DETECTION OF INFLUENTIAL OBSERVATIONS IN LONGITUDINAL MULTIVARIATE MIXED EFFECTS REGRESSION MODELS

by

Yun Ling

M.S., Bioinformatics, University of Pittsburgh, 2007

B.S., Biomedical Engineering and Instrumentation,

Tsinghua University, Beijing, China, 1995

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yun Ling

It was defended on

January 17, 2014

and approved by

Stewart J. Anderson, PhD, Professor, Department of Biostatistics, Graduate School of

Public Health, University of Pittsburgh

Richard A. Bilonick, PhD, Assistent Professor, Department of Ophthalmology, University

of Pittsburgh School of Medicine; Assistent Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Andriy Bandos, PhD, Assistent Professor, Department of Biostatistics, Graduate School of

Public Health, University of Pittsburgh

Yu Cheng, PhD, Associate Professor, Department of Statistics, School of Arts and

Sciences, University of Pittsburgh

Dissertation Director: Stewart J. Anderson, PhD, Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

DETECTION OF INFLUENTIAL OBSERVATIONS IN LONGITUDINAL MULTIVARIATE MIXED EFFECTS REGRESSION MODELS

Yun Ling, PhD

University of Pittsburgh, 2014

ABSTRACT

The purpose of this dissertation is to detect possible influential observations in the multivariate longitudinal data. An influential observation is an observation which has large effect on the parameter estimation of a given model. Influential observations are important because: (1) removal of the observation(s) from the data set can substantially change the values of the estimated parameters; (2) in multivariate longitudinal mixed effect models, influential observations can affect the population and subject-specific trajectories; (3) influential observation(s) of one response may affect the predicted effects of the other response within the same individual; (4) an influential observation may indicate an abnormal or misdiagnosed subject.

This research was motivated by opthalmological clinical research (glaucoma). In many ophthalmology studies, both eyes are repeatedly measured. Sometimes, one eye could be measured by different devices, or, measured for different quantities (retina thickness for different quadrants, OCT, VFI, etc). For example, in one study considered in this dissertation, multivariate measurements (e.g., Retinal Nerve Fiber Layer (RNFL) thickness and Ganglion Cell Complex (GCC) thickness) were repeatedly measured on each eye (subject), within each patient (cluster).

When we detect influential observations for longitudinal ophthamology data, our trajectory model must take into account three kinds of correlations: (1) correlation among different characterisitcs measured at the same time point within the same eye; (2) correlation among different time points; (3) correlation between characteristics in the two eyes.

In the first part of my dissertation, we propose a multivariate conditional version of Cook's distance for multivariate mixed effect models. Some research has shown that, in mixed effect models, influential observations having a large effect on the subject-specific parameters cannot always be detected by the original Cook's distance due to large between-subject variation, the problem of influential observations should be approached conditional on the subjects. Hence, in the multivariate longitudinal setting, the influential observation problem should be approached conditional on their subjects and characteristics. Repeated simulations show that multivariate conditional Cook's distance successfully detected most 92.5% influential observations, but unconditional Cook's distance only detected 7.5%.

In the second part we extended the multivariate conditional Cook's distance to multilevel multivariate mixed effect model. In this model, there are two levels of random effects to handle the subject level and cluster level correlations among different time points, and the residual covariance matrix to handle correlations among different responses. Also, the two-level multivariate conditional Cook's distance can be decomposed into six parts, indicating the influences of fixed effects, 1st and 2nd level of random effects, and the co-variation between them, respectively. Examples are given to illustrate how the influential observation in one characteristic changes the effects of both characteristics.

This research has public health implications because the influence of outliers can bias the results of any longitudinal study in public health. Hence, recognizing observations which have undue influence on study results ensures that reliable conclusions can be obtained in medical and public health research settings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 MULTIVARIATE LONGITUDINAL DATA

Two general types of study designs are used for comparing characteristics among humans or animals in the vast majority of investigations in the health sciences: cross-sectional designs and longitudinal designs. Individual cross-sectional studies allow comparisons between means of different age groups but they cannot provide information that integrates the continuum of time, nor do they provide any information about how individuals or populations change over time. Longitudinal studies provide a powerful tool to address these problems.

In many medical and epidemiological research studies and clincal trials, individuals are measured not only repeatedly, but also with respect to several response variables. Hence, multivariate longitudinal data allow one to study and analyze the joint evolution of multiple response variables over time. Examples of multivariate longitudinal data analysis include: the couple-level growth curve analysis in social sciences [1], the joint modeling of CD4 and CD8 lymphocyte counts in the process of HIV infection [2], and the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) [3]. In the CRISP study, 241 individuals were enrolled and followed for 3 years. MRI was used to monitor the possible change of the renal structure. In this study, three biological markers (characteristics) were repeatedly measured: kidney volume (KVS), cyst volume (CVS) and glomerular filtration rate (GFR). Another example is Xu's longitudinal study for Neonatal-Pediatric brain tissue development [6]. The purpose of that study was to jointly study the growth patterns of gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) volumes segmented from longitudinal brain MR images of neonate-pediatric data from birth to 2 years of age.

1

1.2   MULTIVARIATE MIXED EFFECT MODEL

It is well known that the longitudinal effect of aging or time is an inherently within-individual effect, and the true hidden population growth trajectory should be the average of all individual growth trajectories. However, simple linear regression requires all the subjects and measurements to be independent, which makes simple linear regression methods unsuitable for longitudinal data with repeated measurements, because measurements are almost always dependent if they are within the same subject.

The use of mixed models [47][48][49] thus arises as a strong tool to address this problem by providing a general model (fixed effect) as a function of time to explain the underlying growth process of the population of individuals while at the same time, allowing for each individual to have his/her own trajectory (random effect). This approach can also be easily extended to situations where more than one characteristic is being observed at each time point. The data arising from such situations, referred to as "multivariate longitudinal data", are the primary focus of this dissertation.

