Published in final edited form as:
Biometrika. 2013 March ; 100(1): 203–212. doi:10.1093/biomet/ass050.

Unified inference for sparse and dense longitudinal models

Seonjin Kim and **Zhibiao Zhao**
Department of Statistics, Penn State University, University Park, Pennsylvania 16802, U.S.A.
Seonjin Kim: szk172@psu.edu; Zhibiao Zhao: zuz13@stat.psu.edu

Summary

In longitudinal data analysis, statistical inference for sparse data and dense data could be substantially different. For kernel smoothing estimate of the mean function, the convergence rates and limiting variance functions are different under the two scenarios. The latter phenomenon poses challenges for statistical inference as a subjective choice between the sparse and dense cases may lead to wrong conclusions. We develop self-normalization based methods that can adapt to the sparse and dense cases in a unified framework. Simulations show that the proposed methods outperform some existing methods.

Keywords

Dense longitudinal data; Kernel smoothing; Mixed-effects model; Nonparametric estimation; Self-normalization; Sparse longitudinal data

1. Introduction

Longitudinal models have extensive applications in biomedical, psychometric and environmental sciences (Fitzmaurice et al., 2004; Wu & Zhang, 2006). In longitudinal studies, repeated measurements are recorded over time from subjects, and therefore measurements from the same subject are correlated. One popular framework is to assume that the observations from each subject are noisy discrete realizations of an underlying process $\{\xi(\cdot)\}$:

$$Y_{ij}=\xi_i(X_{ij})+\sigma(X_{ij})\varepsilon_{ij}(i=1,\ldots,n;j=1,\ldots,n_i). \quad (1)$$

Here Y_{ij} is the measurement at time X_{ij} from subject i, $\{\xi_i(\cdot)\}$ are independent realizations of an underlying process $\{\xi(\cdot)\}$, ε_{ij} are errors with $E(\varepsilon_{ij}) = 0$ and $E(\varepsilon_{ij}^2)=1$, n_i is the number of measurements collected on subject i, and n is the total number of subjects.

There are two typical approaches to taking between-subject variation into account: functional principal component analysis (Yao et al., 2005a,b; Yao, 2007; Ma et al., 2012) and the mixed-effects approach (Wu & Zhang, 2002; Zhang & Chen, 2007). The basic idea of the latter is to decompose $\{\xi_i(\cdot)\}$ into a fixed population mean $\mu(\cdot) = E\{\xi_i(\cdot)\}$ and a subject-specific random trajectory $\upsilon_i(\cdot)$ with $E\{\upsilon_i(x)\} = 0$ and covariance function $\gamma(x, x') = \mathrm{cov}\{\upsilon_i(x), \upsilon_i(x')\}$. Then (1) becomes

$$Y_{ij}=\mu(X_{ij})+\upsilon_i(X_{ij})+\sigma(X_{ij})\varepsilon_{ij}(i=1,\ldots,n;j=1,\ldots,n_i). \quad (2)$$

The goal is to estimate the population mean $\mu(\cdot)$ and construct a confidence interval for it.

Depending on the number of measurements within subjects, model (2) has two scenarios: dense and sparse longitudinal data. Dense longitudinal data allow $n_i \to \infty$ and a conventional estimation approach is to smooth each individual curve and then construct an estimator based on the smoothed curves (Ramsay & Silverman, 2005; Hall et al., 2006; Zhang & Chen, 2007). In sparse longitudinal data, the n_i are either bounded or independent and identically distributed with $E(n_i) < \infty$, and due to the sparse observations from individual subjects, it is essential to pool data together (Yao et al., 2005a; Hall et al., 2006; Yao, 2007; Ma et al., 2012).

In practice, the boundary between dense and sparse cases may not always be clear, and such ambiguity could pose challenges for statistical inference, since different researchers may likely classify the same data set differently. To address this issue, Li & Hsing (2010) proposed a unified weighted local linear estimator of $\mu(x)$. However, as shown in Section 2, the latter estimator has different convergence rates and limiting variances under the two scenarios. Therefore, to construct a confidence interval for $\mu(x)$, one should make a subjective decision whether to treat the data as sparse or dense. In Section 2, we show that the constructed confidence intervals based on a sparse or dense assumption could differ substantially, depending on many unknown factors. Another challenging issue is that the limiting variance function contains the unknown functions $\gamma(x, x)$ and $\sigma^2(x)$. As shown by Wu & Zhang (2002), Yao et al. (2005a,b), Müller (2005) and Li & Hsing (2010), covariance estimation requires extra smoothing procedures.

We develop two unified nonparametric approaches that can successfully solve the aforementioned issues. First, we establish a unified convergence theory so that inference can be conducted without deciding whether the data are dense or sparse. Second, the unknown limiting variance is canceled out through a self-normalization technique, and thus the proposed methods do not require estimation of the functions $\gamma(x, x)$ and $\sigma^2(x)$. The first approach introduces a unified self-normalized central limit theorem that can adapt to both cases. The second approach constructs a self-normalizer based on recursive estimates of the mean function. The related methods have been explored mainly under parametric settings for time series data (Lobato, 2001; Kiefer & Vogelsang, 2005; Shao, 2010). In the longitudinal setting, our development of the self-normalization method is more attractive due to the sparse and dense scenario and the more complicated structure such as the within-subject covariance and overall noise variance function. Simulations show that the proposed methods outperform some existing methods.

2. Motivation

For model (2), we consider two scenarios: (i) sparse longitudinal data: $n_1, \ldots, n_n$ are independent and identically distributed positive-integer-valued random variables with $E(n_i) < \infty$; and (ii) dense longitudinal data: $n_i \geq M_n$ for some $M_n \to \infty$ as $n \to \infty$.

