.6), $\mathrm{T}_1/h=O(n\log\log n)$ almost surely. On the other hand, by law of the iterated logarithm for U-statistics in Korolyuk et al (Lemma 6.7), $\mathrm{T}_2=n^{3/2}(h\log\log n)^{1/2}$ *a.s.* Since $c=O(h^3+h\delta^2_\vartheta)$, $(nh^2)^{-1}(\mathrm{T}_1+c\mathrm{T}_2)=O\{h^{-1}\log\log n+(nh^3\log\log n)^{1/2}+\delta^2_\vartheta(n\log\log n/h)^{1/2}\}=o(nh\delta_\vartheta)$.
Proof of (48) can be done in exact the same manner as (54). ■

The proof of (31) consists of the following two Lemmas.

Lemma 6.12 *Let* $R^*_{n2}(\theta)=\sum_{i,j}K^{\vartheta}_{ij}\Big[\rho(Y_i-\hat a_j-\hat b_j\theta^\top X_{ij})-\rho(Y_{ij})-\delta_\theta^\top\varphi(Y_i-a_j-b_jX_{ij}^\top\theta_0)\hat b_jX_{ij}\Big]$. *Then for any fixed* θ*, with probability* 1,

$$(n^2ha^2_{n\vartheta})^{-1}[R^*_{n2}(\theta)-ER^*_{n2}(\theta)]=o(1). \qquad (56)$$

uniformly in ϑ.

Proof Define $X_{ix}=X_i-x,\ \mu_{ix}=(1,X_{ix}^\top)^\top,\ K_{ix}=K(X_{ix}^\top\vartheta/h),\ \beta(x)=[m(\theta_0^\top x),m'(\theta_0^\top x)\theta_0^\top]^\top$ and $\varphi_{ni}(x;t)=\varphi(Y_i;\mu_{ix}^\top\beta(x)+t)$. For any $\alpha,\ \beta\in\mathcal{R}^{d+1}$, let

$$\begin{aligned}
\Phi_{ni}(x;\alpha,\beta) &= K_{ix}\Big[\rho\{Y_i;\mu_{ix}^\top(\alpha+\beta+\beta(x))\}-\rho\{Y_i;\mu_{ix}^\top(\beta+\beta(x))\})-\varphi_{ni}(x;0)\mu_{ix}^\top\alpha\Big] \\
&= K_{ix}\int_{\mu_{ix}^\top\beta}^{\mu_{ix}^\top(\alpha+\beta)}\{\varphi_{ni}(x;t)-\varphi_{ni}(x;0)\}dt
\end{aligned}$$

and $R_{ni}(x;\alpha,\beta) = \Phi_{ni}(x;\alpha,\beta) - E\Phi_{ni}(x;\alpha,\beta)$. Apparently,

$$K_{ij}^{\vartheta}\Big[\rho(Y_i - \hat{a}_j - \hat{b}_j\theta^\top X_{ij}) - \rho(Y_{ij}) - \delta_\theta^\top \varphi(Y_i - a_j - b_j X_{ij}^\top \theta_0)\hat{b}_j X_{ij}\Big] \equiv \Phi_{ni}(X_j;\alpha,\beta)$$

with $\alpha = [0, \hat{b}_j\delta_\theta^\top]^\top$ and $\beta = [\hat{a}_j - a_j, (\hat{b}_j - b_j)\theta_0^\top]^\top$. Let $[a_x, b_x] \equiv [m(\theta_0^\top x), m'(\theta_0^\top x)]$ and $\mathcal{D}$ be any compact subset of the support of X. For any $M > 0$ and $\vartheta \in \Theta_n$, define

$$\begin{aligned}
&M_{n1}^{\vartheta} = Ca_{n\vartheta},\ M_{n2}^{\vartheta} = C\{|\delta_\vartheta| + (nh/\log n)^{-1/2}\},\\
&M_{n3}^{\vartheta} = C\{|\delta_\vartheta| + (nh/\log n)^{-1/2}/h\},\ B_n^{(1)} = \{\alpha \in R^{d+1} | \alpha = [0, \alpha_1^\top]^\top, |\alpha_1| \le M_{n1}^{\vartheta}\}\\
&B_n^{(2)} = \{\beta \in R^{d+1} | \beta = [b_1, b_2\theta_0^\top]^\top, |b_1| \le M_{n2}^{\vartheta}, |b_2| \le M_{n3}^{\vartheta}\}.
\end{aligned}$$

As $|\hat{b}_j\delta_\theta| \le Ca_{n\vartheta}$, $|\hat{a}_j - a_j| = O\{|\delta_\vartheta| + (nh/\log n)^{-1/2}\}$ and $|(\hat{b}_j - b_j)| = O\{|\delta_\vartheta| + (nh/\log n)^{-1/2}/h\}$, (56) will follow if for any $\epsilon > 0$

$$\sup_{x\in\mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n R_{ni}(x;\alpha,\beta)\Big| \le \epsilon d_n \text{ almost surely},\ d_n = nha_{n\vartheta}^2 \tag{57}$$

