Bernoulli **20**(3), 2014, 1532–1559
DOI: 10.3150/13-BEJ532

Asymptotics of nonparametric L-1 regression models with dependent data

ZHIBIAO ZHAO[1,*], YING WEI[2,**] and DENNIS K.J. LIN[1,†]

[1]*Department of Statistics, Penn State University, University Park, PA 16802, USA.*
E-mail: [*]zuz13@stat.psu.edu; [†]DKL5@psu.edu
[2]*Department of Biostatistics, Columbia University, 722 West 168th St., New York, NY 10032, USA.*
E-mail: [**]ying.wei@columbia.edu

We investigate asymptotic properties of least-absolute-deviation or median quantile estimates of the location and scale functions in nonparametric regression models with dependent data from multiple subjects. Under a general dependence structure that allows for longitudinal data and some spatially correlated data, we establish uniform Bahadur representations for the proposed median quantile estimates. The obtained Bahadur representations provide deep insights into the asymptotic behavior of the estimates. Our main theoretical development is based on studying the modulus of continuity of kernel weighted empirical process through a coupling argument. Progesterone data is used for an illustration.

Keywords: Bahadur representation; coupling argument; least-absolute-deviation estimation; longitudinal data; nonparametric estimation; time series; weighted empirical process

1. Introduction

There is a vast literature on the nonparametric location-scale model $Y = \mu(X) + s(X)e$, where X, Y, and e are the covariates, response, and error, respectively. Given observations $\{(X_j, Y_j)\}_{j=1,\ldots,m}$, the latter model has been studied under various settings of data structure. In terms of the dependence structure, there are independent data and time series data scenarios; in terms of the design point X, there are random-design and fixed-design $X_j = j/m$ settings. In these settings, we usually assume that either (X_j, Y_j) are independent observations from subjects $j = 1, \ldots, m$, or $\{(X_j, Y_j)\}_{j=1,\ldots,m}$ is a sequence of time series observations from the same subject. We refer the reader to Fan and Yao [9] and Li and Racine [20] for an extensive exposition of related works.

In this article, we are interested in the following nonparametric location-scale model with serially correlated data from multiple subjects:

$$Y_{i,j} = \mu(x_{i,j}) + s(x_{i,j})e_{i,j}, \qquad 1 \le j \le m_i, 1 \le i \le n, \tag{1.1}$$

where, for each subject i, $\{(x_{i,j}, Y_{i,j})\}_{j=1,\ldots,m_i}$ is the sequence of covariates and responses, and $\{e_{i,j}\}_{j=1,\ldots,m_i}$ is the corresponding error process. We study (1.1) under a general dependence framework for $\{e_{i,j}\}_{j\in\mathbb{N}}$ that allows for both longitudinal data and some spatially correlated data. In typical longitudinal studies, $x_{i,j}$ represents measurement time or covariates at time j, then it is reasonable to assume that $\{e_{i,j}\}_{j\in\mathbb{Z}}$ is a causal time series, that is, the current observation depends only on past but not future observations. In other applications, however, measurements may be dependent on both the left and right neighboring measurements, especially when $x_{i,j}$ represents measurement location. A good example of this type of data is the vertical density profile data in Walker and Wright [27]; see also Section 2.1 for more details. To accommodate this, we propose a general error dependence structure, which can be viewed as an extension of the one-sided causal structure in Wu [32] and Dedecker and Prieur [8] to a two-sided noncausal setting. The proposed dependence framework allows for many linear and nonlinear processes.

We are interested in nonparametric estimation of the location function $\mu(\cdot)$ and the scale function $s(\cdot)$. Least-squares based nonparametric methods have been extensively studied for both time series data (Fan and Yao [9]) and longitudinal data (Hoover *et al.* [16], Fan and Zhang [10], Wu and Zhang [31], Yao, Müller and Wang [35]). While they perform well for Gaussian errors, least-squares based methods are sensitive to extreme outliers, especially when the errors have a heavy-tailed distribution. By contrast, robust estimation methods impose heavier penalty on far-deviated data points to reduce the impact from extreme outliers. For example, median quantile regression uses the absolute loss and the resultant estimator is based on sample local median. Since Koenker and Bassett [19], quantile regression has become popular in parametric and nonparametric inferences and we refer the reader to Yu, Lu and Stander [37] and Koenker [18] for excellent expositions. Recently, He, Fu and Fung [12], Koenker [17] and Wang and Fygenson [28] applied quantile regression techniques to parameter estimation of parametric longitudinal models, He, Zhu and Fung [13] studied median regression for semiparametric longitudinal models, and Wang, Zhu and Zhou [29] studied inferences for a partially linear varying-coefficient longitudinal model. Here we focus on quantile regression based estimation for the nonparametric model (1.1).

We aim to study the asymptotic properties, including uniform Bahadur representations and asymptotic normalities, of the least-absolute-deviation or median quantile estimates for model (1.1) under a general dependence structure. Nonparametric quantile regression estimation has been studied mainly under either the i.i.d. setting (Bhattacharya and Gangopadhyay [4], Chaudhuri [7], Yu and Jones [36]) or the strong mixing setting (Truong and Stone [26], Honda [15], Cai [6]). There are relatively scarce results on Bahadur representations of conditional quantile estimates. Bhattacharya and Gangopadhyay [4] and Chaudhuri [7] obtained point-wise Bahadur representations for conditional quantile estimation of i.i.d. data. For mixing stationary processes, Honda [15] obtained point-wise and uniform Bahadur representations of conditional quantile estimates. For stationary random fields, Hallin, Lu and Yu [11] obtained a point-wise Bahadur representation for spatial quantile regression function under spatial mixing conditions. Due to the nonstationarity and dependence structure, it is clearly challenging to establish Bahadur representations in the context of (1.1).

Our contribution here is mainly on the theoretical side. We establish uniform Bahadur representations for the least-absolute-deviation estimates of $\mu(\cdot)$ and $\sigma(\cdot)$ in (1.1). To derive the uniform Bahadur representations, the key ingredient is to study the modulus of continuity of certain kernel weighted empirical processes of the nonstationary observations $Y_{i,j}$ in (1.1). Empirical processes have been extensively studied under various settings, including the i.i.d. setting (Shorack and Wellner [25]), linear processes (Ho and Hsing [14]), strong mixing setting (Andrews and Pollard [2], Shao and Yu [23]), and general causal stationary processes (Wu [33]). Using a coupling argument to approximate the dependent process by an m-dependent process with a diverging m, we study the modulus of continuity of weighted empirical processes, and the latter result serves as a key tool in establishing our uniform Bahadur representations. These Bahadur representations provide deep insights into the asymptotic behavior of the estimates, and in particular they provide theoretical justification for the profile control chart methodologies in Wei, Zhao and Lin [30]. These technical treatments are also of interest in other nonparametric problems involving dependent data.

The article is organized as follows. In Section 2, we introduce the error dependence structure with examples. In Section 3, we study weighted empirical process through a coupling argument. Section 4 contains uniform Bahadur representations and asymptotic normality. Section 5 contains an illustration using progesterone data. Possible extensions to spatial setting are discussed in Section 6. Proofs are provided in Section 7.

2. Error dependence structure

First, we introduce some notation used throughout this article. For $a, b \in \mathbb{R}$, let $\lfloor a \rfloor$ be the integer part of a, $a \vee b = \max(a, b)$, and $a \wedge b = \min(a, b)$. For a random variable $Z \in \mathcal{L}^q$, $q > 0$, if $\|Z\|_q = [\mathbb{E}(|Z|^q)]^{1/q} < \infty$. Let $\mathcal{C}^r(\mathcal{S})$ be the set of functions with bounded derivatives up to order r on a set $\mathcal{S} \subset \mathbb{R}$.

Assume that, for each i, the error process $\{e_{i,j}\}_{j\in\mathbb{N}}$ in (1.1) is an independent copy from a stationary process $\{e_j\}_{j\in\mathbb{N}}$ which has the representation

$$e_j = G(\varepsilon_j, \varepsilon_{j\pm1}, \varepsilon_{j\pm2}, \ldots), \tag{2.1}$$

where ε_j, $j \in \mathbb{Z}$, are i.i.d. random innovations, and G is a measurable function such that e_j is well defined. We can view (2.1) as an input-output system with $(\varepsilon_j, \varepsilon_{j\pm1}, \varepsilon_{j\pm2}, \ldots)$, G, and e_j being, respectively, the input, filter, and output. Wu [32] considered the causal time series case that e_j depends only on the past innovations $\varepsilon_j, \varepsilon_{j-1}, \ldots$. In contrast, (2.1) allows for noncausal models and is particularly useful for applications that do not have a time structure. For example, if $x_{i,j}$ are locations, then the corresponding measurement $y_{i,j}$ depends on both the left and right neighboring measurements.

Condition 2.1. *Let $\{\varepsilon'_j\}_{j\in\mathbb{Z}}$ be i.i.d. copies of $\{\varepsilon_j\}_{j\in\mathbb{Z}}$. There exist constants $q > 0$ and $\rho \in (0, 1)$ such that*

$$\|e_0 - e_0(k)\|_q = \mathrm{O}(\rho^k), \qquad \textit{where } e_0(k) = G(\varepsilon_0, \varepsilon_{\pm1}, \ldots, \varepsilon_{\pm k}, \varepsilon'_{\pm(k+1)}, \varepsilon'_{\pm(k+2)}, \ldots). \tag{2.2}$$

In (2.2), $e_0(k)$ can be viewed as a coupling process of e_0 with $\{\varepsilon_r\}_{|r|\ge k+1}$ replaced by the i.i.d. copy $\{\varepsilon'_r\}_{|r|\ge k+1}$ while keeping the nearest $2k + 1$ innovations $\{\varepsilon_r\}_{|r|\le k}$. In particular, if e_0 does not depend on $\{\varepsilon_r\}_{|r|\ge k+1}$, then $e_0(k) = e_0$. Thus, $\|e_0 - e_0(k)\|_q$ quantifies the contribution of $\{\varepsilon_r\}_{|r|\ge k+1}$ to e_0, and (2.2) states that the contribution decays exponentially in k. Shao and Wu [24] and Dedecker and Prieur [8] [cf. equation (4.2) therein] considered one-sided causal version of (2.2) where e_0 depends only on $\{\varepsilon_r\}_{r\le0}$.

Propositions 2.1–2.2 below indicate that, if $\{e_i\}$ satisfies (2.2), then its properly transformed process also satisfies (2.2).

Proposition 2.1. *For* $0 < \varsigma \le 1$ *and* $\upsilon \ge 0$, *define the collection of functions* h

$$\mathcal{H}(\varsigma, \upsilon) = \{h: |h(x) - h(x')| \le c|x - x'|^{\varsigma}(1 + |x| + |x'|)^{\upsilon}, x, x' \in \mathbb{R}\}, \tag{2.3}$$

where c *is a constant. Suppose* $\{e_j\}$ *satisfies* (2.2). *Then the transformed process* $e_j^* = h(e_j)$ *satisfies* (2.2) *with* (q, ρ) *replaced by* $q^* = q/(\varsigma + \upsilon)$ *and* $\rho^* = \rho^{\varsigma}$.

In (2.3), $\mathcal{H}(\varsigma, 0)$ is the class of uniformly Hölder-continuous functions with index ς. If $h(x) = |x|^b$, $b > 1$, then $h \in \mathcal{H}(1, b-1)$. Clearly, all functions in $\mathcal{H}(\varsigma, 0)$ are continuous. Interestingly, for noncontinuous transformations, the conclusion may still hold; see Proposition 2.2 below, where $\mathbf{1}$ is the indicator function.

Proposition 2.2. *Let* e_0 *have a bounded density. Suppose* $\{e_j\}$ *satisfies* (2.2). *Then, for any given* x, $\{\mathbf{1}_{e_j \le x}\}$ *satisfies* (2.2) *with* ρ *replaced by* $\rho^* = \rho^{1/(1+q)}$.

Propositions 2.1–2.2 along with the examples below show that the error structure (2.1) and Condition 2.1 are sufficiently general to accommodate many popular linear and nonlinear time series models and their properly transformed processes.

Example 2.1 (m-dependent sequence). Assume that $e_j = G(\varepsilon_j, \varepsilon_{j\pm1}, \dots, \varepsilon_{j\pm m})$ for a measurable function G. Then e_j depends only on the nearest $2m+1$ innovations $\varepsilon_j, \varepsilon_{j\pm1}, \dots, \varepsilon_{j\pm m}$. Clearly, $\{e_j\}_{j\in\mathbb{Z}}$ form a $(2m+1)$-dependent sequence, $\|e_0 - e_0(k)\|_q = 0$ for $k \ge m$, and (2.2) trivially holds. If $m = 0$, then e_j are i.i.d. random variables.

Example 2.2 (Noncausal linear processes). Consider the noncausal linear process $e_j = \sum_{r=-\infty}^{\infty} a_r \varepsilon_{j-r}$. If $\varepsilon_j \in \mathcal{L}^q$ and $a_j = \mathrm{O}(\rho^{|j|})$, then it is easy to see that (2.2) holds.

Example 2.3 (Iterated random functions). Consider random variables e_j defined by

$$e_j = R(e_{j-1}, \dots, e_{j-d}; \varepsilon_j), \tag{2.4}$$

where ε_j, $j \in \mathbb{Z}$, are i.i.d. random innovations, and R is a random map. Many widely time series models are of form (2.4), including threshold autoregressive model $e_j = a\max(e_{j-1}, 0) + b\min(e_{j-1}, 0) + \varepsilon_j$, autoregressive conditional heteroscedastic model $e_j = \varepsilon_j(a^2 + b^2 e_{j-1}^2)^{1/2}$, random coefficient model $e_j = (a + b\varepsilon_j)e_{j-1} + \varepsilon_j$, and exponential autoregressive model $e_j = [a + b\exp(-ce_{j-1}^2)]e_{j-1} + \varepsilon_j$, among others. Suppose there exists z_0 such that $R(z_0; \varepsilon_0) \in \mathcal{L}^q$ and there exist constants $a_1, \dots, a_d$ such that

$$\sum_{j=1}^{d} a_j < 1 \quad \text{and} \quad \|R(z; \varepsilon_0) - R(z'; \varepsilon_0)\|_q^{1\wedge q} \le \sum_{j=1}^{d} a_j |z_j - z_j'|^{1\wedge q}$$

holds for all $z = (z_1, \dots, z_d)$, $z' = (z_1', \dots, z_d')$. By Shao and Wu [24], (2.2) holds.

