HHS Public Access

Author manuscript

Technometrics. Author manuscript; available in PMC 2016 May 01.

Published in final edited form as:

Technometrics. 2015 May ; 57(2): 268–280. doi:10.1080/00401706.2014.914978.

Cook's Distance Measures for Varying Coefficient Models with Functional Responses

Qibing Gao,

Department of Statistics Nanjing Normal University Nanjing 210023, China

Mihye Ahn, and

Department of Biostatistics and Biomedical Research Imaging Center University of North Carolina at Chapel Hill Chapel Hill, NC 27599, USA

Hongtu Zhu*

Department of Biostatistics and Biomedical Research Imaging Center University of North Carolina at Chapel Hill Chapel Hill, NC 27599, USA

Abstract

The aim of this paper is to develop Cook's distance measures for assessing the influence of both atypical curves and observations under varying coefficient model for functional responses. Our Cook's distance measures include Cook's distances for deleting multiple curves and for deleting multiple grid points, and their scaled Cook's distances. We systematically investigate some theoretical properties of these diagnostic measures. Simulation studies are conducted to evaluate the finite sample properties of these Cook's distances under different scenarios. A real diffusion tensor tract data set is analyzed to illustrate the use of our diagnostic measures.

Keywords

Cook's distance; Scaled Cook's distance; Varying coefficient models

1 Introduction

Functional data analysis becomes increasingly popular as modern technology enables relatively easy access to functional data. For instance, with advanced imaging techniques, massive functional data can be observed over both time and space. Such imaging techniques include functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and diffusion tensor imaging (DTI), among many others. In practice, many scientific questions focus on delineating the variability of functional data and their association with a set of covariates of interest, such as diagnostic group. A class of varying-coefficient models (VCM), which allows its regression coefficients to vary over some predictors of interest, is a powerful statistical framework for addressing many scientific questions in biomedical sciences (Hastie and Tibshirani, 1993).

* Address for correspondence and reprints: Hongtu Zhu, Ph.D., hzhu@bios.unc.edu. We thank the Editor, the Associate Editor, and three anonymous referees for valuable suggestions, which greatly helped to improve our presentation. The research of Drs. Zhu and Ahn was partially supported by NIH grants RR025747-01, CA74015, P01CA142538-01, and MH086633..

Since VCM was systematically introduced to statistical literature by Hastie and Tibshirani (1993), many VCMs have been widely studied and developed for longitudinal, time series, and functional data (Yao et al., 2005; Zhang and Chen, 2007; Zhu et al., 2012b). For example, Fan and Zhang (2008) provided a comprehensive review of various statistical procedures proposed for many varying coefficient models. Yao et al. (2005) proposed a nonparametric method to perform functional principal component analysis for sparse longitudinal data without considering any covariates. Zhu et al. (2012b) proposed a multivariate VCM for regularly sampled functional data and developed several statistical inference procedures for MVCM. These statistical methods including both estimation and testing, however, can be very sensitive to atypical curves which do not follow the pattern of the majority of the data. For this reason, influence analysis should be commonly done to detect such atypical curves and to investigate the sensitivity of estimator and test statistic with respect to them.

Although there is a large literature on the development of various case-deletion diagnostic measures (e.g., Cook's distance) for detecting influential observations or clusters under various statistical models for non-functional data (Cook, 1977; Cook and Weisberg, 1982; Davison and Tsai, 1992; Wei, 1998; Zhu et al., 2001, 2012a), little has been done on developing diagnostic measures for VCMs with functional responses. Therefore, it has been common to skip influence analysis completely or to detect outliers visually in functional data analysis (Faraway, 1997; Shen and Faraway, 2004). Recently, a few diagnostic measures have been developed for VCM with functional responses (Shen and Xu, 2007; Chiou and Müller, 2007; Zou et al., 2013). Specifically, Shen and Xu (2007) proposed some diagnostic measures including residual and single-case Cook's distance for VCM. Chiou and Müller (2007) proposed several diagnostic measures including leverage, residual, single-case Cook's distance, and a residual process for a large class of functional regression models. Zou et al. (2013) considered a class of VCM without any covariate information and proposed a novel outlier detection procedure to identify any abnormal profile observations from a baseline dataset.

The aim of this paper is to systematically Cook's distance measures for VCM with functional responses (Zhu et al., 2012b). Compared to the diagnostic methods in Shen and Xu (2007), Chiou and Müller (2007), and Zou et al. (2013), we make several major contributions. First, we introduce local and global Cook's distances for quantifying the e ects of deleting a set of observed curves. We derive the exact form of local and global Cook's distances and then characterize their distributional properties, such as conditional mean and variance. Our results also establish a close link between Cook's distances in standard linear regression and those in VCM. Second, we introduce Cook's distances for quantifying the e ects of deleting observations at multiple grid points. We derive their exact form and their distributional properties, such as conditional mean and variance. Third, we introduce scaled Cook's distance and its associated diagnostic probability in order to evaluate the relatively influential level for different curve subsets across grid points (or subjects).

The rest of this paper is organized as follows. In Section 2, we briefly review VCM and its associated estimation method. In Section 3, we propose Cook's distances and scaled Cook's

distances. In Section 4, we carry out simulation studies and real data analysis using diffusion tensor imaging (DTI) tract data.

2 Methods

2.1 Model Setup

We consider a functional response $y_i(s)$ and a set of $p \times 1$ random covariates $\mathbf{x}_i$ for subject $i = 1, \ldots, n$. It is assumed that $y_i(s)$ is measured at the same L location points $s_1, \ldots, s_L$ and, without loss of generality, s_l's are independently and identically distributed random variables with a common density $f(s)$ on the support interval [0, 1]. Our varying coefficient model (VCM) is defined as

$$y_i(s) = \mathbf{x}_i^T \boldsymbol{\beta}(\mathbf{s}) + \eta_{\mathbf{i}}(\mathbf{s}) + \varepsilon_{\mathbf{i}}(\mathbf{s}) \quad \text{for} \quad \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}, \quad (1)$$

where $\boldsymbol{\beta}(s) = (\beta_1(s), \ldots, \beta_p(s))^T$ is a $p \times 1$ unknown vector of functionals of s, $\varepsilon_i(s)$ denotes the measurement error, and $\eta_i(s)$ characterizes individual curve variations. The $\varepsilon_i(s)$ is an independent and identical copy of a Gaussian stochastic process with mean zero and covariance $\Sigma_\varepsilon(s, t)$, whereas $_i(s)$ is an independent and identical copy of a Gaussian stochastic process with mean zero and covariance $\Sigma_\eta(s, t)$. Moreover, $\varepsilon_i(s)$ and $\eta_i(s)$ are mutually independent and $\varepsilon_i(s)$ and $\varepsilon_i(t)$ are independent for $s \neq t$.

Under some conditions, $\eta_i(s)$ admits the Karhunen-Loève expansion as given below. It is assumed that the covariance function $\Sigma_\eta(s, t)$ can be eigen-decomposed as

$$\Sigma_\eta(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t),$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are the ordered eigenvalues and $\{\varphi_j(s)\}_{j \geq 1}$ are the corresponding orthonormal eigenfunctions satisfying $\int_0^1 \phi_j^2(s)\, ds = 1$ and $\int_0^1 \phi_k(s)\, \phi_j(s)\, ds = 0$ for $k \neq j$. When $\int_0^1 \Sigma_\eta(s, s)\, ds < \infty$, it follows that $\sum_{j=1}^{\infty} \lambda_j < \infty$. If $\Sigma_\eta(s, t)$ is continuous on [0, 1], then we have the Karhunen-Loève expansion as follows:

$$\eta_i(s) = \sum_{j=1}^{\infty} \xi_{ij} \phi_j(s) \quad \text{for} \quad i = 1, \ldots, n,$$

where ξ_{ij} are independent normal random variables with mean zero and variance λ_j for $i = 1, \ldots, n$.

We briefly review some key results of a three-stage estimation procedure proposed by Zhu et al. (2012b). Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector $\mathbf{a}$, $\otimes$ denote the Kronecker product of two matrices, and $\dot{\beta}(s) = (d\beta_1(s)/ds, \ldots, d\beta_p(s)/ds)^T$. The $\boldsymbol{\beta}(s)$ can be estimated by using the local linear regression (LLR) (Fan and Gijbels, 1996) and given by

$$\hat{\boldsymbol{\beta}}(s) = [\mathbf{I}_p \otimes (1,0)]\, \Sigma(s,h_1)^{-1} \sum_{i=1}^{n}\sum_{l=1}^{L} K_{h_1}(s_l - s)\, [\mathbf{x}_i \otimes \boldsymbol{z}_{h_1}(s_l - s)]\, y_i(s_l), \quad (2)$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix, $\boldsymbol{z}_{h1}(s_l - s) = (1, (s_l - s)/h_1)^T$, and $K_{h_1}(s) = h_1^{-1} K(s/h_1)$ is a scaled kernel function with the bandwidth h_1 selected by using a leave-one-curve-out cross-validation criterion. Moreover, $\Sigma(s, h_1)$ is given by $\Sigma(s,h_1) = \Omega_X \otimes \Omega_Z(s,h_1)$, where $\Omega_X = \sum_{i=1}^{n} \mathbf{x}_i^{\otimes 2}$ and $\Omega_Z(s,h_1) = \sum_{l=1}^{L} K_{h_1}(s_l - s)\, [\boldsymbol{z}_{h_1}(s_l - s)]^{\otimes 2}$. Under some conditions, it can be shown that $\sqrt{n}\left\{\left[\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s) - \mathbf{O}\left(\mathbf{h}^{\mathbf{2}}\right)\right] : s \in [0,1]\right\}$ converges weakly to a centered Gaussian process $G(\cdot)$ with covariance matrix $\Sigma_\eta(s,t)\, E\left[\mathbf{x}^{\otimes 2}\right]^{-1}$ (Zhu et al., 2012b).