This is done in a similar style as Lemma 4.2 in Kong et al(2008). Cover $\mathcal{D}$ by a finite number T_n of cubes $\mathcal{D}_k = \mathcal{D}_{n,k}$ with side length $l_n = O\{h(nh/\log n)^{-1/4}\}$ and centers $x_k = x_{n,k}$. Write

$$\begin{aligned}
\sup_{x\in\mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n R_{ni}(x;\alpha,\beta)\Big| \le &\max_{1\le k\le \mathrm{T}_n} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Phi_{ni}(x_k;\alpha,\beta) - E\Phi_{ni}(x_k;\alpha,\beta)\Big|\\
&+ \max_{1\le k\le \mathrm{T}_n} \sup_{x\in\mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{\Phi_{ni}(x_k;\alpha,\beta) - \Phi_{ni}(x;\alpha,\beta)\Big\}\Big|\\
&+ \max_{1\le k\le \mathrm{T}_n} \sup_{x\in\mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{E\Phi_{ni}(x_k;\alpha,\beta) - E\Phi_{ni}(x;\alpha,\beta)\Big\}\Big|\\
\equiv &Q_1 + Q_2 + Q_3.
\end{aligned}$$

In Lemma 6.13, we will prove that $Q_2 = o(d_n),\ a.e.$, whence $Q_3 \le EQ_2 = o(d_n)$. It thus remains to show that $Q_1 \le \epsilon d_n/3\ a.e.$, we follow a similar proof style as in Lemma 4.2 in Kong et al (2008).

Partition $B_n^{(i)},\ i = 1,2$ into a sequence of sub rectangles $D_1^{(i)}, \cdots, D_{J_1}^{(i)},\ i = 1,2$, such that for all $1 \le j_1 \le J_1 \le M^{d+1}$ $(M = \epsilon^{-1})$

$$\begin{aligned}
&\forall\ \alpha, \alpha' \in D_{j_1}^{(1)}, |\alpha - \alpha'| \le M_{n1}^{\vartheta}/M;\\
&\forall\ \beta = [b_1, b_2\theta_0^\top]^\top, \beta' = [b_1', b_2'\theta_0^\top]^\top \in D_{j_1}^{(2)}, |b_1 - b_1'| \le M_{n2}^{\vartheta}/M, |b_2 - b_2'| \le M_{n3}^{\vartheta}/M
\end{aligned}$$

Choose a point $\alpha_{j_1} \in D_{j_1}^{(1)}$ and $b_{k_1} \in D_{k_1}^{(2)}$, $1 \le j_1, k_1 \le J_1$. Then

$$\sup_{\substack{\alpha \in B_n^{(1)} \\ \beta \in B_n^{(2)}}} |\sum_i R_{ix}(\alpha,\beta)| \le \max_{1\le j_1,k_1\le J_1} \sup_{\substack{\alpha \in D_{j_1}^{(1)}, \\ \beta \in D_{k_1}^{(2)}}} |\sum_{i=1}^n \{R_{ix}(\alpha_{j_1}, b_{k_1}) - R_{ix}(\alpha,\beta)\}| + \max_{1\le j_1,k_1\le J_1} |\sum_{i=1}^n R_{ix}(\alpha_{j_1}, \beta_{k_1})| = H_{n1} + H_{n2}. \tag{58}$$

We first show that any $\epsilon > 0$

$$P\Big\{H_{n2} \ge \frac{\epsilon d_n}{2}\Big\} \le J_1^2 P\Big\{|\sum_{i=1}^n R_{ix}(\alpha_{j_1}, \beta_{k_1})| \ge \frac{\epsilon d_n}{2}\Big\} = O(n^{-a}), \tag{59}$$

for some $a > 1$. By Bernstein's Inequality and the fact that $|R_{ix}(\alpha_{j_1}, \beta_{k_1})| \le C a_{n\vartheta}$ and $\text{Var} R_{ix}(\alpha_{j_1}, \beta_{k_1}) = O[nha_{n\vartheta}^2\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}]$, we have

$$T_n J_1^2 P\Big\{|\sum_{i=1}^n R_{ix}(\alpha_{j_1}, \beta_{k_1})| \ge \frac{\epsilon d_n}{2}\Big\} = T_n J_1^2 \exp[-\epsilon^2 nha_{n\vartheta}\{1 + a_{n\vartheta}(nh/\log n)^{1/2})\}] = O(n^{-a}),$$

for some $a > 1$. Therefore, (59) holds.