Throughout we let $f(\cdot)$ denote the density function of X_{ij} and let x be an interior point of the support of $f(\cdot)$. Li & Hsing (2010) proposed a sample-size weighted local linear estimator of $\mu(x)$. For technical convenience, we consider the weighted local constant estimator

$$\hat{\mu}_n(x)=\underset{\theta}{\operatorname{argmin}}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\theta)^2K\left(\frac{X_{ij}-x}{b}\right)=\frac{G_n}{H_n}, \quad (3)$$

where K is a kernel function satisfying $\int_{\mathbb{R}} K(u)du = 1$ and $b > 0$ is a bandwidth, with

$$H_n=\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}K\left(\frac{X_{ij}-x}{b}\right), G_n=\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij}K\left(\frac{X_{ij}-x}{b}\right). \quad (4)$$

The convergence rates and limiting variances are different for sparse and dense longitudinal data. To gain intuition about this, write

$$\hat{\mu}_n(x)-\mu(x)-\frac{1}{H_n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\{\mu(X_{ij})-\mu(x)\}K\left(\frac{X_{ij}-x}{b}\right)=\frac{1}{H_n}\sum_{i=1}^{n}\xi_i, \quad (5)$$

where the right hand side determines the asymptotic distribution of $\hat{\mu_n}(x)$, with

$$\xi_i=\frac{1}{n_i}\sum_{j=1}^{n_i}\xi_{ij}, \xi_{ij}=\{\upsilon_i(X_{ij})+\sigma(X_{ij})\varepsilon_{ij}\}K\left(\frac{X_{ij}-x}{b}\right). \quad (6)$$

Recall $\gamma(x, x') = \operatorname{cov}\{\upsilon_i(x), \upsilon_i(x')\}$. For $j \neq j'$, by $E(\xi_{ij}\xi_{ij'}) = E\{E(\xi_{ij}\xi_{ij'} \mid X_{ij}, X_{ij'})\}$,

$$E(\xi_{ij}\xi_{ij'})=E\left\{\gamma(X_{ij},X_{ij'})K\left(\frac{X_{ij}-x}{b}\right)K\left(\frac{X_{ij'}-x}{b}\right)\right\}\approx b^2f^2(x)\gamma(x,x). \quad (7)$$

Throughout, $c_n \approx d_n$ means that $c_n/d_n \to 1$. Similarly,

$$E(\xi_{ij}^2)=E\{E(\xi_{ij}^2|X_{ij})\}\approx bf(x)\psi_K\{\gamma(x,x)+\sigma^2(x)\}, \psi_K=\int_{\mathbb{R}}K^2(u)du. \quad (8)$$

Applying (7)–(8) to $\operatorname{var}(\xi_i|n_i)=n_i^{-2}\{\sum_{1\le j\neq j'\le n_i}E(\xi_{ij}\xi_{ij'})+\sum_{j=1}^{n_i}E(\xi_{ij}^2)\}$, we obtain

$$\operatorname{var}(\xi_i|n_i)\approx(1-1/n_i)b^2f^2(x)\gamma(x,x)+f(x)\psi_K\{\gamma(x,x)+\sigma^2(x)\}b/n_i. \quad (9)$$

For sparse case with $b \to 0$, $\operatorname{var}(\xi_i \mid n_i) \approx bf(x)\psi_K\{\gamma(x, x) + \sigma^2(x)\}/n_i$; for dense case with $n_i \geq M_n$ and $M_n b \to \infty$, $\operatorname{var}(\xi_i \mid n_i) \approx b^2f^2(x)\gamma(x, x)$.

Theorem 1. *Assume Assumption 1 in the Appendix. Let f(x) be the density of X_{ij}. Write*

$$\psi_K=\int_{\mathbb{R}}K^2(u)du, \rho(x)=\left\{\frac{\mu''(x)}{2}+\frac{\mu'(x)f'(x)}{f(x)}\right\}\int_{\mathbb{R}}u^2K(u)du.$$

i. *Sparse data: Assume $nb \to \infty$ and $\sup_n nb^5 < \infty$. Then*

$$\sqrt{(nb)}\{\hat{\mu}_n(x) - \mu(x) - b^2\rho(x)\} \to N\{0, s^2_{\text{sparse}}(x)\}, \quad (10)$$

where $s^2_{\text{sparse}}(x) = \tau\psi_K\{\gamma(x,x) + \sigma^2(x)\}/f(x)$ *and* $\tau = E(1/n_1)$.

ii. *(ii) Dense data: Assume* $n_i \geq M_n$, $M_n b \to \infty$, $nb \to \infty$ *and* $\sup_n nb^4 < \infty$. *Then*

$$\sqrt{n}\{\hat{\mu}_n(x) - \mu(x) - b^2\rho(x)\} \to N\{0, s^2_{\text{dense}}(x)\}, s^2_{\text{dense}}(x) = \gamma(x,x). \quad (11)$$

It is worth mentioning some related results. Li & Hsing (2010) established the uniform consistency of $\hat{\mu_n}(x)$ with different rates under the sparse and dense cases, but they did not obtain the asymptotic distribution. Wu & Zhang (2002) also showed that the local polynomial mixed-effects estimator has different convergence rates and limiting variances under the two scenarios. Under a Karhunen–Loève representation of longitudinal models, Yao (2007) studied the sparse case by allowing n_i to be dependent on n; see also Ma et al. (2012).