2.1. Some examples

The imposed dependence structure and hence the developed results in Sections 3–4 below are potentially applicable to a wide range of practical data types. We briefly mention some below.

(*Time series data*). In the special case of $n = 1$, $m_1 = m \to \infty$ and $(x_{1,j}, Y_{1,j}, e_{1,j}) = (x_j, Y_j, e_j)$ for a stationary time series $\{e_j\}$, (1.1) becomes the usual nonparametric location-scale model $Y_j = \mu(x_j) + s(x_j)e_j$ with time series data. The latter model has been extensively studied under both the random-design case and the fixed-design case $x_j = j/n$. See Fan and Yao [9] for an excellent introduction to various local least-squares based methods under mixing settings. Quantile regression based estimations have been studied in Truong and Stone [26], Honda [15], and Cai [6] for mixing processes. Despite the popularity of mixing conditions, it is generally difficult to verify mixing conditions even for linear processes. For example, for the autoregressive model $X_i = \rho X_{i-1} + \varepsilon_i$, $\rho \in (0, 1/2]$, where ε_i are i.i.d. Bernoulli random variables $\mathbb{P}(\varepsilon_i = 1) = 1 - \mathbb{P}(\varepsilon_i = 0) = q \in (0, 1)$, the stationary solution is not strong mixing (Andrews [1]). By contrast, as shown above, the imposed Condition 2.1 is easily verifiable for many linear and nonlinear time series models and their proper transformations.

(*Longitudinal data*). For each subject i, if $x_{i,j}$ is the jth measurement time or the covariates at time j, $Y_{i,j}$ is the corresponding response, and $\{e_{i,j}\}_{j\in\mathbb{N}}$ is a stationary causal process [e.g., $e_j = G(\varepsilon_j, \varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)$ in (2.1) depends only on the past], then (1.1) becomes a typical longitudinal data setting. For example, Section 5.2 re-examines the well-studied progesterone data using the proposed methods. Another popular longitudinal data example is the CD4 cell percentage in HIV infection from the Multicenter AIDS Cohort Study. Based on least-squares methods, this data has been studied previously in Hoover *et al.* [16] and Fan and Zhang [10]. We can examine how the response function (CD4 cell percentage) varies with measurement time (age) using the proposed robust estimation method in Section 4.

(*Spatially correlated data*). In the vertical density data of Walker and Wright [27], manufacturers are concerned about engineered wood boards' density, which determines fiberboard's overall quality. For each board, densities are measured at various locations along a designated vertical line. In this example, measurements depend on both the left and right neighboring measurements, and it is reasonable to impose the dependence structure (2.1). See Wei, Zhao and Lin [30] for a detailed analysis. Also, as will be discussed in Section 6, the two-sided framework (2.1) can be extended to spatial lattice settings. We point out that the structure in (1.1) and (2.1) differs from the usual spatial model setting in the sense that (1.1) allows for observations from multiple independent subjects whereas the latter usually assumes that all observations are spatially correlated (see, e.g., Hallin, Lu and Yu [11] for quantile regression of spatial data).

3. Weighted empirical process

In this section, we study weighted empirical processes through a coupling argument. Dependence is the main difficulty in extending results developed for independent data to dependent data. For mixing processes, the widely used large-block-small-block technique partitions the data into asymptotically independent blocks. Here, we adopt a coupling argument which copes well with the dependence structure in Section 2.

We now illustrate the basic idea. By (2.1), the error $e_{i,j}$ in (1.1) has the representation

$$e_{i,j} = G(\varepsilon_{i,j}, \varepsilon_{i,j\pm1}, \varepsilon_{i,j\pm2}, \dots)$$

for i.i.d. innovations $\varepsilon_{i,j}$, $i, j \in \mathbb{Z}$. Thus, $\{e_{i,j}\}_{j\in\mathbb{Z}}$ is a dependent series for each fixed i, whereas $\{e_{i_1,j}\}_{j\in\mathbb{Z}}$ and $\{e_{i_2,j}\}_{j\in\mathbb{Z}}$ are two independent series for $i_1 \neq i_2$. Let $\varepsilon'_{i,j,k}$, $i, j, k \in \mathbb{Z}$, be i.i.d. copies of $\varepsilon_{i,j}$. For $k_n \in \mathbb{N}$, define the coupling process of $e_{i,j}$ as

$$e_{i,j}(k_n) = G\bigl(\varepsilon_{i,j}, \varepsilon_{i,j\pm1}, \dots, \varepsilon_{i,j\pm k_n}, \varepsilon'_{i,j,j\pm(k_n+1)}, \varepsilon'_{i,j,j\pm(k_n+2)}, \dots\bigr) \tag{3.1}$$

by replacing all but the nearest $2k_n + 1$ innovations with i.i.d. copies. We call k_n the coupling lag. Clearly, $e_{i,j}(k_n)$ has the same distribution as $e_{i,j}$.

By construction, for each fixed i, $\{e_{i,j}(k_n)\}_{j\in\mathbb{Z}}$ form $(2k_n+1)$-dependent sequence in the sense that $e_{i,j}(k_n)$ and $e_{i,j'}(k_n)$ are independent if $|j - j'| \geq 2k_n + 1$. Consequently, for each fixed i and s, $\{e_{i,(j-1)(2k_n+1)+s}(k_n)\}_{j\in\mathbb{Z}}$ are i.i.d. The latter property helps us reduce the dependent data to an independent case. On the other hand, under (2.2), $\|e_{i,j} - e_{i,j}(k_n)\|_q = \mathrm{O}(\rho^{k_n})$ is sufficiently small with properly chosen k_n and hence the coupling process is close enough to the original one. Similarly, for $Y_{i,j}$ in (1.1), define its coupling process:

$$\tilde{Y}_{i,j} = \mu(x_{i,j}) + s(x_{i,j})e_{i,j}(k_n). \tag{3.2}$$

First, we present a general result regarding the sum of functions of the coupling process $\tilde{Y}_{i,j}$. Let $\mathcal{V}_n$ be any finite set. For real-valued functions $g_{i,j}(y, v)$, $i, j \in \mathbb{N}$, defined on $\mathbb{R} \times \mathcal{V}_n$ such that $\mathbb{E}[g_{i,j}(\tilde{Y}_{i,j}, v)] = 0$ for all $v \in \mathcal{V}_n$, define

$$H_n(v) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} g_{i,j}(\tilde{Y}_{i,j}, v), \qquad v \in \mathcal{V}_n.$$

Throughout, let $N_n = m_1 + \cdots + m_n$ be the total number of observations.

Theorem 3.1. *Assume that the cardinality $|\mathcal{V}_n|$ of $\mathcal{V}_n$ and the coupling lag k_n grow no faster than a polynomial of N_n. Further assume $|g_{i,j}(y, v)| \leq c$ for a constant $c < \infty$, and for some sequence χ_n,*

$$\max_{v\in\mathcal{V}_n} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \mathbb{E}\bigl[g_{i,j}^2(\tilde{Y}_{i,j}, v)\bigr] \leq \chi_n. \tag{3.3}$$

(i) *If* $\chi_n = \mathrm{O}(1)$, *then* $\max_{v\in\mathcal{V}_n} |H_n(v)| = \mathrm{O}_\mathrm{p}(k_n \log N_n)$.
(ii) *If* $\sup_n \log N_n/\chi_n < \infty$, *then* $\max_{v\in\mathcal{V}_n} |H_n(v)| = \mathrm{O}_\mathrm{p}[k_n(\chi_n \log N_n)^{1/2}]$.

By Theorem 3.1, the magnitude of $\max_{v\in\mathcal{V}_n} |H_n(v)|$ increases with the coupling lag k_n. Intuitively, as k_n increases, there is stronger dependence in the coupling process $\tilde{Y}_{i,j}$ and consequently a larger bound for $H_n(v)$. Therefore, a small k_n is preferred in order to have a tight bound for $H_n(v)$. On the other hand, a reasonably large k_n is needed in order for the coupling process to be sufficiently close to the original process. Under (2.2), for $k_n = \mathrm{O}(\log N_n)$, the coupling process converges to the original one at a polynomial rate, and meanwhile the maximum bound in Theorem 3.1 is optimal up to a logarithm factor. For example, if $\chi_n = \mathrm{O}(1)$, then $\max_{v\in\mathcal{V}_n} |H_n(v)| = \mathrm{O}_{\mathrm{p}}[(\log N_n)^2]$; if $\sup_n \log N_n/\chi_n < \infty$, then $\max_{v\in\mathcal{V}_n} |H_n(v)| = \mathrm{O}_{\mathrm{p}}\{[\chi_n(\log N_n)^3]^{1/2}\}$.

In what follows, we consider the special case of weighted empirical process, which plays an essential role in quantile regression. Let $\varpi_{i,j}(x) \ge 0$ be nonrandom weights that may depend on x. Consider the weighted empirical process

$$F_n(x, y) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \varpi_{i,j}(x)\mathbf{1}_{Y_{i,j}\le y}. \tag{3.4}$$

To study $F_n(x, y)$, recall $\tilde{Y}_{i,j}$ in (3.2) and define the coupling empirical process

$$\tilde{F}_n(x, y) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \varpi_{i,j}(x)\mathbf{1}_{\tilde{Y}_{i,j}\le y}. \tag{3.5}$$

Under mild regularity conditions, Theorem 3.2 below states that $F_n(x, y)$ can be uniformly approximated by $\tilde{F}_n(x, y)$ with proper choice of the coupling lag k_n.

Condition 3.1. (i) *$\varpi_{i,j}(x) \le c$ uniformly for some constant $c < \infty$.* (ii) *$\mu(x_{i,j})$ is uniformly bounded.* (iii) *$s(x_{i,j}) > 0$ is uniformly bounded away from zero and infinity.*

Theorem 3.2. *Assume that Conditions 2.1 and 3.1 hold. In (3.1), let the coupling lag $k_n = \lfloor\lambda \log N_n\rfloor$ for some $\lambda > (q+1)/[q\log(1/\rho)]$, where $N_n = m_1 + \cdots + m_n$. Then*

$$\sup_{x,y\in\mathbb{R}} \bigl|F_n(x, y) - \tilde{F}_n(x, y)\bigr| = \mathrm{O}_{\mathrm{p}}\bigl[(\log N_n)^2\bigr].$$

To study asymptotic Bahadur representations of quantile regression estimates, a key step is to study the modulus of continuity of $F_n(x, y)$, defined by

$$D_n(\delta, x, y) = \bigl\{F_n(x, y+\delta) - \mathbb{E}\bigl[F_n(x, y+\delta)\bigr]\bigr\} - \bigl\{F_n(x, y) - \mathbb{E}\bigl[F_n(x, y)\bigr]\bigr\}. \tag{3.6}$$

Intuitively, $D_n(\delta, x, y)$ measures the oscillation of the centered empirical process $F_n(x, y) - \mathbb{E}[F_n(x, y)]$ in response to a small perturbation δ in y.

The dependence structure in Section 2 along with the coupling argument provides a convenient framework to study $D_n(\delta, x, y)$. Recall $\tilde{F}_n(x, y)$ in (3.5). For $D_n(\delta, x, y)$ in (3.6), define its coupling process

$$\tilde{D}_n(\delta, x, y) = \bigl\{\tilde{F}_n(x, y+\delta) - \mathbb{E}\bigl[\tilde{F}_n(x, y+\delta)\bigr]\bigr\} - \bigl\{\tilde{F}_n(x, y) - \mathbb{E}\bigl[\tilde{F}_n(x, y)\bigr]\bigr\}. \tag{3.7}$$

Notice that $e_{i,j}(k_n)$ and $e_{i,j}$ have the same distribution, so $\mathbb{E}[F_n(x, y)] = \mathbb{E}[\tilde{F}_n(x, y)]$. By Theorem 3.2, it is easy to see that, uniformly over x, y, δ,

$$\bigl|D_n(\delta, x, y) - \tilde{D}_n(\delta, x, y)\bigr| \le 2 \sup_{x,y\in\mathbb{R}} \bigl|F_n(x, y) - \tilde{F}_n(x, y)\bigr| = \mathrm{O}_{\mathrm{p}}\bigl[(\log N_n)^2\bigr]. \tag{3.8}$$

Therefore, the asymptotic properties of $D_n(\delta, x, y)$ are similar to that of $\tilde{D}_n(\delta, x, y)$, which can be studied through Theorem 3.1.