We next consider H_{n1}. For each $j_1 = 1, \cdots, J_1$ and $i = 1, 2$, partition each rectangle $D_{j_1}^{(i)}$ further into a sequence of subrectangles $D_{j_1,1}^{(i)}, \cdots, D_{j_1,J_2}^{(i)}$. Repeat this process recursively as follows. Suppose after the lth round, we get a sequence of rectangles $D_{j_1,j_2,\cdots,j_l}^{(i)}$ with $1 \le j_k \le J_k$, $1 \le k \le l$, then in the $(l+1)$th round, each rectangle $D_{j_1,j_2,\cdots,j_l}^{(i)}$ is partitioned into a sequence of subrectangles $\{D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}, 1 \le j_l \le J_l\}$ such that for all $1 \le j_{l+1} \le J_{l+1}$,

$$\forall\ a, a' \in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}, |a - a'| \le M_{n1}^{\vartheta}/M^{l+1},$$
$$\forall\ \beta = [b_1, b_2\theta_0^\top]^\top, \beta' = [b_1', b_2'\theta_0^\top]^\top \in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(2)}, |b_1 - b_1'| \le \frac{M_{n2}^{\vartheta}}{M^{l+1}}, |b_2 - b_2'| \le \frac{M_{n3}^{\vartheta}}{M^{l+1}}$$

where $J_{l+1} \le M^{d+1}$. End this process after the $(\text{L}_n + 2)$th round, with L_n being the largest integer such that

$$n(2/M)^{L_n} > d_n/M_{n2}^{\vartheta} \tag{60}$$

Let $D_l^{(i)}$, $i = 1, 2$, denote the set of all subrectangles of $D_0^{(i)}$ after the lth round of partition and a typical element $D_{j_1,j_2,\cdots,j_l}^{(i)}$ of $D_l^{(i)}$ is denoted as $D_{(j_l)}^{(i)}$. Choose a point $\alpha_{(j_l)} \in D_{(j_l)}^{(1)}$ and

$\beta_{(j_l)} \in D^{(2)}_{(j_l)}$ and define

$$V_l = \sum_{\substack{(j_{l+1})\\(k_{l+1})}} P\Big\{\Big|\sum_{i=1}^n \{R_{ix}(\alpha_{(j_l)}, \beta_{(k_l)}) - R_{ix}(\alpha_{(j_{l+1})}, \beta_{(k_{l+1})})\}\Big| \geq \frac{\varepsilon d_n}{2^{l+1}}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 1,$$

$$Q_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\sup_{\substack{\alpha\,\in\, D^{(1)}_{(j_l)},\\ \beta\,\in\, D^{(2)}_{(k_l)}}} \Big|\sum_{i=1}^n \{R_{ix}(\alpha_{(j_l)}, \beta_{(k_l)}) - R_{ix}(\alpha, \beta)\}\Big| \geq \frac{\varepsilon d_n}{2^l}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 2.$$

Then $Q_l \leq V_l + Q_{l+1},\ 1 \leq l \leq L_n + 1$. On the other hand, it is easy to see that for any $\alpha \in D^{(1)}_{(j_{L_n+2})}$ and $\beta \in D^{(2)}_{(k_{L_n+2})}$,

$$n|R_{ix}(\alpha_{(j_{L_n+2})}, \beta_{(k_{L_n+2})}) - R_{ix}(\alpha, \beta)| \leq nM^{\vartheta}_{n2}/M^{L_n+2} \leq \epsilon d_n/2^{L_n+2}$$

due to the choice of L_n specified in (60). Therefore, $Q_{L_n+2} = 0$ and it remains to show that

$$T_n P\{H_{n1} \geq \frac{\epsilon d_n}{2}\} \leq T_n J_1^2 Q_1 \leq T_n J_1^2 \sum_{l=1}^{L_n+1} V_l = O(n^{-a}),\ \text{for some } a > 1 \tag{61}$$

To find upper bound for $V_l,\ 1 \leq l \leq L_n + 1$, we again apply Bernstein's inequality. As

$$|R_{ix}(\alpha_{(j_l)}, \beta_{(k_l)}) - R_{ix}(\alpha_{(j_{l+1})}, \beta_{(k_{l+1})})| \leq C\{|\alpha_{(j_l)} - \alpha_{(j_{l+1})}| + |\beta_{(k_l)} - \beta_{(k_{l+1})}|(\delta_\vartheta + h)\} \equiv \frac{M^{\vartheta}_{n2}}{M^l},$$
$$E|R_{ix}(\alpha_{(j_l)}, \beta_{(k_l)}) - R_{ix}(\alpha_{(j_{l+1})}, \beta_{(k_{l+1})})|^2 \leq h(M^{\vartheta}_{n2})^3/M^l,$$

we have

$$V_l \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big) \exp[-\varepsilon^2 nh\{1 + a_{n\vartheta}(nh/\log n)^{1/2}\}],$$

and (61) thus holds. This together with (59) completes the proof. ■

Lemma 6.13 *$Q_2 \leq M d_n$ a.e., for all large enough $M > 0$, where*

$$d_n = nha^2_{n\vartheta} l_n/h\{1 + a^{-1}_{n\vartheta}(nh/\log n)^{-1/2}\} = o(nha^2_{n\vartheta}),$$

if $nh^3/(\log n)^3 \to \infty$.