By Theorem 1, the confidence interval for $\mu(x)$ is different under the two cases. Let $z_{1-\alpha/2}$ be the $1 - \alpha/2$ standard normal quantile. Then an asymptotic $1 - \alpha$ confidence interval for $\mu(x)$ is

$$\hat{\mu}_n(x) - b^2\hat{\rho}(x) \pm z_{1-\alpha/2}(nb)^{-1/2}[\hat{\tau}\psi_K\{\hat{\gamma}(x,x) + \hat{\sigma}^2(x)\}/\hat{f}(x)]^{1/2} \quad (12)$$

for sparse data, or

$$\hat{\mu}_n(x) - b^2\hat{\rho}(x) \pm z_{1-\alpha/2}n^{-1/2}\{\hat{\gamma}(x,x)\}^{1/2} \quad (13)$$

for dense data. Here, $\hat{\tau} = n^{-1}\sum_{i=1}^{n} n_i^{-1}$, $\hat{\gamma}(x, x)$, $\hat{\sigma^2}(x)$, $\hat{f}(x)$ and $\hat{\rho}(x)$ are consistent estimates of τ, $\gamma(x, x)$, $\sigma^2(x)$, $f(x)$ and $\rho(x)$. The ratio of the lengths of the two confidence intervals is $R = [\psi_K \hat{\tau}\{1 + \hat{\sigma^2}(x)/\hat{\gamma}(x, x)\}/\{b\hat{f}(x)\}]^{1/2}$, which depends on the denseness parameter τ, the signal-to-noise ratio $\gamma(x, x)/\sigma^2(x)$, the bandwidth b and the design density $f(\cdot)$. The further away R is from one, the larger the discrepancy between the two constructed confidence intervals.

Remark 1. In the dense case, suppose n_i is proportional to $M_n \to \infty$. Theorem 1 (ii) studies the case $M_n b \to \infty$. If $M_n b \to 0$, then the leading term in (9) is $f(x)\psi_K\{\gamma(x, x) + \sigma^2(x)\}b/n_i$. If $M_n b$ is bounded away from 0 and ∞, then both terms in (9) are of the same order. If b is proportional to $(nM_n)^{-1/5}$, then a sufficient condition for $M_n b \to \infty$ is $M_n^4/n \to \infty$. In many practical problems, n is about 30–200, M_n is about 10–30, and M_n^4/n is sufficiently large.

3. Unified approaches for sparse and dense data

3·1. A unified self-normalized central limit theorem

The discussion in Section 2 suggests a need for a unified approach. For independent and identically distributed random variables $Z_1, \ldots, Z_n$, de la Peña et al. (2009) gave an extensive account of the asymptotic properties of the self-normalized statistic

$\sum_{i=1}^{n} Z_i / \sqrt{(\sum_{i=1}^{n} Z_i^2)}$. In this section, we present a unified self-normalized central limit theorem for $\hat{\mu}_n(x)$. For H_n in (4), define

$$U_n^2(x) = \frac{1}{H_n^2} \sum_{i=1}^{n} \left[\frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - \hat{\mu}(X_{ij})\} K\left(\frac{x - X_{ij}}{b} \right) \right]^2 .$$

Theorem 2. *Assume Assumption 1 in the Appendix. Suppose* $nb / \log n \to \infty$, $\sup_n nb^5 < \infty$ *for sparse data or* $n_i \geq M_n$, $M_n b \to \infty$, $nb^2 / \log n \to \infty$, $\sup_n nb^4 < \infty$ *for dense data. Then* $\{\hat{\mu}_n(x) - \mu(x) - b^2 \rho(x)\} / U_n(x) \to N(0, 1)$ *in both the sparse and dense settings.*

Many papers treat sparse and dense data separately. For example, Yao et al. (2005a,b), Yao (2007) and Ma et al. (2012) studied sparse longitudinal data. For the local polynomial mixed-effects estimator, Wu & Zhang (2002) obtained different central limit theorems under the two scenarios. By contrast, Theorem 2 establishes a unified central limit theorem, which can be used to construct a unified asymptotic pointwise $1 - \alpha$ confidence interval for $\mu(x)$:

$$\hat{\mu}_n(x) - b^2 \hat{\rho}(x) \pm z_{1-\alpha/2} U_n(x). \quad (14)$$

While the confidence intervals (12)–(13) require estimation of the within-subject covariance function $\gamma(x, x)$ and the overall noise variance function $\sigma^2(x)$, (14) avoids such extra smoothing steps and can adapt to the sparse or dense setting through the self-normalizer $U_n(x)$.

To select the bandwidth b, we adopt subject-based cross-validation (Rice & Silverman, 1991). The idea is to leave one subject out in model fitting, validate the fitted model using the left-out subject, and choose the optimal bandwidth by minimizing the prediction error:

$$b^* = \arg\min_b \mathrm{SJCV}(b), \mathrm{SJCV}(b) = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - \hat{\mu}^{(-i)}(X_{ij})\}^2, \quad (15)$$

where $\hat{\mu}^{(-i)}(x)$ represents the estimator of $\mu(x)$ based on data from all but the ith subject.

In practice, it is difficult to estimate the bias $b^2\rho(x)$ due to the unknown derivatives f', μ', μ''. In our simulations, we use $K(u) = 2G(u) - G(u/\sqrt{2})/\sqrt{2}$ with $G(u)$ the standard normal density. Then $\int_{\mathbb{R}} u^2 K(u) du = 0$ and $\rho(x) = 0$. However, this does not solve the bias issue. For example, if f and μ are four times differentiable, then we have the higher order bias term $O(b^4)$. The bias issue is inherently difficult and there is no good solution so far.

