<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Monotone Nonparametric Regression for Functional/Longitudinal Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>SS-2018-0233</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td>10.5705/ss.202018.0233</td>
</tr>
</tbody>
</table>
| **Complete List of Authors** | Ziqi Chen  
Qibing Gao  
Bo Fu and  
Hongtu Zhu |
| **Corresponding Author** | Hongtu Zhu |
| **E-mail** | htzhu@email.unc.edu                                               |

Notice: Accepted version subject to English editing.
Monotone Nonparametric Regression for Functional/Longitudinal Data

Ziqi Chen$^{1,2}$, Qibing Gao$^3$, Bo Fu$^4$ 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

$^5$The University of North Carolina at Chapel Hill

Abstract: Motivated by quantifying a monotonic relationship between gray matter volume and age in the older population, the aim of this article is to propose the 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 based on a general weighting scheme including equal weight per observation (OBS) and equal weight per subject (SUBJ) as two special cases. Most existing methods without a structural constraint can only handle sparse or dense data, thus a subjective choice between the two types may lead to erroneous conclusions for statistical inference. Our proposed method and theories can adapt to the sparse and dense cases on a unified platform under the monotonic constraint. The asymptotic results allow one to categorize functional/longitudinal data into three data types including sparse, dense, and ultra–dense data based on three rel-
ative orders 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. Analysis of gray matter (GM) volume data obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) 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, Monotonicity constraint, Nonparametric estimation, Kernel smoothing, Sparse and dense functional/longitudinal data, weighting schemes.

1. Introduction

This paper 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 to test 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. Aging can generally 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 for the characterization of the structural changes that accompany aging of the healthy brain, such as changes in the volume of grey 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
that is seen with early Alzheimer’s disease. All these results have shown 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 analysis has wide application in the biomedical, psychometric and environmental sciences (Fitzmaurice et al., 2004; Yao et al., 2005; Wu & Zhang, 2006; Wang et al., 2016; Zhu et al., 2018). In this type of analysis, subjects are repeatedly measured 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 given by

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

where \( y_{ij} \) is the response variable of interest for subject \( i \) measured at time \( s_{ij} \), the \( X_i(\cdot) \)'s are independent realizations of the underlying process \( \{ X(\cdot) \} \), and the \( \varepsilon_{ij} \)'s are random errors with zero mean and variance of 1. By 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 zero mean and covariance function \( \gamma(\cdot, t) = \text{cov}\{\eta_i(s), \eta_i(t)\} \). Then, we can rewrite (1.1) as

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

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 estimation methods focus on the estimation of a nonparametric regression function without a structural constraint. See Li & Hsing (2010), Kim & Zhao (2012), Zhang & Wang (2016), and references therein. For instance, local linear and polynomial methods are the most popular methods for unconstrained nonparametric estimation of \( m(\cdot) \) in
the functional/longitudinal data framework (Kim & Zhao, 2012; Zhang & Wang, 2016). However, the mean response function can be a monotonic function of $s$ in some cases. For example, as shown in Section 5, gray matter volume decreases after age 40 (Henkenius et al., 2003). The hippocampus volume decreases very rapidly for those patients suffering from Alzheimer (Dawson & Muller, 2018). To the best of our knowledge, little has been done on the estimation of the monotone mean function for functional/longitudinal data.

