

MONOTONE NONPARAMETRIC REGRESSION FOR FUNCTIONAL/LONGITUDINAL DATA

Ziqi Chen^{1,2}, Qibing Gao³, Bo Fu⁴ and Hongtu Zhu^{1,5}

¹*The University of Texas MD Anderson Cancer Center,*

²*Central South University,* ³*Nanjing Normal University,* ⁴*Fudan University*
and ⁵*The University of North Carolina at Chapel Hill*

Abstract: Motivated by quantifying the monotonic relationship between gray matter (GM) volume and age in the older population, this study proposes a constrained nonparametric estimation and statistical inference for the monotone mean function of functional/longitudinal data. Under some mild conditions, we systematically investigate the asymptotic properties of the proposed estimators, using a general weighting scheme that includes an equal weight per observation (OBS) and an equal weight per subject (SUBJ) as special cases. Most existing methods without a structural constraint can handle sparse or dense data only. Thus, a subjective choice between the two types may lead to erroneous conclusions from statistical inferences. Our proposed method and theories adapt to sparse and dense cases on a unified platform under a monotonic constraint. The asymptotic results enable us to categorize functional/longitudinal data into three data types (i.e., sparse, dense, and ultra-dense), based on the relative order of the number of repeated measurements relative to the total number of subjects. Simulation studies are conducted to examine the finite-sample performance of the estimating and statistical inference procedures. Our analysis of GM volume data, obtained from the Alzheimer’s Disease Neuroimaging Initiative study, confirms the accuracy and rationality of the constrained estimators in characterizing cerebellar GM volume with increasing age.

Key words and phrases: Asymptotic normality, isotone regression, kernel smoothing, monotonicity constraint, nonparametric estimation, sparse and dense functional/longitudinal data, weighting schemes.

1. Introduction

This study is motivated by an analysis of structural brain magnetic resonance imaging (MRI) data extracted from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). The ADNI is an ongoing public–private partnership that tests whether genetic, structural, and functional neuroimaging and clinical data can be combined to measure the progression of mild cognitive impairment and early Alzheimer’s disease. Subjects in the ADNI have been recruited from over 50

sites across the United States and Canada. Our problem of interest is to study the effect of aging on the progression of Alzheimer's disease. In general, aging can be referred to as a progressive deterioration of physiological function, leading to impairments in cognitive function and the ability to execute and learn new movements. Recent methodological advances in MRI allow us to characterize the structural changes that accompany the aging of a healthy brain, such as changes in the volume of gray matter (GM). In addition to devastating cognitive impairment, disorders of degenerative dementia such as Alzheimer's disease are characterized by accelerating cerebral atrophy. MRI is often used to differentiate normal aging from the neurodegeneration evident in early Alzheimer's disease. These results all show that cerebellar GM volume decreases with increasing age in elderly people (Henkenius et al. (2003); Hoogendam et al. (2012)), suggesting a monotonic relationship between GM volume and age.

Functional/Longitudinal data analyses are applied widely in the biomedical, psychometric, and environmental sciences (Fitzmaurice, Laird and Ware (2004); Yao, Muller and Wang (2005); Wu and Zhang (2006); Wang, Chiou and Muller (2016); Zhu et al. (2018)). In this type of analysis, subjects are measured repeatedly over time, and measurements from the same subject are usually highly correlated. Let n_i be the number of repeated measurements for subject i , and n be the total number of subjects. The observations from each subject are assumed to be noisy, discrete realizations of an underlying process $\{X(\cdot)\}$, and are given by

$$y_{ij} = X_i(s_{ij}) + \sigma(s_{ij})\varepsilon_{ij} \quad \text{for } j = 1, \dots, n_i; i = 1, \dots, n, \quad (1.1)$$

where y_{ij} is the response variable of interest for subject i , measured at time s_{ij} , $X_i(\cdot)$ denotes an independent realization of the underlying process $\{X(\cdot)\}$, and ε_{ij} is a random error with mean zero and variance one. Using a mixed effects approach, we decompose $X_i(s_{ij})$ into an unknown population mean $m(\cdot) = E\{X_i(\cdot)\}$ and a subject-specific trajectory $\eta_i(\cdot)$, with mean zero and covariance function $\gamma(s, t) = \text{cov}\{\eta_i(s), \eta_i(t)\}$. Then, we can rewrite (1.1) as

$$y_{ij} = m(s_{ij}) + \eta_i(s_{ij}) + \sigma(s_{ij})\varepsilon_{ij} \quad \text{for } j = 1, \dots, n_i; i = 1, \dots, n. \quad (1.2)$$

Throughout the paper, the density $f(s)$ of time points $\{s_{ij}\}$ is defined on $[0, 1]$.

Estimating the mean function $m(\cdot)$ is an important research topic in functional/longitudinal data analysis. Almost all existing methods estimate a non-parametric regression function without a structural constraint; see Li and Hsing (2010), Kim and Zhao (2012), Zhang and Wang (2016), and the references therein. For instance, local linear and polynomial methods are the most

popular for an unconstrained nonparametric estimation of $m(\cdot)$ in the functional/longitudinal data framework (Kim and Zhao (2012); Zhang and Wang (2016)). However, the mean response function can be a monotonic function of s in some cases. For example, as shown in Section 5, GM volume decreases after age 40 (Henkenius et al. (2003)). The hippocampus volume decreases rapidly for those suffering from Alzheimer's disease (Dawson and Muller (2018)). To the best of our knowledge, few studies have attempted to estimate the monotone mean function for functional/longitudinal data.

However, there is a large amount of literature on estimating monotonic regression functions for cross-sectional data. Please see Gijbels (2005), and the references therein. For instance, Brunk (1955) proposed a modified maximum likelihood estimator of $m(s)$, although the estimator may not be smooth. Mukerjee (1988) and Mammen (1991) proposed a smooth monotonic estimator of $m(s)$, and constrained smoothing spline methods have been proposed for estimating $m(s)$ (Ramsay (1988, 1998); Kelly and Rice (1990); Mammen and Thomas-Agnan (1999); Mammen et al. (2001)). Hall and Huang (2001) proposed monotone-increasing general kernel-type estimators by tilting the empirical distribution. Dette, Neumeyer and Pilz (2006) proposed a nonparametric estimate of the inverse of a monotonic regression function, denoted as $m^{-1}(\cdot)$, and then calculated its numerical inversion.

The aim of this study is to develop a constrained nonparametric estimate of $m(\cdot)$ under model (1.2) and three different types of $\{n_i : i = 1, \dots, n\}$ relative to n . Compared with the existing literature (Zhang and Wang (2016); Gijbels (2005)), we make several important contributions. First, we construct a constrained nonparametric estimator of monotone $m(\cdot)$, denoted as $\widehat{m}_I(\cdot)$, for functional/longitudinal data, based on local kernel methods. Moreover, under a monotonicity constraint, we construct an asymptotic pointwise $1 - \alpha$ confidence interval for the monotone mean function, without estimating the functions $\gamma(s, s)$ and $\sigma^2(s)$. Second, we establish a unified theory of $\widehat{m}_I(\cdot)$ for all three relative orders of $\{n_i\}$ to n under a general weighting scheme. This theory allows us to define three types of functional/longitudinal data: sparse data, dense data, and ultra-dense data. The type depends on whether $\widehat{m}_I(\cdot)$ can achieve the root- n convergence rate and have a negligible asymptotic bias. Our estimation and confidence interval construction methods do not need to distinguish between sparse and dense data; that is, we allow the magnitude of n_i to vary freely, relative to the sample size n . In contrast, most existing methods without a structural

constraint can handle individual sampling design scenarios only (Yao, Muller and Wang (2005); Hall, Muller and Wang (2006); Zhang and Chen (2007)). Third, we consider two commonly used weighting schemes for unconstrained mean function estimations, introduced by Zhang and Wang (2016), and compare them both theoretically and numerically in terms of estimation efficiency under the monotonic constraint. Finally, we have developed companion software, called `monfuncreg`, which is available at <https://github.com/BIG-S2/monfuncreg>.