3·2. Self-normalization based on recursive estimates

In this section we introduce another self-normalization method based on recursive estimates. For $m = 1, \ldots, n$, denote by $\hat{\mu}_m(x)$ the estimator in (3) based on observations from the first m subjects. Then $\hat{\mu}_1(x), \ldots, \hat{\mu}_n(x)$ are estimates of $\mu(x)$ with increasing accuracy. Moreover, $\hat{\mu}_m(x)$ has similar asymptotic normality as in (10)–(11). For example, for each $0 < t \leq 1$, the counterpart of (10) for sparse data is

$\sqrt{(ntb)}\{\hat{\mu}_{\lfloor nt \rfloor}(x) - \mu(x) - b^2\rho(x)\} \to N\{0, s^2_{\text{sparse}}(x)\}$. Throughout, $\lfloor z \rfloor$ is the integer part of z. Therefore, $\hat{\mu}_n(x)$ and $\hat{\mu}_{\lfloor nt \rfloor}(x)$ have proportional convergence rates and the same limiting variance, which motivates us to consider certain ratios between $\hat{\mu}_n(x)$ and $\hat{\mu}_{\lfloor nt \rfloor}(x)$ to cancel out the convergence rates and limiting variance.

Since the above analysis holds for all $0 < t \leq 1$, we consider an aggregated version

$$T_n(x) = \frac{\hat{\mu}_n(x) - \mu(x) - b^2\rho(x)}{V_n(x)}, V_n(x) = n^{-3/2}\left\{\sum_{m=\lfloor cn \rfloor}^{n} m^2 |\hat{\mu}_m(x) - \hat{\mu}_n(x)|^2\right\}^{1/2}.$$

Throughout $c > 0$ is a small constant included to avoid unstable estimation at the boundary. By our simulations, $c = 0{\cdot}1$ works reasonably well. Intuitively, we may interpret $\hat{\mu}_m(x)$, $m = 1, \ldots, n$, as observations from a population with mean $\mu(x)$ and treat $\hat{\mu}_n(x)$ as sample average. Thus, $V_n(x)$ can be viewed as a weighted sample standard deviation with the weight m^2 reflecting the accuracy of $\hat{\mu}_m(x)$, and $V_n(x)$ mimics the usual normalizer in the Student-t distribution.

Theorem 3. *Assume the conditions in Theorem 1. Let $\{B_t\}$ be a standard Brownian motion. Then $T_n(x) \to B_1/\{\int_c^1 (B_t - tB_1)^2 dt\}^{1/2}$ under either the sparse or the dense settings.*

By Theorem 3, an asymptotic pointwise $1 - \alpha$ confidence interval for $\mu(x)$ is $\hat{\mu}_n(x) - b^2\hat{\rho}(x) \pm q_{1-\alpha/2}V_n(x)$, where $q_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the limiting distribution. The latter confidence interval is the same for both scenarios, with the convergence rate and limiting variance being built into the self-normalizer $V_n(x)$ implicitly. Our method can be viewed as an extension of the parametric self-normalization methods in Lobato (2001), Kiefer & Vogelsang (2005) and Shao (2010) for time series data to the nonparametric longitudinal model (2).

In practice, however, subjects have no natural ordering, and we can use the average of multiple copies of $V_n^2(x)$ through permuting the subjects. For a large n, since it is computationally infeasible to enumerate all permutations, we consider only a fixed number, say T, of random permutations. Denote the corresponding $V_n(x)$ by $V_n^1(x), \ldots, V_n^T(x)$. Consider

$$\tilde{T}_n(x) = \frac{\hat{\mu}_n(x) - \mu(x) - b^2\rho(x)}{\tilde{V}_n(x)}, \tilde{V}_n^2(x) = \frac{1}{T}\sum_{r=1}^{T}\{V_n^T(x)\}^2.$$

By the above analysis, the asymptotic distribution of $\tilde{T}_n(x)$ is the same under both the sparse and dense settings. However, it is not clear whether $\tilde{T}_n(x)$ is asymptotically normally distributed. Nevertheless, in light of the asymptotic normality of $\hat{\mu}_n(x)$, the proof of Theorem 3 and $E\{\int_c^1 (B_t - tB_1)^2 dt\} = (1 - 3c^2 + 2c^3)/6$, we propose the pointwise confidence interval

$$\hat{\mu}_n(x) - b^2\hat{\rho}(x) \pm z_{1-\alpha/2}\tilde{V}_n(x)\sqrt{c_1}, c_1 = 6/(1 - 3c^2 + 2c^3). \quad (16)$$

Here $z_{1-\alpha/2}$ is defined in (12). We call it the rule-of-thumb self-normalization based confidence interval. Our quantile-quantile studies show that the empirical quantile of $\tilde{T_n}(x)$ with 200 permutations matches well with that of $N(0, c_1)$ under the settings in Section 4.