However, there is a large amount of literature on the estimation of monotonic regression functions for cross-sectional data. Please see Gijbels (2005) and references therein. For instance, Brunk (1955) proposed a modified maximum likelihood estimator of $m(s)$, but such an estimator may not be smooth. Mukerjee (1988) and Mammen (1991) proposed a smooth monotonic estimator of $m(s)$. Constrained smoothing spline methods have been proposed to estimate $m(s)$ (Ramsay, 1988, 1998; Kelly & Rice, 1990; Mammen & Thomas-Agnan, 1999; Mammen et al., 2001). Hall & Huang (2001) proposed to monotonize general kernel-type estimators by tilting the empirical distribution. Dette et al. (2006) proposed to construct 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 paper is to develop a constrained nonparametric estimate of $m(\cdot)$ under model (1.2) and three different types of $\{n_i : i = 1, \ldots, n\}$ relative to $n$. Compared with the existing literature (Zhang & Wang, 2016; Gijbels, 2005), we make several important contributions. First, we construct a constrained nonparametric estimator of monotone $m(\cdot)$, denoted as $\hat{m}_I(\cdot)$, for functional/longitudinal data based on local kernel methods. Moreover, under monotonicity constraint, we propose to adaptively 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 $\hat{m}_I(\cdot)$ for all the three different relative orders.
of \(\{n_i\}\) to \(n\) under a general weighting scheme. Such theory allows one to clearly define three different types of functional/longitudinal data: sparse data, dense data, and ultra-dense data, which depend on whether \(\hat{m}(\cdot)\) can achieve the root-\(n\) convergence rate and that its asymptotic bias is negligible. Our estimation and confidence interval construction methods do not need to distinguish whether the data are sparse or dense, i.e., allow the magnitude of \(n_i\) relative to the sample size \(n\) to vary freely. In contrast, most existing methods without a structural constraint can only handle individual sampling design scenarios [Yao et al., 2005] [Hall et al., 2006] [Zhang & Chen, 2007]. Third, we consider the two commonly used weighting schemes, which have been introduced for unconstrained mean function estimation by [Zhang & 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, and released it to the public through 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 proposed estimators in Section 2.1. Section 2.3 proposes an adaptive confidence interval for the constraint nonparametric mean function. Section 3 gives simulation studies and Section 4 conducts real ADNI data analysis which demonstrate that the proposed nonparametric estimators perform well and reasonably. Section 5 concludes with some discussions. All assumptions are put 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 only consider the case of isotonic (strictly increasing) regression functions.

Let $\partial_u = d/du$, $K_d(\cdot)$ and $K_r(\cdot)$ be two kernel functions, $h_d$ and $h_r$ be two bandwidths, and $K_{a,h}(v) = h^{-1}K_a(v/h)$ be the rescaled kernel function with a bandwidth $h$ for $a = d$ and $r$.

We start by reviewing several key ideas from Dette et al. (2006). Consider an i.i.d. sample of $N$ uniform random variables, say $U_1, \ldots, U_N \sim U[0,1]$. 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 $\partial_u(m^{-1}(u))1(u \in [m(0), m(1)])$, where $1(A)$ is an indicator function of event $A$. Subsequently, as $h_d \to 0$ and $Nh_d \to \infty$, a consistent estimate of $m^{-1}(t)$ is

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

Moreover, estimator (2.1) is a strictly increasing function almost surely when $N$ is sufficiently large (Dette et al., 2006).

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

$$\hat{\beta} = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \omega_i \{Y_{ij} - \beta_0 - \beta_1(s_{ij} - s)_+\}^2 K_{r,h_r}(s_{ij} - s),$$

where the $\omega_i$’s are weights satisfying $\sum_{i=1}^{n} n_i \omega_i = 1$. We consider two commonly used weighting schemes, OBS for equal weight per observation and SUBJ for equal weight per subject (Yao et al., 2005; Li & Hsing, 2010; Kim & Zhao, 2012; Zhang & 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, $\hat{m}(s)$ is a local...
linear estimator of $m(\cdot)$ \cite{Li & Hsing 2010, Zhang & Wang 2016}. By plugging $\hat{m}(s)$ into (2.1), we obtain

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

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

### 2.2 Theoretical Results

In this subsection, we systematically investigate the asymptotic properties of $\hat{m}_I(s)$. $K_d(v)$ 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 $\kappa_2(K) = \frac{1}{2} \int_{-1}^{1} u^2 K(u) du$. Let $\xrightarrow{d}$ denote the convergence in distribution as $n \to \infty$. We also define $\hat{m}(s) = dm(s)/ds$ and $\tilde{m}(s) = d^2m(s)/ds^2$. Following the arguments in \cite{Zhang & Wang 2016}, we consider three types of data that vary according to the number of repeated measurements $\{n_i\}$:

(i) **Sparse data**: $\bar{n}/n^{1/4} \to 0$ with $\bar{n} = \sum_{i=1}^{n} n_i/n$;

(ii) **Dense data**: $\bar{n}/n^{1/4} \to C$ with $0 < C < \infty$;

(iii) **Ultra–dense data**: $\bar{n}/n^{1/4} \to \infty$.