Condition 3.2. (i) $\varpi_{i,j}(\cdot) = 0$ *outside a common bounded interval for all* i, j. (ii) *There exist* τ_n *and* ϕ_n *such that*

$$\sup_{x \neq x'} \frac{|\varpi_{i,j}(x) - \varpi_{i,j}(x')|}{|x - x'|} \leq \tau_n \quad \textit{and} \quad \sup_x \sum_{i=1}^{n} \sum_{j=1}^{m_i} \varpi_{i,j}^2(x) \leq \phi_n. \tag{3.9}$$

Theorem 3.3. *Assume that Conditions* 2.1 *and* 3.1–3.2 *hold. Further assume* $\delta_n \to 0$, $\sup_n \log N_n/(\delta_n \phi_n) < \infty$, *and that* $1/\delta_n + \tau_n$ *grows no faster than a polynomial of* N_n. *Then*

$$\sup_{|\delta| \leq \delta_n, x, y \in \mathbb{R}} \left|D_n(\delta, x, y)\right| = \mathrm{O_p}\left\{\left[\delta_n \phi_n (\log N_n)^3\right]^{1/2}\right\}. \tag{3.10}$$

4. Quantile regression and Bahadur representations

For a random variable Z, denote by $\mathcal{Q}(Z) = \inf\{z \in \mathbb{R}, \mathbb{P}(Z \leq z) \geq 1/2\}$ the median of Z, and similarly denote by $\mathcal{Q}(\cdot|\cdot)$ the conditional median operator. To ensure identifiability of μ and s in (1.1), without loss of generality we assume $\mathcal{Q}(e_{i,j}) = 0$ and $\mathcal{Q}(|e_{i,j}|) = 1$.

Note that $\mathcal{Q}(Y_{i,j}|x_{i,j} = x) = \mu(x)$. Applying a kernel localization technique, we propose the following least-absolute-deviation or median quantile estimate of $\mu(x)$:

$$\hat{\mu}(x) = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{m_i} |Y_{i,j} - \theta| K_{b_n}(x_{i,j} - x), \qquad \text{where } K_{b_n}(u) = K(u/b_n) \tag{4.1}$$

for a nonnegative kernel function K satisfying $\int_{\mathbb{R}} K(u) = 1$, and $b_n > 0$ is a bandwidth. The estimate $\hat{\mu}_{b_n}(x)$ pools together information across all subjects, an appealing property especially when some subjects have sparse observations. By the Bahadur representation in Theorem 4.1 below, the bias term of $\hat{\mu}(x) - \mu(x)$ is of order $\mathrm{O}(b_n^2)$. Following Wu and Zhao [34], we adopt a jackknife bias-correction technique. In (4.1), denote by $\hat{\mu}(x|b_n)$ and $\hat{\mu}(x|\sqrt{2}b_n)$ the estimates of $\mu(x)$ using bandwidth b_n and $\sqrt{2}b_n$, respectively. The bias-corrected jackknife estimator is

$$\tilde{\mu}_(x) = 2\hat{\mu}(x|b_n) - \hat{\mu}(x|\sqrt{2}b_n), \tag{4.2}$$

which can remove the second-order bias term $\mathrm{O}(b_n^2)$ in $\hat{\mu}(x)$.

After estimating $\mu(\cdot)$, we estimate $s(\cdot)$ based on residuals. Notice that $\mathcal{Q}(|e_{i,j}|) = 1$ implies $\mathcal{Q}(|Y_{i,j} - \mu(x)||x_{i,j} = x) = s(x)$. Therefore, we propose

$$\hat{s}(x) = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left|\left|Y_{i,j} - \tilde{\mu}(x)\right| - \theta\right| K_{h_n}(x_{i,j} - x), \tag{4.3}$$

where $h_n > 0$ is another bandwidth, and $\tilde{\mu}(x)$ is the bias-corrected jackknife estimator in (4.2). As in (4.2), we adopt the following bias-corrected jackknife estimator

$$\tilde{s}(x) = 2\hat{s}(x|h_n) - \hat{s}(x|\sqrt{2}h_n). \tag{4.4}$$

Remark 4.1. By $\mathcal{Q}(|Y_{i,j} - \mu(x_{i,j})||x_{i,j} = x) = s(x)$, an alternative estimator of $s(x)$ is

$$\bar{s}(x) = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \big| |Y_{i,j} - \tilde{\mu}(x_{i,j})| - \theta \big| K_{h_n}(x_{i,j} - x). \tag{4.5}$$

The difference between (4.3) and (4.5) is that (4.3) uses $\tilde{\mu}(x)$ whereas (4.5) uses $\tilde{\mu}(x_{i,j})$. Since K has bounded support, only those $x_{i,j}$'s with $|x_{i,j} - x| = \mathrm{O}(h_n)$ contribute to the summation in (4.5). Thus, as $h_n \to 0$ so that $x_{i,j} \to x$ and $\tilde{\mu}(x_{i,j}) \approx \tilde{\mu}(x)$, the two estimators in (4.3) and (4.5) are expected to be asymptotically close. Our use of (4.3) has some technical and computational advantages. First, the estimation error $\tilde{\mu}(x_{i,j}) - \mu(x_{i,j})$ varies with (i, j), and thus it is technically more challenging to study (4.5). Second, to implement (4.5), we need to compute $\tilde{\mu}(\cdot)$ at each point $x_{i,j}$, which requires solving a large number of optimization problems in (4.1) for a large data set. By contrast, (4.3) only requires estimation of $\tilde{\mu}(\cdot)$ at those grid points x at which we wish to estimate $s(\cdot)$.

To study asymptotic properties, we need to introduce some regularity conditions. Throughout we write $\mathcal{S}_\epsilon([a, b]) = [a + \epsilon, b - \epsilon]$ for an arbitrarily fixed small $\epsilon > 0$. Denote by F_e and $f_e = F_e'$ the distribution and density functions of e_0 in (2.1), respectively. The assumption $\mathcal{Q}(e_0) = 0$ and $\mathcal{Q}(|e_0|) = 1$ implies $F_e(0) = 1/2$ and $F_e(1) - F_e(-1) = 1/2$.

Condition 4.1. *Suppose that all measurement locations $x_{i,j}$ are within an interval $[a, b]$, and order them as $a = \tilde{x}_0 < \tilde{x}_1 < \cdots < \tilde{x}_{N_n} < \tilde{x}_{N_n+1} = b$. Assume that*

$$\max_{0 \le k \le N_n} \left| \tilde{x}_{k+1} - \tilde{x}_k - \frac{b-a}{N_n} \right| = \mathrm{O}(N_n^{-2}), \qquad \text{where } N_n = m_1 + \cdots + m_n. \tag{4.6}$$

Condition 4.1 requires that the pooled covariates $x_{i,j}$ should be approximately uniformly dense on $[a, b]$, which is a natural condition since otherwise it would be impossible to draw inferences for regions with very scarce observations. Pooling all subjects together is an appealing procedure to ensure this uniform denseness even though each single subject may only contain sparse measurements.

In nonparametric regression problems, there are two typical settings on the design points: fixed-design and random-design points. For fixed-design case, it is often assumed that the design points are equally spaced on some interval. For example, for the vertical density profile data of Walker and Wright [27], the density was measured at equispaced points along a designated vertical line of wood boards. Condition 4.1 can be viewed as a generalization of the fixed-design points to allow for approximately fixed-design points. For random-design case, the design points are sampled from a distribution. For example, assumption (a) in Appendix A of

Fan and Zhang [10] imposed the random-design condition. In practice, both settings have different range of applicability. For example, for daily or monthly temperature series, the fixed-design setting may be appropriate; for children's growth curve studies, it may be more reasonable to use the random-design setting since the measurements are usually taken at irregular time points.

Remark 4.2 (Asymptotic results under the random-design case). All our subsequent theoretical results are derived under the approximate fixed-design setting in Condition 4.1, but the same argument also applies to the random-design case. Specifically, assume that the design-points $\{x_{i,j}\}$ are random samples from a density $f_X(\cdot)$ with support $[a,b]$ and that x is an interior point. Then, for the design-adaptive local linear median quantile regression estimates, the subsequent Theorems 4.1–4.2 and Corollaries 4.1–4.2 still hold with $(b-a)$ therein replaced by $1/f_X(x)$. In fact, given the i.i.d. structure of $\{x_{i,j}\}$, the technical argument becomes much easier. For example, to establish Lemma 7.1 (again, with $(b-a)$ therein replaced by $1/f_X(x)$), elementary calculations can easily find the mean and variance for the right-hand side of (7.11). All other proofs can be similarly modified and we omit the details.

Conditions 4.2–4.3 below are standard assumptions in nonparametric estimation.

Condition 4.2. *K is symmetric and has bounded support and bounded derivative. Write*

$$\varphi_K = \int_{\mathbb{R}} K^2(u)\,\mathrm{d}u \quad \textit{and} \quad \psi_K = \frac{1}{2}\int_{\mathbb{R}} u^2 K(u)\,\mathrm{d}u.$$

Condition 4.3. $\mu, s \in \mathcal{C}^4([a,b])$, $\inf_{x\in[a,b]} s(x) > 0$, $f_e \in \mathcal{C}^4(\mathbb{R})$, $f_e(0) > 0$, $f_e(1) + f_e(-1) > 0$.

4.1. Uniform Bahadur representation for $\hat{\mu}(x)$

Theorem 4.1 below provides an asymptotic uniform Bahadur representation for $\hat{\mu}(x)$ in (4.1), and its proof in Section 7.4 relies on the arguments and results in Section 3.

Theorem 4.1. *Let $\hat{\mu}(x)$ be as in (4.1). Assume that Conditions 2.1 and 4.1–4.3 hold. Further assume $b_n \to 0$ and $(\log N_n)^3/(N_n b_n) \to 0$. Then*

(i) *We have the uniform consistency*:

$$\sup_{x\in\mathcal{S}_\epsilon([a,b])} \bigl|\hat{\mu}(x) - \mu(x)\bigr| = \mathrm{O}_\mathrm{p}\left\{b_n^2 + \frac{(\log N_n)^{3/2}}{(N_n b_n)^{1/2}}\right\}. \tag{4.7}$$

(ii) *Moreover, the Bahadur representation*

$$\hat{\mu}(x) - \mu(x) = \psi_K \rho_\mu(x) b_n^2 + \frac{(b-a)s(x)}{f_e(0)} \frac{Q_{b_n}(x)}{N_n b_n} + \mathrm{O}_\mathrm{p}(r_n) \tag{4.8}$$

holds uniformly over $x \in \mathcal{S}_\epsilon([a,b])$, *where*

$$\rho_\mu(x) = \mu''(x) - \left[\frac{\mu'(x) f_e'(0)}{f_e(0)} + 2s'(x)\right]\frac{\mu'(x)}{s(x)},$$

$$Q_{b_n}(x) = -\sum_{i=1}^{n}\sum_{j=1}^{m_i}\bigl\{\mathbf{1}_{Y_{i,j}\le\mu(x)} - \mathbb{E}[\mathbf{1}_{Y_{i,j}\le\mu(x)}]\bigr\}K_{b_n}(x_{i,j}-x),$$

$$r_n = b_n^4 + \frac{b_n^{1/2}(\log N_n)^{3/2}}{N_n^{1/2}} + \frac{(\log N_n)^{9/4}}{(N_n b_n)^{3/4}}.$$

In the Bahadur representation (4.8), $\psi_K \rho_\mu(x) b_n^2$ is the bias term, $Q_{b_n}(x)$ determines the asymptotic distribution of $\hat{\mu}(x) - \mu(x)$, and r_n is the negligible error term. Such a Bahadur representation provides a powerful tool in studying the asymptotic behavior of $\hat{\mu}(x)$. Based on Theorem 4.1, we obtain a Central Limit theorem (CLT) for $\hat{\mu}$ in Corollary 4.1. Clearly, the variance of $Q_{b_n}(x)$ is a linear combination of $K_{b_n}(x_{i,j_1}-x)K_{b_n}(x_{i,j_2}-x)$. The following regularity condition is needed to ensure the negligibility of the cross-term $K_{b_n}(x_{i,j_1}-x)K_{b_n}(x_{i,j_2}-x)$ for $j_1 \ne j_2$.

Condition 4.4. *Assume that, for all given* $x \in \mathcal{S}_\epsilon([a,b])$ *and* $k_n = \mathrm{O}(\log N_n)$, *there exits* ι_n *such that* $k_n\iota_n \to 0$ *and*

$$\sum_{(i,j_1,j_2)\in\mathcal{I}} K_{b_n}(x_{i,j_1}-x)K_{b_n}(x_{i,j_2}-x) = \mathrm{O}\bigl[\min(h, M_n) n b_n k_n \iota_n\bigr], \qquad M_n = \max_{1\le i\le n} m_i \tag{4.9}$$

for all $h \ge (k_n \vee a)$, *where* $\mathcal{I} = \{(i,j_1,j_2)\colon 1 \le i \le n, a \le j_1 < j_2 \le \min(a+h-1, m_i), |j_1 - j_2| \le k_n\}$. *Further assume that* $\max_j \sum_{i=1}^n K_{b_n}^r(x_{i,j}-x) = \mathrm{O}(nb_n)$, $r = 2, 4$.

Condition 4.4 is very mild. Intuitively, we consider $x_{i,j}$, $j \in \mathbb{Z}$, being random locations, then $\mathbb{E}[K_{b_n}(x_{i,j_1}-x)K_{b_n}(x_{i,j_2}-x)] = \mathrm{O}(b_n^2)$ for $j_1 \ne j_2$. Thus, under the mild condition $b_n \log N_n \to 0$, (4.9) holds with $\iota_n = b_n$.