Proof Let $X_{ik} = X_i - x_k,\ \mu_{ik} = (1, X_{ik}^\top)^\top,\ K_{ik} = K(X_{ik}^\top \vartheta/h)$ and write $\Phi_{ni}(x_k; \alpha, \beta) - \Phi_{ni}(x; \alpha, \beta) = \xi_{i1} + \xi_{i2} + \xi_{i3}$, where

$$\xi_{i1} = \Big(K_{ik}\mu_{ik} - K_{ix}\mu_{ix}\Big)^\top \alpha \int_0^1 \Big\{\varphi_{ni}(x_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(x_k; 0)\Big\}\, dt,$$
$$\xi_{i2} = K_{ix}\mu_{ix}^\top \alpha \int_0^1 \Big\{\varphi_{ni}(x_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(x; \mu_{ix}^\top(\beta + \alpha t))\Big\}\, dt,$$
$$\xi_{i3} = K_{ix}\mu_{ix}^\top \alpha\{\varphi_{ni}(x; 0) - \varphi_{ni}(x_k; 0)\}.$$

Then $P(Q_2 > M^{3/2}d_n/3) \le \mathrm{T}_n(P_{n1} + P_{n2} + P_{n3})$, where

$$P_{nj} \equiv \max_{1\le k\le \mathrm{T}_n} P\Big(\sup_{x\in\mathcal{D}_k} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} |\sum_{i=1}^n \xi_{ij}| \ge M^{3/2}d_n/9\Big),\ j = 1, 2, 3.$$

Based on Borel-Cantelli lemma, $Q_2 \le M^{3/2}d_n$ almost surely, if $\sum_n \mathrm{T}_n P_{nj} < \infty,\ j = 1, 2, 3.$ Again this can be accomplished through similar approach in Lemma 5.1 in Kong et al(2008). We only deal with P_{nj} to illustrate.

First note that if $\xi_{i1} \ne 0$,then either $K_{ik} \ne 0$ or $K_{ix} \ne 0$. Without loss of generality, suppose $K_{ik} \ne 0$, i.e. $|X_{ix}^\top \vartheta| \le h$, whence $|X_{ix}^\top \theta_0| \le h + |\delta_\vartheta|$ and $|\mu_{ik}^\top(\beta + \alpha t)| \le C\{M_{n\vartheta}^{(1)} + M_{n\vartheta}^{(2)}\}$.

For any fixed $\alpha \in B_n^{(1)}$ and $\beta \in B_n^{(2)}$, let $I_{ik}^{\alpha,\beta} = 1$, if there exists some $t \in [0, 1]$, such that there are discontinuity points of $\varphi(Y_i - a)$ between $\mu_{ik}^\top(\beta(x_k) + \beta + \alpha t))$ and $\mu_{ik}^\top\beta_p(x_k)$; and $I_{ik}^{\alpha,\beta} = 0$, otherwise. Write $\xi_{i1} = \xi_{i1}I_{ik}^{\alpha,\beta} + \xi_{i1}(1 - I_{ik}^{\alpha,\beta})$. As $|(K_{ik}\mu_{ik} - K_{ix}\mu_{ix})^\top\alpha| \le CM_{n\vartheta}^{(1)}l_n/h$ and $|\mu_{ik}^\top(\beta + \alpha t)| \le CM_{n\vartheta}^{(2)}$, we have

$$|\xi_{i1}(1 - I_{ik}^{\alpha,\beta})| \le CM_{n\vartheta}^1 M_{n\vartheta}^2 l_n/h = o(a_{n\vartheta}^2)$$

uniformly in i, α, β and $x \in \mathcal{D}_k$, if $nh^3/\log n^3 \to \infty$. Let $U_{ik} = I\{|X_{ik}^\top\vartheta| \le 2h\}$. As $\xi_{i1} = \xi_{i1}U_{ik}$ (because $l_n = o(h)$), we have

$$\begin{aligned} P\Big(\sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \sup_{x\in\mathcal{D}_k} \Big|\sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta})\Big| > \frac{Md_n}{18}\Big) &\le P\Big(\sum_{i=1}^n U_{ik} > \frac{Mnh}{18C}\Big) \\ &\le P\Big(|\sum_{i=1}^n U_{ik} - EU_{ik}| > \frac{Mnh}{36C}\Big), \qquad (62) \end{aligned}$$

where the second inequality follows from the fact that $EU_{ik} = O(h)$. We can then apply to (62) Bernstein's inequality for independent data or Lemma 5.4 in Kong et al(2008) for dependent case, to obtain the below result

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta})\Big| > Md_n/18\Big) \text{ is summable over } n, \qquad (63)$$

whence $\sum_n \mathrm{T}_n P_{n1} < \infty$, is equivalent to

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^n \xi_{i1} I_{ik}^{\alpha,\beta}\Big| > Md_n/18\Big) \text{ is summable over } n. \qquad (64)$$