To estimate $\Sigma_\eta(s, t)$ and its associated eigenvalues and eigenfunctions, we employ the local smoothing technique to calculate a local linear estimate of $\eta_i(s)$, denoted by $\hat{\eta}_i(s)$ (Zhu et al., 2012b). After obtaining $\hat{\eta}_i(s)$, we estimate $\Sigma_\eta(s, t)$ by using the empirical covariance of the estimated $\hat{\eta}_i(s)$ as follows:

$$\hat{\Sigma}_\eta(s,t) = (n-p)^{-1} \sum_{i=1}^{n} \hat{\eta}_i(s)\, \hat{\eta}_i(t).$$

Then, we have the spectral decomposition of $\hat{\Sigma}_\eta(s,t)$ as follows:

$$\hat{\Sigma}_\eta(s,t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s)\, \hat{\phi}_j(t),$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq 0$ are the estimated eigenvalues and the $\hat{\phi}_l(t)$'s are the estimates of corresponding principal components. Moreover, it is common to choose the first N eigenvalue-eigenfunction pairs such that the cumulative proportion of eigenvalues, $CPV(N) = \sum_{k=1}^{N} \hat{\lambda}_k / \sum_{k=1}^{\infty} \hat{\lambda}_k$, is greater than a threshold, say 90%. Finally, we use the local constant method to estimate $\sigma^2(s)$ as follows:

$$\hat{\sigma}^2(s) = \frac{1}{n-1} \sum_{i=1}^{n}\sum_{l=1}^{L} \frac{K_{h_2}(s_l - s)\left[\hat{\xi}_i(s_l)\right]^2}{\sum_{k=1}^{L} K_{h_2}(s_k - s)},$$

where $\hat{\xi}_i(s) = y_i(s) - \mathbf{x}_i^T \hat{\beta}(s) - \hat{\eta}_i(s)$ and a bandwidth h_2 is selected by using a cross-validation score method.

2.2 Cook's Distance Measures

We introduce Cook's distance for deleting multiple curves, Cook's distance for deleting observations at multiple grid points, and their scaled Cook's distances in order to assess the influence of deleting a set of observations on the parameter estimates.

2.2.1 Cook's Distance for Deleting Multiple Curves—In practice, outlying curves are frequently found not only in observational studies but also in designed experiments. We propose Cook's distance to quantify the effects of deleting a set of observed curves, denoted by $\mathbf{Y}_I(s)$, on the estimate of $\boldsymbol{\beta}(s)$, where $I = \{i_1, \times\times\times, i_{nI}\}$ denotes a set of indices of n_I subjects. We consider the deletion of a set of curves in order to address the issue of masking and swamping effects when multiple outlying curves are present. It is well known that the single-case Cook's distance is not very efficient for addressing such issue. A subscript '$[I]$' denotes the relevant quantity of all observed curves with I deleted. Let $\mathbf{Y} = (Y_1, \times\times\times, Y_n)^T$, $\mathbf{Y}_I = (Y_{i1}, \times\times\times, Y_{inI})^T$ and $\mathbf{Y}_{[I]}$ be a subsample with $\mathbf{Y}_I$ deleted from $\mathbf{Y}$, where $Y_i = (y_i(s_1), \cdots, y_i(s))^T$ is a $L \times 1$ vector for $i = 1, \cdots, n$. Let $\hat{\boldsymbol{\beta}}_{[I]}(s)$ be the LLR estimator of $\boldsymbol{\beta}(s)$ based on the subsample $\mathbf{Y}_{[I]}$. Similar to (2), it can be shown that

$$\hat{\boldsymbol{\beta}}_{[I]}(s) = [\mathbf{I}_p \otimes (1,0)]\, \Sigma_{[I]}(s,h_1)^{-1} \sum_{i\in[I]}\sum_{l=1}^{L} K_{h_1}(s_l - s)\,[\mathbf{x}_i \otimes \boldsymbol{z}_{h_1}(s_l - s)]\, y_i(s_l),$$

where $\Sigma_{[I]}(s,h_1) = \Omega_{X[I]} \otimes \Omega_Z(s,h_1)$ and $\Omega_{X[I]} = \sum_{i\in[I]} \mathbf{x}_i^{\otimes 2}$. Furthermore, for any s and bandwidth parameter h_1, we introduce a smoothed response given by

$$\tilde{y}_i(s;h_1) = \sum_{l=1}^{L} \tilde{K}_{h_1}(s_l - s)\, y_i(s_l) = \tilde{\mathbf{K}}_{h_1}(s)^T Y_i, \quad (3)$$

where $\tilde{K_{h1}}(s_l - s) = (1, 0)\Omega_Z(s, h)^{-1} K_{h1}(s_l - s)z_{h1}(s_l - s)$ and $\tilde{\mathbf{K}_{h1}}(s) = \tilde{K_{h1}}(s_1) - s), \cdots, \tilde{K_{h1}}(s_L - s))^T$ is an $L \times 1$ vector. Therefore, $\hat{\boldsymbol{\beta}}(s)$ and $\hat{\boldsymbol{\beta}}_{[I]}(s)$ can be, respectively, rewritten as

$$\hat{\boldsymbol{\beta}}(s) = \Omega_X^{-1} \sum_{i=1}^{n} \mathbf{x}_i \tilde{y}_i(s;h_1) \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{[I]}(s) = \Omega_{X[I]}^{-1} \sum_{i\in[I]} \mathbf{x}_i \tilde{y}_i(s;h_1). \quad (4)$$

For a given set I, we introduce a local Cook's distance at each s (denoted by $CD_I(s)$) and a global Cook's distance over [0, 1] (denoted by CD_I), which are defined as, respectively,

$$CD_I(s) = \left[\hat{\boldsymbol{\beta}}_{[I]}(s) - \hat{\boldsymbol{\beta}}(s)\right]^T M_{[I]}(s) \left[\hat{\boldsymbol{\beta}}_{[I]}(s) - \hat{\boldsymbol{\beta}}(s)\right] \quad \text{and} \quad CD_I = \int_0^1 CD_I(s)\, ds, \quad (5)$$

where $M_{[I]}(s)$ is a $p \times p$ known matrix setting to be the inverse of $\Sigma_\eta(s,s)\,\Omega_X^{-1}$. Alternatively, we may use $\mathbf{Y}_{[I]}$ to estimate $\Sigma_\eta(s,s)$ and in turn $M_{[I]}(s)$. More generally, suppose that one is interested in a subset of $\boldsymbol{\beta}(s)$, say p_1 linearly independent combinations of $\boldsymbol{\beta}(s)$ given by $C^T\boldsymbol{\beta}(s)$, in which C is a $p \times p_1$ matrix with rank(C) = p_1 (Cook and Weisberg,

1982). We denote the local and global partial Cook's distances for quantifying the influence of deleting the subset I on $C^T\hat{\boldsymbol{\beta}}$ by $\mathrm{CD}_{I,C}(s)$ and $\mathrm{CD}_{I,C}$, respectively, which can be defined as

$$CD_{I,C}(s)=\left(\left[\hat{\boldsymbol{\beta}}_{[I]}(s)-\hat{\boldsymbol{\beta}}(s)\right]^T C\left\{C^T M_{[I]}(s)\,C\right\}^{-1/2}\right)^{\otimes 2} \quad \text{and} \quad CD_{I,C}=\int_0^1 CD_{I,C}(s)\,ds.$$

Throughout the paper, we do not distinguish $\mathrm{CD}_I(s)$ and CD_I from $\mathrm{CD}_{I,C}(s)$ and $\mathrm{CD}_{I,C}$ for notational simplicity, even though we may focus on a subset of $\boldsymbol{\beta}$.

The local and global Cook's distances can be used to assess the influential level of the subset I at each location s or across [0, 1]. We may regard a subset I as being influential if either the value of $\mathrm{CD}_I(s)$ (or CD_I) is relatively large compared with other Cook's distances or the magnitude of $\mathrm{CD}_I(s)$ (or CD_I) is greater than the critical value of a distribution, which will be derived below.

We define some notation as follows. Let $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{X}n)^T$ be an $n \times p$ matrix, $P_X=\mathbf{X}\Omega_X^{-1}\mathbf{X}^T$ is an $n \times n$ matrix, $\tilde{Y}(s, h_1) = (\tilde{y}_1(s; h_1), \cdots, \tilde{y}_n(s; h_1))^T$ is an $n \times 1$ vector, and $\hat{\mathbf{u}}(s,h_1)=\left(\tilde{y}_1(s;h_1)-\mathbf{x}_1^T\hat{\boldsymbol{\beta}}(s),\cdots,\tilde{y}_n(s;h_1)-\mathbf{x}_n^T\hat{\boldsymbol{\beta}}(s)\right)^T$ is an $n \times 1$ vector. Let $D_I = [\mathbf{e}_{i1}, \cdots, \mathbf{e}_{inI}]$ be an $n \times n_I$ matrix, where $\mathbf{e}_k$ is an $n \times 1$ vector being 1 at the k-th element and 0 elsewhere for $k \in I$, and $\mathbf{S} = (s_1, \cdots, s_L)^T$ denotes a vector of grid points. We obtain the following theorems, whose detailed assumptions and proofs can be found in the Appendix A1.

Theorem 1 *We have the following results:*

(a) $\hat{\boldsymbol{\beta}}(s)-\hat{\boldsymbol{\beta}}_{[I]}(s)+\Omega_X^{-1}X_I^T\left(\mathbf{I}_{n_I}-P_I\right)^{-1}\hat{\mathbf{u}}_I(s,h_1)$, *where* $\hat{\mathbf{u}}_I(s,h_1)=D_I^T\hat{\mathbf{u}}(s,h_1)$, $X_I=D_l^TX$, *and* $P_I=X_I\Omega_X^{-1}X_I^T=D_I^TP_XD_I$.

(b) $CD_I(s)=\tilde{Y}(s,h_1)^T(\mathbf{I}_n-P_X)W_I(\mathbf{I}_n-P_X)\tilde{Y}(s,h_1)/\Sigma_\eta(s,s)$, *where* $W_I=D_I\left(\mathbf{I}_{n_I}-P_I\right)^{-1}P_I\left(\mathbf{I}_{n_I}-P_I\right)^{-1}D_I^T$.

(c) $CD_I=tr\left\{\mathbf{Y}^T(\mathbf{I}_n-P_X)W_I(\mathbf{I}_n-P_X)\mathbf{Y}\mathbf{K}_{h_1,\eta}\right\}$, *where tr(A) is the trace of a matrix A and* $\mathbf{K}_{h_1,\eta}=\int_0^1\tilde{\mathbf{K}}_{h_1}(s)^{\otimes 2}/\Sigma_\eta(s,s)\,ds$.