The rest of this paper is organized as follows. We propose the nonparametric estimating procedure for the monotone regression function $m(\cdot)$ in Section 2.1. Section 2.2 shows the asymptotic properties of the estimators discussed in Section 2.1. Section 2.3 presents an adaptive confidence interval for the constrained nonparametric mean function. Section 3 describes our simulation studies, and Section 4 conducts a real ADNI data analysis to show that the proposed nonparametric estimators perform well and reasonably. Section 5 concludes the paper. All assumptions are provided in the Appendix. All lemmas and detailed proofs are deferred to the Supplementary Material.

2. Estimation Procedure and Theory

2.1. Monotone mean function estimation

We construct a constrained nonparametric estimator of $m(\cdot)$ in model (1.2), as follows. Without loss of generality, we consider the case of isotonic (strictly increasing) regression functions only. Let $\partial_u = d/du$, $K_d(\cdot)$ and $K_r(\cdot)$ be kernel functions, h_d and h_r be bandwidths, and $K_{a,h}(v) = h^{-1}K_a(v/h)$ be the rescaled kernel function with bandwidth h , for $a = d$ and r .

We start by reviewing several key ideas from Dette, Neumeyer and Pilz (2006). Consider an independent and identically distributed (i.i.d.) sample of N uniform random variables, say $U_1, \dots, U_N \sim U[0, 1]$. If $m(\cdot)$ is a strictly increasing function on $[0, 1]$ with a positive derivative, then a density estimator of $m(U)$ for $U \sim U[0, 1]$ is $\sum_{i=1}^N K_{d,h_d}(m(U_i) - u)/N$, which is also the estimator of $\partial_u(m^{-1})(u)\mathbf{1}(u \in [m(0), m(1)])$, where $\mathbf{1}(A)$ is an indicator function of event A . Thus, as $h_d \rightarrow 0$ and $Nh_d \rightarrow \infty$, a consistent estimate of $m^{-1}(t)$ is

$$N^{-1} \int_{-\infty}^t \sum_{i=1}^N K_{d,h_d}(m(U_i) - u) du, \quad \text{for any point } t \in (m(0), m(1)). \quad (2.1)$$

Moreover, estimator (2.1) is a strictly increasing function, almost surely, when N is sufficiently large (Dette, Neumeyer and Pilz (2006)).

To obtain an estimator of $m^{-1}(t)$ in (2.1), we need an unconstrained estimator of $m(t)$, that is, $\widehat{m}(s) = \widehat{\beta}_0$, where

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \omega_i \sum_{j=1}^{n_i} \{Y_{ij} - \beta_0 - \beta_1(s_{ij} - s)\}^2 K_{r, h_r}(s_{ij} - s),$$

and ω_i is a weight satisfying $\sum_{i=1}^n n_i \omega_i = 1$. We consider two commonly used weighting schemes: equal weight per observation (OBS) and equal weight per subject (SUBJ) (Yao, Muller and Wang (2005); Li and Hsing (2010); Kim and Zhao (2012); Zhang and Wang (2016)). Specifically, we set $\omega_i = 1/(\sum_{i=1}^n n_i)$ for OBS, whereas we set $\omega_i = 1/(nn_i)$ for SUBJ. Moreover, $\widehat{m}(s)$ is a local linear estimator of $m(\cdot)$ (Li and Hsing (2010); Zhang and Wang (2016)). By substituting $\widehat{m}(s)$ into (2.1), we obtain

$$\widehat{m}_I^{-1}(t) = N^{-1} \int_{-\infty}^t \sum_{i=1}^N K_{d, h_d} \left(\widehat{m} \left(\frac{i}{N} \right) - u \right) du.$$

Our constrained estimator of $m(s)$, denoted as $\widehat{m}_I(s)$, is then calculated using a numerical inversion.

2.2. Theoretical results

In this subsection, we systematically investigate the asymptotic properties of $\widehat{m}_I(s)$. $K_d(v)$ and $K_r(v)$ are symmetric kernels, with compact support $[-1, 1]$, and are twice continuously differentiable on $[-1, 1]$. For a specific kernel function K , we define $\kappa_2(K) = (1/2) \int_{-1}^1 u^2 K(u) du$. Let \xrightarrow{d} denote the convergence in distribution as $n \rightarrow \infty$. We also define $\dot{m}(s) = dm(s)/ds$ and $\ddot{m}(s) = d^2m(s)/ds^2$. Following the arguments in Zhang and Wang (2016), we consider three types of data, which vary according to the number of repeated measurements $\{n_i\}$:

- (i) Sparse data: $\bar{n}/n^{1/4} \rightarrow 0$, with $\bar{n} = \sum_{i=1}^n n_i/n$;
- (ii) Dense data: $\bar{n}/n^{1/4} \rightarrow C$, with $0 < C < \infty$;
- (iii) Ultra-dense data: $\bar{n}/n^{1/4} \rightarrow \infty$.

To establish the asymptotic properties of general weighted estimators, we treat n_i and ω_i as fixed quantities, while allowing them to vary with n . When n_i is random, the theory can be regarded as conditional on the value of n_i .

Define

$$\Gamma_0^A(t) := \left\{ \int K_r^2(u) du \right\} \frac{\gamma(m^{-1}(t), m^{-1}(t)) + \sigma^2(m^{-1}(t))}{\{\dot{m}(m^{-1}(t))\}^2 f(m^{-1}(t))}$$

$$\text{and } \Gamma_0^B(t) := \frac{\gamma(m^{-1}(t), m^{-1}(t))}{\{\dot{m}(m^{-1}(t))\}^2}.$$

The assumptions for the following theorem are provided in the Appendix. The theorem is obtained by Lemmas 1 and 5, given in the Supplementary Material.

Theorem 1. *Suppose that Assumptions (A)–(C) in the Appendix hold. In addition, assume*

- (i) $\min[h_r/(\sum_{j=1}^n \omega_j^2 n_j), 1/\{\sum_{j=1}^n \omega_j^2 n_j(n_j - 1)\}]h_r^6 \rightarrow 0$;
- (ii) $h_r \sum_{j=1}^n \omega_j^2 n_j(n_j - 1)/\sum_{j=1}^n \omega_j^2 n_j \rightarrow C_0 \in [0, \infty]$;
- (iii) $Nh_d h_r^2 \rightarrow \infty$;
- (v) $m(s)$ is strictly increasing.

Then, for a fixed interior point $t \in (m(0), m(1))$ satisfying $\dot{m}(m^{-1}(t)) > 0$, we have

$$\Gamma(t)^{-1/2} \left\{ \hat{m}_I^{-1}(t) - m^{-1}(t) + h_r^2 \kappa_2(K_r) \left(\frac{\ddot{m}}{\dot{m}} \right) (m^{-1}(t)) \right\} \xrightarrow{d} N(0, 1), \quad (2.2)$$

where $\Gamma(t)$ is equal to the sum of $\Gamma^A(t) = \Gamma_0^A(t)(\sum_{j=1}^n \omega_j^2 n_j)/h_r$ and $\Gamma^B(t) = \Gamma_0^B(t) \sum_{j=1}^n \omega_j^2 n_j(n_j - 1)$.