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

Define

$$
\Gamma^A_0(t) := \left\{ \int K^2_r(\omega) d\omega \right\} \gamma(m^{-1}(t), m^{-1}(t)) + \sigma^2(m^{-1}(t)) \frac{f(m^{-1}(t))}{\{m(m^{-1}(t))\}^2}
$$

and

$$
\Gamma^B_0(t) := \frac{\gamma(m^{-1}(t), m^{-1}(t))}{\{m(m^{-1}(t))\}^2}.
$$

We state the following theorem, the assumptions of which can be found in the Appendix. It can be obtained by Lemmas 1 and 5 in the Supplementary Material.
Theorem 1. Suppose that Assumptions (A)–(C) in the Appendix hold. Also, assume

- (i) \(\min\left[h_r/\left(\sum_{j=1}^{n} \omega_j^2 n_j\right), 1/\left(\sum_{j=1}^{n} \omega_j^2 n_j(n_j - 1)\right)\right] h^6_r \to 0;\)

- (ii) \(h_r \sum_{j=1}^{n} \omega_j^2 n_j(n_j - 1)/\sum_{j=1}^{n} \omega_j^2 n_j \to C_0 \in [0, \infty];\)

- (iii) \(N h_d h^2_r \to \infty;\)

- (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

\[
\Gamma(t)^{-1/2} \left\{ \tilde{m}_{i-1}^{-1}(t) - m^{-1}(t) + h_r^2 \kappa_r^2(K_r) \left( \frac{\tilde{n}}{\tilde{m}} \right) (m^{-1}(t)) \right\} \overset{d}{\to} N(0, 1), \tag{2.2}
\]

where \(\Gamma(t)\) equals the sum of \(\Gamma^A(t) = \Gamma^A_0(t) (\sum_{j=1}^{n} \omega_j^2 n_j)/h_r\) and \(\Gamma^B(t) = \Gamma^B_0(t) \sum_{j=1}^{n} \omega_j^2 n_j(n_j - 1).\)

Theorem 1 can be regarded as a generalization of Theorem 3.1 in Dette et al. (2006). Specifically, the asymptotic bias is the same as that for cross-sectional studies (Dette et al., 2006), whereas the variance term \(\Gamma\) is more complex. Specifically, \(\Gamma^A\) characterizes the variances of all observations and \(\Gamma^B\) mainly comes from the correlations among 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 \(\min\{\tilde{n} h_r, n(\tilde{n})^2/(\tilde{n} S^2 - \tilde{n})\} h^6_r \to 0\) and \(h_r(\tilde{n} S^2 - \tilde{n})/\tilde{n} \to C_0 \in [0, \infty],\) where \(\tilde{n} S^2 = \sum_{i=1}^{n} n_i^2/n,\) then the asymptotic normality \(2.2\) holds with \(\Gamma(t)\) for OBS, denoted as \(\Gamma_{\text{obs}}(t),\) equal to the sum of \(\Gamma^A_{\text{obs}}(t)\) and \(\Gamma^B_{\text{obs}}(t)\) given by

\[
\Gamma^A_{\text{obs}}(t) = \Gamma^A_0(t)/(n \tilde{n} h_r) \quad \text{and} \quad \Gamma^B_{\text{obs}}(t) = \frac{(\tilde{n} S^2 - \tilde{n})}{n \tilde{n}^2} \Gamma^B_0(t).
\]
(b) SUBJ: If \( \min\{n\bar{h}_H, n/(1 - 1/\bar{h}_H)\}\) \(h_r \to 0\) and \(h_r(\bar{h}_H - 1) \to C_0 \in [0, \infty]\), where \(\bar{h}_H = (n^{-1}\sum_{i=1}^{n} n_i^{-1})^{-1}\), then the asymptotic normality holds with \(\Gamma(t)\) for SUBJ, denoted as \(\Gamma_{subj}(t)\), equal to the sum of \(\Gamma^A_{subj}(t)\) and \(\Gamma^B_{subj}(t)\) given by

\[
\Gamma^A_{subj}(t) = \Gamma^A(t)/(n\bar{h}_H h_r) \quad \text{and} \quad \Gamma^B_{subj}(t) = n^{-1}(1 - \bar{h}_H^{-1})\Gamma^B_0(t).
\]

For either the OBS or SUBJ scheme, three types of asymptotic normality emerge from Corollary depending on the order of \(\bar{n}\) and \(\bar{h}_H\) relative to \(n\).