Theorem 1 establishes the exact form of $\hat{\boldsymbol{\beta}}(s)-\hat{\boldsymbol{\beta}}_{[I]}(s)$, $\mathrm{CD}_I(s)$, and CD_I for an arbitrary set I. These results for our VCM are close to those for the classical linear regression model (Cook and Weisberg, 1982) except that the smoothed response is used. When $n_I = 1$ and $I = \{i\}$, $\hat{\boldsymbol{\beta}}(s)-\hat{\boldsymbol{\beta}}_{[I]}(s)$ and $\mathrm{CD}_I(s)$, respectively, reduce to $\Omega_X^{-1}\mathbf{x}_i\hat{u}_i(s,h_1)/(1-P_i)$ and $\hat{u}_i(s, h_1)^2P_i/\{\Sigma_\eta(s, s)(1 - P_i)^2\}$, $\hat{u}_i(s,h_1)=\tilde{y}_i(s;h_1)-\mathbf{x}_i^T\hat{\boldsymbol{\beta}}(s)$. Moreover, for $I = \{i\}$, CD_I reduces to

$\int_0^1 \hat{u}_i(s,h_1)^2 ds P_i/(1-P_i)^2 = \mathbf{e}_i^T (\mathbf{I}_n - P_X)\, \mathbf{Y}\mathbf{K}_{h_1,\eta}\mathbf{Y}^{\mathrm{T}} (\mathbf{I}_n - P_X)\, \mathbf{e}_i P_i/(1-P_i)^2$. Large $\mathrm{CD}_{\{i\}}$ values correspond to curves with large residual line $\{\hat{u}_i(s, h_1) : s \in [0, 1]\}$ and/or high leverage P_i, which have a large effect on the estimate of $\boldsymbol{\beta}(s)$. Furthermore, inspecting $\mathrm{CD}_{\{i\}}(s_l)$ can reveal the influence level of the i-th curve at location s_l.

To determine the magnitude of $\mathrm{CD}_I(s)$ and CD_I, we need to characterize the distributional properties of $\mathrm{CD}_I(s)$ and CD_I as follows.

Theorem 2 *We have the following results:*

(a) Let Σ_Y be the covariance matrix of Y_i. Then, $\mathrm{CD}_I(s)$ *and* CD_I *can be rewritten as*

$$CD_I(s) = G^T \left(\left(\Sigma_Y^{1/2} \frac{\tilde{\mathbf{K}}_{h_1}(s)^{\otimes 2}}{\Sigma_\eta(s,s)} \Sigma_Y^{1/2} \right) \otimes \left[(\mathbf{I}_n - P_X)\, W_I\, (\mathbf{I}_n - P_X) \right] \right) G$$

and

$$CD_I = G^T \left(\left(\Sigma_Y^{1/2} \mathbf{K}_{h_1,\eta} \Sigma_Y^{1/2} \right) \otimes \left[(\mathbf{I}_n - P_X)\, W_I\, (\mathbf{I}_n - P_X) \right] \right) G,$$

where $G = Vec\left((\mathbf{Y} - \mathbf{X}\mathbf{B})\, \Sigma_Y^{1/2} \right)$ *is an* $nL \times 1$ *vector with zero mean,* $\mathbf{I}_{nL}$ *covariance matrix,* $Vec(\cdot)$ *denotes the matrix vectorization, and* $\mathbf{B} = [\boldsymbol{\beta}(s_1), \ldots, \boldsymbol{\beta}(s_L)]$ *is a* $p \times L$ *matrix of unknown true parameters.*

(b) The conditional means of $CD_I(s)$ *and* CD_I *are, respectively, given by*

$$E\left[CD_I(s) \,|\mathbf{X}, \mathbf{S}\right] = tr\left\{ \left(\mathbf{I}_{n_I} - P_I \right)^{-1} P_I \right\} \Sigma_Y(s, h_1) / \Sigma_\eta(s,s)$$

and

$$E\left[CD_I | \mathbf{X}, \mathbf{S}\right] = tr\left\{ \left(\mathbf{I}_{n_I} - P_I \right)^{-1} P_I \right\} \int_0^1 \Sigma_Y(s, h_1) / \Sigma_\eta(s,s)\, ds,$$

where $\Sigma_Y(s, h_1) = \sum_{l,l'=1}^{L} \tilde{K}_{h_1}(s_l - s)\, \tilde{K}_{h_1}(s_{l'} - s) \left\{ \Sigma_\eta(s_l, s_{l'}) + \Sigma_\epsilon(s_l, s_{l'}) \right\}$.

(c) The conditional variances of $CD_I(s)$ *and* CD_I *are, respectively, given by*

$$Var\left[CD_I(s) \,|\mathbf{X}, \mathbf{S}\right] = 2tr\left\{ \left(\mathbf{I}_{n_I} - P_I \right)^{-1} P_I \left(\mathbf{I}_{n_I} - P_I \right)^{-1} P_I \right\} \Sigma_Y(s, h_1)^2 \Sigma_\eta(s,s)^2$$

and

$$Var\left[CD_I|\mathbf{X},\mathbf{S}\right]=2tr\left\{\left\{\left(\mathbf{I}_{n_I}-P_I\right)^{-1}P_I\right\}^2\right\}tr\left(\mathbf{K}_{h_1,\eta}\Sigma_Y\mathbf{K}_{h_1,\eta}\Sigma_Y\right).$$

Theorem 2 characterizes the distributional properties of $CD_I(s)$ and CD_I. Theorem 2 (a) shows that both $CD_I(s)$ and CD_I can be represented as a quadratic form of G. When $G \sim N(0, \mathbf{I}_{nL})$, both $CD_I(s)$ and CD_I follow a weighted χ^2 distribution. Theorem 2 (b) and (c) give the conditional means and variances of $CD_I(s)$ and CD_I given covariates. Although the means of Cook's distances do not depend on the distribution of G, their variances depend greatly on the distribution of G. When $h_1 = o(1)$ and $Lh_1 \to \infty$, it can be shown that

$$\Sigma_Y\left(s,h_1\right)/\Sigma_\eta\left(s,s\right)\to 1 \quad \text{and} \quad tr\left(\mathbf{K}_{h_1,\eta}\Sigma_Y\mathbf{K}_{h_1,\eta}\Sigma_Y\right)\to\int_0^1\int_0^1\rho_\eta\left(s,s'\right)^2dsds',$$

where $\rho_\eta\left(s,s'\right)=\Sigma_\eta\left(s,s'\right)/\sqrt{\Sigma_\eta\left(s,s\right)\Sigma_\eta\left(s',s'\right)}$ is the correlation function of $\eta_i(s)$. Moreover, for $I = \{i\}$, we have

$$\mathrm{E}\left[CD_i\left(s\right)|\mathbf{X},\mathbf{S}\right]=\frac{\Sigma_Y\left(s,h_1\right)P_i}{\Sigma_\eta\left(s,s\right)\left(1-P_i\right)} \quad \text{and} \quad Var\left[CD_i\left(s\right)|\mathbf{X},\mathbf{S}\right]=2\frac{\Sigma_Y(s,h_1)^2P_i^2}{\Sigma_\eta(s,s)^2(1-P_i)^2}.$$

2.2.2 Cook's Distance for Deleting Observations at Multiple Grid Points—In practice, observations at multiple grid points may be abnormal due to varying amounts of noise of diverse origins, such as numerous physiological processes, rigid body motion, and nonrigid motion. Hence, we propose Cook's distance for quantifying the effects of deleting a set of observations at multiple grid points on the estimate of $\boldsymbol{\beta}(s)$. Let $I_S = \{l_1, \cdots, l_{LI}\}$ denote a set of indices of L_I grid points. A subscript '$[I_S]$' denotes the relevant quantity of all observations at the grid points with I_S deleted. Let $\mathbf{Y}_{[I_S,n]}$ be a subsample of $\mathbf{Y}$ with $\mathbf{Y}_{I_S,n} = \{Y(s_{lk}) = (y_1(s_{lk}), \ldots, y_n(s_{lk}))^T : l_k \in I_S\}$ deleted. Let $\hat{\boldsymbol{\beta}}_{\left[I_S,n\right]}(s)$ be the LLR estimator of $\boldsymbol{\beta}(s)$ based on the subsample $\mathbf{Y}_{[I_S,n]}$. Similar to (2), it can be shown that

$$\hat{\boldsymbol{\beta}}\left[I_S,n\right](s)=\left[\mathbf{I}_p\otimes(1,0)\right]\Sigma_{\left[I_{S,n}\right]}(s,h_1)^{-1}\sum_{i=1}^{n}\sum_{l\in\left[I_S\right]}K_{h_1}\left(s_l-s\right)\left[\mathbf{x}_i\otimes\boldsymbol{z}_{h_1}\left(s_l-s\right)\right]y_i\left(s_l\right), \quad (6)$$

where $\Sigma_{\left[I_S\right]}(s,h_1)=\Omega_X\otimes\Omega_{Z,\left[I_S\right]}(s,h_1)$ and $\Omega_{Z,\left[I_S\right]}(s,h_1)=\sum_{l\in\left[I_S\right]}K_{h_1}\left(s_l-s\right)\left[\boldsymbol{z}_{h_1}\left(s_l-s\right)\right]^{\otimes 2}$.

For a given set I_S, we introduce a local Cook's distance at each s (denoted by $CD_{I_S,n}(s)$) and a global Cook's distance over [0, 1] (denoted by $CD_{I_S,n}$) defined as

$$CD_{I_S,n}(s) = \left[\hat{\boldsymbol{\beta}}_{[I_{S,n}]}(s) - \hat{\boldsymbol{\beta}}(s)\right]^T \Omega_X \left[\hat{\boldsymbol{\beta}}_{[I_{S,n}]}(s) - \hat{\boldsymbol{\beta}}(s)\right] / \Sigma_\eta(s,s) \quad (7)$$

and

$$CD_{I_S,n} = \int_0^1 CD_{I_S,n}(s)\, ds, \quad (8)$$

respectively.

Furthermore, for any s and h_1, we introduce a smoothed response given by

$$\tilde{y}_{i[I_S]}(s;h_1) = \sum_{i \in [I_S]} \tilde{K}_{h_1,[I_S]}(s_l - s)\, y_i(s_l), \quad (9)$$

where $K_h\tilde{1}_{[IS]}(s_l - s) = (1, 0)\Omega_{Z,[IS]}(s, h_1)^{-1}K_{h1}(s_l - s)\mathbf{z}_{h1}(s_l - s)$. Therefore, $\hat{\boldsymbol{\beta}}(s)$ and $\hat{\boldsymbol{\beta}}_{[I_S,n]}(s)$ can be, respectively, rewritten as

$$\hat{\boldsymbol{\beta}}(s) = \Omega_X^{-1}\sum_{i=1}^n \mathbf{x}_i \tilde{y}_i(s;h_1) \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{[I_S,n]}(s) = \Omega_X^{-1}\sum_{i=1}^n \mathbf{x}_i \tilde{y}_{i[I_S]}(s;h_1). \quad (10)$$

We need some notation as follows. Let $Z_{h1}(s) = [\mathbf{z}_{h1}(s_1 - s)^T, \cdots, \mathbf{z}_{h1}(s_L - s)^{TT}$ be an $L \times 2$ matrix, $K_D(s, h_1) = \text{diag}(K_{h1}(s_1 - s), \times \times \times, K_{h1}(s_L - s))$ be an $L \times L$ diagonal matrix, $P_Z(s, h_1) = K_D(s, h_1)^{1/2}Z_{h1}(s)\Omega_Z(s, h_1)^{-1}Z_{h1}(s)^T K_D(s, h_1)^{1/2}$ be an $L \times L$ matrix, and $D_{IS} = [\mathbf{e}_{l1}, \cdots, \mathbf{e}_{lLI}]$ be an $L \times L_I$ matrix.