The random coefficient mixed effect model longitudinal study assumes that the regression coefficients are a random sample from the whole population of possible coefficients and allow one to model variations between study units [46]. Also, for multivariate longitudinal studies, it is common that not all characteristics are measured at all time points. Furthermore, the assessments are not always equally spaced. The use of multivariate mixed effect models allows one to model a longitudinal process where multiple outcomes are measured at each time point, and allows one to easily handle missing data. However, the analysis of multivariate longitudinal process is complicated because (1) the variances of residuals may be different for different variables; (2) the residuals may be correlated for the same characteristic measured at different time points (within-characteristic correlation); (3) the residuals are also correlated among different characteristics measured at a given time point (inter-characteristic correlation); (4) often, not all variables are measured at all time points, or the assessments are not always equally spaced, even within one subject [4].

In recent years, many papers have extended the longitudinal (mixed effect) model to the multivariate case. Reinsel [7][8] developed models for balanced multivariate longitudinal

data using a multivariate random-effect model. Shah, et al. [9] suggested an extension of Reinsel's work [7][8] to the case of unbalanced data to accomodate the case of arbitrary measurement time points. In this paper Shah used the EM algorithm for maximum likelihood estimation of parameters. Morrel, et al., [10], had utilized Shah's [9] algorithm in a Bayesian framework to investigate the feasibility of predicting hypertension based on a longitudinal study of body mass index (BMI), systolic blood pressure (SBP) and triglyceride levels from the Baltimore longitudinal study of aging[14]. Ibrahim, et al. [15], Xu, et al. [16], Song, et al. [18] and Tsiatis [19] all extended the longitudinal model to the joint modeling of both longitudinal and time to event data. Other variants of mixed effect model approaches include the stochastic mixed effect, state-space model [11], self-modeling regression (SEMOR) [12] and pairwise fitting [13].

1.3   INFLUENTIAL OBSERVATIONS

Some observations can be very influential on parameter estimation and the removal of the observation from the data set can substantially change the regression equation. Such an observation is called an "influential observation" by Belsley, et al [45].

Influential observations are important in data analysis. It is well known that not all observations in a dataset play an equal role in determining parameter estimates. Suppose we have a simple linear regression analysis using a dataset that contains N data pairs. We may think that each pair contributes a weight of $1/N$ towards the estimation of the model parameters. But this is not always correct. Sometimes the magnitude of the estimates in the model may be determined only by a few cases while most other data are essentially ignored. The reason is that the ordinary least squares (OLS) regression minimizes the sum of squared error (SSE) so it gives more weight to those observations in which changes in the parameters produce the largest reductions in the SSE. Those observations have the most influence. If the observations have a strong influence, then, when they are removed, the estimated parameters will be changed substantially.

Therefore, it is very important to be aware of particular observations that have unusually

large influence on the parameter estimates of the analysis. These observations (1) may be appropriate and retained in the analysis, (2) may be identified as inappropriate data and have to be removed from the analysis, (3) may suggest that we need additional data, (4) may suggest that current model is inadequate, or (5) may indicate data input or data entry mistakes. However, when interpreting the results, their influence has to be taken into consideration. Regardless of the final assessment concerning those observations, identifying them is necessary before intelligent, subject-matter-based conclusions can be made.

In the detection of univariate influential observations, an approach called 'case deletion' studies the effect of deleting an observation on the parameter estimates. Cook (1977) [21] defines a measure of distance between two maximum likelihood estimates, where one is calculated based on the complete set of observations and the other based on deletion of a specific observation. Cook (1986) [22] also developed a local influence method for the analysis of the simple linear regression model. Beckman, Nachtsheim and Cook (1987) [23] developed the local influence method of Cook (1986) [22] for the analysis of the linear mixed effect model. Zhu and Ibrahim [36] developed a Bayesian local influence measure method for joint models for longitudinal and survival data. Although assessments of the influence of a model perturbation are generally regarded as being useful, a practical and well established approach to influence analysis in statistical modeling is still based on case deletion methods, as Lawrance [51] pointed out.

Influence measure extensions to a multivariate case have been suggested by many researchers. Hossain and Naik [31] and Naik [32] extended deletion of single observation in univariate regression models to the multivariate case. Srivastava and von Rosen [33] developed a formal test for detecting a single influential observation for a multivariate linear regression model. Hadi and Simonoff [39] proposed procedures and tests for detection of multiple influential observations in univariate linear models. However, such methods have the problem of masking and swamping, where the effect of one outlier masks the effect of other outliers. Examples of this phenomenon were given by Barnett and Lewis [50]. Weiss [28] developed a goodness of fit test for multivariate outliers, and the difference between the multivariate outliers and multiple univariate outliers. Barrett and Ling [24] proposed general classes of influence measures for multivariate regression based on analogous forms of

4

univariate Cook's distance. Diaz-Garcia, et al. [30], proposed a generalized Cook's distance for elliptical distributions. Altunkaynak and Ekni [26] proposed a three-stage method. In the first and second stage, the author introduces a linear restriction and a transformation of the multivariate linear regression model into a restricted multivariate linear regression model. The restricted model and the full model were used to facilitate the calculation of the difference between parameter estimates of the multivariate linear regression model and that of the restricted multivariate linear regression model. The third stage contains the assessment of the influential observations using the generalized Cook's distance. Chi and Ibrahim [38] proposed a joint model for multivariate linear mixed effect model and multivariate survival model. Cerioli [29] developed multivariate outlier tests based on the "high-breakdown" minimum covariance determinant estimator. This test uses robust estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and the outliers in $\mathbf{y}$ are revealed by their large distance from the robust fit. The author claims the method has good performance under the null hypothesis of no outliers in the data. However, their method is only applicable for non-longitudinal multivariate data, and does not take into account any random effects.