Corollary 4.1. *Let the conditions in Theorem* 4.1 *be fulfilled and Condition* 4.4 *hold. Further assume that* $(\log N_n)^9/(N_n b_n) + N_n b_n^9 \to 0$ *and* $nM_n = \mathrm{O}(N_n)$, $nb_n \to \infty$, $\log N_n = \mathrm{O}(\sqrt{M_n})$, *where* M_n *is defined as in* (4.9). *Then, for any* $x \in \mathcal{S}_\epsilon([a,b])$, *we have*

$$(N_n b_n)^{1/2}\bigl[\hat{\mu}(x) - \mu(x) - \psi_K \rho_\mu(x) b_n^2\bigr] \Rightarrow N\left(0, \frac{\varphi_K (b-a) s^2(x)}{4 f_e^2(0)}\right). \tag{4.10}$$

The proof of Corollary 4.1, given in Section 7.5, uses the coupling argument in Section 3. The condition $nM_n = \mathrm{O}(N_n)$ is in line with the classical CLT Lindeberg condition that none of the subjects dominates the others. If b_n is of the order $N_n^{-\beta}$, then the bandwidth condition in Corollary 4.1 holds if $\beta \in (1/9, 1)$. By Corollary 4.1, the optimal bandwidth minimizing the asymptotic mean squared error is

$$b_n = \left[\frac{\varphi_K (b-a) s^2(x)}{4\psi_K^2 \rho_\mu^2(x) f_e^2(0)}\right]^{1/5} N_n^{-1/5}. \tag{4.11}$$

For this optimal bandwidth, the bias term is of order $\mathrm{O}(N_n^{-2/5})$ and contains the derivatives s', μ', μ'' and f_e' that can be difficult to estimate. Based on the Bahadur representation (4.8), we can correct the bias term $\psi_K \rho_\mu(x) b_n^2$ via the jackknife estimator $\tilde{\mu}(x)$ in (4.2). Then the bias term for $\tilde{\mu}(x)$ becomes $2\psi_K \rho_\mu(x) b_n^2 - \psi_K \rho_\mu(x)(\sqrt{2} b_n)^2 = 0$. By (4.8), following the proof of Corollary 4.1, we have

$$(N_n b_n)^{1/2}\big[\tilde{\mu}(x) - \mu(x)\big] \Rightarrow N\left(0, \frac{\varphi_{K^*}(b-a)s^2(x)}{4 f_e^2(0)}\right), \tag{4.12}$$

where $K^*(u) = 2K(u) - 2^{-1/2} K(u/\sqrt{2})$.

4.2. Uniform Bahadur representation for $\hat{s}(x)$

Theorem 4.2 below provides a uniform Bahadur representation for $\hat{s}(x)$ in (4.3).

Theorem 4.2. *Let $\hat{s}(x)$ be as in* (4.3). *Assume that the conditions in Theorem* 4.1 *hold. Further assume $h_n + (\log N_n)^3/(N_n h_n) \to 0$. Then*

(i) *We have the uniform consistency*:

$$\sup_{x \in \mathcal{S}_\epsilon([a,b])} \big|\hat{s}(x) - s(x)\big| = \mathrm{O}_\mathrm{p}\left\{b_n^2 + h_n^2 + \frac{(\log N_n)^{3/2}}{(N_n b_n)^{1/2}} + \frac{(\log N_n)^{3/2}}{(N_n h_n)^{1/2}}\right\}. \tag{4.13}$$

(ii) *Moreover, the Bahadur representation*

$$\hat{s}(x) - s(x) = \psi_K \rho_s(x) h_n^2 + (b-a)s(x)\left[\frac{W_{h_n}(x)}{N_n h_n \kappa_+} - \frac{\kappa T_{b_n}(x)}{N_n b_n f_e(0)}\right] + \mathrm{O}_\mathrm{p}(\tilde{r}_n), \tag{4.14}$$

holds uniformly over $x \in \mathcal{S}_\epsilon([a,b])$, where $\kappa_+ = f_e(-1) + f_e(1), \kappa = [f_e(1) - f_e(-1)]/\kappa_+$, $Q_{b_n}(x)$ is defined as in Theorem 4.1,

$$T_{b_n}(x) = 2Q_{b_n}(x) - 2^{-1/2} Q_{\sqrt{2}b_n}(x),$$

$$\rho_s(x) = s''(x) - \frac{2s'(x)^2}{s(x)} + \kappa\left[\mu''(x) - \frac{2\mu'(x)s'(x)}{s(x)}\right] - \frac{f_e'(1)[s'(x)+\mu'(x)]^2 - f_e'(-1)[s'(x)-\mu'(x)]^2}{\kappa_+ s(x)},$$

$$W_{h_n}(x) = -\sum_{i=1}^{n}\sum_{j=1}^{m_i}\big\{\mathbf{1}_{|Y_{i,j}-\mu(x)| \le s(x)} - \mathbb{E}[\mathbf{1}_{|Y_{i,j}-\mu(x)| \le s(x)}]\big\} K_{h_n}(x_{i,j} - x),$$

$$\tilde{r}_n = b_n^4 + h_n^4 + \frac{h_n^{1/2}(\log N_n)^{3/2}}{N_n^{1/2}} + \frac{(\log N_n)^{9/4}}{(N_n h_n)^{3/4}} + \frac{(\log N_n)^{9/4}}{N_n^{3/4} b_n^{1/4} h_n^{1/2}} + \frac{b_n(\log N_n)^{3/2}}{(N_n h_n)^{1/2}}.$$

As in Corollary 4.1, we can use the Bahadur representation (4.14) to obtain a CLT for $\hat{s}(x) - s(x)$. However, the convergence rate depends on the ratio h_n/b_n. If $h_n/b_n \to \infty$, then the term $T_{b_n}(x)/(N_n b_n)$ dominates and we have $(N_n b_n)^{1/2}$-convergence; if $h_n/b_n \to 0$, then the term $W_{h_n}(x)/(N_n h_n)$ dominates and we have $(N_n h_n)^{1/2}$-convergence; if $h_n/b_n \to c$ for a constant $c \in (0, \infty)$, then both terms contribute.

Corollary 4.2. *Let the conditions in Theorem 4.2 be fulfilled and Condition 4.4 and its counterpart version with b_n being replaced by h_n hold. Further assume that*

$$N_n(b_n \vee h_n)^9 + \frac{(\log N_n)^9}{N_n(b_n \wedge h_n)} \to 0,$$

and $nM_n = \mathrm{O}(N_n)$, $n(b_n \wedge h_n) \to \infty$, $\log N_n = \mathrm{O}(\sqrt{M_n})$, where M_n is defined as in (4.9). Recall $K^*(u) = 2K(u) - 2^{-1/2}K(u/\sqrt{2})$ in (4.12) and κ, κ_+ in Theorem 4.2. Let $x \in \mathcal{S}_\epsilon([a, b])$ be a fixed point. Suppose $h_n/b_n \to c$.

(i) *If $\kappa \neq 0$ and $c = \infty$, then*

$$(N_n b_n)^{1/2}\big[\hat{s}(x) - s(x) - \psi_K \rho_s(x) h_n^2\big] \Rightarrow N\left(0, \frac{\varphi_{K^*}\kappa^2(b-a)s^2(x)}{4 f_e^2(0)}\right).$$

(ii) *If $\kappa \neq 0$ and $c \in [0, \infty)$, then*

$$(N_n h_n)^{1/2}\big[\hat{s}(x) - s(x) - \psi_K \rho_s(x) h_n^2\big] \Rightarrow N\big(0, \sigma_c^2\big), \tag{4.15}$$

where

$$\sigma_c^2 = \frac{(b-a)s^2(x)}{4}\left\{\frac{\varphi_K}{\kappa_+^2} + \frac{c^2\kappa^2\varphi_{K^*}}{f_e^2(0)} - \frac{2c\kappa[1 - 4F_e(-1)]}{\kappa_+ f_e(0)} \int_{\mathbb{R}} K(u)K^*(cu)\,\mathrm{d}u\right\}.$$

(iii) *If $\kappa = 0$, then for all $c \in [0, \infty]$, (4.15) holds with $\sigma_c^2 = \varphi_K(b-a)s^2(x)/(4\kappa_+^2)$.*

One can similarly establish CLT results for $\tilde{s}(x)$ in (4.4). We omit the details.

5. An illustration using real data

5.1. Bandwidth selection

For least-squares based estimation of longitudinal data, Rice and Silverman [21] suggested the subject-based cross-validation method. The basic idea is to use all but one subject to do model fitting, validate the fitted model using the left-out subject, and finally choose the optimal bandwidth by minimizing the overall prediction error:

$$b_{\mathrm{LS}}^* = \operatorname*{argmin}_b \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{Y_{i,j} - \tilde{\mu}^{(-i)}(x_{i,j})\}^2, \tag{5.1}$$

where $\tilde{\mu}^{(-i)}(x)$ represents the estimator of $\mu(x)$ based on data from all but ith subject. As in Wei, Zhao and Lin [30], we replace the square loss by absolute deviation:

$$b_{\mathrm{LAD}}^* = \operatorname*{argmin}_b \sum_{i=1}^{n} \sum_{j=1}^{m_i} |Y_{i,j} - \tilde{\mu}^{(-i)}(x_{i,j})|. \tag{5.2}$$

5.2. An illustration using progesterone data

Urinary metabolite progesterone levels are measured daily, around the ovulation day, over 22 conceptive and 69 nonconceptive women's menstrual cycles so that each curve has about 24 design points; see the left panel of Figure 1 for a plot of the trajectories of the 22 conceptive women. Previous studies based on least-squares (LS) methods include Brumback and Rice [5], Fan and Zhang [10], and Wu and Zhang [31]. Here we reanalyze the conceptive group using our least-absolute-deviation (LAD) estimates.

From the left plot in Figure 1, subject 14 (dashed curve) has two sharp drops in progesterone levels at days -3 and 9. Similarly, subject 13 (dotted curve) has unusually low levels on days $-1, 0, 1$. While such sharp drops or "outliers" may be caused by incorrect measurements or other unknown reasons, we investigate the impact of such "outliers" on the LS and LAD estimates. In the right plot of Figure 1, the thick solid and thin solid curves are the LAD and LS estimates of $\mu(\cdot)$. The two estimates are reasonably close except during the periods $[-4, 1]$ and $[8, 15]$. Notice that the latter periods contain the "outliers" from subjects 13, 14.

To understand the impact of such possible "outliers", we consider two scenarios of perturbing the data below.

(i) Scenario I: remove subjects 13 and 14 and estimate $\mu(\cdot)$ using the remaining subjects. The thick dotted and thin dotted curves are the corresponding LAD and LS estimates. Clearly, the discrepancy is largely diminished.

(ii) Scenario II: make the two outlier subjects 13 and 14 even more extreme by shifting their curves three units down. We see that the discrepancy between the LAD (thick dashed) and LS (thin dashed) estimates becomes even more remarkable.

Compared with the estimate based on the original data, the LS estimates under the two perturbation scenarios differ significantly. By contrast, the LAD estimates under the three cases are similar, indicating the robustness in the presence of outliers. We conclude that, for the progesterone data with several possible outliers, the proposed LAD estimate offers an attractive alternative over the well-studied LS estimates. In practice, we recommend the LAD estimate if the data has suspicious, unusual observations or extreme outliers.

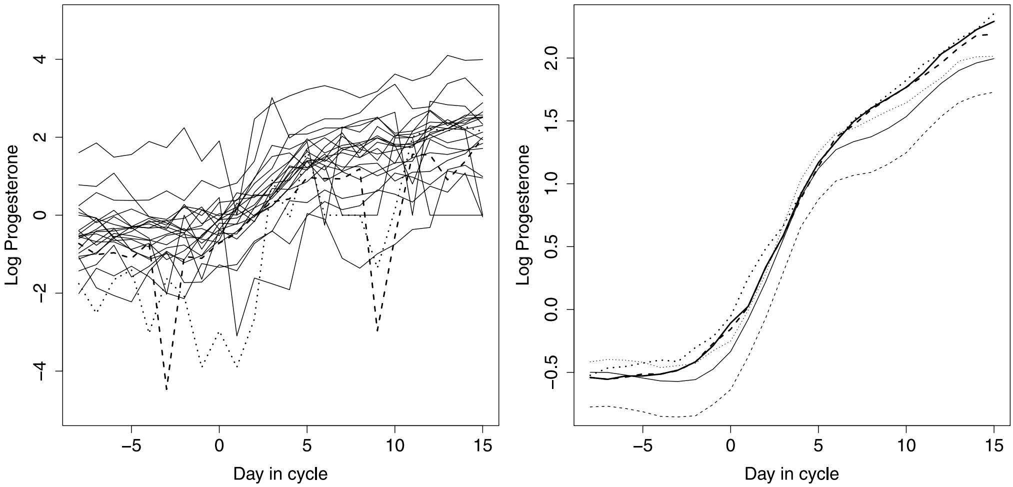

Figure 1. Left: Trajectories of the measurements from 22 conceptive women. Right: Estimates of $\mu(\cdot)$ using both the original data and perturbed data. Thin solid, dotted, and dashed curves are the least-squares estimates of $\mu(\cdot)$ based on the original data, perturbation scenario I (remove subjects 13 and 14), and perturbation scenario II (shift subjects 13 and 14 down by three units), respectively. Similarly, thick solid, dotted, and dashed curves are least-absolute-deviation estimates.

6. Conclusion and extension to spatial setting

This paper studies robust estimations of the location and scale functions in a nonparametric regression model with serially dependent data from multiple subjects. Under a general error dependence structure that allows for many linear and nonlinear processes, we study uniform Bahadur representations and asymptotic normality for least-absolute-deviation estimations of a location-scale longitudinal model. In the large literature on nonparametric estimation of longitudinal models, most existing works use least-squares based methods, which are sensitive to extreme observations and may perform poorly in such circumstances. Despite the popularity of quantile regression methods in linear models and nonparametric regression models, little research has been done in quantile regression based estimations for nonparametric longitudinal models, partly due to difficulties in dealing with the dependence. Therefore, our work provides a solid theoretical foundation for quantile regression estimations in longitudinal models.

The study of asymptotic Bahadur representations is a difficult area and has focused mainly on the i.i.d. setting or stationary time series setting. For longitudinal data, deriving Bahadur representations is more challenging due to the nonstationarity and dependence. To obtain our Bahadur representations, we develop substantial theory for kernel weighted empirical processes via a coupling argument.