Theorem 1 can be regarded as a generalization of Theorem 3.1 in Dette, Neumeyer and Pilz (2006). Specifically, the asymptotic bias is the same as that for cross-sectional studies (Dette, Neumeyer and Pilz (2006)), whereas the variance term Γ is more complex. Specifically, Γ^A characterizes the variances of all observations, and Γ^B comes mainly from the correlations between repeated measures across all subjects. Corollary 1 follows directly from Theorem 1.

Corollary 1. *Suppose that the assumptions of Theorem 1 hold. Let t be a fixed interior point in $(m(0), m(1))$ satisfying $\dot{m}(m^{-1}(t)) > 0$.*

(a) *OBS: If $\min\{n\bar{n}h_r, n(\bar{n})^2/(\bar{n}_{S_2} - \bar{n})\}h_r^6 \rightarrow 0$ and $h_r(\bar{n}_{S_2} - \bar{n})/\bar{n} \rightarrow C_0 \in [0, \infty]$, where $\bar{n}_{S_2} = \sum_{i=1}^n n_i^2/n$, then the asymptotic normality (2.2) holds with $\Gamma(t)$ for OBS, denoted as $\Gamma_{obs}(t)$, equal to the sum of $\Gamma_{obs}^A(t)$ and $\Gamma_{obs}^B(t)$, given by*

$$\Gamma_{obs}^A(t) = \frac{\Gamma_0^A(t)}{(n\bar{n}h_r)} \quad \text{and} \quad \Gamma_{obs}^B(t) = \frac{(\bar{n}_{S_2} - \bar{n})}{n\bar{n}^2} \Gamma_0^B(t).$$

(b) *SUBJ: If $\min\{n\bar{n}_H h_r, n/(1 - 1/\bar{n}_H)\}h_r^6 \rightarrow 0$ and $h_r(\bar{n}_H - 1) \rightarrow C_0 \in [0, \infty]$, where $\bar{n}_H = (n^{-1} \sum_{i=1}^n n_i^{-1})^{-1}$, then the asymptotic normality (2.2) holds with $\Gamma(t)$ for SUBJ, denoted as $\Gamma_{subj}(t)$, equal to the sum of $\Gamma_{subj}^A(t)$ and $\Gamma_{subj}^B(t)$,*

given by

$$\Gamma_{subj}^A(t) = \frac{\Gamma_0^A(t)}{(n\bar{n}_H h_r)} \quad \text{and} \quad \Gamma_{subj}^B(t) = n^{-1}(1 - \bar{n}_H^{-1})\Gamma_0^B(t).$$

Three types of asymptotic normality emerge for the two schemes from Corollary 1, depending on the order of \bar{n} and \bar{n}_H relative to that of n .

Corollary 2. *Suppose that the assumptions of Theorem 1 hold.*

(a) *OBS: Assume $\limsup_n \bar{n}_{S_2}/\bar{n}^2 < \infty$.*

Case 1 (Sparse data) When $\bar{n}/n^{1/4} \rightarrow 0$ and $h_r \asymp (n\bar{n})^{-1/5}$, we have

$$\sqrt{n\bar{n}h_r} \left\{ \widehat{m}_{I(OBS)}^{-1}(t) - m^{-1}(t) + h_r^2 \kappa_2(K_r) \left(\frac{\ddot{m}}{\dot{m}} \right) (m^{-1}(t)) \right\} \rightarrow_d N(0, \Gamma_0^A(t)).$$

Case 2 (Dense data) When $\bar{n}/n^{1/4} \rightarrow C$ and $h_r \bar{n}_{S_2}/\bar{n} \rightarrow C_1$, for $0 < C, C_1 < \infty$, we have

$$\begin{aligned} & \sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} \left\{ \widehat{m}_{I(OBS)}^{-1}(t) - m^{-1}(t) + h_r^2 \kappa_2(K_r) \left(\frac{\ddot{m}}{\dot{m}} \right) (m^{-1}(t)) \right\} \\ & \rightarrow_d N\left(0, \frac{\Gamma_0^A(t)}{C_1} + \Gamma_0^B(t)\right). \end{aligned}$$

Case 3 (Ultra-dense data) When $\bar{n}/n^{1/4} \rightarrow \infty, h_r n^{1/4} \rightarrow 0$, and $h_r \bar{n} \rightarrow \infty$, we have

$$\sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} \left(\widehat{m}_{I(OBS)}^{-1}(t) - m^{-1}(t) \right) \rightarrow_d N\left(0, \Gamma_0^B(t)\right).$$

(b) *SUBJ: We can obtain similar asymptotic normality results corresponding to sparse, dense, and ultra-dense data for $\widehat{m}_{I(SUBJ)}^{-1}(t)$ by replacing $\bar{n}, \bar{n}_{S_2}/\bar{n}$, and \bar{n}^2/\bar{n}_{S_2} in (a) with \bar{n}_H, \bar{n}_H , and 1, respectively.*

Corollary 1 indicates that $\widehat{m}_{I(OBS)}^{-1}(t)$ and $\widehat{m}_{I(SUBJ)}^{-1}(t)$ have an identical asymptotic bias. However, their asymptotic variances are different.

We observed from Corollary 2 that, for sparse data, Γ_{obs}^A and Γ_{subj}^A dominate Γ_{obs}^B and Γ_{subj}^B , respectively, and we obtain $\Gamma_{obs}^A \leq \Gamma_{subj}^A$ using arguments similar to Corollary 3.3 in Zhang and Wang (2016). Thus, the OBS scheme achieves a more efficient estimator of $m^{-1}(t)$ than the SUBJ scheme does. Intuitively, for sparse data, the bandwidth satisfies $\bar{n}h_r \rightarrow 0$ or $\bar{n}_H h_r \rightarrow 0$. A special and simple case is $n_i h_r \rightarrow 0$, for $i = 1, \dots, n$, where each subject contributes only one observation for estimating $m^{-1}(t)$. That is, for a given t , the data for estimating $m^{-1}(t)$ are i.i.d. Therefore, the OBS scheme, which assigns the same weight to each observation, yields a more efficient estimator.

In contrast, for ultra-dense data, Γ_{obs}^B and Γ_{subj}^B dominate Γ_{obs}^A and Γ_{subj}^A , respectively, and $\Gamma_{obs}^B \geq \Gamma_{subj}^B$. Thus, the SUBJ scheme is preferable to the OBS scheme. The conclusions here are consistent with those of Zhang and Wang (2016) for unconstrained nonparametric estimates of $m(\cdot)$ in model (1.2). Intuitively, for a given t , there exist subjects who contribute infinitely many observations to the estimation of $m^{-1}(t)$, because $\bar{n}h_r \rightarrow \infty$ or $\bar{n}_H h_r \rightarrow \infty$. The observations from the same subject are correlated; thus, the covariances within one subject tend to dominate the variances, and could have an undue influence on the variance of $\hat{m}^{-1}(t)$. The SUBJ scheme avoids this situation by assigning the weight $1/n_i$ to subject i and, thus, yields a more efficient estimator than the OBS scheme does. Essentially, for ultra-dense data, the SUBJ scheme is equivalent to the so-called “smooth-first-then-estimate” approach (Hall, Muller and Wang (2006); Zhang and Chen (2007)), which first preprocesses the discrete functional data for subject i ($i = 1, \dots, n$) using smoothing, and then adopts a sample mean of the smoothed functional data.