1.4   INFLUENTIAL OBSERVATION FOR LONGITUDINAL DATA ANALYSIS

Detecting influential observations in longitudinal data analysis is more complicated. In the multivariate influential observation detection methods described above, influence measures are constructed to detect influential subjects and observations for the fixed regression parameters. However, in mixed effect models, these statistics may fail to, or incorrectly detect observations influential due to their omission of variances and covariances of associated random effects [20]. Banerjee [34][35] noticed that the effectiveness of Cook's distance is limited in longitudinal data analysis because it was designed for independent observations and hence cannot be directly used in the longitudinal setting. Tan, et al., [20] and Ouwens, et al., [5] showed the advantage of using observation-oriented influence measures instead of subject-oriented influence measures because the subject-oriented influence measures may fail

to or incorrectly detect influential subjects or influential observations, owing to the relative position of the observations within and across subjects. Tan [20] proposed a conditional version of Cook's distance by conditioning on the subjects. Zhu and Ibrahim [37] developed scaled Cook's distance to address a fundamental issue, that is, deleting subsets with different numbers of observations introduces different degrees of perturbation to the current model fitted to the data and the magnitude of Cook's distance associated with the degree of the perturbation. They also proposed a general parametric model for many complex data structures, including longitudinal data.

1.5   OVERVIEW OF DISSERTATION

There has clearly been a great deal of work done for the detection of multivariate influential observations and the detection of influential observations in univariate longitudinal data analysis. However, for the detection of influential observations in multivariate mixed effect model, no rigorous approach has been developed.

This dissertation will address two of these areas. First, we propose a quantity to measure the degree of influence by removing a set of observations. If the removal of observations from the dataset results in a significant change to the parameter estimates for the current statistical model, and equivalently, generates a large value of this quantity, then the set of observations should be flagged as influential. In a multivariate longitudinal dataset, "a set of" observations could be one observation, one characteristic in an individual, or an individual's complete set of observations. This quantity will be an extension of Tan's conditional Cook's distance [20] to the multivariate longitudinal case. Second, we will extend this method to the case of the multilevel multivariate longitudinal dataset. There are two levels of random effects to handle the subject level and cluster level correlations among different time points. Also, the multilevel multivariate conditional Cook's distance can be decomposed into six parts, indicating the influences of fixed effects, first and second levels of random effects, and the co-variation between them, respectively. We will derive the influence functions for the parameter estimates. Influence functions are functions to assess the effect (or "influence")

of removing an observation (or a subset of observations) on the value of a statistic without having to re-compute that statistic. This part will be an extension to Christensen's result [25] to the multilevel and multivariate longitudinal case.

2.0 MULTIVARIATE LONGITUDINAL EXTENSION OF COOK'S DISTANCE

2.1 MODELING TRAJECTORIES OF MULTIVARIATE LONGITUDINAL OUTCOMES THROUGH MIXED EFFECT MODEL

2.1.1 Univariate Mixed Effect Model

A typical univariate ($m = 1$) mixed effect model for individual i can be written as:

$$
\begin{aligned}
\boldsymbol{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \\
\boldsymbol{y}_i &\sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + R_{n_i})
\end{aligned}
\tag{2.1}
$$

where $\boldsymbol{y}_i$ is a vector of observations, $\boldsymbol{\beta}$ is the vector of the fixed effect parameters, $\boldsymbol{b}_i$ is the vector of the random effect parameters.

This model and more general univariate models were proposed by Laird and Ware[42], Jennrich and Schluchter[40], Laird, Lange and Stram[41], Lindstrom and Bates[43], and others. Many software packages were developed to fit those linear mixed effect models, such as SAS PROC MIXED and R packages nlme and lme4, STATA. Those programs are able to analyze unbalanced longitudinal datasets, in which the measurements are repeatedly taken at arbitrary set of time points for each individual. Missing data (outcomes) are ignored along with the corresponding rows of $\mathbf{X}_i$ and $\mathbf{Z}_i$.

2.1.2 Multivariate Mixed Effect Model

2.1.2.1 Model Structure A multivariate normal mixed effects model is proposed to model the multivariate longitudinal response. The model for the kth characteristic at the jth time point of ith individual is given as:

$$y_{ijk} = y_{ijk}(t_{ij}) = \underbrace{\boldsymbol{x}_{ijk}}_{1 \times p_k} \underbrace{\boldsymbol{\beta}_k}_{p_k \times 1} + \underbrace{\boldsymbol{z}_{ijk}}_{1 \times q_k} \underbrace{\boldsymbol{b}_{ik}}_{q_k \times 1} + \epsilon_{ijk} \tag{2.2}$$

Based on 2.2, the model for i^{th} individual at j^{th} time point for all characteristics is given as:

$$\underbrace{\boldsymbol{y}_{ij}}_{m \times 1} = \underbrace{\mathbf{X}_{ij}}_{m \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\mathbf{Z}_{ij}}_{m \times q} \underbrace{\boldsymbol{b}_i}_{q \times 1} + \underbrace{\boldsymbol{\epsilon}_{ij}}_{m \times 1} \tag{2.3}$$

and the model for ith individual for all variables and time points is given as:

$$\underbrace{\boldsymbol{y}_i}_{mn_i \times 1} = \underbrace{\mathbf{X}_i}_{mn_i \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\mathbf{Z}_i}_{mn_i \times q} \underbrace{\boldsymbol{b}_i}_{q \times 1} + \underbrace{\boldsymbol{\epsilon}_i}_{mn_i \times 1} \tag{2.4}$$

where the quantities of models 2.2, 2.3 and 2.4 are as follows:

- $i = 1, 2, ..., N$, number of subjects, total N subjects;
- $j = 1, 2, ..., n_i$, j indicates the indices of repeated measurements for the ith subject. That is, the ith subject is repeatedly measured n_i times.
- $k = 1, 2, ..., m$, k indicates the indices of characteristics for the ith individual, m indicates the number of characteristics measured for each individual.
- $y_{ijk} = y_{ijk}(t_{ijk})$ is the assessment of the kth characteristic of ith subject measured at j^{th} time point, time t_{ijk};
- $\boldsymbol{y}_{ij} = [y_{ij1} \ y_{ij2}, \ ... \ y_{ijm}]^T = [y_{ij1}(t_{ij}) \ y_{ij2}(t_{ij}) \ ... \ y_{ijm}(t_{ij})]^T$, is an $m \times 1$ vector;
- $\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_{i1}^T \ \boldsymbol{y}_{i2}^T \ \cdots \ \boldsymbol{y}_{in_i}^T \end{bmatrix}^T$, $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1^T \ \boldsymbol{y}_2^T \ \cdots \ \boldsymbol{y}_n^T \end{bmatrix}^T$
- $\underbrace{\boldsymbol{x}_{ijk}}_{1 \times p_k} = \underbrace{\boldsymbol{x}_{ijk}(t_{ijk})}_{1 \times p_k}$ is the vector of fixed effect covariates (could be time varing) for the kth characteristic of the ith individualat at j^{th} time point (time t_{ijk}). p_k is the number of the fixed effect parameters for characteristic k.
- $\underbrace{\boldsymbol{\beta}_k}_{p_k \times 1}$, is a vector of the fixed effects parameters for k^{th} characteristic. In most cases, $p_1 = p_2 = ... = p_m$.

- $\underbrace{\mathbf{X}_{ij}}_{m \times p} = \mathrm{diag}\left(\boldsymbol{x}_{ij1}^T \; \boldsymbol{x}_{ij2}^T \ldots \; \boldsymbol{x}_{ijm}^T\right)$ is a fixed effects design matrix for $\boldsymbol{y}_{ij}$.

- $\mathbf{X}_i = \left[\mathbf{X}_{i1}^T \; \mathbf{X}_{i2}^T \; \ldots \; \mathbf{X}_{in_i}^T\right]^T$, $\mathbf{X} = \left[\mathbf{X}_1^T \; \mathbf{X}_2^T \; \ldots \; \mathbf{X}_N^T\right]^T$ is the fixed effects design matrix.

- $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1^T \; \boldsymbol{\beta}_2^T \; \ldots \; \boldsymbol{\beta}_m^T\right]^T$;

- $\underbrace{\boldsymbol{z}_{ijk}}_{1 \times q_k} = \underbrace{\boldsymbol{z}_{ijk}(t_{ijk})}_{1 \times q_k}$ is a potentially time-varyng vector of random effect covariates for the kth characteristic at time t_{ijk}, of the ith individual. q_k is the number of the random effect parameters for characteristic k.

- $\underbrace{\mathbf{Z}_{ij}}_{m \times q} = \mathrm{diag}\left(\boldsymbol{z}_{ij1}^T \; \boldsymbol{z}_{ij2}^T \ldots \; \boldsymbol{z}_{ijm}^T\right)$ is a random effects design matrix for $\boldsymbol{y}_{ij}$.

- $\underbrace{\boldsymbol{b}_{ik}}_{q_k \times 1}$ vector, is the vector of the random effects parameters for k^{th} characteristic of ith individual. In most cases, $q_1 = q_2 = ... = q_m$.

- $\mathbf{Z}_i = \left[\mathbf{Z}_{i1}^T \; \mathbf{Z}_{i2}^T \; \ldots \; \mathbf{Z}_{in_i}^T\right]$, $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1^T \; \mathbf{Z}_2^T \; \ldots \; \mathbf{Z}_N^T)^T$ is the random effects design matrix.

- $\boldsymbol{b}_i = \left[\boldsymbol{b}_{i1}^T \; \boldsymbol{b}_{i2}^T \; \ldots \; \boldsymbol{b}_{im}^T\right]^T$, $\boldsymbol{b} = \left[\boldsymbol{b}_1^T \; \boldsymbol{b}_2^T \; \ldots \; \boldsymbol{b}_N^T\right]^T$; and

- $\boldsymbol{\epsilon}_{ij} = \left[\epsilon_{ij1} \; \epsilon_{ij2}, \; \ldots \; \epsilon_{ijm}\right]^T$, $\boldsymbol{\epsilon}_i = \left[\boldsymbol{\epsilon}_{i1}^T \; \boldsymbol{\epsilon}_{i2}^T \; \ldots \; \boldsymbol{\epsilon}_{in_i}^T\right]^T$.

2.1.2.2 Assumptions   The purpose of the model given above is to analyze the multivariate longitudinal data in which a set of m characteristics are repeatedly measured n_i times on the ith individual. To do this, the model has to account for three sources of correlation: (1) inter-source (different measures at the same visit), (2) intra-source (same measurement at different visits), and cross correlation (different measurement at different visits).

In our model 2.2, 2.3 and 2.4, we assume $\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{m \times m})$, where $\boldsymbol{\Sigma}_{m \times m}$ is an unstructured variance-covariance matrix. In this case, $\mathrm{var}(\boldsymbol{\epsilon}_i) \sim N(\mathbf{0}, \; (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}))$. The random effects are distributed as $\boldsymbol{b}_i \sim N_q(\mathbf{0}, \mathbf{G})$ independently for $i = 1, ..., N$. Depending on the application, we may allow $\mathbf{G}$ to be either unstructured or block diagonal with m non-zero blocks of size $q_k \times q_k$ corresponding to the m characteristics.

Without conditioning on each individual and characteristic, the marginal model of $\boldsymbol{y}$ is:

$$\boldsymbol{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \;\; \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}))$$

The multivariate mixed effect models have received much attention in the literature. For models similar to 2.2, 2.3 and 2.4, Reinsel [8] derived a closed-form estimate with completely

observed $\boldsymbol{y}_i$ and balanced designs. Shah, Laird and Schoenfeld[9] developed an EM algorithm for bivariate $(m = 2)$ setting, which also took into account the unbalanced design and missing responses. After Shah, Laird and Schoenfeld's work, Schafer and Yucel [44] developed a new EM algorithm for the multivariate mixed effect model with unbalanced missing data using multiple imputation methods, which is implemented in R package *mlmmm*. In certain situations, it may be possible to recast the multivariate model as a univariate one and apply existing software such as SAS PROC MIXED with a user-specified covariance structure[44]. Fieuws and Verbeke have proposed a pairwise fitting approach for the multivariate mixed model with a large number of characteristics[13].