4. Numerical results

Following Li & Hsing (2010), we consider the model

$$Y_{ij} = \mu(X_{ij}) + \sum_{k=1}^{3} \alpha_{ik}\Phi_k(X_{ij}) + \sigma\varepsilon_{ij} (i = 1, \ldots, n; j = 1, \ldots, n_i),$$

where $\alpha_{ik} \sim N(0, \omega_k)$ and $\varepsilon_{ij} \sim N(0, 1)$. Let $\mu(x) = 5\,(x - 0.6)^2$, $\Phi_1(x) = 1$, $\Phi_2(x) = \sqrt{2}\sin(2\pi x)$, $\Phi_3(x) = \sqrt{2}\cos(2\pi x)$, $(\omega_1, \omega_2, \omega_3) = (0{\cdot}6, 0{\cdot}3, 0{\cdot}1)$, and $n = 200$. Then the variance function $\gamma(x, x) = 0{\cdot}6 + 0{\cdot}6\sin^2(2\pi x) + 0{\cdot}2\cos^2(2\pi x)$. Two noise levels $\sigma = 1, 2$ are considered. The design points X_{ij} are uniformly distributed on [0, 1]. For the vector $N = (n_1, \ldots, n_n)$ of the number of measurements on individual subjects, we consider four cases

$$N_1 : n_i \sim U[\{2, 3, \ldots, 8\}]; N_2 : n_i \sim U[\{15, 16, \ldots, 35\}]; \quad (17)$$

$$N_3 : n_i \sim U[\{30, 31, \ldots, 70\}]; N_4 : n_i \sim U[\{150, 151, \ldots, 250\}]. \quad (18)$$

Here $U[\mathscr{D}]$ stands for the discrete uniform distribution on a finite set $\mathscr{D}$.

We compare six confidence intervals: the two self-normalization based confidence intervals in (14) and (16) with 200 permutations, the asymptotic normality based confidence intervals (12)–(13) assuming sparse and dense data, respectively, the bootstrap confidence interval with 200 bootstrap replications from sampling subjects with replacement, and the confidence interval

$$\hat{\mu}_n(x) - b^2\hat{\rho}(x) \pm z_{1-\alpha/2}n^{-1/2}\left\{(1 - \hat{\tau})\gamma(x, x) + \hat{\tau}\psi_K \frac{\gamma(x, x) + \sigma^2(x)}{bf(x)}\right\}^{1/2}. \quad (19)$$

The confidence interval (19) is practically infeasible as we need to estimate the unknown functions. Nevertheless, by using the true theoretical limiting variance function in (9), (19) serves as a standard against which we can measure the performance of other confidence intervals. When using the local linear method in Li & Hsing (2010) to estimate $\gamma(x, x)$, we found that negative estimates of $\gamma(x, x)$ occur frequently, especially when the noise level σ is high. For the purpose of comparison, we use the true functions $\gamma(x, x)$, $\sigma^2(x)$ and $f(x)$ to implement (12)–(13).

We consider two criteria: empirical coverage probabilities and lengths of confidence intervals. Let $x_1 < \cdots < x_{20}$ be 20 grid points evenly spaced on [0·1, 0·9]. For each x_j and a given nominal level, we construct confidence intervals for $\mu(x_j)$, and compute the empirical

coverage probabilities based on 1000 replications. For each of the six confidence intervals, we average their empirical coverage probabilities and lengths at 20 grid points. To facilitate computations in bandwidth selection, instead of using (15) for each replication, we set b to be the average of 20 optimal bandwidths in (15) based on 20 replications from each set of parameter choices.

The results are presented in Table 1. The performance of the confidence intervals (12)–(13) depends on whether the data are sparse or dense. As we increase the number of measurements on each subject from the sparse setting N_1 to the dense setting N_4, (12) under sparse assumption performs increasingly worse whereas (13) under dense assumption performs increasingly better. The simulation study further confirms the theoretical results in Theorem 1 that the confidence intervals (12)–(13) perform well only under their corresponding sparse or dense assumption. By contrast, the self-normalization based confidence intervals (14) and (16) deliver robust and superior performance: (i) they have similar widths but slightly better coverage probabilities than the bootstrap confidence interval; and (ii) they perform similar to the infeasible confidence interval (19) with true functions. Finally, (14) and (16) have comparable performance.

Acknowledgments

We are grateful to Editor Anthony Davison and three referees for their constructive comments. This work was supported by a National Institute on Drug Abuse grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.

Appendix: Regularity conditions and proofs

Assumption 1. (i) $K(\cdot)$ is bounded, symmetric, and has bounded support and bounded derivative. (ii) $\{\upsilon_i(\cdot)\}_i$, $\{X_{ij}\}_{ij}$, $\{\varepsilon_{ij}\}_{ij}$ are independent and identically distributed and mutually independent. Furthermore, the density function $f(\cdot)$ of X_{ij} is twice continuously differentiable in a neighborhood of x and $f(x) > 0$. (iii) In a neighborhood of x, $\mu(\cdot)$ is twice continuously differentiable, $\sigma^2(\cdot)$ is continuously differentiable; in a neighborhood of (x, x), $\gamma(x, x') = \mathrm{cov}\{\upsilon_i(x), \upsilon_i(x')\}$ is continuously differentiable. Moreover, $\gamma(x, x) > 0$, $\sigma^2(x) > 0$. (iv) $E\{|\upsilon_i(\cdot) + \sigma(\cdot)\varepsilon_{ij}|^4\}$ is continuous in a neighborhood of x and $E\{|\upsilon_i(x) + \sigma(x)\varepsilon_{ij}|^4\} < \infty$.