To this end, first note that $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in S_{i;k}^{\alpha,\beta}\}$, where

$$\begin{aligned} S_{i;k}^{\alpha,\beta} &= \bigcup_{j=1}^{m} \bigcup_{t\in[0,1]} [a_j - A(X_i, x_k) + \mu_{ik}^\top(\beta+\alpha t), a_j - A(X_i, x_k)] \\ &\subseteq \bigcup_{j=1}^{m} [a_j - CM_{n\vartheta}^{(2)}, a_j + CM_{n\vartheta}^{(2)}] \equiv D_n, \quad \text{for some } C > 0, \\ A(x_1, x_2) &= m(x_1^\top\theta_0) - m(x_2^\top\theta_0) - m'(x_1^\top\theta_0)(x_1 - x_2)^\top\theta_0 \end{aligned}$$

where in the derivation of $S_{i;k}^{\alpha,\beta} \subseteq D_n$, we have used the fact that $|X_{ik}| \le 2h$, $\mu_{ik}^\top(\beta+\alpha t) = O(M_n^{(2)})$ and $A(X_i, x_k) = O(h^2 + |\delta_\vartheta|^2) = o(M_n^{(2)})$ uniformly in i. As $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in D_n\}$, we have $|\xi_{i1}|I_{ik}^{\alpha,\beta} \le |\xi_{i1}|U_{ni}$, where $U_{ni} \equiv I(|X_{ik}| \le 2h)I\{\varepsilon_i \in D_n\}$, which is independent of the choice of α and β. Therefore,

$$\begin{aligned} P\Big(\sup_{\substack{\alpha\in B_n^{(1)},\\ \beta\in B_n^{(2)}}} \Big|\sum_{i=1}^{n} \xi_{i1} I_{ik}^{\alpha,\beta}\Big| > Md_n/18\Big) &\le P\Big(\sum_{i=1}^{n} U_{ni} > MnhM_n^{(2)}/(18C)\Big) \\ &\le P\Big(\sum_{i=1}^{n}(U_{ni} - EU_{ni}) > \frac{MnhM_n^{(2)}}{36C}\Big), \end{aligned} \tag{65}$$

where the first inequality is because $|\xi_{i1}| \le CMa_{n\vartheta}l_n/h$ and the second one because $EU_{ni} = O(hM_n^{(4)})$. Similar to (62), we could apply either Bernstein's inequality for independent data or in dependent case Lemma 5.4 in Kong et al(2008) to see that (64) indeed holds. ■

Lemma 6.14 $\sum_{i,j} Z_{ij} - EZ_{ij} = o(n^2 h a_{n\vartheta})$, *where*

$$Z_{ij} = K_{ij}[\varphi(Y_i - a_j - b_j\theta_0^\top X_{ij}) - \varphi(Y_i - \hat{a}_j - \hat{b}_j\theta_0^\top X_{ij})]\hat{b}_j X_{ij} \tag{66}$$

Proof As $\hat{a}_j - a_j = O(a_{n\vartheta})$, $(\hat{b}_j - b_j) = O\{a_{n\vartheta} + (nh/\log n)^{1/2}/h\}$ and for any $\epsilon > 0$,

$$P\Big\{|\sum_{i,j} Z_{ij} - EZ_{ij}| \ge \epsilon n^2 h a_{n\vartheta}\Big\} \le nP\Big\{|\sum_{i} Z_{ij} - EZ_{ij}| \ge \epsilon nh a_{n\vartheta}\Big\}$$

then (66) would follow if we could show that for any x,

$$P\{\sup_{\substack{\alpha\in B_n^{(1)}\\ \beta\in B_n^{(2)}}} |\sum_i R_{ix}(a,b)| \ge \epsilon nh a_{n\vartheta}\} = O(n^{-a}) \text{ for some } a > 2, \tag{67}$$

where $B_n^{(1)} = \{a \in R : |a - a_x| \le ca_{n\vartheta}\}$, $B_n^{(2)} = \{b \in R : |b - b_x| \le c\{a_{n\vartheta} + (nh/\log n)^{1/2}/h\}\}$, $a_x = m(\theta_0^\top x)$, $b_x = m'(\theta_0^\top x)$, $R_{ix}(a,b) = Z_{ix}(a,b) - EZ_{ix}(a,b)$ and

$$Z_{ix}(a,b) = K_{ix}X_{ix}[\varphi(Y_i - a_x - b_x\theta_0^\top X_{ix}) - \varphi(Y_i - a - b\theta_0^\top X_{ix})], \; K_{ix} = K(X_{ix}^\top\vartheta/h)$$

To this end, partition $B_n^{(i)},\ i=1,2$ into a sequence of sub rectangles $D_1^{(i)},\cdots,D_{J_1}^{(i)},\ i=1,2$ such that

$$|D_{j_1}^{(i)}|=\sup\Big\{|a-a'|:a,a'\in D_{j_1}^{(i)}\Big\}\le M_n^{(i)}/M,\quad 1\le j_1\le J_1;\ M\equiv\epsilon^{-1}$$

where $M_n^{(1)}=ca_{n\vartheta}$, $M_n^{(2)}=c\{a_{n\vartheta}+(nh/\log n)^{1/2}/h\}$ and $J_1\le M$. Choose a point $a_{j_1}\in D_{j_1}^{(1)}$ and $b_{k_1}\in D_{k_1}^{(2)}$. Then

$$\begin{aligned}\sup_{\substack{a\in B_n^{(1)}\\ b\in B_n^{(2)}}}|\sum_i R_{ix}(a,b)| \quad\le\quad & \max_{1\le j_1,k_1\le J_1}\sup_{\substack{a\in D_{j_1}^{(1)},\\ b\in D_{k_1}^{(2)}}}|\sum_{i=1}^n\{R_{ix}(a_{j_1},b_{k_1})-R_{ix}(a,b)\}| \\ &+\max_{1\le j_1,k_1\le J_1}|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|=H_{n1}+H_{n2}.\end{aligned}\tag{68}$$