2.2   COOK'S DISTANCE

Cook's distance is based on the concept of the influence function introduced by Hampel [27]. The influence function of the distance measurement was developed by Cook [21]. The concept of Cook's distance is as follows: Suppose there is a probability density function of a random vector $\mathbf{Y}$, denoted by $p(\mathbf{Y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of the parameters of the probability density function. Cook's distance measures the distance between the maximum likelihood estimators (MLE) of $\boldsymbol{\theta}$ with and without the subset of the data. Let A denotes the subset of the data to be removed. The new probability density function is denoted by $p(\mathbf{Y}_{(A)}|\boldsymbol{\theta})$. The MLE of $\boldsymbol{\theta}$ based on the full dataset $\mathbf{Y}$ is $\hat{\boldsymbol{\theta}}$, and the MLE of $\boldsymbol{\theta}$ based on the subsample dataset with A removed, that is, $\mathbf{Y}_{(A)}$, is $\hat{\boldsymbol{\theta}}_{(A)}$, respectively. Hence, the Cook's distance for the subset A, denoted by CD(A), is defined as follows:

$$CD(A) = (\hat{\boldsymbol{\theta}}_{(A)} - \boldsymbol{\theta})^T \mathbf{B} (\hat{\boldsymbol{\theta}}_{(A)} - \boldsymbol{\theta})$$

where $\mathbf{B}$ is a positive definite matrix to be estimated but does not change when the subset of data is removed.

For multivariate data, longitudinal data, or multivariate longitudinal data, the within subject observations are correlated. The likelihood function $p(\mathbf{Y}|\boldsymbol{\theta})$ has to be able to model

the correlation structure. Here we set

$$\mathbf{B} = I(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log(p(\mathbf{Y}|\boldsymbol{\theta})) \tag{2.5}$$

which incorporates the correlation structure [37]. Hence, $I(\boldsymbol{\theta})$ denotes the Fisher information for $\boldsymbol{\theta}$.

In this model $\boldsymbol{\theta}$ is the vector of the parameters of the probability density function, including both fixed effects and random effects. Multivariate influence measures for models with and without random effects will be developed and compared.

2.2.1 Example: Simple linear regression model

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

we have $\mathbf{B} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$, where $\mathbf{V} = \text{cov}(\boldsymbol{y})$. The Cook's distance [21] is:

$$C_j = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(j)})^T\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(j)})}{p} = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(j)})^T\mathbf{V}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(j)})}{p}$$

where p is the number of predictors.

2.2.2 Multivariate Cook's Distance

Barrett and Ling [24] generalized the univariate measures of influence to the multivariate regression model:

$$D_i = [\boldsymbol{b} - \boldsymbol{b}_{(i)}]^T[\mathbf{S}^{-1} \otimes (\mathbf{X}^T\mathbf{X})][\boldsymbol{b} - \boldsymbol{b}_{(i)}]/p$$

where $\mathbf{S} = \mathbf{e}^T\mathbf{e}/(n - p)$, $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, and p is the number of predictors. "$\otimes$" indicates the direct product or Kronecker product of two matrices.

2.2.3 Conditional Cook's Distance

In the longitudinal context, Cook's distance is used to identify observations with unusually large influence on the fixed effects, $\mathbf{X}\boldsymbol{\beta}$. Unfortunately, influential observations on the subject-specific (random) effects, $\boldsymbol{b}$, cannot always be correctly detected by Cook's distance. For example, non-influential observations are sometimes incorrectly identified as being influential [20]. Tan, Ouwens and Berger [20] suggested an extension to the Cook's distance for univariate longitudinal data. To do so, they condition on removing observations at time point j which can be written as follows:

$$C_{cond_j} = \sum_{i=1}^{N} \frac{((\mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i) - (\mathbf{X}_i\hat{\boldsymbol{\beta}}_{(j)} + \mathbf{Z}_i\hat{\mathbf{b}}_{i(j)}))^T((\mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i) - (\mathbf{X}_i\hat{\boldsymbol{\beta}}_{(j)} + \mathbf{Z}_i\hat{\mathbf{b}}_{i(j)}))}{\sigma^2\{(N-1)q+p\}}$$

where p is the number of fixed effect predictors, q is the number of random effect predictors, N is the number of total individuals.

2.3 MULTIVARIATE LONGITUDINAL EXTENSION OF COOK'S DISTANCE

Using a concept similar to Cook's distance[21] and the conditional Cook's distance[20], we propose a multivariate longitudinal extension. Conditioning on all of the individuals and each characterisitc of the individuals, we have the following likelihood:

$$L(\boldsymbol{\phi}) = |\mathbf{S}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^T\mathbf{S}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})\right]$$

and its corresponding log-likelihood:

$$l(\boldsymbol{\phi}) = -\frac{1}{2}\log|\mathbf{S}| - \frac{1}{2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^T\mathbf{S}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})$$

where $\mathbf{S} = \text{diag}\left[\mathbf{I}_{n_1} \otimes \boldsymbol{\Sigma}, \ldots, \mathbf{I}_{n_N} \otimes \boldsymbol{\Sigma}\right]$.

Here we use $\boldsymbol{\phi}$ to denote the vector containing all the fixed and random effects parameters to be estimated, that is, $\boldsymbol{\phi} = (\boldsymbol{\beta}^T, \boldsymbol{b}^T)^T$.