We show the distributional properties of $CD_{IS,n}(s)$ and $CD_{IS,n}$ as follows.

Theorem 3 *We have the following results:*

(a) $\hat{\boldsymbol{\beta}}(s) - \hat{\boldsymbol{\beta}}_{[I_S,n]}(s) = \Omega_X^{-1}\sum_{i=1}^n \mathbf{x}_i R_{I_S}(s,h_1)^T Y_i$, *where*

$P_{\tilde{Z},I_S}(s,h_1) = D_{I_S}{}^T P_{\tilde{Z}}(s,h_1) D_{I_S}$ and

$$R_{I_S}(s,h_1)^T = (1,0)\,\Omega_Z(s,h_1)^{-1} Z_{h_1}(s)^T K_D(s,h_1)^{1/2} D_{I_S} \left\{\mathbf{I}_{L_I} - P_{\tilde{Z},I_S}(s,h_1)\right\}^{-1} D_{I_S}{}^T \left\{\mathbf{I}_L - P_{\tilde{Z}}(s,h_1)\right\} K_D(s,h_1)^{1/2}$$

.

(b) $CD_{I_S,n}(s) = Vec(\boldsymbol{Y})^T \left\{R_{I_S}(s,h_1)^{\otimes 2}/\Sigma_\eta(s,s) \otimes \mathbf{X}\Omega_X^{-1}\mathbf{X}^T\right\} Vec(\boldsymbol{Y})$.

(c) $CD_{I_S,n} = Vec(\boldsymbol{Y})^T \left(\mathbf{R}_{I_S,h_1,\eta} \otimes \mathbf{X}\Omega_X^{-1}\mathbf{X}^T\right) Vec(\boldsymbol{Y})$, *where*

$\mathbf{R}_{I_S,h_1,\eta} = \int_0^1 R_{I_S}(s,h_1)^{\otimes 2}/\Sigma_\eta(s,s)\, ds$.

(d) The conditional means of $CD_{IS,n}(s)$ and $CD_{IS,n}$ are, respectively, given by

$$E\left[CD_{I_S,n}(s)\,|\mathbf{X},\mathbf{S}\right]=R_{I_S}(s,h_1)^T\left\{p\Sigma_Y+\mathbf{B}^T\Omega_X\mathbf{B}\right\}R_{I_S}(s,h_1)/\Sigma_\eta(s,s)$$

and

$$E\left[CD_{I_S,n}|\mathbf{X},\mathbf{S}\right]=tr\left(\left\{p\Sigma_Y+\mathbf{B}^T\Omega_X\mathbf{B}\right\}\mathbf{R}_{I_S,h_1,\eta}\right).$$

(e) The conditional variances of $CD_{IS,n}(s)$ and $CD_{IS,n}$ are, respectively, given by

$$\begin{aligned}&Var\left[CD_{I_S,n}(s)\,|\mathbf{X},\mathbf{S}\right]\\&=2ptr\left(\left\{R_{I_S}(s,h_1)^{\otimes 2}\Sigma_Y\right\}^2\right)\\&/\Sigma_\eta(s,s)^2+tr\left(R_{I_S}(s,h_1)^{\otimes 2}\mathbf{B}^T\Omega_X\mathbf{B}R_{I_S}(s,h_1)^{\otimes 2}\Sigma_Y\right)\\&/\Sigma_\eta(s,s)^2\end{aligned}$$

and

$$Var\left[CD_{I_S,n}|\mathbf{X},\mathbf{S}\right]=2ptr\left\{\mathbf{R}_{I_S,h_1,\eta}\Sigma_Y\mathbf{R}_{I_S,h_1,\eta}\Sigma_Y\right\}+tr\left(\mathbf{B}^T\Omega_X\mathbf{B}\mathbf{R}_{I_S,h_1,\eta}\Sigma_Y\mathbf{R}_{I_S,h_1,\eta}\right).$$

Theorem 3 (a) characterizes the effect of deleting the observations in $\mathbf{Y}_{IS,n}$ on the *i*-th smoothed curve. This result can be regarded as an extension of an existing result for local polynomial regression (Kim et al., 2001). Theorem 3 (b) and (c) show that both $CD_{IS,n}(s)$ and $CD_{IS,n}$ can be written as a quadratic form of Vec(**Y**) and follow a noncentral weighted χ^2 distribution. Compared with *G* for $CD_I(s)$ and CD_I, Vec(**Y**) in $CD_{IS,n}(s)$ and $CD_{IS,n}$ is not mean-centered, since we only delete the observations in *Y* (s_{lk}), not covariates. Theorem 3 (d) and (e) give the conditional means and variances of $CD_{IS,n}(s)$ and $CD_{IS,n}$ given covariates. Although the means of Cook's distances do not depend on the Gaussianity of Vec(**Y**), their variances depend on the Gaussian distribution of Vec(**Y**).

2.2.3 Scaled Cook's Distances—A major size issue regarding Cook's distance is that the magnitude of Cook's distance is positively associated with the amount of perturbation to VCM introduced by deleting a subset of observations. Specifically, a large value of Cook's distance can be caused by deleting a subset with a larger number of observations and/or other causes, such as the presence of influential observations in the deleted subset (Zhu et al., 2012a). To delineate the cause of a large Cook's distance, Zhu et al. (2012a) proposed several types of scaled Cook's distances and their associated diagnostic probabilities to account for different degrees of perturbation introduced by deleting subsets with different numbers of observations.

Following Zhu et al. (2012a), we introduce a scaled Cook's distance by matching a pair of features, mean and standard deviation, for all Cook's distance measures. By matching the mean and standard deviation, we can at least ensure that the centers and scales of the scaled

Cook's distances for different subsets are the same when the proposed VCM is the true data generator. Specifically, we can define the scaled Cook's distances for deleting multiple curves as follows:

$$SCD_I(s) = \frac{CD_I(s) - E[CD_I(s)|\mathbf{X},\mathbf{S}]}{Std[CD_I(s)|\mathbf{X},\mathbf{S}]} \quad \text{and} \quad SCD_I = \frac{CD_I - E[CD_I|\mathbf{X},\mathbf{S}]}{Std[CD_I|\mathbf{X},\mathbf{S}]},$$

where 'E' and 'Std' denote the expectation and standard deviation, respectively. Furthermore, we can define the scaled Cook's distances for deleting observations at multiple grid points and denote them by $SCD_{I_S,n}(s)$ and $SCD_{I_S,n}$. The scaled Cook's distances measure the standardized influential level of the set I when VCM is true. Moreover, it is possible for $SCD_I(s)$ to be negative unlike $CD_I(s)$. Since the means and standard deviations of Cook's distances have been derived in Theorems 2 and 3, it is computationally straightforward to calculate the scaled Cook's distances.

We use $SCD_I(s)$ (or SCD_I) to evaluate the relatively influential level for different curve subsets I across grid points (or subjects). A large value of $SCD_I(s)$ (or SCD_I) indicates that the curves in subset I are relatively influential. Therefore, for any two curve subsets I_1 and I_2, the probability of observing the event $\{SCD_{I1}(s) > SCD_{I2}(s)\}$ should be reasonably close to that of observing the event $\{SCD_{I1}(s) < SCD_{I2}(s)\}$. Similar comments are applicable to SCD_I. Thus, the SCD_I and $SCD_I(s)$ are roughly comparable.

We can employ diagnostic probabilities associated with the Cook's distances as an alternative approach to deal with the size issue. Let $\mathscr{M}$ be the VCM model in (1) proposed to fit the data. Specifically, we define the local and global diagnostic probabilities for deleting multiple curves as follows:

$$p_I(s) = P\left(CD_{I,\mathscr{M}}(s) < CD_I(s)|\mathscr{M},\mathbf{X}\right) \quad \text{and} \quad p_I = P\left(CD_{I,\mathscr{M}} < CD_I|\mathscr{M},\mathbf{X}\right),$$

where $CD_{I,\mathscr{M}}(s)$ and $CD_{I,\mathscr{M}}$ denote local and global random Cook's distances, respectively, and $P(\cdot|\mathscr{M},\mathbf{X})$ denotes the conditional probability when $\mathscr{M}$ is the true generator and $\mathbf{X}$ is fixed. Furthermore, we can define local and global diagnostic probabilities for deleting observations at multiple grid points and denote them by $p_{IS,n}(s)$ and $p_{IS,n}$, respectively. Compared with the scaled Cook's distances, since the diagnostic probabilities vary in a fixed interval [0, 1], it is much easier to assess their magnitude. We regard a subset I as being influential if the value of $p_I(s)$ (or p_I) is close to 1. However, the diagnostic probabilities strongly depend on the distributions of $\eta_i(s)$ and $\in_i(s)$ in model (1). In practice, based on Theorem 2, $p_I(s)$ and p_I can be approximated by simulation method. For example, we generate B samples from $N_{nL}(0, \mathbf{I})$ and then use Theorem 2 (a) to compute $CD_I^{(b)}(s)$ and $CD_I^{(b)}$ for $b = 1, \ldots, B$. Finally, $p_I(s)$ and p_I are approximated by

$$\hat{p}_I(s) = \frac{\sum_{b=1}^{B} \mathbf{1}\left[\mathbf{CD}_{\mathbf{I}}^{(\mathbf{b})}(\mathbf{s}) < \mathbf{CD}_{\mathbf{I}}(\mathbf{s})\right]}{B} \quad \text{and} \quad \hat{p}_I = \frac{\sum_{b=1}^{B} \mathbf{1}\left[\mathbf{CD}_{\mathbf{I}}^{(\mathbf{b})} < \mathbf{CD}_{\mathbf{I}}\right]}{B},$$

respectively, where $1(\cdot)$ is an indicator function. Similarly, $p_{I_S,n}(s)$ and $p_{I_S,n}$ can be approximated by using the parametric bootstrap random method which is described in Appendix A2.

3 Simulations and Real Data Analysis

In this section, we evaluate the finite sample performance of our diagnostic methods through the simulated datasets and a real data example.