Remark 1. The asymptotic normality for “sparse data” is consistent with Theorem 3.1 in Dette, Neumeyer and Pilz (2006), which is established for independent data. Moreover, the convergence rate of $\hat{m}_{I(OBS)}^{-1}(t)$ is $(n\bar{n})^{2/5}$ or $(n\bar{n}_H)^{2/5}$, and both are of the order $o_p(n^{1/2})$.

Remark 2. The convergence rate of $\hat{m}_{I(OBS)}^{-1}(t)$ for both “dense data” and “ultra-dense data” is $O_p(n^{1/2})$. Furthermore, “ultra-dense data” fall within the parametric paradigm, where the limiting normal distribution has a zero mean.

Remark 3. An explicit partition of functional/longitudinal data can be concluded from Corollary 2 based on the asymptotic properties of the proposed monotonicity-constraint nonparametric regression. Specifically, the functional/longitudinal data can be divided into three types: sparse data, dense data, and ultra-dense data. The type is based on the relative order of \bar{n} and \bar{n}_H to $n^{1/4}$ under the OBS or the SUBJ scheme, respectively. This partition is consistent with that of Zhang and Wang (2016).

The functions \hat{m}_I^{-1} and m_N^{-1} are strictly increasing, regardless of the monotonicity of the “true” regression function m , for sufficiently large n and N . The following theorem states that the corresponding inverse function \hat{m}_I of \hat{m}_I^{-1} also satisfies an asymptotic normal distribution. Its proof is similar to that of Theorem 3.2 in Dette, Neumeyer and Pilz (2006) and, hence, is omitted here.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold. For a fixed in-*

terior point $s \in (0, 1)$, with $\dot{m}(s) > 0$, we have

$$\Gamma_*(s)^{-1/2} \{ \widehat{m}_I(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \xrightarrow{d} N(0, 1),$$

where Γ_* is given by

$$\Gamma_*(s) = h_r^{-1} \sum_{j=1}^n (\omega_j^2 n_j) \left\{ \int K_r^2(u) du \right\} \frac{\{ \gamma(s, s) + \sigma^2(s) \}}{f(s)} + \sum_{j=1}^n \omega_j^2 n_j (n_j - 1) \gamma(s, s).$$

We can see that our proposed constrained estimator and the unconstrained estimator of Zhang and Wang (2016) have the same asymptotic distribution.

Similarly to Corollary 2, we can show the following results for $\widehat{m}_{I(OBS)}$ and $\widehat{m}_{I(SUBJ)}$.

Corollary 3. *Suppose that the assumptions of Theorem 1 hold and s is a fixed interior point in $(0, 1)$ satisfying $\dot{m}(s) > 0$.*

(a) *OBS: Assume $\limsup_n \bar{n}_{S_2} / \bar{n}^2 < \infty$.*

Case 1 (Sparse data) When $\bar{n} / n^{1/4} \rightarrow 0$ and $h_r \asymp (n\bar{n})^{-1/5}$, we have

$$\begin{aligned} & \sqrt{n\bar{n}h_r} \{ \widehat{m}_{I(OBS)}(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \\ & \rightarrow_d N \left(0, \left\{ \int K_r^2(u) du \right\} \frac{\{ \gamma(s, s) + \sigma^2(s) \}}{f(s)} \right). \end{aligned}$$

Case 2 (Dense data) When $\bar{n} / n^{1/4} \rightarrow C$ and $h_r \bar{n}_{S_2} / \bar{n} \rightarrow C_1$, where $0 < C, C_1 < \infty$, we have

$$\begin{aligned} & \sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} \{ \widehat{m}_{I(OBS)}(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \\ & \rightarrow_d N \left(0, \left\{ \int K_r^2(u) du \right\} \frac{\{ \gamma(s, s) + \sigma^2(s) \}}{\{ C_1 f(s) \}} + \gamma(s, s) \right). \end{aligned}$$

Case 3 (Ultra-dense data) When $\bar{n} / n^{1/4} \rightarrow \infty$, $h_r n^{1/4} \rightarrow 0$, and $h_r \bar{n} \rightarrow \infty$, we have

$$\sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} (\widehat{m}_{I(OBS)}(s) - m(s)) \rightarrow_d N(0, \gamma(s, s)).$$

(b) *SUBJ: We can obtain similar asymptotic normality results corresponding to sparse, dense, and ultra-dense data for $\widehat{m}_{I(SUBJ)}(s)$ by replacing \bar{n} , \bar{n}_{S_2} / \bar{n} , and $\bar{n}^2 / \bar{n}_{S_2}$ in (a) with \bar{n}_H , \bar{n}_H , and 1, respectively.*

Remark 4. An important implication of Corollary 3 is that the partition of functional/longitudinal data for estimating $m(\cdot)$ is the same as that discussed in Remark 3. Consistent with the results of $\widehat{m}_{I(OBS)}^{-1}$ for sparse data, OBS outperforms SUBJ by achieving a smaller asymptotic variance of $\widehat{m}_{I(OBS)}(s)$. For

ultra-dense data, however, SUBJ outperforms OBS.

Remark 5. Corollary 3 indicates that the asymptotic normality results corresponding to the sparse, dense, or ultra-dense data vary significantly under the monotonicity constraint. For instance, their corresponding asymptotic variances are different. To construct a confidence interval, we may need to estimate $\sigma^2(s)$ and $\gamma(s, s)$ for each of the three types. Next, we show how to construct an adaptive confidence interval for all three data types.

2.3. Adaptive confidence interval

In this subsection, we construct an adaptive confidence interval for a monotone mean function, that can be adapted to the three data types (Kim and Zhao (2012)). Recall that the asymptotic distribution of the constrained mean function estimator is the same as that of the unconstrained one. Specifically, for s a fixed interior point in $(0, 1)$ satisfying $\dot{m}(s) > 0$, we estimate the variance of $\hat{m}_I(s)$ using

$$U_n^2(s) = H_n^{-2}(s) \sum_{i=1}^n \left[\omega_i \sum_{j=1}^{n_i} \{y_{ij} - \hat{m}(s_{ij})\} K_{h_r}(s - s_{ij}) \right]^2,$$

where $H_n(s) = \sum_{i=1}^n \omega_i \sum_{j=1}^{n_i} K_{h_r}(s - s_{ij})$. We define

$$\bar{U}_n^2(s) = H_n^{-2}(s) \sum_{i=1}^n \left[\omega_i \sum_{j=1}^{n_i} \{y_{ij} - m(s_{ij})\} K_{h_r}(s - s_{ij}) \right]^2.$$

It follows from Lemma 3 in the Supplementary Material that $U_n^2(s) = \bar{U}_n^2(s)\{1 + o_p(1)\}$. After some calculation, we have $\bar{U}_n^2(s) = \Gamma_*(s)\{1 + o_p(1)\}$. Thus, it follows from Slutsky's theorem that we have

$$U_n(s)^{-1} \{ \hat{m}_I(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \xrightarrow{d} N(0, 1). \quad (2.3)$$

Therefore, (2.3) can be used to construct a unified asymptotic pointwise $1 - \alpha$ confidence interval for the monotone mean function $m(s)$, which can then be adapted for each type of data.

Remark 6. We can also use

$$U_n(s)^{-1} \{ \hat{m}(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \xrightarrow{d} N(0, 1) \quad (2.4)$$

to construct a good asymptotic pointwise $1 - \alpha$ confidence interval for the mean function $m(s)$ (Zhang and Wang (2016)) using the unconstrained estimator. However, $\hat{m}(s)$ does not satisfy the monotonic constraint. Thus, we show in

the simulation studies that the standard deviation of $\widehat{m}_I(s)$ may be smaller than that of $\widehat{m}(s)$ in the finite-sample performance. If true, the average coverage rate of the $1 - \alpha$ confidence interval based on (2.3) may be better than that based on (2.4) in terms of the finite-sample performance.