We first consider H_{n2}.

$$P\Big\{H_{n2}\ge\frac{\varepsilon nha_{n\vartheta}}{2}\Big\}\le J_1^2P\Big\{|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|\ge\frac{\epsilon nha_{n\vartheta}}{2}\Big\}$$

As $R_{ix}(a_{j_1},b_{k_1})$ is bounded and $\mathrm{Var}R_{ix}(a_{j_1},b_{k_1})=O\{h(a_{n\vartheta}+(nh/\log n)^{-1/2}\}$, then by Bernstein's Inequality

$$J_1^2P\Big\{|\sum_{i=1}^n R_{ix}(a_{j_1},b_{k_1})|\ge\frac{\epsilon nha_{n\vartheta}}{2}\Big\}\le CJ_1^2\exp\{-\epsilon^2n^{1/2}h^{3/2}\}=O(n^{-a}),$$

for some $a>2$.

We next consider H_{n1}. For each $j_1=1,\cdots,J_1$ and $i=1,2$, partition each rectangle $D_{j_1}^{(i)}$ further into a sequence of subrectangles $D_{j_1,1}^{(i)},\cdots,D_{j_1,J_2}^{(i)}$. Repeat this process recursively as follows. Suppose after the lth round, we get a sequence of rectangles $D_{j_1,j_2,\cdots,j_l}^{(i)}$ with $1\le j_k\le J_k,\ 1\le k\le l$, then in the $(l+1)$th round, each rectangle $D_{j_1,j_2,\cdots,j_l}^{(i)}$ is partitioned into a sequence of subrectangles $\{D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)},1\le j_l\le J_l\}$ such that

$$|D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}|=\sup\Big\{|a-a'|:a,a'\in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}\Big\}\le M_n^{(i)}/M^{l+1},\ 1\le j_{l+1}\le J_{l+1},$$

where $J_{l+1}\le M$. End this process after the (L_n+2)th round, with L_n being the smallest integer such that

$$(2/M)^{L_n}>a_{n\vartheta}/M_{n\vartheta}^{(2)}\ [\text{which means } 2^{L_n}\le\{M_{n\vartheta}^{(2)}/a_{n\vartheta}\}^{\log(M/2)/\log 2}]\tag{69}$$

Let $D_l^{(i)},\ i=1,2$, denote the set of all subrectangles of $D_0^{(i)}$ after the lth round of partition and a typical element $D_{j_1,j_2,\cdots,j_l}^{(i)}$ of $D_l^{(i)}$ is denoted as $D_{(j_l)}^{(i)}$. Choose a point $a_{(j_l)}\in D_{(j_l)}^{(1)}$ and

$b_{(j_l)} \in D^{(2)}_{(j_l)}$ and define

$$V_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\Big|\sum_{i=1}^{n}\{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a_{j_{l+1}}, b_{k_{l+1}})\}\Big| \geq \frac{\epsilon nh a_{n\vartheta}}{2^{l+1}}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 1,$$

$$Q_l = \sum_{\substack{(j_l)\\(k_l)}} P\Big\{\sup_{\substack{a\,\in\, D^{(1)}_{(j_l)},\\ b\,\in\, D^{(2)}_{(k_l)}}}\Big|\sum_{i=1}^{n}\{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}\Big| \geq \frac{\epsilon nh a_{n\vartheta}}{2^{l}}\Big\},\ 1 \leq l \leq \mathrm{L}_n + 2.$$

Then $Q_l \leq V_l + Q_{l+1},\ 1 \leq l \leq L_n + 1$. We first give a bound for $V_l,\ 1 \leq l \leq L_n + 1$. As $R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a_{j_{l+1}}, b_{k_{l+1}})$ is bounded and

$$E|R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a_{j_{l+1}}, b_{k_{l+1}})|^2 \leq h\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}/M^{l+1}$$

applying Bernstein's Inequality and using (69), we have

$$V_l \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big)\exp[-\epsilon^2 nh\min\{a_{n\vartheta}, a_{n\vartheta}^2(nh/\log n)^{1/2}\}] \leq \Big(\prod_{j=1}^{l+1} J_j^2\Big)\exp(-\epsilon^2 n^{1/2}h^{3/2}) \tag{70}$$

We now focus on Q_{L_n+2}. Recall the definition of $Z_{ix}(a,b)$

$$Z_{ix}(a,b) = K_{ix}[\varphi(Y_i - a_x - b_x\theta_0^\top X_{ix}) - \varphi(Y_i - a - b\theta_0^\top X_{ix})]X_{ix}.$$