3.1 Simulations

The goal of our simulations is to examine the finite sample performance of our diagnostic measures for detecting influential curves and/or influential observations at multiple grid points in functional data. Following Zhu et al. (2012b), we generated data sets according to model (1). We set $n = 50$, $L = 40$, $\eta_i(s) = \xi_{i1}\varphi_1(s) + \xi_{i2}\varphi_2(s)$, $s_l \sim U[0, 1]$, $\varepsilon_i(s_l) \sim N(0, 0.4^2)$, and $\mathbf{x}_i = (1, x_{i1}, x_{i2})^T$, where $x_{i1} \sim Bernoulli(0.5)$, $x_{i2} \sim N(0, 1)$, $\xi_{i1} \sim N(0, 0.36)$, and $\xi_{i2} \sim N(0, 0.25)$. Moreover, we set the functional coefficients and eigenfunctions as follows:

$$\begin{aligned} \beta_1(s) &= 2s^2, \quad \beta_2(s) = 3(1-s)^2, \quad \beta_3(s) = 4s(1-s), \\ \phi_1(s) &= \sqrt{2}\sin(2\pi s), \quad \phi_2(2) = \sqrt{2}\cos(2\pi s). \end{aligned}$$

We used a normal kernel function and $B = 100$ for computing diagnostic probabilities.

To add some influential curves or grid points, we modified the simulated data set according to four different scenarios, including (i) a single outlying curve; (ii) three outlying curves; (iii) a single curve with several outlying observations; and (iv) a single outlying time point. Figure 1 shows a graphical illustration of Scenarios (i)-(iv). For Scenario (i), we changed the curve $y_{10}(s)$ into $y_{10}(s)+5\sin(\pi s)$ in order to create a single atypical curve. For Scenario (ii), we changed the curves $y_{10}(s)$, $y_{20}(s)$, and $y_{30}(s)$ into $y_{10}(s) + 5\sin(\pi s)$, $y_{20}(s) - 4\cos(\pi s)$, and $y_{30}(s) + 15s(1 - s)$, respectively. For Scenario (iii), we added $5\sin(\pi s)$ to the values of $y_{10}(s)$ at two segments of grid points including $\{s_{11}, \cdots, s_{20}\}$ and $\{s_{31}, \cdots, s_{35}\}$ in order to generate an atypical curve with influential observations at several segments of consecutive grid points, but not for the whole curve. For Scenario (iv), we added 1 to all observations at the 10th grid point in order to result in influential observations at a single grid point.

Figures 2-4, present the diagnostic measures for Scenarios (i)-(iii), respectively. In each figure, we include the global Cook's distance, scaled global Cook's distance, and diagnostic probabilities for deleting a single case or two cases. It reveals that our diagnostic measures are quite effective in all scenarios. Specifically, for Scenarios (i): The maximum global Cook's distance is CD_{10} with the value of 32.3661, whereas the second largest is CD_{25} with the value of 7.1788. Thus, the global Cook's distance for the outlying curve is significantly larger those for the non-outlier curves (Figure 2 (a)). To delineate the cause of such large

Cook's distance, scaled global Cook's distances are also computed and presented in Figure 2 (c) such that $SCD_{10} = 17.3043$ is much larger than the second largest $SCD_{25} = 2.6105$. Moreover, the diagnostic probabilities presented in Figure 2 (e) lead to the same conclusion. Therefore, the 10th outlying curve can be effectively identified by our diagnostic measures. We also included the diagnostic results associated with the deletion of two cases for Scenarios (i) in Figure 2 (b), (d), and (f), respectively. Similarly, we presented diagnostic results for Scenarios (ii) and (iii) in Figures 3 and 4. Moreover, we computed the local Cook's distances, scaled local Cook's distances, and diagnostic probabilities for deleting a single case or two cases under the three Scenarios. For the sake of space, we only present the local Cook's distances and their scaled local Cook's distances in Figure 5. Figure 5 shows that our diagnostic methods perform quite well under the above three Scenarios by computing the Cook's distances at each grid point. Figure 6 presents that both global and local diagnostic measures can detect an influential grid point for Scenario (iv). Moreover, it can be seen that the 10th grid point also a ects nearby observations by the kernel smoothing method used in the estimation procedure.

3.2 Real Data Analysis

To apply our diagnostic measures to a real example, we analyze a diffusion tensor imaging (DTI) tract data which was reported in Zhu et al. (2010) and Zhu et al. (2011). Since diffusion tensor imaging (DTI) can track the effective diffusion of water in the human brain in vivo, it can map the structure and orientation of the white matter fiber tracts of the brain. From DTI tracts, we often calculate a 3 × 3 matrix, called a diffusion tensor (DT), and its tensor-derived quantities to quantify the degree of diffusivity and the directional dependence of water diffusion for each voxel. Such diffusion quantities commonly include three eigenvalue-eigenvector pairs of diffusion tensor and other related parameters, such as mean diffusivity (MD). A wealth of neuroimaging studies has used these diffusion quantities as a marker of white matter tract maturation and integrity for neuropsychiatric and neuro-degenerative disorders.

The data set consists of 263 healthy full-term infants. All infants were younger than one year old, and the written informed consent was obtained from their parents before imaging acquisition. The gestational ages at MR scanning were 298 ± 17.>6 (Mean±Std) days and their range was 262-433. For details about statistical results and imaging processing procedure, see Zhu et al. (2010) and Zhu et al. (2011).

We fitted the VCM (1) to the MD values along a Genu tract with 45 grid points obtained from all subjects. We used gender and gestational age as covariates. Then, we obtained $\hat{\boldsymbol{\beta}}(s), \hat{\boldsymbol{\eta}}(s)$, and the eigenvalues and eigenfunctions of $\hat{\Sigma}_{\eta}(s,t)$. Subsequently, we calculated various Cook's distance measures for a signal-case deletion. In Figure 7 (a), the global Cook's distances detected Case 65 as the most influential subject. Figure 7 (c) shows some other influential cases including 28, 31, 64, 199, and 200. In Figure 7 (e), the diagnostic probabilities for the global Cook's distances confirm that Cases 28, 31, 64, 65, 199, and 200 are relatively influential.

In addition, we present various diagnostic measures for two-case deletion in Figure 7 (b), (d), and (f). The combinations of Case 65 with all other cases have relatively larger Cook's distances and diagnostic probabilities and Cases (31, 65) is the most influential pair. Moreover, almost all combinations of Case 28 (or 31) with all other cases have quite larger Cook's distances and diagnostic probabilities than others. All results reveal that Case 65 is the most influential, while Cases 28, 31, 64, 199, and 200 are comparably influential.

We re-estimated $\beta(s)$ after deleting each of Cases 1, 28, 31, and 65. Here, Case 1 was selected from the un-influential curves. Figure 8 displays the estimated coefficient functions before and after deleting each of the four Cases. The $\beta_1(s)$, $\beta_2(s)$, and $\beta_3(s)$ correspond to an intercept, gender, and gestational age, respectively. Figure 8 shows that deleting Cases 28, 31, and 65 have more effects on the estimates of coefficient functions than deleting Case 1.

4 Discussion

We have developed various Cook's distance measures for assessing the influence of both atypical curves and observations under varying coefficient model for functional responses. Our Cook's distance measures include Cook's distances for deleting multiple curves and for deleting multiple grid points, and their scaled Cook's distances. We have systematically investigated some theoretical properties of the proposed diagnostic measures. Simulation studies and a real DTI tract data have confirmed that our Cook's distances perform well under various scenarios.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Chiou J, Müller HG. Diagnostics for functional regression via residual processes. Computational Statistics and Data Analysis. 2007; 51(10):4849–4863.

Cook RD. Deletion of influential observation in linear regression. Techno-metrics. 1977; 19:15–18.

Cook, RD.; Weisberg, S. Residuals and Influence in Regression. Chapman & Hall; London: 1982.

Davison A, Tsai C. Regression model diagnostics. International Statistical Review. 1992; 60:337–353.

Fan, J.; Gijbels, I. Local Polynomial Modelling and Its Applications. Chapman & Hall; London: 1996.

Fan J, Zhang W. Statistical methods with varying coefficient models. Statistics and Its Interface. 2008; 2:179–195. [PubMed: 18978950]

Faraway J. Regression analysis for a functional response. Technometrics. 1997; 39:254–261.

Hastie TJ, Tibshirani RJ. Varying-coefficient models. J. Roy.Statist. Soc. B. 1993; 55:757–796.

Kim C, Lee Y, Park B. Cook's distance in local polynomial regression. Statistics and Probability Letters. 2001; 54:33–40.

Shen Q, Faraway J. An F test for linear models with functional responses. Statistica Sinica. 2004; 14:1239–1257.

Shen Q, Xu H. Diagnostics for linear models with functional responses. Technometrics. 2007; 49(1): 26–33.

Wei, B. Exponential Family Nonlinear Models. Lecture Notes in Statist. Vol. 130. Springer; Singapore: 1998.

Yao F, Müller HG, Wang J. Functional data analysis for Sparse longitudianl data. Journal of the American Statistical Association. 2005; 100:577–590.

Zhang J, Chen J. Statistical inference for functional data. The Annals of Statistics. 2007; 35:1052–1079.

Zhu H, Ibrahim JG, Cho H. Pertubation and scaled cook's distance. The Annals of Statistics. 2012a; 40:785–811. [PubMed: 23152646]

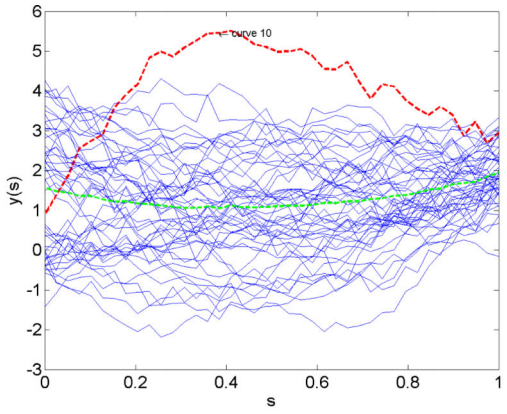
Zhu H, Lee S, Zhou J. Case-deletion measures for models with incomplete data. Biometrika. 2001; 88:727–737.