Remark 7. Define

$$U_{In}^2(s) = \frac{1}{H_n^2(s)} \sum_{i=1}^n \left[\omega_i \sum_{j=1}^{n_i} \{y_{ij} - \widehat{m}_I(s_{ij})\} K_{h_r}(s - s_{ij}) \right]^2.$$

We can also show that

$$U_{In}(s)^{-1} \{ \widehat{m}_I(s) - m(s) - h_r^2 \kappa_2(K_r) \ddot{m}(s) \} \xrightarrow{d} N(0, 1) \tag{2.5}$$

holds as $n \rightarrow \infty$. Similarly, we may use (2.5) to construct an asymptotic point-wise $1 - \alpha$ confidence interval for $m(s)$. Because the unconstrained estimator is constructed by minimizing the difference between the responses y_{ij} and the estimated mean values $\widehat{m}(s_{ij})$, $U_n^2(s)$ may be slightly smaller than $U_{In}^2(s)$. Thus, the average length of the confidence interval based on (2.3) may be slightly shorter than that based on (2.5), in which case, the confidence interval based on (2.3) may be a better one.

3. Simulation Studies

We carried out the following simulation studies to examine the finite-sample performance of the proposed estimation method. The data were simulated according to

$$y_{ij} = m(s_{ij}) + \sum_{k=1}^3 \alpha_{ik} \Phi(s_{ij}) + \sigma \varepsilon_{ij}, \quad \text{for } j = 1, \dots, n_i; \ i = 1, \dots, n,$$

where $m(s) = \sin(s)$, for $s \in [0, 1]$, $\alpha_{ik} \sim N(0, \omega_k)$, and ε_{ij} is i.i.d. Let $\Phi_1(s) = 1$, $\Phi_2(s) = \sqrt{2} \sin(2\pi s)$, $\Phi_3(s) = \sqrt{2} \cos(2\pi s)$, $(\omega_1, \omega_2, \omega_3) = (0.6, 0.3, 0.1)$, and $n = 150$. The design points s_{ij} were uniformly simulated on $[0, 1]$. Two distributions of ε_{ij} were considered: $N(0, 0.5)$, and a T distribution with three degrees of freedom. For each distribution, we considered the following three cases for the vector $\mathbf{n} = (n_1, \dots, n_n)^T$:

$$\begin{aligned} (\text{Case 1}) \ \mathbf{n}_1 : n_i &\sim U[\{1, 2, \dots, 5\}], & (\text{Case 2}) \ \mathbf{n}_2 : n_i &\sim U\left[\left\{\frac{n}{10}, \dots, \frac{n}{5}\right\}\right], \\ (\text{Case 3}) \ \mathbf{n}_3 : n_i &\sim U\left[\left\{\frac{n}{3}, \dots, \frac{2n}{3}\right\}\right], \end{aligned}$$

where $U[\mathcal{D}]$ denotes a discrete uniform distribution on a finite set \mathcal{D} . Here, \mathbf{n}_1 can be regarded as the case of sparse data, and \mathbf{n}_3 denotes cases of ultra-dense data. For each case, the simulation was repeated $Q = 500$ times. We used the commonly used Gaussian kernel function $K_r(u) = K_g(u) = \phi(x)$, where $\phi(x)$ is the standard normal density. Let $S = \{s_{ij}, j = 1, \dots, n_i; i = 1, \dots, n\}$. We followed Silverman's rule of thumb to choose the bandwidth parameters by setting $h_r = 1.06(\sum_{i=1}^n n_i)^{-1/5} \min\{\hat{\sigma}_S, (S_{[0.75]} - S_{[0.25]})/1.34\}$, where $\hat{\sigma}_S$ is the standard deviation of S , and $S_{[0.25]}$ and $S_{[0.75]}$ are the 25% and 75% sample quantiles of S , respectively. We set $h_d = h_r^3/4$ (Dette, Neumeier and Pilz (2006)). The N used to estimate $m^{-1}(t)$ is set to 500.

We compared our proposed estimator with the unconstrained estimator of (Zhang and Wang (2016)). Let $S_l = 0.04 + l \times 0.01$, for $l = 1, \dots, E = 91$. We calculated the bias and standard deviation (SD) at each of the 91 points $\{S_l\}$, based on 500 replications, to obtain the average bias and SD. Here, "SD" can be viewed as the true standard deviation of the resulting estimates and, thus, can be used to measure the efficiency of the methods. We define the empirical mean integrated squared error (EMISE) and empirical mean supremum absolute error (EMSAE) as follows:

$$\text{EMISE}(\hat{m}_I) = \sum_{q=1}^Q \sum_{l=1}^E \frac{\{\hat{m}_I^{(q)}(S_l) - m(S_l)\}^2}{(QE)},$$

$$\text{EMSAE}(\hat{m}_I) = \sum_{q=1}^Q \frac{\max_l |\hat{m}_I^{(q)}(S_l) - m(S_l)|}{Q},$$

where $\hat{m}_I^{(q)}(\cdot)$ is the estimator of $m(\cdot)$ using the q th data set, for $q = 1, \dots, Q$. We use the EMISE and EMSAE to measure the estimation accuracy and efficiency of all estimates of $m(\cdot)$. Furthermore, we built confidence intervals based on (2.3) and (2.4), and measured their accuracy by using the average empirical coverage probability (AECP) and average length (AL) of the confidence intervals. For each S_l and the given nominal level 95%, we built a confidence interval for $m(S_l)$ for each data set, and then calculated the empirical coverage probabilities and average lengths based on 500 replications. We then averaged the empirical coverage probabilities and lengths across all 91 points S_l .

Table 1 presents the simulation results. We have three important observations. First, our constrained estimator is more efficient and accurate than the unconstrained estimator. Second, the confidence intervals of our constrained estimator have the same width, but obviously better coverage probabilities com-

Table 1. The Bias, SD, EMISE, EMSAE, AECP and AL for the unconstrained estimators (UE) and the proposed estimators (PE) for either the OBS or the SUBJ scheme.

Scheme			N(0, 0.5)			T(3)		
			Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
UE	OBS	Bias	-0.005	-0.005	0.002	-0.003	-0.002	-0.001
		SD	0.133	0.092	0.086	0.196	0.114	0.095
		EMISE	0.018	0.009	0.007	0.039	0.013	0.009
		EMSAE	0.252	0.158	0.135	0.393	0.207	0.165
		AECP	93.54	94.35	94.76	93.19	93.93	94.73
		AL	0.499	0.360	0.337	0.712	0.427	0.366
	SUBJ	Bias	-0.003	-0.006	0.002	0.007	-0.002	-0.001
		SD	0.141	0.091	0.084	0.216	0.114	0.094
		EMISE	0.020	0.008	0.007	0.048	0.013	0.009
		EMSAE	0.279	0.156	0.133	0.448	0.208	0.165
		AECP	93.29	94.57	94.85	93.54	93.92	94.78
		AL	0.524	0.357	0.332	0.782	0.427	0.362
PE	OBS	Bias	-0.005	-0.004	0.002	-0.003	-0.002	0.000
		SD	0.119	0.089	0.085	0.156	0.104	0.091
		EMISE	0.014	0.008	0.007	0.025	0.011	0.008
		EMSAE	0.216	0.154	0.134	0.302	0.182	0.156
		AECP	96.17	95.05	95.01	97.12	95.79	95.58
		AL	0.499	0.360	0.337	0.712	0.427	0.366
	SUBJ	Bias	-0.003	-0.005	0.002	0.006	-0.002	0.001
		SD	0.124	0.088	0.083	0.164	0.103	0.090
		EMISE	0.015	0.007	0.006	0.029	0.011	0.008
		EMSAE	0.227	0.152	0.132	0.328	0.183	0.156
		AECP	96.44	95.23	95.05	97.40	95.80	95.69
		AL	0.524	0.357	0.332	0.782	0.427	0.362

pared with those of the unconstrained estimator. Third, compared with the SUBJ scheme, the OBS scheme produces more efficient and accurate estimators for sparse data and less efficient and accurate estimators for ultra-dense data, in general. These conclusions hold even when the normal distribution assumption of the random error ε is violated.