For any $a \in D^{(1)}_{(j_l)}$ and $b \in D^{(2)}_{(k_l)}$, let $I_i^{a,b} = 1$, if there is a discontinuity point of $\varphi(.)$ between $Y_i - a_{j_l} - b_{k_l}\theta_0^\top X_{ix}$ and $Y_i - a - b\theta_0^\top X_{ix}$ and $I_i^{a,b} = 0$ otherwise. Write

$$R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b) = \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}I_i^{a,b} + \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}(1 - I_i^{a,b}).$$

Then we have $|\{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}(1 - I_i^{a,b})| \leq C\{a_{n\vartheta} + (nh/\log n)^{-1/2}\}/M^l$ and specifically for $l = \mathrm{L}_n + 2$

$$\begin{aligned} &P\Big\{\sup_{\substack{a\,\in\, D^{(1)}_{(j_l)},\\ b\,\in\, D^{(2)}_{(k_l)}}}\Big|\sum_{i=1}^{n}\{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}(1 - I_i^{a,b})\Big| \geq \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}}\Big\} \\ &\leq P\Big\{\sum_{i=1}^{n} U_i \geq \frac{1}{8}Mnh\Big\} \leq P\Big\{\sum_{i=1}^{n} U_i - EU_i \geq \frac{Mnh}{16}\Big\} \end{aligned}$$

where $U_i = I\{|X_{ix}^\top\vartheta| \leq h\}$ and the first inequality is due to (69). By Bernstein's inequality, this in turn implies that for $l = L_n + 2$

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big)P\Big\{\sup_{\substack{a\,\in\, D^{(1)}_{(j_l)},\\ b\,\in\, D^{(2)}_{(k_l)}}}\Big|\sum_{i=1}^{n}\{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\}(1 - I_i^{a,b})\Big| \geq \frac{\epsilon nh a_{n\vartheta}}{2^{\mathrm{L}_n+3}}\Big\} = O(n^{-a}), \tag{71}$$

for some $a > 2$. Now we have to show similar result for

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big) P\Big\{ \sup_{\substack{a \in D^{(1)}_{(j_l)}, \\ b \in D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\},\ l = L_n + 2.$$

Note that for any $a \in D^{(1)}_{(j_l)}$ and $b \in D^{(2)}_{(k_l)}$, $I_i^{a,b} \leq I\{Y_i \in S_i\}$, where

$$S_i = [a_{j_l} + b_{k_l}\theta_0^\top X_{ix} - CM_n^{(2)}/M^l, a_{j_l} + b_{k_l}\theta_0^\top X_{ix} + CM_n^{(2)}/M^l],$$

which is independent of a, b. Let $U_i = I\{|X_{ix}^\top \vartheta| \leq h\} I\{Y_i \in S_i\}$. As $R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)$ is bounded, we have for $l = L_n + 2$,

$$\begin{aligned} & P\Big\{ \sup_{\substack{a \in D^{(1)}_{(j_l)}, \\ b \in D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\} \\ & \leq P\Big\{ \sum_{i=1}^n U_i \geq \frac{\epsilon n h a_{n\vartheta}}{C2^{\mathrm{L}_n+2}} \Big\} \leq P\Big\{ \sum_{i=1}^n U_i - EU_i \geq \frac{\epsilon n h a_{n\vartheta}}{C2^{\mathrm{L}_n+4}} \Big\} \end{aligned} \tag{72}$$

where the second inequality is due to (69). Applying Bernstein's inequality to the right hand side of (Bern) and observing (69) lead to

$$\Big(\prod_{j=1}^{l+1} J_j^2\Big) P\Big\{ \sup_{\substack{a \in D^{(1)}_{(j_l)}, \\ b \in D^{(2)}_{(k_l)}}} \Big| \sum_{i=1}^n \{R_{ix}(a_{j_l}, b_{k_l}) - R_{ix}(a,b)\} I_i^{a,b} \Big| \geq \frac{\epsilon n h a_{n\vartheta}}{2^{\mathrm{L}_n+3}} \Big\} = O(n^{-a}),\ \text{for } l = L_n + 2$$

for some $a > 2$. This together with (71) implies that $Q_{L_n+2} = O(n^{-a})$ for some $a > 2$. Therefore, based on (70), we have

$$P\Big\{ H_{n2} \geq \frac{\epsilon n h a_{n\vartheta}}{2} \Big\} \leq Q_1 \leq \sum_{l=1}^{L_n+1} V_l + Q_{L_n+2} = O(n^{-a}),$$

for some $a > 2$. ■

Lemma 6.15 *All eigenvalues of $S_2^{-1}\Omega_1$ are nonnegative and strictly smaller than* 1; *ϑ is the only eigenvector of $S_2^{-1}\Omega_1$ corresponding to eigenvalue* 0.