The proposed error dependence structure and coupling argument provide a flexible and powerful framework for asymptotics from dependent data, such as time series data, longitudinal data and spatial data, whereas similar problems have been previously studied mainly for either independent data or stationary time series. In (2.1), e_j depends on the innovations or shocks

$\varepsilon_j, \varepsilon_{j\pm 1}, \dots,$ indexed by integers on a line. A natural extension is the function of innovations indexed by bivariate integers on a square:

$$e_j = G(\varepsilon_{j,j}, \varepsilon_{j,j\pm 1}, \varepsilon_{j\pm 1,j}, \varepsilon_{j\pm 1,j\pm 1}, \dots), \qquad j \in \mathbb{Z}.$$

The coupling argument still holds by replacing the innovations $\varepsilon_{j\pm r, j\pm s}$, $r, s \geq k+1$, outside the k nearest squares with i.i.d. copies. As in Condition 2.1, we can assume that the impact of perturbing the distant innovations decays exponentially fast (or polynomially fast with slight modifications of the proof). More generally, the coupling argument holds for function of innovations indexed by multivariate spatial lattice, and such setting may be useful in studying asymptotics for spatial data.

7. Technical proofs

Throughout $c, c_1, c_2, \dots,$ are generic constants. First, we give an inequality for the indicator function. Let Z, Z' be two random variables and $y \in \mathbb{R}$. For $\alpha > 0$, we have

$$\mathbf{1}_{Z \leq y < Z'} = \mathbf{1}_{Z \leq y < Z', |Z-Z'| \geq \alpha} + \mathbf{1}_{Z \leq y < Z', |Z-Z'| < \alpha} \leq \mathbf{1}_{|Z-Z'| \geq \alpha} + \mathbf{1}_{y < Z' < y+\alpha},$$

similarly, $\mathbf{1}_{Z' \leq y < Z} \leq \mathbf{1}_{|Z-Z'| \geq \alpha} + \mathbf{1}_{y-\alpha < Z' \leq y}$. Therefore,

$$|\mathbf{1}_{Z \leq y} - \mathbf{1}_{Z' \leq y}| = \mathbf{1}_{Z \leq y < Z'} + \mathbf{1}_{Z' \leq y < Z} \leq 2\mathbf{1}_{|Z-Z'| \geq \alpha} + \mathbf{1}_{y-\alpha < Z' < y+\alpha}. \tag{7.1}$$

7.1. Proof of Propositions 2.1–2.2

Proof of Proposition 2.1. Let $q^* = q/(\varsigma + \upsilon)$, $p_1 = \upsilon/\varsigma + 1$, and $p_2 = \varsigma/\upsilon + 1$ so that $\varsigma q^* p_1 = q$, $\upsilon q^* p_2 = q$, and $1/p_1 + 1/p_2 = 1$. For convenience, write $e_0' = e_0(k)$. By assumption, $\|e_0 - e_0'\|_q = \mathrm{O}(\rho^k)$. By (2.3) and the Hölder inequality $\mathbb{E}|Z_1 Z_2| \leq \|Z_1\|_{p_1} \|Z_2\|_{p_2}$,

$$\begin{aligned}
\|h(e_0') - h(e_0)\|_{q^*}^{q^*} &\leq \mathrm{O}(1)\mathbb{E}\bigl[|e_0' - e_0|^{\varsigma q^*}\bigl(1 + |e_0| + |e_0'|\bigr)^{\upsilon q^*}\bigr] \\
&\leq \mathrm{O}(1)\bigl\{\mathbb{E}\bigl[|e_0 - e_0'|^{\varsigma q^* \cdot p_1}\bigr]\bigr\}^{1/p_1}\bigl\{\mathbb{E}\bigl[\bigl(1 + |e_0| + |e_0'|\bigr)^{\upsilon q^* \cdot p_2}\bigr]\bigr\}^{1/p_2} \\
&= \mathrm{O}(1)\|e_0 - e_0'\|_q^{q/p_1}\|e_0\|_q^{q/p_2} = \mathrm{O}\bigl(\rho^{kq/p_1}\bigr).
\end{aligned}$$

The above expression gives $\|h(e_0') - h(e_0)\|_{q^*} \leq \mathrm{O}(1)[\rho^{q/(p_1 q^*)}]^k = \mathrm{O}(\rho^{k\varsigma})$. □

Proof of Proposition 2.2. Let $\alpha = \rho^{kq/(1+q)}$. By (7.1) and the triangle inequality,

$$\begin{aligned}
\|\mathbf{1}_{e_0 \leq x} - \mathbf{1}_{e_0(k) \leq x}\|_q &\leq 2\|\mathbf{1}_{|e_0 - e_0(k)| \geq \alpha}\|_q + \|\mathbf{1}_{x-\alpha \leq e_0 \leq x+\alpha}\|_q \\
&= 2\bigl[\mathbb{P}\bigl\{|e_0 - e_0(k)| \geq \alpha\bigr\}\bigr]^{1/q} + \bigl[\mathbb{P}\{x - \alpha \leq e_0 \leq x + \alpha\}\bigr]^{1/q}.
\end{aligned}$$

By the Markov inequality, $\mathbb{P}\{|e_0 - e_0(k)| \geq \alpha\} \leq \mathbb{E}[|e_0 - e_0(k)|^q]/\alpha^q = \mathrm{O}(\rho^{kq}/\alpha^q)$. Since e_0 has a bounded density, $\mathbb{P}\{x - \alpha \leq e_0 \leq x + \alpha\} = \mathrm{O}(\alpha)$. The result then follows. □

7.2. Proof of Theorems 3.1–3.3

Proof of Theorem 3.1. For $r = 1, 2, \ldots, 2k_n + 1$, let

$$\mathcal{I}_r = \big\{(i, j)\colon 1 \le i \le n, 1 \le j \le \big\lfloor (m_i - r)/(2k_n + 1)\big\rfloor + 1\big\}. \tag{7.2}$$

Using the identity $\sum_{j=1}^m a_j = \sum_{r=1}^k \sum_{j=1}^{\lfloor (m-r)/k\rfloor+1} a_{(j-1)k+r}$ for all $k, m \in \mathbb{N}, a_1, \ldots, a_m \in \mathbb{R}$, we can rewrite $H_n(v)$ as

$$H_n(v) = \sum_{r=1}^{2k_n+1} \sum_{(i,j)\in\mathcal{I}_r} g_{i,(j-1)(2k_n+1)+r}(\tilde{Y}_{i,(j-1)(2k_n+1)+r}, v) := \sum_{r=1}^{2k_n+1} H_n(v, r). \tag{7.3}$$

Now we consider $H_n(v, r)$. By the discussion in Section 3, the summands in $H_n(v, r)$ are independent. By (3.3),

$$\begin{aligned} \operatorname{Var}\big[H_n(v, r)\big] &= \sum_{(i,j)\in\mathcal{I}_r} \mathbb{E}\big[g^2_{i,(j-1)(2k_n+1)+r}(\tilde{Y}_{i,(j-1)(2k_n+1)+r}, v)\big] \\ &\le \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbb{E}\big[g^2_{i,j}(\tilde{Y}_{i,j}, v)\big] \le \chi_n, \end{aligned} \tag{7.4}$$

uniformly over v, r.

(i) Consider the case $\chi_n = \mathrm{O}(1)$. Recall the condition $|g_{i,j}(y, v)| \le c$. By Berstein's exponential inequality (Bennett [3]) for bounded and independent random variables, for any given $c_1 > 0$, when N_n is sufficiently large,

$$\mathbb{P}\big\{\big|H_n(v, r)\big| \ge c_1 \log N_n\big\} \le 2\exp\left\{-\frac{(c_1 \log N_n)^2}{2\operatorname{Var}[\Lambda_n(r, h)] + cc_1 \log N_n}\right\} \le 2N_n^{-c_1/(3c)}, \tag{7.5}$$

uniformly over r and h. Here the second inequality follows from $\operatorname{Var}[H_n(v, r)] \le \chi_n = \mathrm{O}(1) \le cc_1 \log N_n$ for large enough N_n. Thus,

$$\begin{aligned} \mathbb{P}\Big\{\max_{v\in\mathcal{V}_n, 1\le r\le 2k_n+1} \big|H_n(v, r)\big| \ge c_1 \log N_n\Big\} &\le \sum_{v\in\mathcal{V}_n, 1\le r\le 2k_n+1} \mathbb{P}\big\{\big|H_n(v, r)\big| \ge c_1 \log N_n\big\} \\ &\le 2|\mathcal{V}_n| k_n N_n^{-c_1/(3c)}. \end{aligned}$$

By the assumption that both $|\mathcal{V}_n|$ and k_n grow no faster than a polynomial of N_n, we can make the above probability go to zero by choosing a large enough c_1. Therefore, $\max_{v\in\mathcal{V}_n, 1\le r\le 2k_n+1} |H_n(v, r)| = \mathrm{O}_\mathrm{p}(\log N_n)$. By (7.3), the desired result follows from

$$\max_{v\in\mathcal{V}_n} \big|H_n(v)\big| \le (2k_n + 1) \max_{v\in\mathcal{V}_n, 1\le r\le 2k_n+1} \big|H_n(v, r)\big|.$$

(ii) Consider the case $\sup_n \log N_n/\chi_n < \infty$. As in (7.5),

$$\mathbb{P}\{|H_n(v,r)| \geq c_1\sqrt{\chi_n \log N_n}\} \leq 2\exp\left\{-\frac{(c_1\sqrt{\chi_n \log N_n})^2}{2\chi_n + cc_1\sqrt{\chi_n \log N_n}}\right\} = \mathrm{O}\big[N_n^{-c_1^2/(2+cc_1c_2)}\big],$$

uniformly over r and h, where $c_2 = \sup_n[\log N_n/\chi_n]^{1/2} < \infty$. The rest of the proof follows from the same argument as in the case (i) by choosing a sufficiently large c_1. □

Proof of Theorem 3.2. Let $\alpha = 1/N_n$. Since $\varpi_{i,j}(x) \leq c$, applying (7.1), we obtain

$$\begin{aligned}
\big|F_n(x,y) - \tilde{F}_n(x,y)\big| &\leq \sum_{i=1}^{n}\sum_{j=1}^{m_i} \varpi_{i,j}(x)|\mathbf{1}_{Y_{i,j}\leq y} - \mathbf{1}_{\tilde{Y}_{i,j}\leq y}| \\
&\leq 2c\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{1}_{|Y_{i,j}-\tilde{Y}_{i,j}|\geq\alpha} + \sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{1}_{y-\alpha<\tilde{Y}_{i,j}<y+\alpha}\right] \\
&:= 2c\big[\Omega_n + \Lambda_n(y)\big].
\end{aligned} \tag{7.6}$$

Notice that, $|Y_{i,j} - \tilde{Y}_{i,j}| = \mathrm{O}(1)|e_{i,j} - e_{i,j}(k_n)|$. By (2.2) and the Markov inequality,

$$\mathbb{E}(\mathbf{1}_{|Y_{i,j}-\tilde{Y}_{i,j}|\geq\alpha}) \leq \frac{\|Y_{i,j} - \tilde{Y}_{i,j}\|_q^q}{\alpha^q} = \mathrm{O}(1)\frac{\|e_{i,j} - e_{i,j}(k_n)\|_q^q}{\alpha^q} = \mathrm{O}\big(N_n^q\rho^{qk_n}\big).$$

Thus, $\Omega_n = \mathrm{O}_{\mathrm{p}}(N_n^{1+q}\rho^{qk_n}) = \mathrm{O}_{\mathrm{p}}[N_n^{1+q}\rho^{q\lambda\log(N_n)}] = \mathrm{o}_{\mathrm{p}}(1)$ for $\lambda > (q+1)/[q\log(1/\rho)]$.

For $\Lambda_n(y)$ over $y \in \mathbb{R}$, consider two cases: $|y| > N_n^{1/q}$ and $|y| \leq N_n^{1/q}$. For $|y| > N_n^{1/q}$, since $\alpha = 1/N_n \to 0$, $\mu(x_{i,j})$ and $s(x_{i,j})$ are bounded, $\{y-\alpha < \tilde{Y}_{i,j} < y+\alpha\} \subset \{|e_{i,j}(k_n)| \geq c_1N_n^{1/q}\}$ for some constant $c_1 > 0$. Therefore, by $e_{i,j}(k_n) \in \mathcal{L}^q$ and the Markov inequality,

$$\begin{aligned}
\mathbb{E}\Big[\sup_{|y|>N_n^{1/q}}\Lambda_n(y)\Big] &\leq \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{1}_{|e_{i,j}(k_n)|>c_1N_n^{1/q}}\right] \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\|e_{i,j}(k_n)\|_q^q}{(c_1N_n^{1/q})^q} = \mathrm{O}(1).
\end{aligned} \tag{7.7}$$

We conclude that $\sup_{|y|>N_n^{1/q}}\Lambda_n(y) = \mathrm{O}_{\mathrm{p}}(1)$.