4. ADNI Real-Data Analysis

The structural brain MRI data and corresponding clinical and genetic data from the baseline and follow-up observations were downloaded from the publicly available ADNI database (<http://adni.loni.ucla.edu/>). The structural MRI data were collected across a variety of 1.5 Tesla MRI scanners, with protocols individualized for each scanner. The data include standard T1-weighted images

obtained using volumetric three-dimensional sagittal MPRAGE (or equivalent) protocols, with varying resolutions. The settings for the typical protocol were as follows: repetition time = 2,400 ms, inversion time = 1,000 ms, flip angle = 8° , and field of view = 24 cm, with a $256 \times 256 \times 170$ acquisition matrix in the x -, y -, and z -dimensions, yielding a voxel size of $1.25 \times 1.26 \times 1.2$ mm³. The MRI data were preprocessed by standard steps, including anterior commissure and posterior commissure corrections, skull-stripping, cerebellum removal, intensity inhomogeneity correction, segmentation, and registration (Shen and Davatzikos (2004)). Automatic regional labeling was then carried out by labeling the template, and then transferring the labels following the deformable registration of the subject images. We were able to compute volumes for each region of interest for each subject after labeling 93 regions of interest.

We wish to estimate the monotonic relationship between the mean of the GM volume and age. The ADNI data set considered here includes 562 subjects with longitudinal measurements of GM volumes. Among them, 39 subjects have one observation, 56 subjects have two observations, 117 subjects have three observations, 255 subjects have four observations, 88 subjects have five observations, and seven subjects have six observations. We scaled the age s_{ij} to $\{s_{ij} - \min_s\}/(\max_s - \min_s) \in [0, 1]$, where s_{ij} denotes the j th observation time point (age) for the i th subject, and \min_s and \max_s denote the minimum and the maximum values of $\{s_{ij}, j = 1, \dots, n_i; i = 1, \dots, 562\}$, respectively. We standardized the GM volume y_{ij} to $(y_{ij} - \bar{y})/s_y$, where y_{ij} denotes the GM volume for the i th subject, measured at the j th time point, and \bar{y} and s_y denote the sample mean and sample standard deviation for $\{y_{ij}, j = 1, \dots, n_i; i = 1, \dots, 562\}$, respectively. The longitudinal trajectories of the standard GM volumes at different scaled ages are shown in Figure 1 (A).

Figure 1 (A) and Figure 2 (A) and (B) show that GM volume tends to decrease as age increases, at the individual level, indicating that the proposed constrained estimation method may be a good choice for establishing such a monotonic relationship. In contrast, the local linear regression method (Zhang and Wang (2016)) cannot automatically impose a monotonic relationship between age and GM volume. Inspecting Figure 1 (A), Figure 2 (A) and (B), and Table 2 reveals that: (i) female subjects tend to have a lower GM volume than male subjects do; (ii) the number of female subjects is higher than that of male subjects, and the age of females corresponding to the first visiting time is younger than that of males before 63 years old; and (iii) the number of male subjects is higher than that of female subjects around age 63. This characteristic of the data

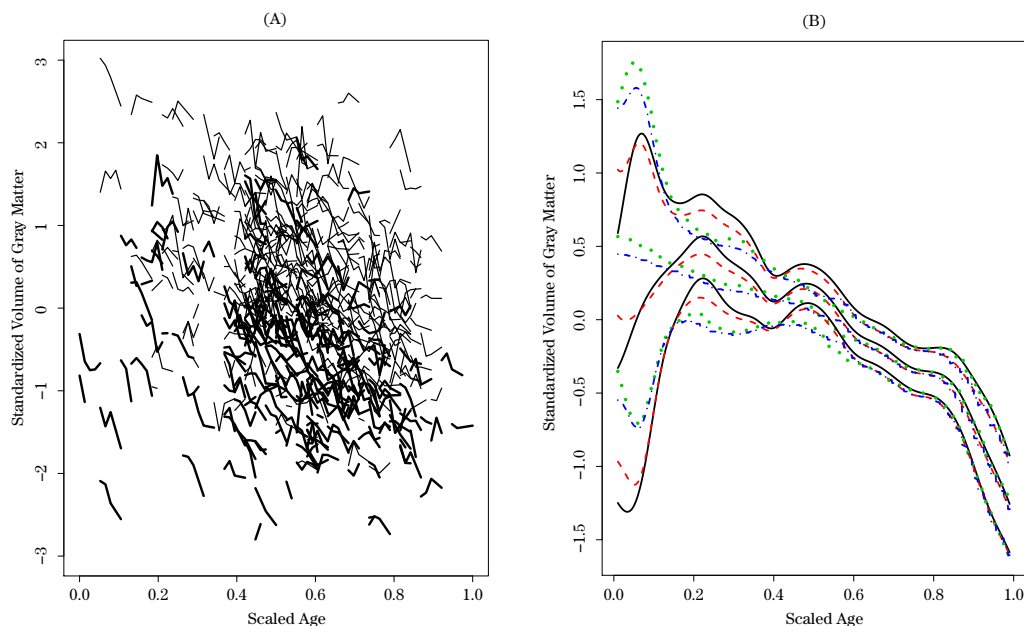


Figure 1. ADNI data analysis. Panel (A) is the standardized volume of gray matter versus scaled age, where thin denotes male and bold denotes female. Panel (B) shows the unconstrained estimators $\hat{m}_{(OBS)}(s)$ and the corresponding lower and upper bounds of the 95% confidence intervals (solid). Those for the unconstrained estimators $\hat{m}_{(SUBJ)}(s)$ and the constrained estimators $\hat{m}_{I(OBS)}(s)$ and $\hat{m}_{I(SUBJ)}(s)$ are dashed, dotted, and dot-dash, respectively.

Table 2. ADNI data analysis: Demographic information for subjects. “Number of Subj” means the number of subjects with the age of first visit in a certain range, “Average Visit” is the average number of visits for the subjects in a certain range.

Age range (years)	[55, 63]	(63, 73]	(73, 83]	(83, 93]
Number of Subj	32	174	281	75
Male/Female	14/18	102/72	162/119	50/25
Average Visit	3.6875	3.5862	3.6192	3.2667

creates an illusion that the GM volume increases before age 63, and decreases after age 63. Therefore, as shown in Figure 1 (B), the local linear method does not ensure the monotonicity of the GM volume curve with a turning age around 0.22, which corresponds to a true age of around 63.