Proof of Theorem 1. Let ξ_i be defined in (6). Recall the decomposition (5). Write

$$H_n = \sum_{i=1}^{n} \nu_i, \nu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \nu_{ij}, \nu_{ij} = K\left(\frac{X_{ij} - x}{b}\right), \quad (20)$$

$$I_n = \sum_{i=1}^{n} \zeta_i, \zeta_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \zeta_{ij}, \zeta_{ij} = \{\mu(X_{ij}) - \mu(x)\}K\left(\frac{X_{ij} - x}{b}\right). \quad (21)$$

By the symmetry of K and Taylor's expansion, $E(\nu_{ij}) = \{1 + O(b^2)\}bf(x)$, $\mathrm{var}(\nu_{ij}) = O(b)$, $E(\zeta_{ij}) = b^3 f(x)\rho(x) + o(b^3)$, $\mathrm{var}(\zeta_{ij}) = O(b^3)$. In either the sparse or the dense case, $E(\nu_i \mid n_i) =$

$E(\nu_{ij})$ is non-random. Thus, $\mathrm{var}(\nu_i) = E\{\mathrm{var}(\nu_i \mid n_i)\} = \mathrm{var}(\nu_{ij})E(1/n_i)$ and $\mathrm{var}(H_n)=\sum_{i=1}^{n}\mathrm{var}(\nu_i)=O(b)\sum_{i=1}^{n}E(1/n_i)$. Write $\tau_n=n^{-1}\sum_{i=1}^{n}E(1/n_i)$. Then

$$H_n=E(H_n)+O_p\{\sqrt{\mathrm{var}(H_n)}\}=[1+O_p\{b^2+(nb/\tau_n)^{-1/2}\}]nbf(x). \quad (22)$$

Similarly, $I_n = nb^3 f(x)\rho(x) + o(nb^3) + O_p\{\sqrt{(nb^3\tau_n)}\}$. Thus,

$$I_n/H_n=b^2\rho(x)+\delta_n,\ \delta_n=o_p(b^2)+O_p\{\sqrt{(b\tau_n/n)}\}. \quad (23)$$

Dense case: Under given conditions, $\delta_n = o_p(n^{-1/2})$ and $\{nb^2f^2(x)\}^{-1}\mathrm{var}(\sum_{i=1}^{n}\xi_i) \to \gamma(x,x)$. For distinct j, r, s, k, by the argument in (7), $E(\xi_{ij}\xi_{ir}\xi_{is}\xi_{ik}) = O(b^4)$, $E(\xi_{ij}^2\xi_{ir}\xi_{is})=O(b^3), E(\xi_{ij}^2\xi_{ir}^2)=O(b^2), E(\xi_{ij}^3\xi_{ir})=O(b^2), E(\xi_{ij}^4)=O(b)$. Thus, $\sum_{i=1}^{n}E(\xi_i^4)=O(nb^4)=o\{(b\sqrt{n})^4\}$. By the Lyapunov central limit theorem, $\sum_{i=1}^{n}\xi_i/\{b\sqrt{nf(x)}\} \to N\{0,\gamma(x,x)\}$.

Sparse case: In (5), $\xi_1, \ldots, \xi_n$ are independent and identically distributed. The result follows from $\delta_n = o_p\{(nb)^{-1/2}\}$ and $\mathrm{var}(\xi_i) = E\{\mathrm{var}(\xi_i \mid n_i)\} \approx b\tau\psi_K f(x)\{\gamma(x,x) + \sigma^2(x)\}$.

Proof of Theorem 2. By Theorem 1, it suffices to show $nU_n^2(x) \to s^2_{\mathrm{dense}}(x)$ for dense data or $nbU_n^2(x) \to s^2_{\mathrm{sparse}}(x)$ for sparse data. For convenience, write $K_{ij} = K\{(X_{ij} - x)/b\}$. Let

$$S_n=\sum_{i=1}^{n}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\{Y_{ij}-\hat{\mu}_n(X_{ij})\}K_{ij}\right]^2=\sum_{i=1}^{n}(\xi_i^2+\eta_i^2+2\xi_i\eta_i), \quad (24)$$

where ξ_i is defined in (6) and $\eta_i=n_i^{-1}\sum_{j=1}^{n_i}\{\mu(X_{ij})-\hat{\mu}(X_{ij})\}K_{ij}$. By Theorem 3.1 in Li & Hsing (2010), $|\hat{\mu}(z) - \mu(z)| = O_p(\ell_n)$ uniformly for z in the neighborhood of x, where $\ell_n = b^2 + (n/\log n)^{-1/2}$ for dense data or $\ell_n = b^2 + (nb/\log n)^{-1/2}$ for sparse data. Then $\eta_i=O_p(\ell_n)n_i^{-1}\sum_{j=1}^{n_i}|K_{ij}|$. Using $\xi_i=n_i^{-1}\sum_{j=1}^{n_i}\xi_{ij}$, where ξ_{ij} is defined in (6), we obtain

$$\begin{aligned}\sum_{i=1}^{n}|\eta_i^2 &\\ &+2\xi_i\eta_i| \\ &=O_p(\ell_n^2)\sum_{i=1}^{n}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}|K_{ij}|\right)^2 \\ &+O_p(\ell_n)\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j=1}^{n_i}|\xi_{ij}|\sum_{j=1}^{n_i}|K_{ij}| \le O_p(\ell_n)J_n, J_n \\ &=\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}K_{ij}^2 \\ &+\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{j'=1}^{n_i}(K_{ij}^2+\xi_{ij'}^2).\end{aligned}$$

Here we have used $\ell_n^2=o(\ell_n), (\sum_{j=1}^{n_i}|K_{ij}|)^2 \le n_i\sum_{j=1}^{n_i}K_{ij}^2, 2|K_{ij}\xi_{ij'}| \le K_{ij}^2+\xi_{ij'}^2$. By $E(K_{ij}^2)=O(b)\text{and}E(\xi_{ij}^2)=O(b)$, $E(J_n) = O(nb)$. Thus, $\sum_{i=1}^{n}|\eta_i^2+2\xi_i\eta_i|=O_p(nb\ell_n)$. By (24) and the independence of $\xi_1, \ldots, \xi_n$,

$$S_n=\sum_{i=1}^{n}E(\xi_i^2)+O_p(\chi_n), \chi_n=\left\{\sum_{i=1}^{n}\text{var}(\xi_i^2)\right\}^{1/2}+nb\ell_n.$$