Using model 2.2, 2.3, 2.4 and from 2.5, we have the score function:

$$\frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{\beta}} \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} -(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})\boldsymbol{\beta} + (\boldsymbol{y} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})^T\mathbf{S}^{-1}\mathbf{X} \\ -(\mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z})\boldsymbol{b} + (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\boldsymbol{c})^T\mathbf{S}^{-1}\mathbf{Z} \end{bmatrix}$$

and Fisher Information:

$$\mathbf{B} = I(\boldsymbol{\phi}) = -\frac{\partial^2}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^T} l(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z} \end{bmatrix}$$

Then, the conditional Cook's distance is written as:

$$
\begin{aligned}
CD(A) &= \frac{(\hat{\boldsymbol{\phi}}_{(A)} - \hat{\boldsymbol{\phi}})^T \mathbf{B}(\hat{\boldsymbol{\phi}}_{(A)} - \hat{\boldsymbol{\phi}})}{c} \\
&= \frac{1}{c} \begin{bmatrix} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T & (\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \end{bmatrix} \begin{bmatrix} \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}} \end{bmatrix} \\
&= \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c} + \frac{(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c} \\
&\quad + \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c} \\
&= C_{A1} + C_{A2} + C_{A3}
\end{aligned}
\tag{2.6}
$$

where

$$c = (Nm - 1)\, q + p$$

Here p is the total number of fixed effect predictors, that is, $p = \sum_{k=1}^{m} p_k$; q is the total number of random effect predictors, $q = \sum_{k=1}^{m} q_k$; m is the number of characteristics for each individual; and A indicates a subset of the observations. Other quatities include $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$, and $\hat{\boldsymbol{\beta}}_{(A)}$, $\hat{\boldsymbol{b}}_{(A)}$ which are the fitted values of $\boldsymbol{\beta}$, $\boldsymbol{b}$ using the samples with and without the subset of observations, respectively. The subset, A, could be one data point, all observations in one individual, or some other specified subset.

From Equation 2.6, we see that $CD(A)$ can be decomposed to three parts: C_{A1}, C_{A2}, and C_{A3}, which are as follows:

$$C_{A1} = \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{S}^{-1} \mathbf{X}(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{X}_i (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{X}_i (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_{ij} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

C_{A1} is the total distance measurement for the fixed (marginal) effect between the complete dataset and the data with subset A removed. The term $\frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_{ij} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$ is actually the overall marginal Cook's distance for the ith subject at the jth time point. It is the total distance measurement of m characteristics, but only normalizing the residual variance-covariance matrix, without normalizing the random variance-covariance matrices[20]. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be extracted to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \mathbf{X}_{ij} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c\sigma_k^2}$, which is the simple summation of the distance measurements for all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the fixed (marginal) effect also takes into account the correlations among all the m characteristics.

Continuing with C_{A2}, we have:

$$C_{A2} = \frac{(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \mathbf{Z}^T \mathbf{S}^{-1} \mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_i^T (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{Z}_i (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{Z}_i (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

C_{A2} is the total distance measurement for the first level (individual level) random effect parameters between the complete dataset and the data with subset A removed. The term $\frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$ is actually the overall distance measurement of the random effects

for the ith subject at the jth time point. It is the total distance measurement of m characteristics, normalizing the residual variance-covariance matrix. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c\sigma_k^2}$, which is a simple summation of the distance measurements for all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the individual level random effect also takes into account the correlations among all the m characteristics.

Continuing with C_{A3},

$$
\begin{aligned}
C_{A3} &= \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{S}^{-1} \mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c} \\
&= 2\sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{Z}_i (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c} \\
&= 2\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}
\end{aligned}
$$

C_{A3} is the distance measure of covariation between the change in the population average profile and the change in the subject-specific profile relative to the population average profile. The term $\frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$ is actually the overall distance measurement of the covariation between the change in the population average profile and the change in the subject-specific profile relative to the population average profile for the ith subject at the jth time point. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c\sigma_k^2}$, which is a simple summation of the distance measurements for the covariance of all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the covariance also takes into account the correlations among all the m characteristics.

Note that the situation here is different from that presented by Tan[20], because we have multiple characteristics per individual at each time point whereas Tan had only one characteristic per individual at each time point.

2.4 SIMULATION STUDY

The purpose of our simulation study is two fold: (1) To demonstrate the conditional and naive multivariate Cook's distance for a single realization; and (2) To investigate the ability of each method to detect a "known" influential observation.

2.4.1 The Model used in the simulation study

The goal of our simulation study is to examine the performance of the multivariate conditional Cook's distance for detecting influential observations in a multivariate longitudinal dataset, and compare our results to that of the naive Cook's distance.

We generated a bivariate longitudinal dataset for our simulation study. The dataset contains n individuals and each individual has two characteristics, y_{ij1} and y_{ij2}, which are repeatedly measured. The bivariate mixed effect model is:

$$y_{ij1} = \beta_{10} + \beta_{11}u_{i1} + \beta_{12}t_{ij} + b_{1i} + \epsilon_{ij1}$$

$$y_{ij2} = \beta_{20} + \beta_{21}u_{i2} + \beta_{22}t_{ij} + b_{2i} + \epsilon_{ij2}$$

(2.7)

where i indicates the individual, $i = 1, \ldots, N$; j indicates the time point, $j = 1, \ldots, n_i$; n_i is randomly sampled from $\{1, 2, \ldots, 9\}$. The random effects $\boldsymbol{b}_i = [b_{1i}, b_{2i}]^T$, are generated from a bivariate normal distribution $\boldsymbol{b}_i = \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}\right)$. The fixed effects design matrix $\mathbf{X}_{ij} = \begin{bmatrix} 1 & u_{i1} & t_{ij} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & u_{i2} & t_{ij} \end{bmatrix}$, where t_{ij} is the jth time point for i^{th} individual, and u_{i1} and u_{i2} denote baseline covariates for the two characteristics. The random variables u_{i1} and u_{i2} were generated from a bivariate normal distribution, $\boldsymbol{u}_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$. $t_{ij} = \log(j)$, $\boldsymbol{\epsilon}_{ij} = \begin{bmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$. The true components of $\boldsymbol{\beta}$ are $[\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}]^T = [1, 1, 1, 1, 1, 1]^T$.

Since our goal of the simulation is only to compare the multivariate conditional Cook's distance and the naive multivariate Cook's distance, the model for this simulation is only

17

one-level, and the random effects only contain random intercepts. We generated 50 individuals ($N = 50$). Without loss of generality, we set the number of measurements of the 50th individual to be 9. Then we reset $b_{1,50} = 6$ for time point 5. Thus, the observation of y_1 at time point 5 of the 50th individual has strong influence due to extreme values of b_{1i}.