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

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

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

\[
\sqrt{n\bar{h}_r}\left\{\hat{\beta}^{-1}_{i(OBS)}(t) - m^{-1}(t) + h_r^2 \kappa_2(K_r) \left(\frac{\hat{m}}{m}\right) (m^{-1}(t))\right\} \to_d N(0, \Gamma^A_0(t)).
\]

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

\[
\sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} \left\{\hat{\beta}^{-1}_{i(OBS)}(t) - m^{-1}(t) + h_r^2 \kappa_2(K_r) \left(\frac{\hat{m}}{m}\right) (m^{-1}(t))\right\} \to_d N(0, \Gamma^A_0(t)/C_1 + \Gamma^B_0(t)).
\]

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

\[
\sqrt{\frac{n\bar{n}^2}{\bar{n}_{S_2}}} \left\{\hat{\beta}^{-1}_{i(OBS)}(t) - m^{-1}(t)\right\} \to_d N(0, \Gamma^B_0(t)).
\]

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

Corollary indicates that \(\hat{\beta}^{-1}_{i(OBS)}(t)\) and \(\hat{\beta}^{-1}_{i(SUBJ)}(t)\) have identical asymptotic bias. But their asymptotic variances are different.

We observed from Corollary that, for sparse data, \(\Gamma^A_{obs}\) and \(\Gamma^A_{subj}\) dominate \(\Gamma^B_{obs}\) and \(\Gamma^B_{subj}\), respectively, and we obtain \(\Gamma^A_{obs} \leq \Gamma^A_{subj}\) by using arguments similar to Corollary 3.3 in
Thus, the OBS scheme achieves a more efficient estimator of \( m^{-1}(t) \) than the SUBJ scheme. We give the intuitive explanation: for sparse data, the bandwidth satisfies \( \bar{n}h_r \to 0 \) or \( \bar{n}_Hh_r \to 0 \). A special and simple case is \( n_ih_r \to 0 \) for \( i = 1, \ldots, n \), in which each subject contributes only one observation for estimating \( m^{-1}(t) \). That is, for 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, can yield a more efficient estimator.

In contrast, for ultra–dense data, \( \Gamma^B_{\text{obs}} \) and \( \Gamma^B_{\text{subj}} \) dominate \( \Gamma^A_{\text{obs}} \) and \( \Gamma^A_{\text{subj}} \), respectively and \( \Gamma^B_{\text{obs}} \geq \Gamma^B_{\text{subj}} \). Thus, the SUBJ scheme is preferable to the OBS scheme. The conclusions here are consistent with those in Zhang & Wang (2016) for unconstrained nonparametric estimates of \( m(\cdot) \) in model (1.2). Intuitively, for given \( t \), there exist subjects contribute infinitely many observations to the estimation of \( m^{-1}(t) \), because \( \bar{n}h_r \to \infty \) or \( \bar{n}_Hh_r \to \infty \). The observations from the same subject are correlated, thus the covariances within one subject tend to dominate the variances and could have undue influence to the variance of \( \hat{m}^{-1}(t) \). The SUBJ scheme can avoid such a circumstance by assigning weight \( 1/n_i \) to subject \( i \) and thus yields a more efficient estimator than OBS. Essentially, for ultra-dense data, the SUBJ scheme is equivalent to the so-called “smooth-first-then-estimate” approach (Hall et al., 2006; Zhang & Chen, 2007), which first preprocesses the discrete functional data by smoothing for subject \( i \) \( (i = 1, \ldots, n) \), 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 et al. (2006), which is established for independent data. Moreover, the convergence rate of \( \hat{m}^{-1}_{I(OBS)}(t) \) 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 \( \hat{m}^{-1}_{I(OBS)}(t) \) for both “dense data” and “ultra-dense data” is \( O_p(n^{1/2}) \). Furthermore, “ultra-dense data” fall in the parametric paradigm, where the limiting normal distribution has zero mean.
Remark 3. An explicit partition of functional/longitudinal data can be concluded from Corollary 2 based on the different characteristics of the asymptotic properties for the proposed monotonicity-constraint nonparametric regression. Specifically, the functional/longitudinal data can be divided into three types including sparse data, dense data, and ultra–dense data, based on the relative order of $n$ and $n_H$ to $n^{1/4}$ under either the OBS or SUBJ scheme, respectively. This partition is consistent with that in Zhang & Wang (2016).