From the proof of Theorem 1, $\{nb^2f^2(x)\}^{-1}\sum_{i=1}^{n}E(\xi_i^2) \to s_{\text{dense}}^2(x)$ for dense data or $\{nbf^2(x)\}^{-1}\sum_{i=1}^{n}E(\xi_i^2) \to s_{\text{sparse}}^2(x)$ for sparse data. By (22), $H_n = \{1 + o_p(1)\}nbf(x)$. Thus, it remains to show $\chi_n = o(nb^2)$ for dense data or $\chi_n = o(nb)$ for sparse data. In the dense case, by the proof of Theorem 1, $\sum_{i=1}^{n}\text{var}(\xi_i^2) \le \sum_{i=1}^{n}E(\xi_i^4)=O(nb^4)$, and consequently $\chi_n = O\{\sqrt{}nb^2 + nb^3 + b\sqrt{}(n \log n)\} = o(nb^2)$. In the sparse case, by the proof of the dense case in Theorem 1, $E(\xi_i^4|n_i)=O(1)n_i^{-4}(n_i^4b^4+n_i^3b^3+n_i^2b^2+n_ib), E(\xi_i^4)=E\{E(\xi_i^4|n_i)\}=O(b)$, and thus $\chi_n = O\{\sqrt{}(nb) + nb^3 + \sqrt{}(nb \log n)\} = o(nb)$.

Proof of Theorem 3. Recall $s_{\text{sparse}}(x)$ and $s_{\text{dense}}(x)$ in Theorem 1. Let $\Gamma_n = nbf(x)/\Lambda_n$, $\Lambda_n = \sqrt{}(nb)f(x)s_{\text{sparse}}(x)$ for sparse data or $\Lambda_n = b\sqrt{}nf(x)s_{\text{dense}}(x)$ for dense data. Suppose we can show the weak convergence

$$\{\Gamma_n t\{\hat{\mu}_{\lfloor nt\rfloor}(x) - \mu(x) - b^2\rho(x)\}\}_{c\le t\le 1} \to \{B_t\}_{c\le t\le 1}. \quad (25)$$

For convenience, we write $\mathscr{L}_2(g)=\{\int_c^1|g(t)|^2dt\}^{1/2}$ and suppress the argument x. By (25) and the continuous mapping theorem, $(\hat{\mu}_n - \mu - b^2\rho)\mathscr{L}_2\{t(\hat{\mu}_{\lfloor nt\rfloor} - \hat{\mu}_n)\} \to B_1/\mathscr{L}_2(B_t - tB_1)$. By $|n^{-1}\lfloor nt\rfloor - t| \le n^{-1}$ for $t \in [c, 1]$, $\mathscr{L}_2\{t(\hat{\mu}_{\lfloor nt\rfloor} - \hat{\mu}_n)\}$ is asymptotically equivalent to $\mathscr{L}_2\{n^{-1}\lfloor nt\rfloor(\hat{\mu}_{\lfloor nt\rfloor} - \hat{\mu}_n)\} = V_n(x)$, where $V_n(x)$ is defined in $T_n(x)$. This completes the proof.

It remains to show (25). Recall ν_i and ζ_i in (20)–(21). As in (3) and (5),

$$\hat{\mu}_{\lfloor nt\rfloor}(x) - \mu(x) - \frac{1}{H_{\lfloor nt\rfloor}}\sum_{i=1}^{\lfloor nt\rfloor}\zeta_i = \frac{W_n(t)}{H_{\lfloor nt\rfloor}}, H_{\lfloor nt\rfloor} = \sum_{i=1}^{\lfloor nt\rfloor}\nu_i, W_n(t) = \sum_{i=1}^{\lfloor nt\rfloor}\xi_i.$$

By Kolmogorov's maximal inequality for independent random variables,

$$\sup_{c\le t\le 1}|H_{\lfloor nt\rfloor} - E(H_{\lfloor nt\rfloor})| = \max_{\lfloor cn\rfloor\le m\le n}|H_m - E(H_m)| = O_p\left[\left\{\sum_{i=1}^{n}\mathrm{var}(\nu_i)\right\}^{1/2}\right].$$

Thus, similar to (22), $H_{\lfloor nt\rfloor} = [1 + O_p\{b^2 + (nb/\tau_n)^{-1/2}\}]\lfloor nt\rfloor bf(x)$ uniformly in $c \le t \le 1$.

Applying the same argument to (23) gives $\sum_{i=1}^{\lfloor nt\rfloor}\zeta_i/H_{\lfloor nt\rfloor} = b^2\rho(x) + \delta_n$ uniformly, where δ_n is defined in (23). Thus it suffices to show $\{W_n(t)/\Lambda_n\}_{c\le t\le 1} \to \{B_t\}_{c\le t\le 1}$. The finite-dimensional convergence follows from the same argument in Theorem 1 and the Cramér–Wold device. It remains to prove the tightness. Let $c \le t < t' \le 1$. By independence,

$$\Delta_n(t,t') = E\left\{\frac{W_n(t)}{\Lambda_n} - \frac{W_n(t')}{\Lambda_n}\right\}^4 = \frac{1}{\Lambda_n^4}\left\{\sum_{i=\lfloor nt\rfloor+1}^{\lfloor nt'\rfloor}E(\xi_i^1) + 6\sum_{\lfloor nt\rfloor+1\le i<k}^{\lfloor nt'\rfloor}E(\xi_i^2)E(\xi_k^2)\right\}.$$

By the argument in the proof of Theorem 1, in the dense case, $E(\xi_i^2) = O(b^2), E(\xi_i^4) = O(b^4)$, and thus $\Delta_n(t, t') = O\{|t - t'|/n + |t - t'|^2\}$; in the sparse case, $E(\xi_i^4) = O(b), E(\xi_i^2) = O(b)$, and thus $\Delta_n(t, t') = O\{|t - t'|/(nb) + |t - t'|^2\}$. This proves the tightness.