The unconstrained estimator $\hat{m}(s)$ yields a monotonically decreasing curve based on male observations, as shown in Figure 2 (C), but it produces monotonically increasing estimated curves before 63 years old, as shown in Figure 2 (D),

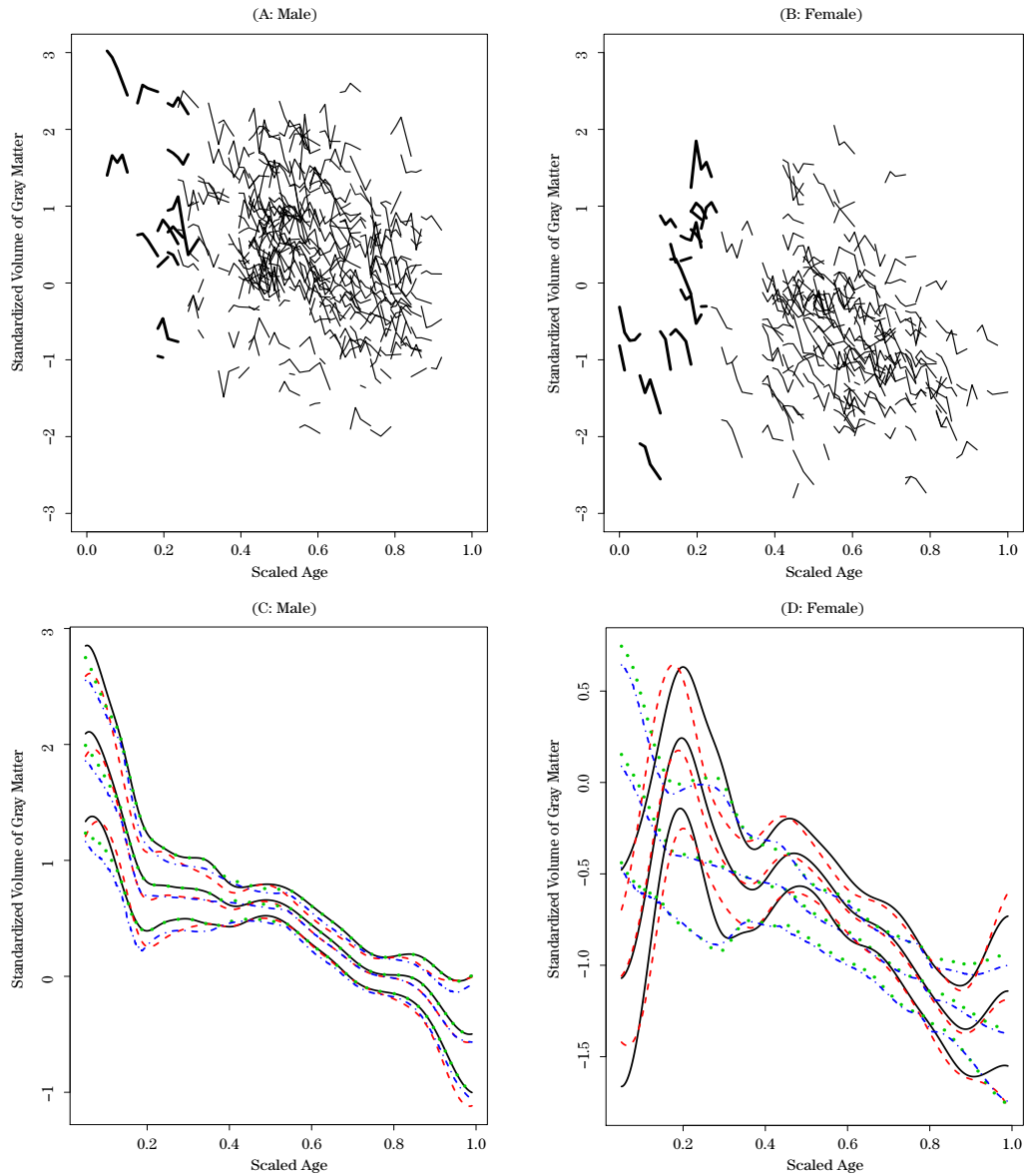


Figure 2. ADNI data analysis. Panels (A) and (B) are the standardized volume of gray matter versus scaled age for male and female subjects, respectively. The bold lines are subjects with scaled ages of first visits less than 0.22. Panels (C) and (D), for male and female, respectively, show the unconstrained OBS estimators and their corresponding lower and upper bounds of the 95% confidence intervals (solid). Those for the unconstrained estimators $\hat{m}_{(SUB,J)}(s)$ and the constrained estimators $\hat{m}_{I(OBS)}(s)$ and $\hat{m}_{I(SUB,J)}(s)$ are dashed, dotted, and dot-dash, respectively.

based on the female observations. This is caused by possible data sparsity and biased data sampling for earlier ages. Specifically, only 32 subjects have a scaled age of the first visit less than 63 years old, including 14 males and 18 females (Table 2). Although the GM volume tends to decrease for each female subject, as shown in Figure 2 (B), it shows an opposite trend before 63 years old, caused by biased sampling. That is, subjects who are much younger than 63 have much smaller GM volumes than subjects aged about 63. In contrast, our proposed method $\hat{m}_I(s)$ produces more reasonable results based on the male observations, female observations, and all observations.

5. Discussion

We have estimated a mean regression function with a monotonicity constraint for functional/longitudinal data. We proposed a two-stage estimating procedure. The first stage obtains a local linear estimator of the mean function, without a constraint. The second stage refines the unconstrained estimator based on Equation (2.1), and then performs a numerical inversion. The theoretical results, simulations, and real-data analysis confirm the good performance of our proposed estimating procedure. The asymptotic normality properties are consistent with those of the unconstrained estimator (i.e., the local linear estimator). However, when the true mean regression function is known to be monotone, our proposed method can take the monotonicity constraint into account. Thus, more information is incorporated, making the proposed estimator more efficient and accurate than the local linear estimator in terms of the finite-sample performance. Furthermore, we compare two commonly used weighting schemes (OBS and SUBJ; see Zhang and Wang (2016)), both theoretically and numerically, for our proposed estimators.

Theorem 1 shows that the correlated structure from the same subject plays a key role in the asymptotic variance of the asymptotic distribution. However, our estimating procedure does not take into account such correlation. In the first estimating stage, we can use the idea in Chen et al. (2011); Chen, Tang and Gao (2018) to incorporate the correlation. That is, we can regress the error on its predecessors and implement the local linear procedure based on the prediction error. In practice, we can test the monotonicity assumption (Ghosal, Sen and van der Vaart (2000); Wang and Meyer (2011); Ahkim, Gijbels and Verhasselt (2017)) before performing our proposed method. Recently, Dawson and Muller (2018) developed an estimation based on conditional quantile trajectories under the

monotonicity constraint of the underlying processes. The proposed estimation method can be extended to estimate conditional quantile functions with the monotonicity constraint. However, these topics are beyond the scope of this study, and thus are left to future research.

Supplementary Material

All lemmas and technical proofs are included in the Supplementary Material.

Acknowledgements

We thank the two reviewers, the Associate Editor, and the Joint Editor for their constructive comments. The research of Dr. Chen was supported by NSFC grant 11871,477 and Hunan Provincial Natural Science Foundation of China 2016JJ3138. The research of Dr. Zhu was supported by NIH grants MH116527 and MH086633, a grant from the Cancer Prevention Research Institute of Texas, and the endowed Bao-Shan Jing Professorship in Diagnostic Imaging. The research of Dr. Gao was supported by the National Social Science Fund of China (Grant No.18BTJ040). The research of Dr. Fu was supported by a UK MRC grant (MR/M025152/1). The data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data, but did not participate in the subsequent analysis or in writing this manuscript. A complete list of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Appendix

Assumptions

We present all the assumptions as follows.