Zhu H, Li R, Kong L. Multivariate varying coefficient model for functional responses. The Annals of Statistics. 2012b; 40:785–811. [PubMed: 23152646]

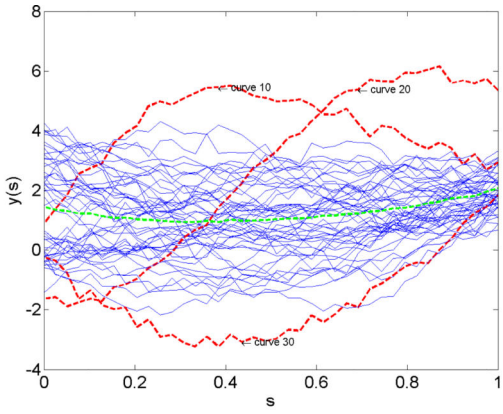
Zhu H, Styner M, Tang N, Liu Z, Lin W, Gilmore J. Frats: functional regression analysis of DTI tract statistics. IEEE Trans. Med. Imaging. 2010; 29:1039–1049. [PubMed: 20335089]

Zhu H, Styner M, Tang N, Liu Z, Lin W, Gilmore J. FADTT: Functional analysis of diffusion tensor tract statistics. NeuroImage. 2011; 56:1412–1425. [PubMed: 21335092]

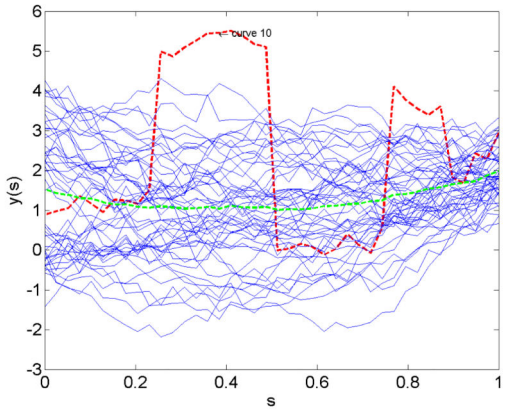
Zou C, Tseng S, Wang Z. Outlier detection in general profiles using penalized regression method. IIE Transactions. 2013; 46:2, 106–117.



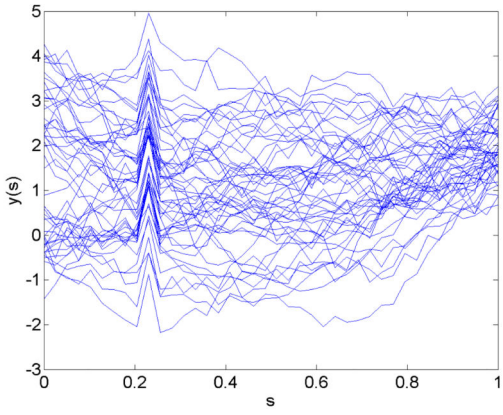
(a) Scenario (i): a single red outlying curve for case 10.



(b) Scenario (ii): three red outlying curves for cases 10, 20 and 30.

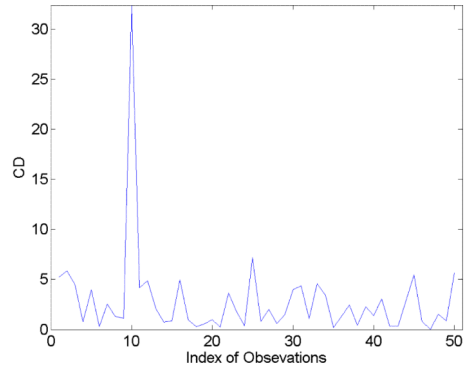


(c) Scenario (iii): multiple outlying grid points on a red single curve for case 10.

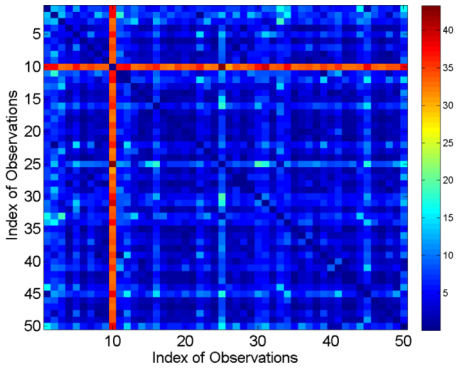


(d) Scenario (iv): a single outlying grid point for the 10th grid point on all curves.

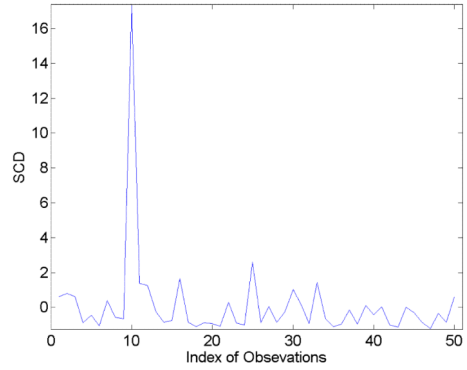
Figure 1.
Graphical illustration of four simulation scenarios.



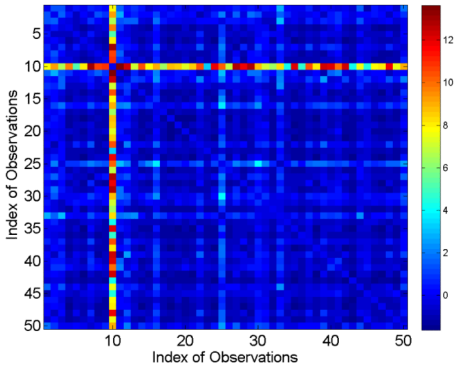
(a) Global CD : single-case



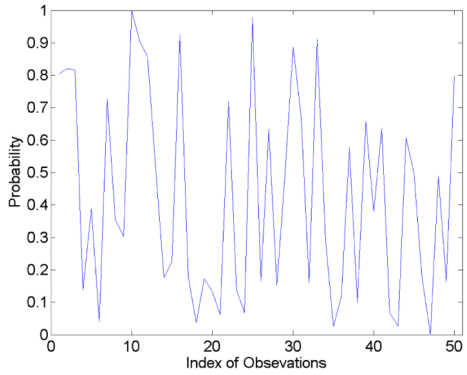
(b) Global CD : two-case



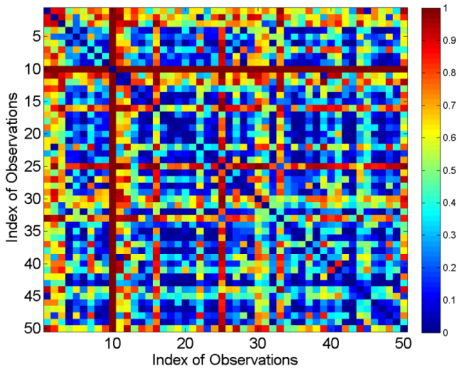
(c) Scaled global CD : single-case



(d) Scaled global CD : two-case



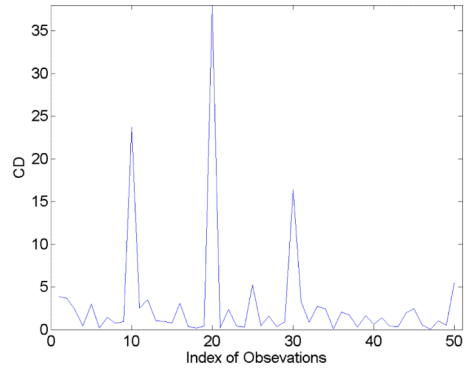
(e) Global diagnostic probability : single-case



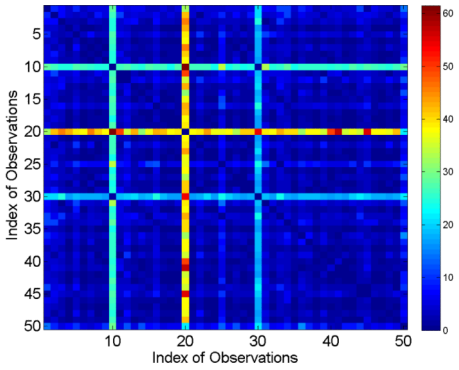
(f) Global diagnostic probability : two-case

Figure 2.

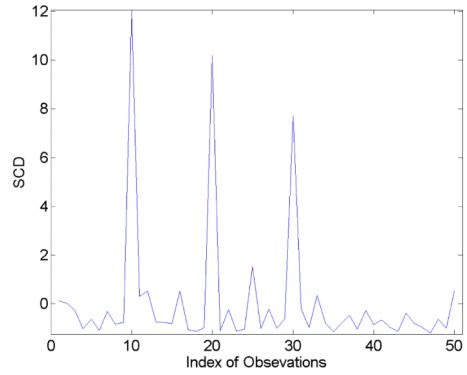
Cook's distance measures for Scenario (i) with a single outlying curve. Top: global CD (i.e., CD_I); middle: scaled global CD (i.e., SCD_I); and bottom: global di-agnostic probability (i.e., $\hat{p_I}$). Left: single-case deletion and Right: two-case deletion. Index of observation stands for the index of case deleted and the case 10 is an outlying curve.