References

De la Peña, VH.; Lai, TL.; Shao, QM. Self-Normalized Processes. New York: Springer; 2009.

Fitzmaurice, GM.; Laird, NM.; Ware, JM. Applied Longitudinal Analysis. New Jersey: Wiley; 2004.

Hall P, Müller HG, Wang JL. Properties of principal component methods for functional and longitudinal data analysis. Ann. Statist. 2006; 34:1493–1517.

Kiefer NM, Vogelsang TJ. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Economet. Theory. 2005; 21:1130–1164.

Li Y, Hsing T. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Ann. Statist. 2010; 38:3321–3351.

Lobato IN. Testing that a dependent process is uncorrelated. J. Am. Statist. Assoc. 2001; 96:1066–1076.

Ma S, Yang L, Carroll RJ. A simultaneous confidence band for sparse longitudinal regression. Statist. Sinica. 2012; 22:95–122.

Müller HG. Functional modeling and classification of longitudinal data. Scand. J. Statist. 2005; 32:223–240.

Ramsay, JO.; Silverman, BW. Functional Data Analysis. New York: Springer; 2005.

Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. J. R. Statist. Soc. 1991; B 53:233–243.

Shao X. A self-normalized approach to confidence interval construction in time series. J. R. Statist. Soc. 2010; B 72:343–366.

Wu H, Zhang JT. Local polynomial mixed-effects for longitudinal data. J. Am. Statist. Assoc. 2002; 97:883–897.

Wu, H.; Zhang, JT. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. New Jersey: Wiley; 2006.

Yao F. Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. J. Mult. Anal. 2007; 98:40–56.

Yao F, Müller HG, Wang JL. Functional linear regression analysis for longitudinal data. Ann. Statist. 2005a; 33:2873–2903.

Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. J. Am. Statist. Assoc. 2005b; 100:577–590.

Zhang JT, Chen J. Statistical inferences for functional data. Ann. Statist. 2007; 35:1052–1079.

Table 1

Average empirical coverage percentages and lengths, in brackets, of six confidence intervals.

$1-\alpha$	σ	N	SN1	SN2	NS	ND	NSD	BS
90%	1	N_1	88·1 (0·381)	88·6 (0·386)	82·8 (0·331)	66·5 (0·236)	89·2 (0·389)	87·4 (0·377)
		N_2	88·9 (0·288)	89·2 (0·290)	68·0 (0·178)	81·0 (0·236)	89·4 (0·292)	88·1 (0·284)
		N_3	89·8 (0·262)	89·8 (0·263)	57·8 (0·126)	86·3 (0·236)	90·3 (0·265)	89·1 (0·258)
		N_4	88·4 (0·246)	88·5 (0·247)	37·3 (0·076)	86·9 (0·236)	88·6 (0·247)	87·5 (0·242)
	2	N_1	88·8 (0·528)	89·3 (0·534)	86·5 (0·497)	51·7 (0·236)	89·4 (0·537)	87·8 (0·523)
		N_2	88·6 (0·330)	88·7 (0·332)	75·8 (0·243)	74·3 (0·236)	89·3 (0·335)	87·9 (0·326)
		N_3	89·5 (0·293)	89·4 (0·294)	69·5 (0·183)	81·1 (0·236)	90·1 (0·297)	88·8 (0·289)
		N_4	88·4 (0·257)	88·6 (0·257)	48·6 (0·106)	85·1 (0·236)	88·7 (0·258)	87·6 (0·252)
95%	1	N_1	93·6 (0·454)	94·0 (0·460)	89·7 (0·394)	75·2 (0·281)	94·6 (0·464)	92·9 (0·446)
		N_2	94·1 (0·343)	94·3 (0·345)	76·4 (0·212)	88·1 (0·281)	94·7 (0·347)	93·4 (0·335)
		N_3	95·0 (0·312)	95·1 (0·314)	66·0 (0·150)	92·1 (0·281)	95·3 (0·316)	94·0 (0·305)
		N_4	94·2 (0·293)	94·3 (0·294)	43·7 (0·090)	92·9 (0·281)	94·3 (0·294)	93·1 (0·286)
	2	N_1	94·2 (0·629)	94·4 (0·636)	92·6 (0·592)	59·7 (0·281)	94·8 (0·640)	93·2 (0·618)
		N_2	93·9 (0·393)	94·0 (0·395)	83·6 (0·289)	82·3 (0·281)	94·2 (0·399)	93·0 (0·385)
		N_3	94·7 (0·349)	94·8 (0·351)	77·9 (0·219)	88·0 (0·281)	95·1 (0·354)	93·8 (0·341)
		N_4	94·1 (0·306)	94·1 (0·307)	56·4 (0·127)	91·6 (0·281)	94·2 (0·308)	93·0 (0·298)

SN1 and SN2: the self-normalized confidence intervals in (14) and (16) with 200 permutations, respectively; NS and ND: the asymptotic normality based confidence intervals (12)–(13) assuming sparse and dense data, respectively; NSD: the infeasible confidence interval in (19); BS: bootstrap confidence interval; N_1–N_4: the numbers of measurements on individual subjects in (17)–(18).