In what follows, we use a chain argument to prove $\sup_{y\in[-N_n^{1/q},N_n^{1/q}]}\Lambda_n(y) = \mathrm{O}_{\mathrm{p}}[(\log n)^2]$. Without loss of generality, consider $y \in [0, N_n^{1/q}]$. Write $\ell_n = \lfloor N_n^{1+1/q}\rfloor$ and let $\mathcal{V}_n = \{y_v = vN_n^{1/q}/\ell_n, v = 0,1,\ldots,\ell_n\}$ be the set of $\ell_n + 1$ grid points uniformly spaced over $[0, N_n^{1/q}]$. Partition $[0, N_n^{1/q}]$ into intervals $I_v = [y_{v-1}, y_v], v = 1,\ldots,\ell_n$. For any $y \in I_v$, we have $\mathbf{1}_{y-\alpha<\tilde{Y}_{i,j}<y+\alpha} \leq \mathbf{1}_{y_{v-1}-\alpha<\tilde{Y}_{i,j}<y_v+\alpha}$. Since $s(x_{i,j})$ is bounded away from zero, $\sup_u f_e(u) < \infty$, and $|y_v - y_{v-1}| = \mathrm{O}(1/N_n)$, we have $\mathbb{E}(\mathbf{1}_{y_{v-1}-\alpha<\tilde{Y}_{i,j}<y_v+\alpha}) \le c_2/N_n$ uniformly for some constant $c_2 < \infty$. Consequently, for any $y \in I_v$, we have

$$\Lambda_n(y) \le \sum_{i=1}^{n}\sum_{j=1}^{m_i}\big[\{\mathbf{1}_{y_{v-1}-\alpha<\tilde{Y}_{i,j}<y_v+\alpha} - \mathbb{E}(\mathbf{1}_{y_{v-1}-\alpha<\tilde{Y}_{i,j}<y_v+\alpha})\} + c_2/N_n\big] = \Lambda_n^*(v) + c_2.$$

We apply Theorem 3.1 to $\Lambda_n^*(v)$. For χ_n in (3.3), using $\mathbb{E}(\mathbf{1}_{y_{v-1}-\alpha<\tilde{Y}_{i,j}<y_v+\alpha}) \le c_2/N_n$, we have $\chi_n = \mathrm{O}(1)$ and thus $\max_{v\in\mathcal{V}_n} |\Lambda_n^*(v)| = \mathrm{O_p}[(\log N_n)^2]$, completing the proof. $\square$

Proof of Theorem 3.3. Recall the coupling process $\tilde{D}_n(\delta, x, y)$ in (3.7). Under the assumption $\sup_n \log N_n/(\delta_n\phi_n) < \infty$, $(\log N_n)^2 = \mathrm{O}\{[\delta_n\phi_n(\log N_n)^3]^{1/2}\}$. Thus, by (3.8), it suffices to show $\sup_{|\delta|\le\delta_n,x,y\in\mathbb{R}} |\tilde{D}_n(\delta, x, y)| = \mathrm{O_p}\{[\delta_n\phi_n(\log N_n)^3]^{1/2}\}$.

Without loss of generality, assume $\delta \in [0, \delta_n]$. Recall $\tilde{Y}_{i,j}$ in (3.5). Rewrite

$$\tilde{D}_n(\delta, x, y) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \varpi_{i,j}(x)\big\{\tilde{\xi}_{i,j}(\delta, y) - \mathbb{E}\big[\tilde{\xi}_{i,j}(\delta, y)\big]\big\}, \qquad \tilde{\xi}_{i,j}(\delta, y) = \mathbf{1}_{y<\tilde{Y}_{i,j}\le y+\delta}.$$

As in the proof of Theorem 3.2, consider $|y| > N_n^{1/q}$ and $|y| \le N_n^{1/q}$.

For $|y| > N_n^{1/q}$, since $\mu(x_{i,j})$ and $s(x_{i,j})$ are bounded and $|\delta| \le \delta_n \to 0$, $\{y < \tilde{Y}_{i,j} \le y + \delta\} \subset \{|e_{i,j}(k_n)| \ge c_1 N_n^{1/q}\}$ for some $c_1 > 0$. Therefore, by the boundedness of $\varpi_{i,j}(\cdot)$, the same argument in (7.7) shows $\tilde{D}_n(\delta, x, y) = \mathrm{O_p}(1)$ uniformly over $x \in \mathbb{R}$, $|y| > N_n^{1/q}$, $|\delta| \le \delta_n$.

Next, we consider $|y| \le N_n^{1/q}$. Since $\varpi_{i,j}(x)$ vanishes for x outside a bounded interval, without loss of generality we only consider $x \in [0, b]$ for some $b > 0$, $y \in [0, N_n^{1/q}]$, and $\delta \in [0, \delta_n]$. As in the proof of Theorem 3.2, we use the chain argument. Let $\ell_n = \lfloor N_n^{1/q}/\delta_n + N_n\tau_n + N_n^{1+1/q}\rfloor$, and

$$\mathcal{V}_n = \left\{(x_{v_1}, y_{v_2}, t_{v_3}), x_{v_1} = \frac{v_1 b}{\ell_n}, y_{v_2} = \frac{v_2 N_n^{1/q}}{\ell_n}, t_{v_3} = \frac{v_3\delta_n}{\ell_n}, v_1, v_2, v_3 = 0, 1, \dots, \ell_n\right\}$$

be uniformly spaced grid points. Partition $[0, b] \times [0, N_n^{1/q}] \times [0, \delta_n]$ into intervals $I_{v_1,v_2,v_3} = [x_{v_1-1}, x_{v_1}] \times [y_{v_2-1}, y_{v_2}] \times [t_{v_3-1}, t_{v_3}]$, $v_1, v_2, v_3 = 1, \dots, \ell_n$. Let

$$\underline{\xi}_{i,j}(v_2, v_3) = \mathbf{1}_{y_{v_2}<\tilde{Y}_{i,j}\le y_{v_2-1}+t_{v_3-1}} \quad\text{and}\quad \overline{\xi}_{i,j}(v_2, v_3) = \mathbf{1}_{y_{v_2-1}<\tilde{Y}_{i,j}\le y_{v_2}+t_{v_3}}.$$

Clearly, for any $(x, y, \delta) \in I_{v_1,v_2,v_3}$, we have $\underline{\xi}_{i,j}(v_2, v_3) \le \tilde{\xi}_{i,j}(\delta, y) \le \overline{\xi}_{i,j}(v_2, v_3)$. Since $N_n \to \infty$ and $\delta_n \to 0$, there exists a constant $c_2 < \infty$ such that $0 \le \mathbb{E}[\overline{\xi}_{i,j}(v_2, v_3)] - \mathbb{E}[\underline{\xi}_{i,j}(v_2, v_3)] \le c_2 N_n^{1/q}/\ell_n$. Additionally, for $x \in [x_{v_1-1}, x_{v_1}]$, by Condition 3.2, $|\varpi_{i,j}(x) - \varpi_{i,j}(x_{v_1})| \le \tau_n|x -$

$x_{v_1}| \leq \tau_n b/\ell_n$. Thus, there exists a constant $c_3 < \infty$ such that

$$\begin{aligned}
&\varpi_{i,j}(x)\{\tilde{\xi}_{i,j}(\delta, y) - \mathbb{E}[\tilde{\xi}_{i,j}(\delta, y)]\} \\
&\quad \leq \varpi_{i,j}(x_{v_1})\{\overline{\xi}_{i,j}(v_2, v_3) - \mathbb{E}[\underline{\xi}_{i,j}(v_2, v_3)]\} + \tau_n b/\ell_n \\
&\quad \leq \varpi_{i,j}(x_{v_1})\{\overline{\xi}_{i,j}(v_2, v_3) - \mathbb{E}[\overline{\xi}_{i,j}(v_2, v_3)]\} + c_3(\tau_n + N_n^{1/q})/\ell_n,
\end{aligned} \tag{7.8}$$

uniformly over i, j, and $(x, y, \delta) \in I_{v_1,v_2,v_3}$. Similarly,

$$\begin{aligned}
&\varpi_{i,j}(x)\{\tilde{\xi}_{i,j}(\delta, y) - \mathbb{E}[\tilde{\xi}_{i,j}(\delta, y)]\} \\
&\quad \geq \varpi_{i,j}(x_{v_1})\{\underline{\xi}_{i,j}(v_2, v_3) - \mathbb{E}[\underline{\xi}_{i,j}(v_2, v_3)]\} - c_3(\tau_n + N_n^{1/q})/\ell_n.
\end{aligned} \tag{7.9}$$

Combining (7.8) and (7.9) and using $N_n(\tau_n + N_n^{1/q})/\ell_n = \mathrm{O}(1)$, we have

$$\sup_{x,y,\delta} |\tilde{D}_n(\delta, x, y)| \leq \max_{v \in \mathcal{V}_n} \{|\underline{\Delta}_n(v)| + |\overline{\Delta}_n(v)|\} + \mathrm{O}(1), \tag{7.10}$$

where $v = (v_1, v_2, v_3)$,

$$\underline{\Delta}_n(v) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \varpi_{i,j}(x_{v_1})\{\underline{\xi}_{i,j}(v_2, v_3) - \mathbb{E}[\underline{\xi}_{i,j}(v_2, v_3)]\},$$

$$\overline{\Delta}_n(v) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \varpi_{i,j}(x_{v_1})\{\overline{\xi}_{i,j}(v_2, v_3) - \mathbb{E}[\overline{\xi}_{i,j}(v_2, v_3)]\}.$$

We now apply Theorem 3.1 to $\underline{\Delta}_n(v)$ and $\overline{\Delta}_n(v)$. For χ_n in (3.3), with ϕ_n in (3.9) and $\mathbb{E}[\overline{\xi}_{i,j}(h_2, h_3)] = \mathrm{O}(\delta_n + N_n^{1/q}/\ell_n) = \mathrm{O}(\delta_n)$, we can take $\chi_n = \mathrm{O}(\delta_n \phi_n)$. By Theorem 3.1(ii), $\max_{v \in \mathcal{V}_n} |\overline{\Delta}_n(v)| = \mathrm{O}_{\mathrm{p}}\{[\delta_n \phi_n (\log N_n)^3]^{1/2}\}$. The latter bound also holds for $\max_{v \in \mathcal{V}_n} |\underline{\Delta}_n(v)|$. The desired result then follows from (7.10). $\square$

7.3. Asymptotic expansions

Throughout the proofs, we use the following notation:

$$\begin{aligned}
L_\mu(\delta_1, x) &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{b_n}(x_{i,j} - x) \mathbf{1}_{Y_{i,j} \leq \mu(x) + \delta_1}, \\
L_\mu(x) &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{b_n}(x_{i,j} - x), \\
J_\mu(\delta_1, x) &= \mathbb{E}[L_\mu(\delta_1, x)],
\end{aligned}$$

$$L_s(\delta_1,\delta_2,x)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_{h_n}(x_{i,j}-x)\mathbf{1}_{|Y_{i,j}-\mu(x)-\delta_1|\le s(x)+\delta_2},$$

$$L_s(x)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_{h_n}(x_{i,j}-x),$$

$$J_s(\delta_1,\delta_2,x)=\mathbb{E}\big[L_s(\delta_1,\delta_2,x)\big].$$

Lemma 7.1. *Assume that Conditions 4.1–4.2 hold. Then, we have*

(i) *Uniformly over $x\in\mathcal{S}_\epsilon[a,b]$,*

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\frac{x_{i,j}-x}{b_n}\right)^r K\left(\frac{x_{i,j}-x}{b_n}\right)=\frac{N_n b_n}{b-a}\int_{\mathbb{R}}u^r K(u)\,\mathrm{d}u+\mathrm{O}(1). \tag{7.11}$$

(ii) *Let $g(x,v)$ be a measurable bivariate function on $[a,b]^2$. Define*

$$\mathcal{G}_g(x)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}g(x,x_{i,j})K_{b_n}(x_{i,j}-x). \tag{7.12}$$

Further assume that $\sup_{x\in[a,b]}|\partial^s(x,v)/\partial v^s|<\infty, s=0,1,\ldots,r$ for some $r\in\mathbb{N}$. Then uniformly over $x\in\mathcal{S}_\epsilon[a,b]$,

$$\mathcal{G}_g(x)=\sum_{s=0}^{r-1}\frac{\partial^s g(x,v)}{\partial v^s}\Big|_{v=x}\frac{N_n b_n^{s+1}}{(b-a)s!}\int_{\mathbb{R}}u^s K(u)\,\mathrm{d}u+\mathrm{O}\big(1+N_n b_n^{r+1}\big). \tag{7.13}$$

Proof. (i) Recall the ordered locations $\tilde{x}_k$ in Condition 4.1. Define

$$S_n(x)=\sum_{k=1}^{N_n}\left(\frac{\tilde{x}_k-x}{b_n}\right)^r K\left(\frac{\tilde{x}_k-x}{b_n}\right), \tag{7.14}$$

$$I_n(x)=\sum_{k=0}^{N_n}(\tilde{x}_{k+1}-\tilde{x}_k)\left(\frac{\tilde{x}_k-x}{b_n}\right)^r K\left(\frac{\tilde{x}_k-x}{b_n}\right), \tag{7.15}$$

$$\varrho_n=\max_{0\le k\le N_n}\big|\tilde{x}_{k+1}-\tilde{x}_k-(b-a)/N_n\big|=\mathrm{O}\big(N_n^{-2}\big), \tag{7.16}$$

$$\mathcal{I}(x)=\big\{1\le k\le N_n\colon \tilde{x}_k-x\in\big[-b_n-(b-a)/N_n-\varrho_n,b_n\big]\big\}. \tag{7.17}$$

Assume without loss of generality that K has support $[-1,1]$. Condition (4.6) implies that $\sup_{x\in[a,b]}|\mathcal{I}(x)|=\mathrm{O}(N_n b_n)$, where and hereafter $|\mathcal{I}|$ is the cardinality of a set $\mathcal{I}$. Because K has support $[-1,1]$, $K_{b_n}(\tilde{x}_k-x)=0$ for $k\notin\mathcal{I}(x)$. Additionally, for $k\in\mathcal{I}(x)$, the summands in

$S_n(x)$ are uniformly bounded. Thus,

$$S_n(x) = \sum_{k \in \mathcal{I}(x)} \left(\frac{\tilde{x}_k - x}{b_n}\right)^r K\left(\frac{\tilde{x}_k - x}{b_n}\right) = \mathrm{O}\big[\big|\mathcal{I}(x)\big|\big] = \mathrm{O}(N_n b_n), \tag{7.18}$$

uniformly over $x \in [a, b]$.