The functions $\hat{m}_I^{-1}$ and $m_N^{-1}$ are strictly increasing independent of the monotonicity of the ‘true’ regression function $m$ for sufficiently large $n$ and $N$. We know that the corresponding inverse function $\hat{m}_I$ of $\hat{m}_I^{-1}$ also satisfies an asymptotic normal distribution by the following result. Its proof is similar to that of Theorem 3.2 in Dette et al. (2006) and hence is omitted here.

Theorem 2. Suppose that all the assumptions of Theorem 1 hold. For a fixed interior point $s \in (0,1)$ with $\hat{m}(s) > 0$, we have

$$\Gamma_*(s)^{-1/2} \left\{ \hat{m}_I(s) - m(s) - h_i^2 \sigma_2(K_r) \hat{m}(s) \right\} \xrightarrow{d} N(0,1),$$

where $\Gamma_*$ is given by

$$\Gamma_*(s) = h_i^{-1} \sum_{j=1}^n \left\{ \int K^2_r(u) du \right\} \left\{ \gamma(s, s) + \sigma^2(s) \right\} / f(s) + \sum_{j=1}^n \omega_j^2 \hat{n}_j (\hat{n}_j - 1) \gamma(s, s).$$

We can see that our proposed estimator under constraint and the unconstrained one in Zhang & Wang (2016) have the same asymptotic distribution.

Similar to Corollary 2, we can show the following results for $\hat{m}_I(OBS)$ and $\hat{m}_I(SUBJ)$.

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

(a) OBS: Assume $\limsup_n \bar{n}_S^2 / \hat{n}^2 < \infty$. 

11
Case 1 (Sparse data) When \( \bar{n}/n^{1/4} \to 0 \) and \( h_r \asymp (n\bar{n})^{-1/5} \), we have

\[
\sqrt{n\bar{n}}h_r \left\{ \hat{m}(s) - m(s) - h_r^2\kappa_2(K_r)\bar{m}(s) \right\} \to_d N \left( 0, \left\{ \int K_r^2(u)du \right\} \left\{ \gamma(s,s) + \sigma^2(s) \right\} / f(s) \right).
\]

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

\[
\sqrt{n\bar{n}/n^{1/2}} \left\{ \hat{m}(s) - m(s) - h_r^2\kappa_2(K_r)\bar{m}(s) \right\} \to_d N \left( 0, \left\{ \int K_r^2(u)du \right\} \left\{ \gamma(s,s) + \sigma^2(s) \right\} / \{C_1f(s)\} + \gamma(s,s) \right).
\]

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

\[
\sqrt{n\bar{n}/n^{1/2}} \left\{ \hat{m}(s) - m(s) \right\} \to_d N \left( 0, \gamma(s,s) \right).
\]

(b) SUBJ: We can obtain similar asymptotic normality results corresponding to sparse, dense, and ultra-dense data for

\[ \hat{m}(\text{SUBJ}) \]

by replacing \( n, n\bar{n}/n, \) and \( n^2/n\bar{n} \) in (a) by \( n_H, n_H, \) 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 1. Consistent with the results of \( \hat{m}(\text{OBS}) \) for sparse data, OBS outperforms SUBJ in terms of smaller asymptotic variance of \( \hat{m}(\text{OBS})(s) \). 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 differ substantially from each other under the monotonicity constraint. For instance, their corresponding asymptotic variances are different. To construct an confidence interval, one might need to estimate \( \sigma^2(s) \) and \( \gamma(s,s) \) for each of the three types. We will show how to construct an adaptive confidence interval for all data types.
2.3 Adaptive Confidence Interval

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

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

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

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

It follows from Lemma 3 in the Supplementary Material that $U_n^2(s) = \overline{U}_n^2(s) \{1 + o_p(1)\}$ holds. By using some calculations, we have $\overline{U}_n^2(s) = \Gamma_+(s) \{1 + o_p(1)\}$. Thus, it follows from Slusky’s Theorem that we have

$$U_n(s)^{-1} \left\{ \hat{m}(s) - m(s) - h_r^2 \kappa_2(K_r) \dot{m}(s) \right\} \overset{d}{\rightarrow} N(0, 1). \quad (2.3)$$

Therefore, (2.3) can be used to construct a unified asymptotic point-wise $1 - \alpha$ confidence interval for the monotone mean function $m(s)$, which can adapt to all the three types of data.