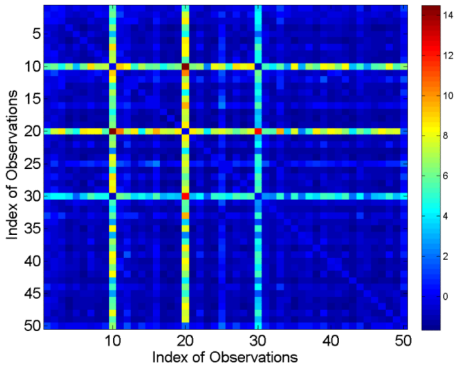
(a) Global CD : single-case



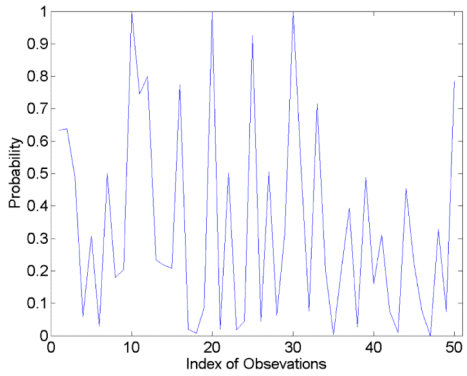
(b) Global CD : two-case



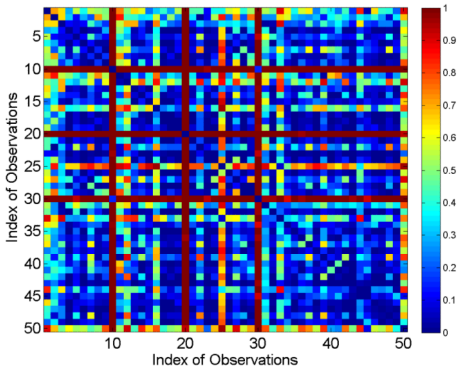
(c) Scaled global CD : single-case



(d) Scaled global CD : two-case



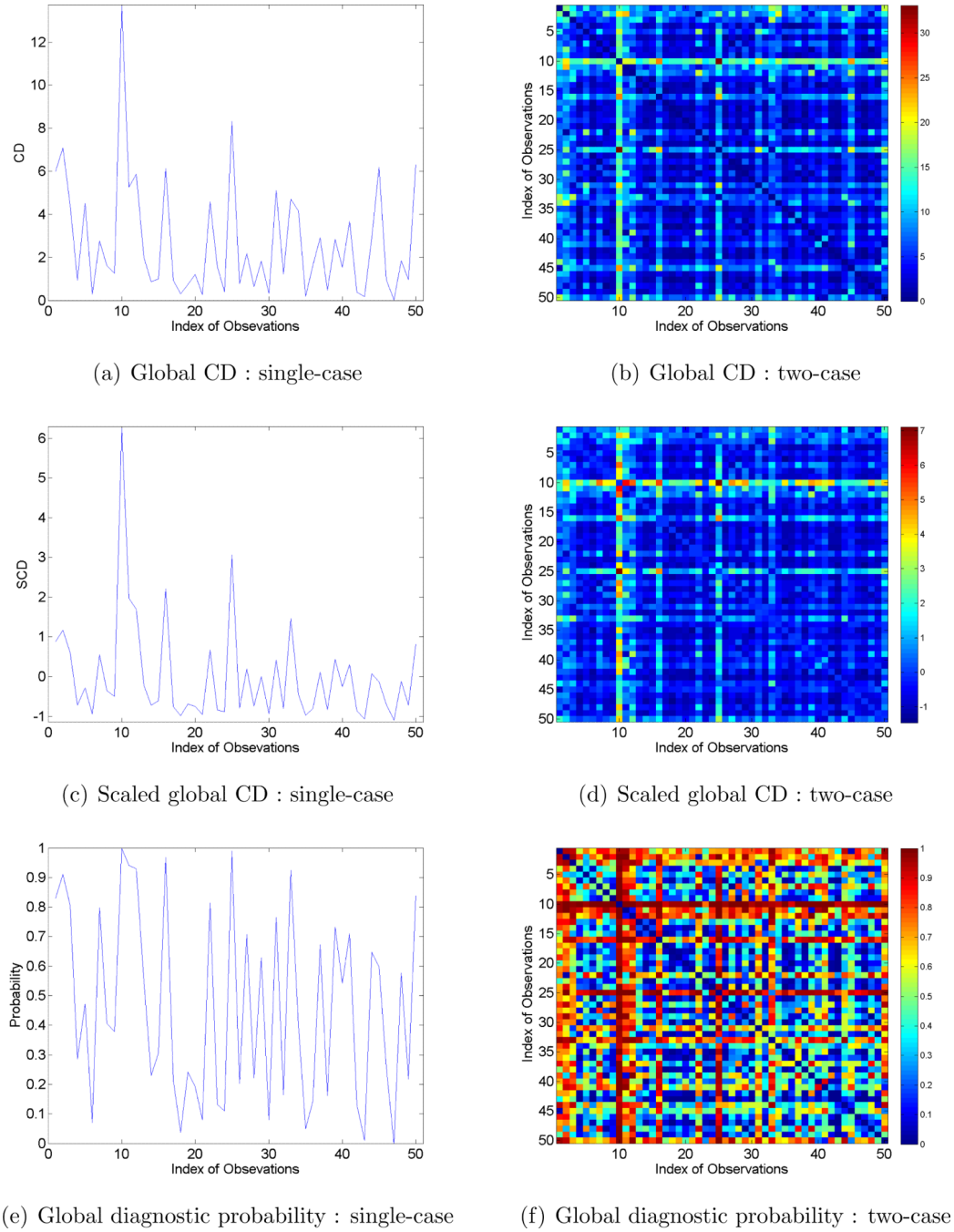
(e) Global diagnostic probability : single-case



(f) Global diagnostic probability : two-case

Figure 3.

Cook's distance measures for Scenario (ii) with three outlying curves. Top: global CD (i.e., CD_I); middle: scaled global CD (i.e., SCD_I); and bottom: global di-agnostic probability (i.e., p_I). Left: single-case deletion and Right: two-case deletion. Index of observation stands for index of subject deleted and the cases 10, 20 and 30 are outlying curves.



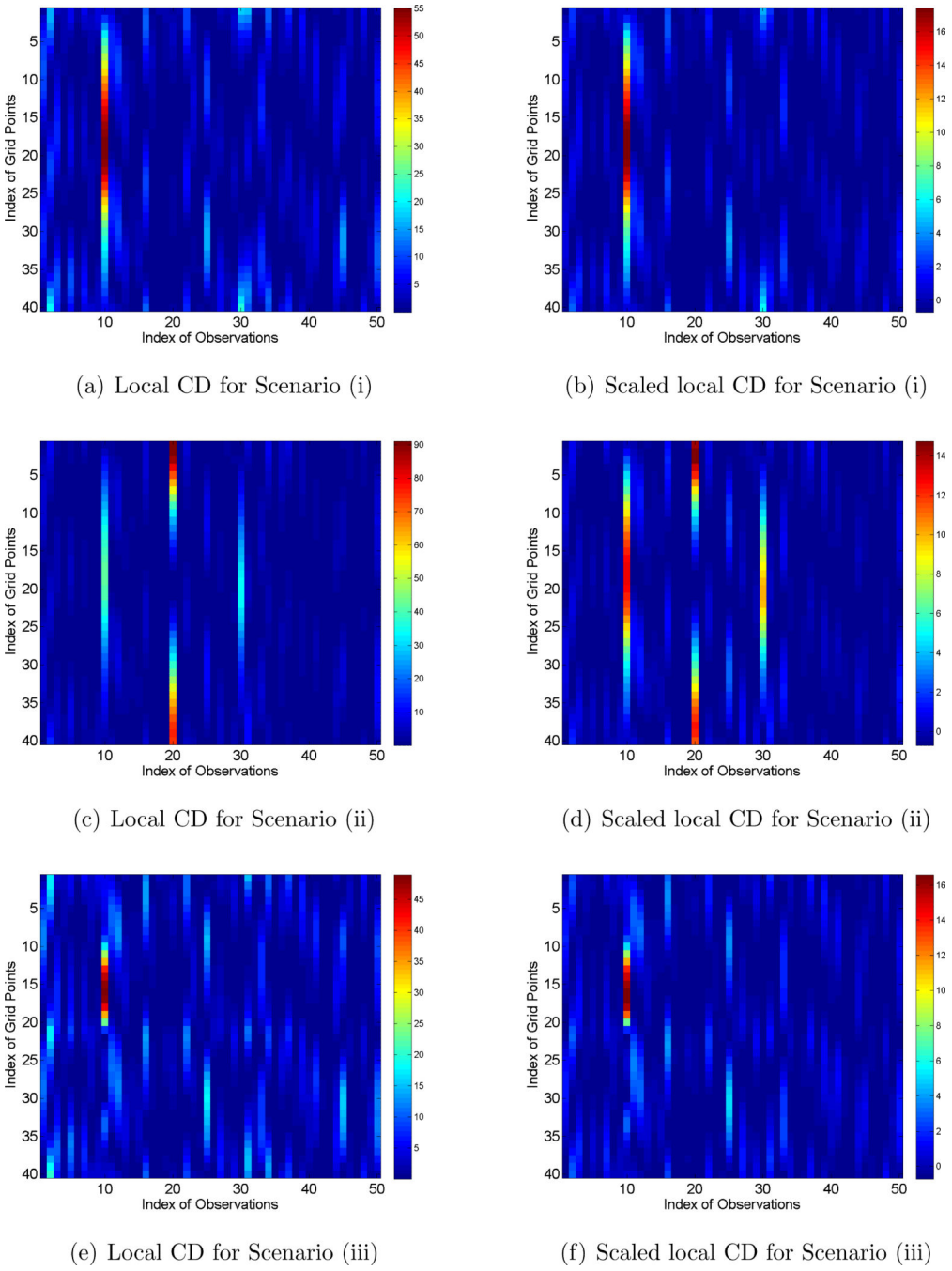
(a) Global CD : single-case (b) Global CD : two-case

(c) Scaled global CD : single-case (d) Scaled global CD : two-case

(e) Global diagnostic probability : single-case (f) Global diagnostic probability : two-case

Figure 4.

Cook's distance measures for Scenario (iii) with multiple outlying grid points on a single outlying curve. Top: global CD (i.e., CD_I); middle: scaled global CD (i.e., SCD_I); and bottom: global diagnostic probability (i.e., $\hat{p_I}$). Left: single-case deletion and Right: two-case deletion. Index of observation stands for case index deleted and the case 10 is an outlying curve.



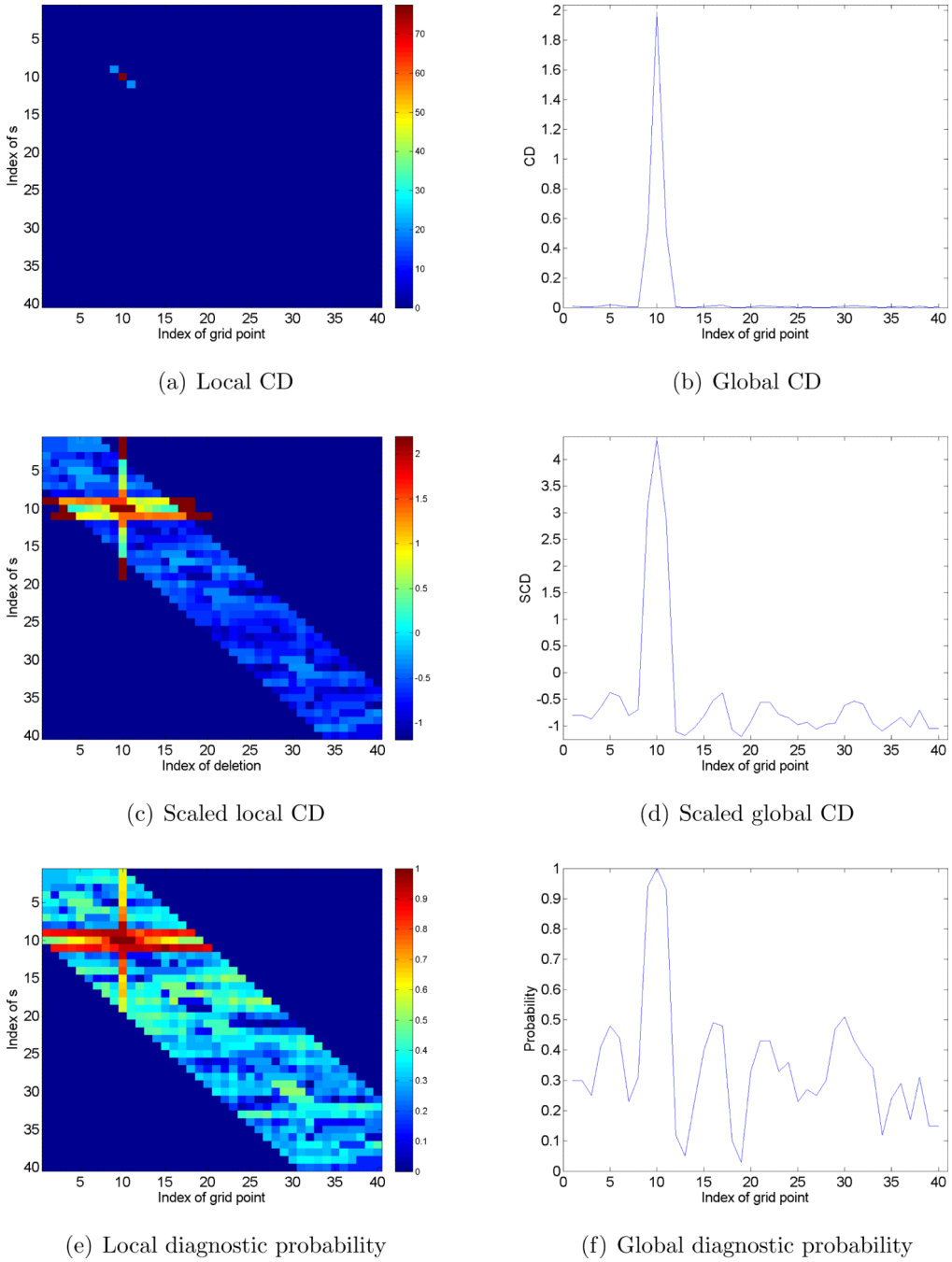
(a) Local CD for Scenario (i) (b) Scaled local CD for Scenario (i)

(c) Local CD for Scenario (ii) (d) Scaled local CD for Scenario (ii)

(e) Local CD for Scenario (iii) (f) Scaled local CD for Scenario (iii)

Figure 5.