Proof By the definition of S_2 and Ω_1 and the Cauchy-Schwarz Inequality that

$$\begin{aligned} & E\{g(X)(X-x)|X^\top\vartheta = x^\top\vartheta\} E\{g(X)(X-x)|X^\top\vartheta = x^\top\vartheta\}^\top \\ & \leq E\{g(X)|X^\top\vartheta = x^\top\vartheta\} E\{g(X)(X-x)(X-x)^\top|X^\top\vartheta = x^\top\vartheta\}, \end{aligned}$$

we know for any ϑ_1, if $\vartheta_1^\top(S_2-\Omega_1)\vartheta_1=0$, then for any x, there exists some C, such that

$$\{g(X)\}^{1/2}\vartheta_1^\top(X-x)\equiv C\{g(X)\}^{1/2},\ \text{for all}\ X^\top\vartheta=x^\top\vartheta\Rightarrow\vartheta_1\equiv\vartheta \tag{73}$$

We need the following assumptions. For any $\vartheta_1,\vartheta_2\in\Theta$,

$$E\{g(X)\vartheta_2^\top(X-x)(X-x)^\top\vartheta_2|\vartheta_1^\top X=\vartheta_1^\top x\}=0,\ \text{for any}\ x\in R^d\Rightarrow\vartheta_1\equiv\vartheta_2. \tag{74}$$

$$E\{g(X)\vartheta_2^\top(X-x)|\vartheta_1^\top X=\vartheta_1^\top x\}=0,\ \text{for any}\ x\in R^d\Rightarrow\vartheta_1\equiv\vartheta_2. \tag{75}$$

For any nonzero eigenvalue λ and corresponding eigenvector $x(\neq\vartheta)$

$$S_2^{-1}\Omega_1x=\lambda x\Rightarrow\Omega_1x=\lambda S_2x\Rightarrow x^\top\Omega_1x=\lambda x^\top S_2x\Rightarrow\lambda>0$$

Next we show that $\lambda_{max}<1$ by contradiction. If not, suppose x is the corresponding eigenvector,

$$\begin{aligned}
&S_2^{-1}\Omega_1x=\lambda_{max}x\Rightarrow\Omega_1x=\lambda_{max}S_2x\Rightarrow x^\top\Omega_1x=\lambda_{max}x^\top(\Omega_1+S_2-\Omega_1)x\\
&\Rightarrow(1-\lambda_{max})x^\top\Omega_1x=x^\top(S_2-\Omega_1)x\le0\overset{(73)}{\Rightarrow}x\equiv\vartheta\Rightarrow\lambda_{max}=0\\
&(S_2+\theta_0\theta_0^\top)^{-1}(\Omega_1+\theta_0\theta_0^\top)x=\lambda_{max}x\Rightarrow(\Omega_1+\theta_0\theta_0^\top)x=\lambda_{max}(S_2+\theta_0\theta_0^\top)x\\
&\Rightarrow x^\top(\Omega_1+\theta_0\theta_0^\top)x=\lambda_{max}x^\top(S_2+\theta_0\theta_0^\top)x\Rightarrow x^\top\Omega_1x\ge\lambda_{max}x^\top S_2x(\because\lambda_{max}x\ge1)
\end{aligned}$$

which contradicts the fact that $S_2-\Omega_1>0$ if $x\neq\theta_0$. ■

REFERENCES

Andrews, D.W.K. (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62**, 43-72.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer, New York.

Chaudhuri, P., Doksum, K. and Samarov, A. (1997) On average derivative quantile regression. *Ann. Statist.* **25**, 715-44.

Fan, Y. and Li, Q. (1999) Central Limit Theorems for Degenerate U-statistics of Absolutely Regular Processes With Applications to Model Specification Testing. *Journal of Nonparametric Statistics* **10** 245-71.

Giné, E. ,Kwapień, S., Latala, R. and Zinn, J. (2001) The LIL for canonical U-statistics of order 2. *Ann. Probab.* **29**, 520-57.

Ibragimov, I.A. and Linnik, Y.V. (1971) Independent and stationary sequences of random variables. Wolters-Noordhoff Publishing: Groningen

Koenker, R. (2005) *Quantile Regression.* Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978) Regression Quantiles. Econometrica. **46** (1), 3350.

Kong, E., Linton, O. and Xia, Y. (2008) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model, *Econometric Theory* (to appear)

Korolyuk, V. S. and Borovskikh, Yu. V. (1989) Law of the iterated logarithm for U-statistics. *Ukrainian Mathematical Journal* **1**, 89-92.

Lange, T. Rahbek, A. and Jensen, S.T. (2006) Estimation and Asymptotic Inference in the First Order AR-ARCH Model. Preprint

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186-99.

Sen, P.K. (1972) Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes *Z. Wahrischeinlichkeitstheorie verw. Geb.* **25**, 71-82.

Welsh, A.H. (1996) Robust estimation of smooth regression and spread functions and their derivatives *Statistica Sinica* **6**, 347-366.

Xia, Y. and Tong, H. (2006) On efficiencies of estimations for single index models. pp. 63-86 in *Frontiers in Statistics,* Eds. Fan, J. and Hourl, K. World Scientific Publishing: New York.