Remark 6. We can also use

$$U_n(s)^{-1} \left\{ \hat{m}(s) - m(s) - h_r^2 \kappa_2(K_r) \dot{m}(s) \right\} \overset{d}{\rightarrow} N(0, 1) \quad (2.4)$$

to construct a good asymptotic point-wise $1 - \alpha$ confidence interval for the mean function $m(s)$ \cite{Zhang & Wang, 2016} based on the unconstrained estimator. However, $\hat{m}(s)$ does not satisfy the monotonic constraint and we will show in the simulation studies that the standard deviation
of $\hat{m}_I(s)$ may be smaller than that of $\hat{m}(s)$ in the finite-sample performance. So the average coverage rate of the $1 - \alpha$ confidence interval based on (2.3) may be better than that based on (2.4) in the finite-sample performance.

**Remark 7.** Define

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

We can also show that

$$U_{1n}(s)^{-1} \left\{ \hat{m}_I(s) - m(s) - h^2_r \kappa_2(K_r) \hat{m}(s) \right\} \overset{d}{\rightarrow} N(0, 1) \quad (2.5)$$

holds as $n \rightarrow \infty$. Similarly, we may use (2.5) to construct an asymptotic point-wise $1 - \alpha$ confidence interval for $m(s)$. Since the unconstrained estimator is constructed by minimizing the difference between the responses $y_{ij}$'s and the estimated mean values $\hat{m}(s_{ij})$'s, $U^2_{1n}(s)$ may be slightly smaller than $U^2_{1n}(s)$ and thus, the average length of the confidence interval based on (2.3) may be slightly shorter than that based on (2.5). So 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

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

where $m(s) = \sin(s)$ for $s \in [0, 1]$, $\alpha_{ik} \sim N(0, \omega_k)$, and $\varepsilon_{ij}$'s are independent and identically distributed. 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 \( \varepsilon_{ij} \) were considered: \( N(0, 0.5) \) and T distribution with degree of freedom being 3.

For each distribution, we considered the following three cases for the vector \( \mathbf{n} = (n_1, \ldots, n_n)^T \):

(Case 1) \( \mathbf{n}_1 : n_i \sim U[1, 2, \ldots, 5] \),  
(Case 2) \( \mathbf{n}_2 : n_i \sim U[n/10, \ldots, n/5] \),  
(Case 3) \( \mathbf{n}_3 : n_i \sim U[n/3, \ldots, 2n/3] \),

where \( U[\mathcal{D}] \) stands for the discrete uniform distribution on a finite set \( \mathcal{D} \). The \( \mathbf{n}_1 \) can be regarded as the case of sparse data, and \( \mathbf{n}_3 \) may be treated as cases of ultra-dense data. For each case, the simulation was repeated for \( Q = 500 \) 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 \( S = \{s_{ij}, j = 1, \ldots, n; i = 1, \ldots, n\} \). 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 et al., 2006). The \( N \) used to estimate \( m^{-1}(t) \) is set to 500.

We compared our proposed estimator with the the unconstrained one (Zhang & Wang (2016)). Let \( S_l = 0.04 + l \times 0.01 \) for \( l = 1, \ldots, E = 91 \). We calculated the bias and standard deviation (SD) at each of the 91 points \( \{S_l\}'s \) based on 500 replications and then got the average bias and SD. The “SD” can be viewed as the true standard deviation of the resulting estimates and thus it can be used to measure the efficiency of the methods. We defined the empirical mean integrated squared error (EMISE) and empirical mean supremum absolute error (EMSAE) as follows:

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

\[
\text{EMSAE}(\hat{m}_I) = \frac{\sum_{q=1}^{Q} \max_l |\hat{m}_I^{(q)}(S_l) - m(S_l)|}{Q},
\]
where $\hat{m}_i^{(q)}(\cdot)$ is the estimator of $m(\cdot)$ using the $q$-th dataset for $q = 1, \ldots, Q$. We adopted 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 probabilities (AECP) and average lengths (AL) of confidence intervals. For each $S_i$ and the given nominal level 95%, we built confidence interval for $m(S_i)$ for each dataset 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 the 91 points $S_i$’s.

Table 1 presents the simulation results. We have three important observations. First, our constrained estimator is more efficient and accurate than the unconstrained one. Second, the confidence intervals of our constrained estimator have the same width, but obviously better coverage probabilities compared with those of the unconstrained one. Third, compared with SUBJ, the OBS scheme generally produces more efficient and accurate estimators for sparse data and less efficient and accurate estimators for ultra-dense data. These conclusions also hold even when the normal distribution assumption of the random error $\varepsilon$ is violated.