Simulation data results for Scenarios (i)-(iii): local CDs (left panel, i.e., $CD_I(s)$) and scaled local CDs (right panel, i.e., $SCD_I(s)$) for a single-case deletion. Index of observations stands for the index of subject deleted.



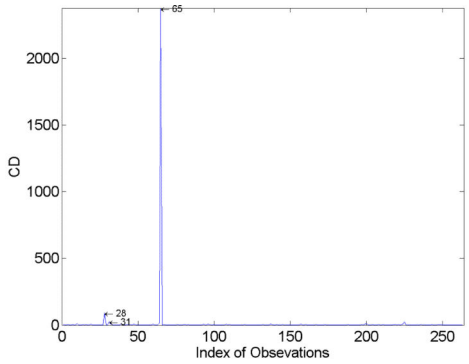
(a) Local CD (b) Global CD

(c) Scaled local CD (d) Scaled global CD

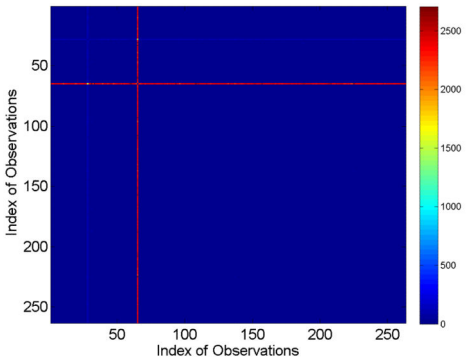
(e) Local diagnostic probability (f) Global diagnostic probability

Figure 6.

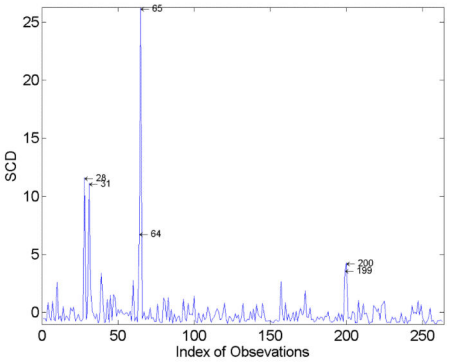
Simulation data results for Scenario (iv). Local (global) CD, scaled local (global) CD and local (global) diagnostic probability are denoted as $CD_{IS,n}(s)$, $SCD_{IS,n}(s)$, and $p_{\widehat{IS},n}(s)$ ($CD_{IS,n}$, $SCD_{IS,n}$, and $p_{\widehat{IS},n}$), respectively.



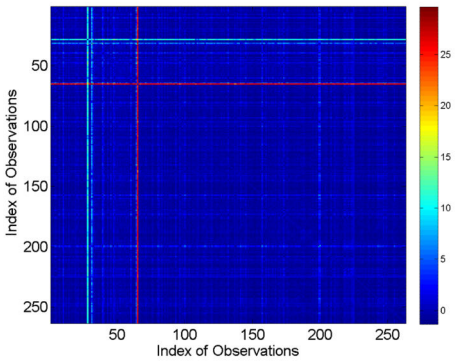
(a) Global CD for single-case deletion



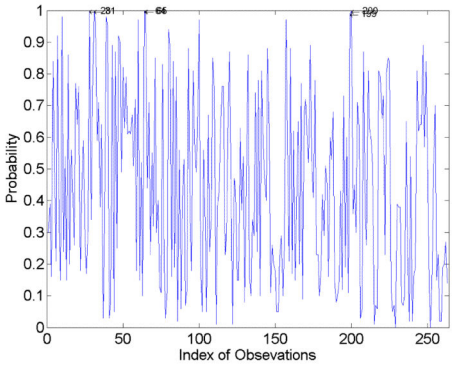
(b) Global CD for two-case deletion



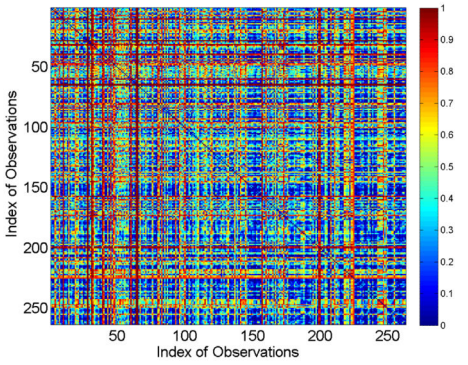
(c) Scaled global CD for single-case deletion



(d) Scaled global CD for two-case deletion



(e) Global diagnostic probability for single-case deletion



(f) Global diagnostic probability for two-case deletion

Figure 7.

Diagnostic measures from real DTI data for single-case and two-case deletions: top: global CD (i.e., CD_I); middle: scaled global CD (i.e., SCD_I); and bottom: global diagnostic probability (i.e., $\hat{p_I}$). Left: single-case deletion and Right: two-case deletion.

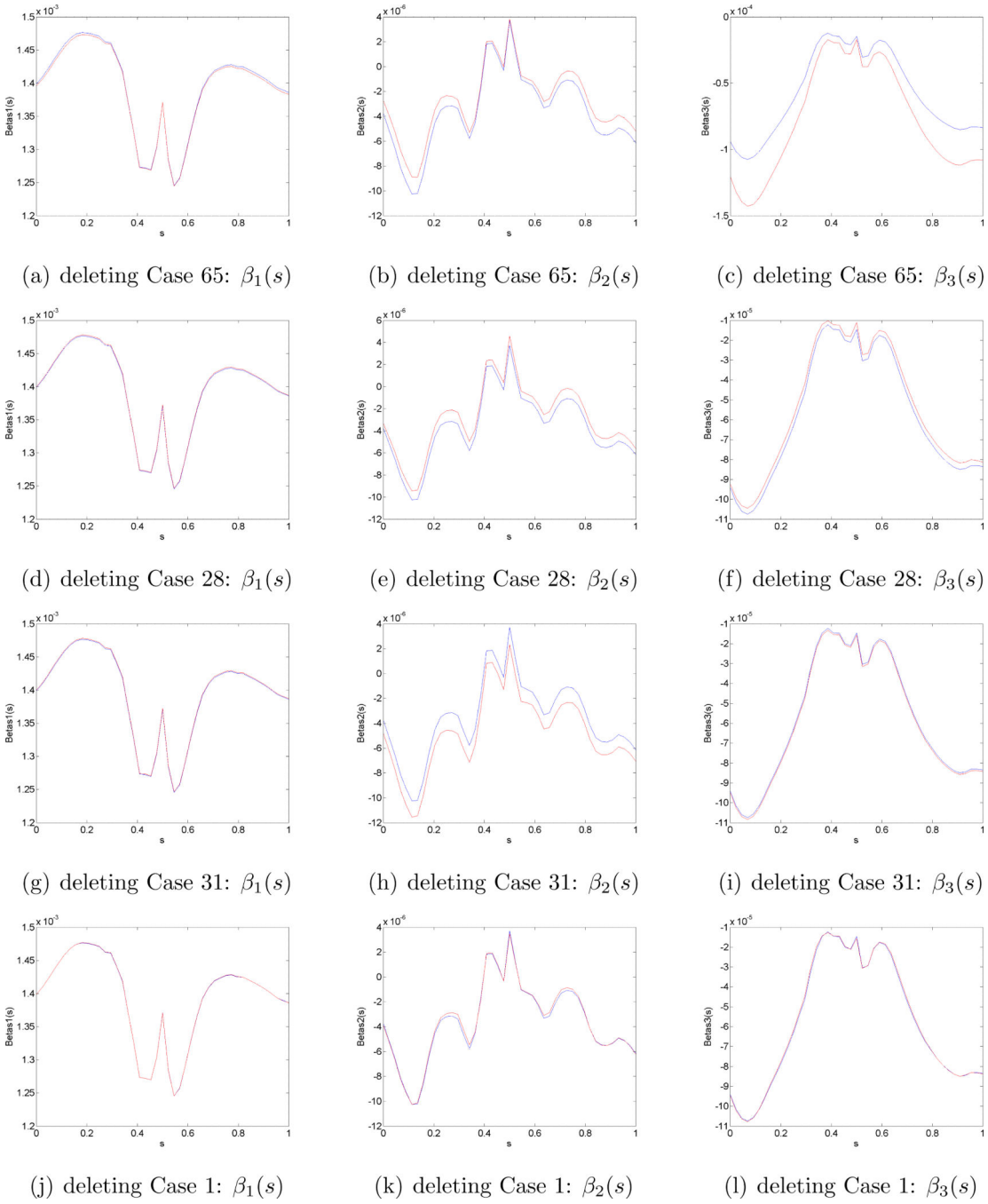


Figure 8.

Estimated coefficient functions after (Red dash line) and before (Blue solid line) deleting some cases from real DTI data. The $\boldsymbol{\beta}_1(s)$, $\boldsymbol{\beta}_2(s)$, and $\boldsymbol{\beta}_3(s)$ correspond to intercept, gender, and gestational age, respectively. The labels of *y*-axis Beta1(s), Beta2(s) and Beta3(s) stand for $\boldsymbol{\beta}_1(s)$, $\boldsymbol{\beta}_2(s)$, and $\boldsymbol{\beta}_3(s)$, respectively.