(A) Kernel function.

$K_r(\cdot)$ is assumed to be a symmetric probability density function on $[-1, 1]$ and K_r is twice continuously differentiable on its support such that

$$\kappa_2(K_r) < \infty, \quad \int K_r^2(u) du < \infty.$$

The assumptions on K_d are the same as those on K_r .

(B) Time points and true functions

(B1) $\{s_{ij} : i = 1, \dots, n; j = 1, \dots, n_i\}$ are i.i.d. copies of a random variable S defined on $[0, 1]$. The density $f(\cdot)$ of S is bounded from below and above with $0 < m_f \leq \min_{s \in [0,1]} f(s) \leq \max_{s \in [0,1]} f(s) \leq M_f < \infty$ and $\ddot{f}(s)$, the second derivative of $f(\cdot)$, is continuous on $[0, 1]$.

(B2) $\ddot{m}(s)$, the second derivative of $m(s)$, is continuous on $[0, 1]$.

(B3) $\ddot{\sigma}(s)$, the second derivative of $\sigma(\cdot)$, is continuous on $[0, 1]$.

(B4) $\{\eta_i(\cdot)\}_i$ are i.i.d. copies of $\eta(\cdot)$ and $\{\varepsilon_{ij}\}_{ij}$ are i.i.d. copies of ε . Furthermore, $E(\varepsilon) = 0$, $E(\varepsilon^2) = 1$.

(B5) X is independent of S and ε is independent of S and η .

(B6) $\partial^2\gamma(s, t)/\partial s^2$, $\partial^2\gamma(s, t)/\partial s\partial t$ and $\partial^2\gamma(s, t)/\partial t^2$ are continuous on $[0, 1]^2$.

(C) Bandwidths and moments

(C1) $h_r \rightarrow 0$, $h_d \rightarrow 0$, $h_r^2/h_d \rightarrow 0$, $h_r/h_d \rightarrow \infty$, $h_d/\log(n)^2 = O(1)$,

$h_d^2 h_r^{-8} \max \left\{ \sum_{i=1}^n n_i \omega_i^2 / h_r, \sum_{i=1}^n \omega_i^2 n_i (n_i - 1) \right\} \rightarrow \infty$,

$\log(n)^2 h_r^{-2} h_d^{-1} \max \left\{ \sum_{i=1}^n n_i \omega_i^2 / h_r, \sum_{i=1}^n \omega_i^2 n_i (n_i - 1) \right\} \rightarrow 0$,

$\sum_{j=1}^n \omega_j^4 (n_j^4 + n_j^3/h_d + n_j^2/h_d^2 + n_j/h_d^3) \{ \sum_{j=1}^n \omega_j^2 n_j / h_r + \sum_{j=1}^n \omega_j^2 n_j (n_j - 1) \}^{-2} \rightarrow 0$.

(C2) $E(\varepsilon^5) < \infty$, $E \sup_{s \in [0,1]} \eta^5(s) < \infty$, and $E\eta^5(s)$ is continuous on $[0, 1]$.

(C3)

$$n \left\{ \sum_{i=1}^n n_i \omega_i^2 h_r + \sum_{i=1}^n n_i (n_i - 1) \omega_i^2 h_r^2 \right\} \left\{ \frac{\log(n)}{n} \right\}^{-3/5} \rightarrow \infty.$$

(C4) $\sup_n (n \max_i n_i \omega_i) \leq B < \infty$.

References

Ahkim, M., Gijbels, I. and Verhasselt, A. (2017). Shape testing in varying coefficient models. *Test* **26**, 429–450.

Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26**, 607–616.

Chen, Z., Shi, N. Z., Gao, W. and Tang, M. L. (2011). Efficient semiparametric estimation via cholesky decomposition for longitudinal data. *Comput. Statist. Data Anal.* **55**, 3344–3354.

Chen, Z., Tang, M. L. and Gao, W. (2018). A profile likelihood approach for longitudinal data analysis. *Biometrics* **74**, 220–228.

Dawson, M. and Muller, H. G. (2018). Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data. *J. Amer. Statist. Assoc.* **113**, 1612–1624.

Dette, H., Neumeyer, N. and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12(3)**, 469–490.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. M. (2004). *Applied Longitudinal Analysis*. Wiley, New Jersey.

Ghosal, S., Sen, A. and van der Vaart, A. (2000). Testing monotonicity of regression. *Ann.*

- Statist.* **28**, 1054–1092.
- Gijbels, I. (2005). Monotone regression. In *The Encyclopedia of Statistical Sciences* (Edited by N. Balakrishnan, S. Kotz, C. B. Read and B. VadaKovic). 2nd Edition. Wiley, Hoboken, NJ.
- Hall, P. and Huang, L. S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29**, 624–647.
- Hall, P., Muller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493–1517.
- Henkenius, A., Peterson, B. and et al. (2003). Mapping cortical change across the human life span. *Nature Neuroscience* **6**, 309–315.
- Hoogendam, Y. Y., van der Geest, J. N., van der Lijn, F., van der Lugt, A., Niessen, W. J., Krestin, G. P., Hofman, A., Vernooij, M. W., Breteler, M. M. and Ikram, M. A. (2012). Determinants of cerebellar and cerebral volume in the general elderly population. *Neurobiol. Aging* **33**, 2774–2781.
- Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism. *Biometrics* **46**, 1071–1085.
- Kim, S. and Zhao, Z. B. (2012). Unified inference for sparse and dense longitudinal models. *Biometrika* **100**, 203–212.
- Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38**, 3321–3351.
- Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19**, 724–740.
- Mammen, E., Marron, J. S., Turlach, B. A. and Wand, M. P. (2001). A general projection framework for constrained smoothing. *Statist. Sci.* **16**, 232–248.
- Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions. *Scand. J. Stat.* **26**, 239–252.
- Mukerjee, H. (1988). Monotone nonparametric regression. *Ann. Statist.* **16**, 741–750.
- Ramsay, J. O. (1988). Monotone regression splines in action (with comments). *Statist. Sci.* **3**, 425–461.
- Ramsay, J. O. (1998). Estimating smooth monotone functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60**, 365–375.
- Shen, D. G. and Davatzikos, C. (2004). Measuring temporal morphological changes robustly in brain mr images via 4-dimensional template warping. *Neuroimage* **21**, 1508–1517.
- Wang, J. and Meyer, M. (2011). Testing the monotonicity or convexity of a function using regression splines. *Canad. J. Stat.* **39**, 89–107.
- Wang, J. L., Chiou, J. M. and Muller, H. G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3**, 257–295.
- Wu, H. and Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. Wiley, New Jersey.
- Yao, F., Muller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577–590.
- Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35**, 1052–1079.
- Zhang, X. and Wang, J. L. (2016). From sparse to dense functional data and beyond. *Ann.*

Statist. **44**, 2281–2321.

Zhu, H., Chen, K., Luo, X., Yuan, Y. and Wang, J. L. (2018). Fmem: Functional mixed effects models for longitudinal functional responses. *Statistica Sinica* Accepted, in press.

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

School of Mathematics and Statistics, Central South University, Changsha, China.

E-mail: chenzq453@gmail.com

School of Mathematical Sciences, Nanjing Normal University, Nanjing, China.

E-mail: gaoqibing@njnu.edu.cn

School of Data Science, Fudan University, Shanghai, China.

E-mail: fu@fudan.edu.cn

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

E-mail: htzhu@email.unc.edu

(Received June 2018; accepted February 2019)