(Table 1 about here.)

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 3-dimensional sagittal MPRAGE or equivalent.
protocols with varying resolutions. The settings for the typical protocol were repetition time \( t_r = 2400 \) ms, inversion time \( t_i = 1000 \) ms, flip angle \( \alpha = 8^\circ \), 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 \( 1.25 \times 1.26 \times 1.2 \) mm\(^3\). The MRI data were preprocessed by standard steps including anterior commissure and posterior commissure correction, skull-stripping, cerebellum removing, intensity inhomogeneity correction, segmentation, and registration (Shen & Davatzikos, 2004). Automatic regional labeling was then carried out by labeling the template and by 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 region of interests.

We are interested in estimating 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 1 observation, 56 subjects have 2 observations, 117 subjects have 3 observations, 255 subjects have 4 observations, 88 subjects have 5 observations, and 7 subjects have 6 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 value and the maximum value of \( \{s_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, 562\} \), 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 \( \{y_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, 562\} \), respectively. The longitudinal trajectories of the standard GM volumes at different scaled ages are shown in Figure 1 (A).

Figure 1 (A) shows that GM volume tends to decrease as age increases at individual level, indicating that the constrained estimation method proposed here may be a good choice for establishing such monotonic relationship. In contrast, the local linear regression method
cannot automatically impose such monotonic relationship between age and the GM volume. Inspecting Figure 1 (A), Figure 2 (A) and (B) and Table 2 reveals that: (i) female subjects tend to have less GM volumes than male subjects, (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 data characteristic causes an illusion that the GM volume increases before 63 years old and decreases after 63 years old. 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 true age around 63.

The unconstrained estimator $\hat{m}(s)$ yields 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) based on the female observations. This is caused by possible data sparsity and biased data sampling at an early age. Specifically, the number of subjects with the scaled age of the first visit earlier than 63 years old is only 32, 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 compared with those subjects aging around 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 considered the estimation of mean regression function with monotonicity constraint for functional/longitudinal data in this article. We proposed a two-stage estimating procedure. The first stage is to get a local linear estimator of the mean function without constraint. The second stage is to refine the unconstrained estimator based on Equation (2.1) and then a numerical inversion. The theoretical results, simulations and real data analysis have confirmed the good performances of our proposed estimating procedure. The asymptotical 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, i.e., more information is incorporated, thus our proposed estimator is more efficient and more accurate than the local linear estimator in the finite-sample performances. Further, the two commonly used weighting schemes (OBS and SUBJ, see, Zhang & Wang (2016)) are compared 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, 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. Further, in practice, we can test the monotonicity assumption (Ghosal et al. 2000, Wang & Meyer 2011, Ahkim et al. 2017) before performing our proposed method. Recently, Dawson & Muller (2018) developed conditional quantile trajectories estimation under the monotonicity
constraint of the underlying processes. One may extend our estimation method to estimate the conditional quantile functions with monotonicity constraint. These topics are beyond the scope of the current article and will be pursued elsewhere.

Acknowledgement

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 11871477 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 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). 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 ADNI and/or provided data but did not participate in the analysis herein nor 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) $\{s_{ij} : i = 1, \ldots, n; j = 1, \ldots, n_i\}$ 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 $\bar{f}(s)$, the second derivative of $f(\cdot)$, is continuous on $[0,1]$.

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

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

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

(B5) $X$ is independent of $S$ and $\varepsilon$ is independent of $S$ and $\eta$.

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

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

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

$\sum_{j=1}^{n} \omega^2_i (n_i^j + n_j^j) / h_d + n_i^2 / h_d^2 + n_j^2 / h_d^2) (\sum_{j=1}^{n} \omega^2_j n_j / h_r + \sum_{j=1}^{n} \omega^2_j n_j (n_j - 1))^{-2} \to 0.$

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

(C4) $\sup_{u} (n \max_{i} n_i \omega_i) \leq B < \infty.$

21
References


Ziqi Chen
Department of Biostatistics,
The University of Texas MD Anderson Cancer Center, Houston, USA
School of Mathematics and Statistics, Central South University, Changsha, China
E-mail: (chenzq453@gmail.com)

Qibing Gao
School of Mathematical Sciences, Nanjing Normal University, Nanjing, China
E-mail: (gaoqibing@njnu.edu.cn)