We repeated the simulation for 1000 times, we generated 1000 datasets according to 2.7, and then use our method to detect the "known" influential observation. The result of the simulation will be discussed in subsections 2.4.2 and 2.4.3.

Since the model of this simulation is a single-level multivariate mixed effect model, we used R package `mlmmm` [52] to fit the model. The single-level multivariate mixed effect model with missing value and correlated error term can also be fitted using SAS PROC MIXED (SAS version 9.2 or later), the following repeated statement allows one to fit the desired error structure:

```
random int_b1 int_b2 /subject=id type=un g gcorr;
repeated var_type /subject=id*visit_order type=un r rcorr;
```

If we want to fit the single level multivariate mixed effect model with independent errors structure ($\boldsymbol{\Sigma}$ is diagonal), just simply change the option to `type=vc` in the repeated statement above.

2.4.2 Demonstration of method for one dataset

In this section we illustrate one of the 1000 realizations. Figure 2.1 shows the scattergram of the relationship between the response $[y_{ij1}, y_{ij2}]^T$ and the time points. Note that the blue points indicates y_1 and the red points indicates y_2. It can be seen that the fifth observation of y_1 (blue point) of individual 50 is an extremely high value. In this simulation dataset the individuals have at least 1 measurement and 9 at most.
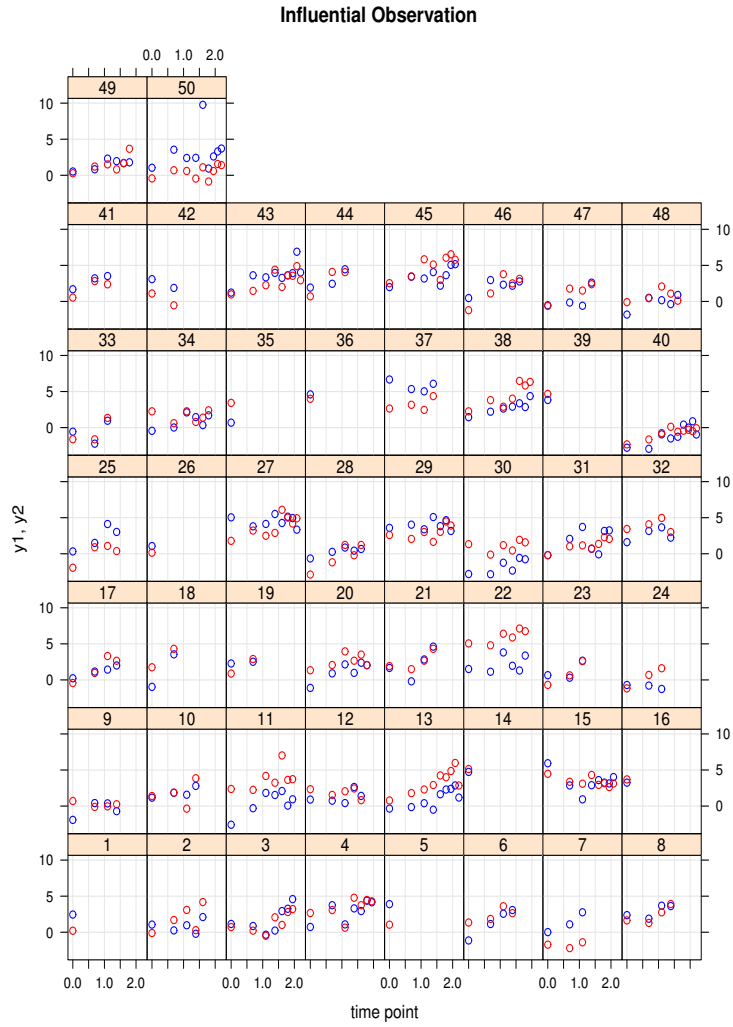
Figure 2.1: Scattergram of one simulated dataset

Figure 2.2 shows the multivariate conditional Cook's distance for all observations. Clearly the y_1 value of the fifth measurement of individual 50 was detected.
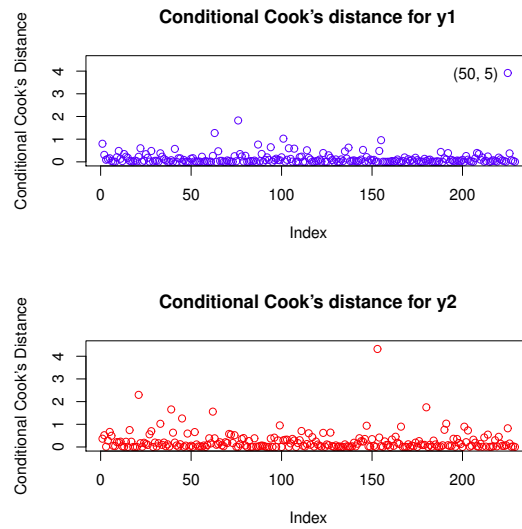
Figure 2.2: Conditional Cook's Distance for all observations

Figure 2.3 shows the multivariate naive Cook's distance for all observations. Clearly the y_1 value of the fifth measurement of individual 50 was NOT detected.
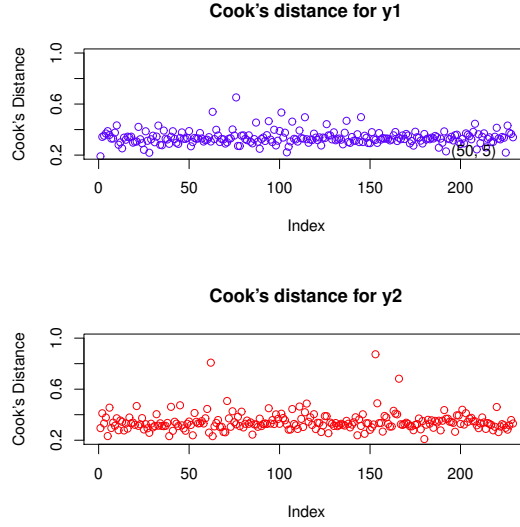
Figure 2.3: Cook's Distance for all observations