By (4.6), elementary calculation shows that, uniformly over $x \in \mathcal{S}_\epsilon[a, b]$,

$$\begin{aligned}\frac{b-a}{N_n} S_n(x) - I_n(x) &= -\sum_{k=1}^{N_n} \left(\tilde{x}_{k+1} - \tilde{x}_k - \frac{b-a}{N_n}\right)\left(\frac{\tilde{x}_k - x}{b_n}\right)^r K\left(\frac{\tilde{x}_k - x}{b_n}\right) \\ &= \mathrm{O}(\varrho_n) \sup_{x \in [a,b]} \big|S_n(x)\big| = \mathrm{O}(b_n/N_n).\end{aligned} \tag{7.19}$$

Write $u_k = (\tilde{x}_k - x)/b_n$. Observe that $I_n(x) = \sum_{k=0}^{N_n} \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} u_k^r K(u_k)\,\mathrm{d}v$. Thus, by the triangle inequality, we have

$$\begin{aligned}&\left|I_n(x) - \int_{\tilde{x}_0}^{\tilde{x}_{N_n+1}} \left(\frac{v-x}{b_n}\right)^r K\left(\frac{v-x}{b_n}\right)\mathrm{d}v\right| \le \sum_{k=0}^{N_n} V_k, \\ &\text{where } V_k = \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} \left|u_k^r K(u_k) - \left(\frac{v-x}{b_n}\right)^r K\left(\frac{v-x}{b_n}\right)\right|\mathrm{d}v.\end{aligned} \tag{7.20}$$

Since K has bounded derivative, $|y^r K(y) - z^r K(z)| = \mathrm{O}(|y-z|)$ for $y, z \in [-1, 1]$. Also, $|u_k - (v-x)/b_n| = |v - \tilde{x}_k|/b_n$. Thus, under Condition 4.1,

$$|V_k| = \mathrm{O}(1) \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} \frac{v - \tilde{x}_k}{b_n}\,\mathrm{d}v = \frac{\mathrm{O}[(\tilde{x}_{k+1} - \tilde{x}_k)^2]}{b_n} = \frac{\mathrm{O}(1)}{N_n^2 b_n}. \tag{7.21}$$

Furthermore, it is easily seen that, for $k \notin \mathcal{I}(x)$, $\min(|\tilde{x}_k - x|, |\tilde{x}_{k+1} - x|) > b_n$, which implies $K(u_k) = 0$, $K\{(v-x)/b_n\} = 0$ for $v \in [\tilde{x}_k, \tilde{x}_{k+1}]$, and consequently $V_k = 0$. Thus, by (7.20) and (7.21),

$$\left|I_n(x) - \int_{\tilde{x}_0}^{\tilde{x}_{N_n+1}} \left(\frac{v-x}{b_n}\right)^r K\left(\frac{v-x}{b_n}\right)\mathrm{d}v\right| \le \sum_{k \in \mathcal{I}(x)} V_k = \mathrm{O}(1/N_n), \tag{7.22}$$

uniformly over $x \in \mathcal{S}_\epsilon[a, b]$,

Notice that $\sum_{i=1}^n \sum_{j=1}^{m_i} [(x_{i,j} - x)/b_n]^r K_{b_n}(x_{i,j} - x) = S_n(x)$. Recall that $\tilde{x}_0 = a$ and $\tilde{x}_{N_n+1} = b$. The desired result then follows from (7.19) and (7.22) in view of

$$\int_{\tilde{x}_0}^{\tilde{x}_{N_n+1}} \left(\frac{v-x}{b_n}\right)^r K\left(\frac{v-x}{b_n}\right)\mathrm{d}v = b_n \int_{(a-x)/b_n}^{(b-x)/b_n} u^r K(u)\,\mathrm{d}u = b_n \int_{-1}^{1} u^r K(u)\,\mathrm{d}u$$

for all $x \in \mathcal{S}_\epsilon[a, b]$ and large enough n.

(ii) The expression (7.13) easily follows from (i) in view of the Taylor expansion $g(x, x_{i,j}) = \sum_{s=0}^{r-1} \partial^s g(x, v)/\partial v^s|_{v=x}(x_{i,j} - x)^s/s! + \mathrm{O}(b_n^r)$ for $|x_{i,j} - x| \le b_n$. □

Lemma 7.2. *Assume that Conditions 4.1–4.2 hold. Let $\rho_\mu(x), \rho_s(x), \kappa, \kappa_+$ be as in Theorems 4.1–4.2. Then, for $\delta_1 \to 0, \delta_2 \to 0$, we have uniformly over $x \in \mathcal{S}_\epsilon[a, b]$,*

$$J_\mu(0, x) = L_\mu(x)/2 - N_n b_n^3 \rho_\mu(x) f_e(0)\psi_K/\big[(b-a)s(x)\big] + \mathrm{O}\big(1 + N_n b_n^5\big),$$
$$J_\mu(\delta_1, x) = J_\mu(0, x) + N_n b_n \delta_1\big\{f_e(0)/\big[(b-a)s(x)\big] + \mathrm{O}\big[(N_n b_n)^{-1} + b_n^2 + \delta_1\big]\big\},$$
$$J_s(\delta_1, 0, x) = L_s(x)/2 - N_n h_n \kappa_+\big\{\big[h_n^2 \psi_K \rho_s(x) - \delta_1\kappa\big]/\big[(b-a)s(x)\big] + \mathrm{O}\big(h_n^4 + \delta_1^2\big)\big\},$$
$$J_s(\delta_1, \delta_2, x) = J_s(\delta_1, 0, x) + N_n h_n \delta_2\big\{\kappa_+/\big[(b-a)s(x)\big] + \mathrm{O}\big(h_n^2 + \delta_1 + \delta_2\big)\big\}.$$

Proof. Recall that F_e and f_e are the distribution and density functions of $e_{i,j}$. The assumption $\mathcal{Q}(e_{i,j}) = 0$ implies that $F_e(0) = 1/2$. Notice that

$$\begin{aligned} J_\mu(0, x) - L_\mu(x)/2 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} K_{b_n}(x_{i,j} - x)\big[\mathbb{P}\big\{Y_{i,j} \le \mu(x)\big\} - 1/2\big] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} K_{b_n}(x_{i,j} - x) g(x, x_{i,j}), \end{aligned}$$

where $g(x, v) = F_e\{[\mu(x) - \mu(v)]/s(v)\} - F_e(0)$. The symmetry of K entails $\int u^s K(u)\,\mathrm{d}u = 0, s = 1, 3$. The first expression then follows from Lemma 7.1(ii) with $r = 4$.

Similarly, we can show $J'_\mu(0, x) := \partial J_\mu(\delta_1, x)/\partial\delta_1|_{\delta_1=0} = N_n b_n f_e(0)/[(b-a)s(x)] + \mathrm{O}(1 + N_n b_n^3)$ and $J''_\mu(\delta_1, x) := \partial^2 J_\mu(\delta_1, x)/\partial\delta_1^2 = \mathrm{O}(N b_n)$ uniformly over δ_1, x. So, the second expression follows from the Taylor expansion $J_\mu(\delta_1, x) - J_\mu(0, x) = \delta_1 J'_\mu(0, x) + \mathrm{O}(N b_n \delta_1^2)$. The other two expressions can be similarly treated. We omit the details. □

7.4. Proof of Theorems 4.1–4.2

Let $L_\mu(x)$, $L_\mu(\delta_1, x)$, $J_\mu(\delta_1, x)$, $L_s(x)$, $L_s(\delta_1, \delta_2, x)$ and $J_s(\delta_1, \delta_2, x)$ be as in Section 7.3.

Proof of Theorem 4.1. Let $\delta_n = [(\log N_n)^3/(N_n b_n)]^{1/2} + b_n^2 \to 0$. Let $l_n \uparrow \infty$ be a positive sequence satisfying $\delta_n l_n \to 0$. First, we show $\hat\Delta_\mu(x) := \hat\mu(x) - \mu(x) = \mathrm{O_p}(l_n\delta_n)$ uniformly over $x \in \mathcal{S}_\epsilon([a, b])$. Since $\hat\mu(x)$ is a solution to (4.1), by Koenker ([18], pages 32–33),

$$\big|L_\mu\big(\hat\Delta_\mu(x), x\big) - L_\mu(x)/2\big| \le \sum_{i,j} K_{b_n}(x_{i,j} - x)\mathbf{1}_{Y_{i,j}=\hat\mu(x)} = \mathrm{O_p}(1), \tag{7.23}$$

uniformly over x. Let

$$\Omega_n(x) = \big[L_\mu(l_n\delta_n, x) - J_\mu(l_n\delta_n, x)\big] - \big[L_\mu(0, x) - J_\mu(0, x)\big].$$

We can apply Theorem 3.3 with $\varpi_{i,j}(x) = K_{b_n}(x_{i,j} - x)$ to $\Omega_n(x)$. For τ_n and ϕ_n in Condition 3.2, $\tau_n = \mathrm{O}(1/b_n)$ and $\phi_n = \mathrm{O}(N_n b_n)$ (see Lemma 7.1). By Theorem 3.3, $\sup_{x\in[a,b]} |\Omega_n(x)| = \mathrm{O_p}\{[N_n b_n l_n \delta_n (\log N_n)^3]^{1/2}\}$. By the same argument, we can show

$$\sup_{x\in[a,b]} \bigl|L_\mu(0,x) - J_\mu(0,x)\bigr| = \mathrm{O_p}\bigl\{\bigl[N_n b_n (\log N_n)^3\bigr]^{1/2}\bigr\}. \tag{7.24}$$

Hence, by (7.24) and Lemma 7.2, uniformly over $x \in \mathcal{S}_\epsilon([a,b])$,

$$\begin{aligned} L_\mu(l_n\delta_n, x) - L_\mu(x)/2 &= \bigl[J_\mu(l_n\delta_n, x) - J_\mu(0,x)\bigr] + \bigl[J_\mu(0,x) - L_\mu(x)/2\bigr] \\ &\quad + \bigl[L_\mu(0,x) - J_\mu(0,x)\bigr] + \Omega_n(x) \\ &= N_n b_n l_n \delta_n f_e(0)/\bigl[(b-a)s(x)\bigr]\bigl[1+\mathrm{o}(1)\bigr] + \mathrm{O_p}(\nu_n), \end{aligned} \tag{7.25}$$

where $\nu_n = N_n b_n^3 + 1 + [N_n b_n (\log N_n)^3]^{1/2} + [N_n b_n l_n \delta_n (\log N_n)^3]^{1/2}$. Because $l_n \to \infty$ and $l_n\delta_n \to 0$, it is easy to see that $\nu_n = \mathrm{o}(N_n b_n l_n \delta_n)$ and $N_n b_n l_n \delta_n \to \infty$, which implies $L_\mu(l_n\delta_n, x) - L_\mu(x)/2 \to \infty$ uniformly over $x \in \mathcal{S}_\epsilon[a,b]$ in view of $\sup_x s(x) < \infty$. Since $L_\mu(\delta_1, x)$ is nondecreasing in δ_1, (7.23) and (7.25) entail $\mathbb{P}\{\sup_x \hat{\Delta}_\mu(x) \le l_n\delta_n\} \to 1$. Similarly, $\mathbb{P}\{\inf_x \hat{\Delta}_\mu(x) \ge -l_n\delta_n\} \to 1$. So, $\sup_x |\hat{\Delta}_\mu(x)| = \mathrm{O_p}(l_n\delta_n)$. Since the rate of $l_n \to \infty$ can be arbitrarily slow, $\sup_x |\hat{\Delta}_\mu(x)| = \mathrm{O_p}(\delta_n)$.

Again, by (7.23) and Lemma 7.2, uniformly over $x \in \mathcal{S}_\epsilon([a,b])$,

$$\begin{aligned} L_\mu(0,x) - J_\mu(0,x) &= L_\mu\bigl(\hat{\Delta}_\mu(x), x\bigr) - J_\mu\bigl(\hat{\Delta}_\mu(x), x\bigr) + \mathrm{O_p}\Bigl[\sqrt{N_n b_n \delta_n (\log N_n)^3}\Bigr] \\ &= \bigl[L_\mu\bigl(\hat{\Delta}_\mu(x), x\bigr) - L_\mu(x)/2\bigr] + \bigl[L_\mu(x)/2 - J_\mu(0,x)\bigr] \\ &\quad - \bigl[J_\mu\bigl(\hat{\Delta}_\mu(x), x\bigr) - J_\mu(0,x)\bigr] + \mathrm{O_p}\Bigl[\sqrt{N_n b_n \delta_n (\log N_n)^3}\Bigr] \\ &= \mathrm{O_p}(1) + N_n b_n^3 \rho_\mu(x) f_e(0)\psi_K/\bigl[(b-a)s(x)\bigr] + \mathrm{O}\bigl(1 + N_n b_n^5\bigr) \\ &\quad - N_n b_n \hat{\Delta}_\mu(x)\bigl\{f_e(0)/\bigl[(b-a)s(x)\bigr] + \mathrm{O}(\delta_n)\bigr\} \\ &\quad + \mathrm{O_p}\Bigl[\sqrt{N_n b_n \delta_n (\log N_n)^3}\Bigr]. \end{aligned}$$

The representation (4.8) then follows by solving $\hat{\Delta}_\mu(x)$ from the above equation. □

Proof of Theorem 4.2. We use the argument in Theorem 4.1 and only sketch the outline. Let

$$D_s(\delta_1, \delta_2, x) = \bigl[L_s(\delta_1, \delta_2, x) - J_s(\delta_1, \delta_2, x)\bigr] - \bigl[L_s(0,0,x) - J_s(0,0,x)\bigr].$$

Using Theorem 3.3, we can show that

$$\sup_{x\in[a,b]} \bigl|L_s(0,0,x) - J_s(0,0,x)\bigr| = \mathrm{O_p}\bigl\{\bigl[N_n h_n (\log N_n)^3\bigr]^{1/2}\bigr\}, \tag{7.26}$$

$$\sup_{|\delta_1|+|\delta_2|\le\delta_n, x\in[a,b]} \bigl|D_s(\delta_1, \delta_2, x)\bigr| = \mathrm{O_p}\bigl\{\bigl[N_n h_n (\log N_n)^3\bigr]^{1/2}\bigr\}, \tag{7.27}$$

hold for all $b_n \to 0$, $h_n \to 0$ and $\delta_n \to 0$ satisfying $\sup_n \log N_n/[N_n \min(b_n, h_n)\delta_n] < \infty$.