Bo Fu
School of Data Science, Fudan University, Shanghai, China
E-mail: (fu@fudan.edu.cn)

Hongtu Zhu
Department of Biostatistics,
The University of Texas MD Anderson Cancer Center, Houston, USA
The University of North Carolina at Chapel Hill, Chapel Hill, USA
E-mail: (htzhu@email.unc.edu)
Figure 1: ADNI data analysis. Panel (A) is the standardized volume of grey matter versus the scaled age, the black for male and red for female. Panel (B) shows the unconstrained estimators $\hat{m}_{OBS}(s)$ and their corresponding lower and upper bounds of 95% confidential intervals (black and solid). Those for the unconstrained estimators $\hat{m}_{SUBJ}(s)$, the constrained estimators $\hat{m}_{I(OBS)}(s)$ and $\hat{m}_{I(SUBJ)}(s)$ are red and dashed, green and dotted, blue and dot-dash, respectively.
Figure 2: ADNI data analysis. Panels (A) and (B) are the standardized volume of grey matter versus the scaled age for male and female subjects, respectively. The red lines are the subjects with scaled ages of first visits being less than 0.22. Panels (C) and (D), for male and female, respectively, show the unconstrained OBS estimators and its corresponding lower and upper bounds of 95% confidential intervals (black and solid). Those for the unconstrained estimators $\hat{m}_{\text{SUBJ}}(s)$, the constrained estimators $\hat{m}_{\text{OBS}}(s)$ and $\hat{m}_{\text{SUBJ}}(s)$ are red and dashed, green and dotted, blue and dot-dash, respectively.
Table 1: The Biases, SDs, EMISEs, EMSAEs, AECPs and ALs for the unconstrained estimators (UE) and the proposed estimators (PE) for either OBS or SUBJ scheme.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>N(0,0.5)</th>
<th>T(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case 1</td>
<td>Case 2</td>
</tr>
<tr>
<td>UE OBS Bias</td>
<td>−0.005</td>
<td>−0.005</td>
</tr>
<tr>
<td>SD</td>
<td>0.133</td>
<td>0.092</td>
</tr>
<tr>
<td>EMISE</td>
<td>0.018</td>
<td>0.009</td>
</tr>
<tr>
<td>EMSAE</td>
<td>0.252</td>
<td>0.158</td>
</tr>
<tr>
<td>AECP</td>
<td>93.54</td>
<td>94.35</td>
</tr>
<tr>
<td>AL</td>
<td>0.499</td>
<td>0.360</td>
</tr>
<tr>
<td>SUBJ Bias</td>
<td>−0.003</td>
<td>−0.006</td>
</tr>
<tr>
<td>SD</td>
<td>0.141</td>
<td>0.091</td>
</tr>
<tr>
<td>EMISE</td>
<td>0.020</td>
<td>0.008</td>
</tr>
<tr>
<td>EMSAE</td>
<td>0.279</td>
<td>0.156</td>
</tr>
<tr>
<td>AECP</td>
<td>93.29</td>
<td>94.57</td>
</tr>
<tr>
<td>AL</td>
<td>0.524</td>
<td>0.357</td>
</tr>
<tr>
<td>PE OBS Bias</td>
<td>−0.005</td>
<td>−0.004</td>
</tr>
<tr>
<td>SD</td>
<td>0.119</td>
<td>0.089</td>
</tr>
<tr>
<td>EMISE</td>
<td>0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>EMSAE</td>
<td>0.216</td>
<td>0.154</td>
</tr>
<tr>
<td>AECP</td>
<td>96.17</td>
<td>95.05</td>
</tr>
<tr>
<td>AL</td>
<td>0.499</td>
<td>0.360</td>
</tr>
<tr>
<td>SUBJ Bias</td>
<td>−0.003</td>
<td>−0.005</td>
</tr>
<tr>
<td>SD</td>
<td>0.124</td>
<td>0.088</td>
</tr>
<tr>
<td>EMISE</td>
<td>0.015</td>
<td>0.007</td>
</tr>
<tr>
<td>EMSAE</td>
<td>0.227</td>
<td>0.152</td>
</tr>
<tr>
<td>AECP</td>
<td>96.44</td>
<td>95.23</td>
</tr>
<tr>
<td>AL</td>
<td>0.524</td>
<td>0.357</td>
</tr>
</tbody>
</table>
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.

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