Let $\delta_n = b_n^2 + h_n^2 + [(\log N_n)^3/(N_n b_n)]^{1/2} + [(\log N_n)^3/(N_n h_n)]^{1/2}$ and $l_n \to \infty$ be a sequence such that $l_n\delta_n \to 0$. By Theorem 4.1, $\tilde{\Delta}_\mu(x) := \tilde{\mu}(x) - \mu(x) = \mathrm{O_p}(\delta_n)$. Using (7.27) and Lemma 7.2, we can derive the following counterpart of (7.25)

$$\begin{aligned}
L_s\bigl(\tilde{\Delta}_\mu(x), l_n\delta_n, x\bigr) - L_s(x)/2 &= \bigl[J_s\bigl(\tilde{\Delta}_\mu(x), l_n\delta_n, x\bigr) - J_s\bigl(\tilde{\Delta}_\mu(x), 0, x\bigr)\bigr] \\
&\quad + \bigl[J_s\bigl(\tilde{\Delta}_\mu(x), 0, x\bigr) - L_s(x)/2\bigr] + L_s(0,0,x) - J_s(0,0,x) \\
&\quad + \mathrm{O_p}\bigl\{\bigl[N_n h_n l_n \delta_n (\log N_n)^3\bigr]^{1/2}\bigr\} \\
&= N_n h_n l_n \delta_n \kappa_+ / \bigl[(b-a)s(x)\bigr]\bigl[1 + \mathrm{o_p}(1)\bigr] \to \infty.
\end{aligned}$$

Let $\hat{\Delta}_s(x) = \hat{s}(x) - s(x)$. By the same argument in (7.23), $\sup_x |L_s(\tilde{\Delta}_\mu(x), \hat{\Delta}_s(x), x) - L_s(x)/2| = \mathrm{O_p}(1)$. Notice that $L_s(\tilde{\Delta}_\mu(x), \delta_2, x)$ is nondecreasing in x. Thus, $\mathbb{P}\{\sup_x \hat{\Delta}_s(x) \le l_n k_n\} \to 1$. Similarly, $\mathbb{P}\{\inf_x \hat{\Delta}_s(x) \ge -l_n k_n\} \to 1$. Then $\sup_x |\hat{\Delta}_s(x)| = \mathrm{O_p}(\delta_n)$.

Write $\varpi_n = [N_n h_n \delta_n (\log N_n)^3]^{1/2}$. To derive the Bahadur representation (4.14), we use (7.27) and Lemma 7.2 to obtain

$$\begin{aligned}
&L_s(0,0,x) - J_s(0,0,x) \\
&\quad = \bigl[L_s\bigl(\tilde{\Delta}_\mu(x), \hat{\Delta}_s(x), x\bigr) - L_s(x)/2\bigr] + \bigl[L_s(x)/2 - J_s\bigl(\hat{\Delta}_\mu(x), 0, x\bigr)\bigr] \\
&\qquad - \bigl[J_s\bigl(\tilde{\Delta}_\mu(x), \hat{\Delta}_s(x), x\bigr) - J_s\bigl(\tilde{\Delta}_\mu(x), 0, x\bigr)\bigr] + \mathrm{O_p}(\varpi_n) \\
&\quad = \mathrm{O_p}(1) + N_n h_n \kappa_+ \bigl\{\bigl[h_n^2 \psi_K \rho_s(x) - \kappa \tilde{\Delta}_u(x)\bigr]/\bigl[(b-a)s(x)\bigr] + \mathrm{O}\bigl(h_n^4 + \delta_n^2\bigr)\bigr\} \\
&\qquad - N_n h_n \hat{\Delta}_s(x)\bigl\{\kappa_+/\bigl[(b-a)s(x)\bigr] + \mathrm{O}(\delta_n)\bigr\} + \mathrm{O_p}(\varpi_n).
\end{aligned}$$

Solving $\hat{\Delta}_s(x)$ from the above equation, we obtain the Bahadur representation (4.14). □

7.5. Proof of Corollaries 4.1–4.2

Again we use the coupling argument to convert the dependent data to m-dependent case. Theorem 7.1 below presented a CLT for m-dependent sequence with unbounded m.

Theorem 7.1 (Romano and Wolf [22]). *Let $Z_{n,j}$, $1 \le j \le d_n$, be a triangular array of mean zero k_n-dependent random variables. Define*

$$S_n = \sum_{j=1}^{d_n} Z_{n,j}, \qquad B_n^2 = \mathrm{Var}(S_n), \qquad S_{n,h,a} = \sum_{j=a}^{a+h-1} Z_{n,j}, \qquad B_{n,h,a}^2 = \mathrm{Var}(S_{n,h,a}).$$

Assume that there exist some $\delta > 0$, $-1 \le \gamma < 1$, $C_{n,1}, C_{n,2}, C_{n,3} > 0$ such that

(a)	$\mathbb{E}\bigl(\lvert Z_{n,j}\rvert^{2+\delta}\bigr) = \mathrm{O}(C_{n,1})$;	(b)	$B_{n,h,a}^2/h^{1+\gamma} = \mathrm{O}(C_{n,2})$	*for all $h \ge k_n, a$;*
(c)	$B_n^2/\bigl(d_n C_{n,2}^{\gamma}\bigr) \ge C_{n,3}$;	(d)	$C_{n,2}/C_{n,3} = \mathrm{O}(1)$;	
(e)	$C_{n,1}/C_{n,3}^{(2+\delta)/2} = \mathrm{O}(1)$;	(f)	$k_n^{1+(1-\gamma)(1+2/\delta)}/d_n \to 0$.	

Then $S_n/B_n \Rightarrow N(0,1)$.

Proof of Corollaries 4.1–4.2. We only prove Corollary 4.1 since Corollary 4.2 can be similarly treated. By the Bahadur representation (4.8), under the specified condition, $r_n\sqrt{N_n b_n} \to 0$. Thus, it suffices to show $(N_n b_n)^{-1/2} Q_{b_n}(x) \Rightarrow N(0, \varphi_K/[4(b-a)])$. Recall $e_{i,j}(k_n)$ and $\tilde{Y}_{i,j}$ in (3.1) and (3.5). Define the coupling process

$$\tilde{Q}_{b_n}(x) = -\sum_{i=1}^{n}\sum_{j=1}^{m_i}\bigl\{\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)} - \mathbb{E}[\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)}]\bigr\}K_{b_n}(x_{i,j}-x).$$

Let the coupling lag $k_n = \lfloor c\log N_n\rfloor$ be chosen as in Theorem 3.2. By Theorem 3.2, $Q_{b_n}(x) - \tilde{Q}_{b_n}(x) = \mathrm{O_p}[(\log N_n)^2] = \mathrm{o_p}[(N_n b_n)^{1/2}]$. It remains to show $(N_n b_n)^{-1/2}\tilde{Q}_{b_n}(x) \Rightarrow N(0, \varphi_K/[4(b-a)])$. Recall $M_n = \max_{1\le i\le n} m_i$. Set $\tilde{Y}_{i,j} = 0$ for $m_i < j \le M_n$. Define

$$Z_{n,j} = \sum_{i=1}^{n}\zeta_{i,j}, \qquad \text{where } \zeta_{i,j} = (N_n b_n)^{-1/2}\bigl\{\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)} - \mathbb{E}[\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)}]\bigr\}K_{b_n}(x_{i,j}-x).$$

Then we can write $-(N_n b_n)^{-1/2}\tilde{Q}_{b_n}(x) = \sum_{j=1}^{M_n} Z_{n,j}$. Notice that $Z_{n,j}$, $j = 1, 2, \dots$, are $(2k_n+1)$-dependent, and $\zeta_{i,j}$, $i = 1, 2, \dots$, are independent for each fixed j.

Let S_n, B_n^2, $S_{n,h,a}$ and $B_{n,h,a}^2$ be defined in Theorem 7.1. We shall verify the conditions in Theorem 7.1. By the independence of the summands $\zeta_{i,j}$ in $Z_{n,j}$,

$$\begin{aligned}\mathbb{E}\bigl(|Z_{n,j}|^4\bigr) &= \sum_{i=1}^{n}\mathbb{E}\bigl(|\zeta_{i,j}|^4\bigr) + 6\sum_{i_1\ne i_2}\mathbb{E}\bigl(|\zeta_{i_1,j}|^2\bigr)\mathbb{E}\bigl(|\zeta_{i_2,j}|^2\bigr)\\ &= \frac{\mathrm{O}(1)}{(N_n b_n)^2}\left\{\sum_{i=1}^{n}K_{b_n}^4(x_{i,j}-x) + \left[\sum_{i=1}^{n}K_{b_n}^2(x_{i,j}-x)\right]^2\right\} = \mathrm{O}\bigl(1/M_n^2\bigr),\end{aligned}$$

in view of $nM_n = \mathrm{O}(N_n)$. Since $\tilde{Y}_{i,j}$ and $Y_{i,j}$ have same distribution, we have $g(x, x_{i,j}) := \mathrm{Var}(\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)}) = F_e\{[\mu(x)-\mu(x_{i,j})]/s(x_{i,j})\} - F_e^2\{[\mu(x)-\mu(x_{i,j})]/s(x_{i,j})\}$. Recall $F_e(0) = 1/2$. Then $g(x,x) = 1/4$. Thus, by (4.9) and the $(2k_n+1)$-dependence of $\tilde{Y}_{i,j}$, $j\in\mathbb{Z}$, applying Lemma 7.2(ii) with $r = 1$ produces

$$\begin{aligned}B_n^2 &= \frac{1}{N_n b_n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\mathrm{Var}(\mathbf{1}_{\tilde{Y}_{i,j}\le\mu(x)})K_{b_n}^2(x_{i,j}-x)\\ &\quad + \frac{\mathrm{O}(1)}{N_n b_n}\sum_{i=1}^{n}\sum_{1\le j_1<j_2\le n_i, |j_1-j_2|\le 2k_n} K_{b_n}(x_{i,j_1}-x)K_{b_n}(x_{i,j_2}-x)\\ &= \frac{1}{N_n b_n}\left[\frac{N_n b_n\varphi_K}{4(b-a)} + \mathrm{O}\bigl(N_n b_n^2\bigr)\right] + \frac{\mathrm{O}(nM_n k_n b_n \iota_n)}{N_n b_n} \to \frac{\varphi_K}{4(b-a)},\end{aligned}$$

in view of $nM_n = \mathrm{O}(N_n)$ and $k_n \iota_n \to 0$. Similarly, we can show $B^2_{n,h,a} = \mathrm{O}(nh/N_n) = \mathrm{O}(h/M_n)$. Therefore, it is easy to see that the conditions in Theorem 7.1 hold with $\delta = 2$, $\gamma = 0$, and straightforward choices of $C_{n,1}, C_{n,2}, C_{n,3}$, completing the proof. $\square$

Acknowledgements

We are grateful to an Associate Editor and three anonymous referees for their insightful comments. Wei's research was supported by the National Science Foundation (DMS-09-06568) and a career award from NIEHS Center for Environmental Health in Northern Manhattan (ES-009089). Zhao's research was supported by a NIDA Grant P50-DA10075-15. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

References

[1] Andrews, D.W.K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830

[2] Andrews, D.W.K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int. Stat. Rev.* **62** 119–132.

[3] Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[4] Bhattacharya, P.K. and Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415. MR1062716

[5] Brumback, B.A. and Rice, J.A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.* **93** 961–994. MR1649194

[6] Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory* **18** 169–192. MR1885356

[7] Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. MR1105843

[8] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** 203–236. MR2199291

[9] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Series in Statistics*. New York: Springer. MR1964455

[10] Fan, J. and Zhang, J.T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 303–322. MR1749541

[11] Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli* **15** 659–686. MR2555194

[12] He, X., Fu, B. and Fung, W.K. (2003). Median regression for longitudinal data. *Stat. Med.* **22** 3655–3669.

[13] He, X., Zhu, Z.Y. and Fung, W.K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. MR1929164

[14] Ho, H.C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

[15] Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470. MR1794246

[16] Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699
[17] Koenker, R. (2004). Quantile regression for longitudinal data. *J. Multivariate Anal.* **91** 74–89. MR2083905
[18] Koenker, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge: Cambridge Univ. Press. MR2268657
[19] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644
[20] Li, Q. and Racine, J.S. (2007). *Nonparametric Econometrics*: *Theory and Practice*. Princeton, NJ: Princeton Univ. Press. MR2283034
[21] Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **53** 233–243. MR1094283
[22] Romano, J.P. and Wolf, M. (2000). A more general central limit theorem for m-dependent random variables with unbounded m. *Statist. Probab. Lett.* **47** 115–124. MR1747098
[23] Shao, Q.M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24** 2098–2127. MR1415243
[24] Shao, X. and Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105
[25] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838963
[26] Truong, Y.K. and Stone, C.J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97. MR1150335
[27] Walker, E. and Wright, S.P. (2002). Comparing curves using additive models. *J. Qual. Technol.* **34** 118–129.
[28] Wang, H.J. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *Ann. Statist.* **37** 756–781. MR2502650
[29] Wang, H.J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37** 3841–3866. MR2572445
[30] Wei, Y., Zhao, Z. and Lin, D.K.J. (2012). Profile control charts based on nonparametric $L-1$ regression methods. *Ann. Appl. Stat.* **6** 409–427. MR2951543
[31] Wu, H. and Zhang, J.T. (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc.* **97** 883–897. MR1941417
[32] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic). MR2172215
[33] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990
[34] Wu, W.B. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 391–410. MR2323759
[35] Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106
[36] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628
[37] Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research areas. *The Statistician* **52** 331–350. MR2011179

Received December 2011 and revised May 2013