Figure 2.4 shows, for each observation, the relative changes (in percentage) in the estimated fixed effects of y_1, $\hat{\beta}_{10}$, $\hat{\beta}_{11}$, $\hat{\beta}_{12}$, the estimated fixed effects of y_2, $\hat{\beta}_{20}$, $\hat{\beta}_{21}$, $\hat{\beta}_{22}$, and the estimated random intercept of the 50th individual, $\hat{b}_{50,1}$ and $\hat{b}_{50,2}$. Note that the percentages of change for $\hat{b}_{50,1}$ and $\hat{b}_{50,2}$ were divided by 10 so it can be shown more clearly in the plot. The actual relative change of $\hat{b}_{50,1,y_2}$ is around 51%, not 5%. The blue points indicates y_1 and the red points indicates y_2.

Relative changes of all parameters
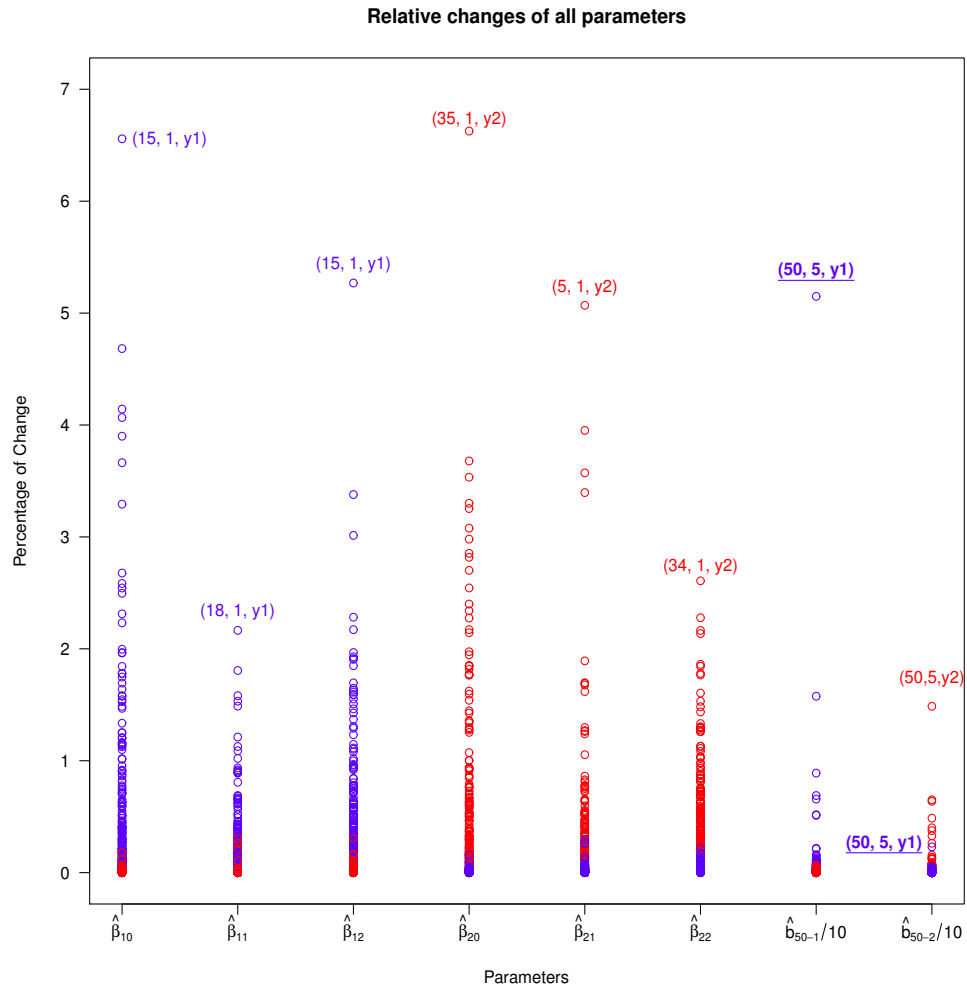
Figure 2.4: Relative changes for all parameters estimated (in percentage)

Figure 2.4 indicates that the fifth observation of y_1 of the 50th individual (the extreme observation we made) does not have the largest effect on any of the six fixed effects parameters. But it shows that the value of $\hat{b}_{50,1}$ (the random intercept of y_1 of the 50th individual) is strongly influenced by the fifth observation of y_1 of the 50th individual.

Also, Figure 2.4 shows that observations of y_1 have much stronger influence on y_1's parameters (both fixed effects and random effects) than those of y_2, and similarly for y_2. This is understandable.

For the random intercept of y_2 of the 50th individual, of course one of the observations of y_2 has the largest influence. But it is noticable that, among the observations of y_1, the fifth observation of the 50th individual (the extreme observation we made) has the largest effect. That is because the two characteristics y_1 and y_2 are correlated (estimated correlation coefficient is 0.3904).

2.4.3 Comparing Performance of Methods (1000 realizations)

In order to compare our extended conditional Cook's distance to that of the unconditional (original) Cook's distance, we repeated the simulation 1000 times. Accordingly, we generated 1000 datasets using to the Model 2.7, and then use our method to detect a "known" influential observation in the 1000 datasets.

For each of the 1000 datasets, a bivariate linear mixed effect model was fitted, and the model parameters, variance-covariance matrices were calculated. The averages of the 1000 estimated model parameters, random effect variance-covariance matrix ($\mathbf{G}$ matrix) and residual variance-covariance matrix ($\mathbf{\Sigma}$ matrix) are listed below. Table 2.1 shows the average estimation of the fixed effect parameters and the standard deviations for 1000 simulations.

Table 2.1: The average estimation of the fixed effect parameters for 1000 simulations

Parameters	Estimated value	Standard Deviation
$\hat{\beta}_{10}$	1.0021	0.0341
$\hat{\beta}_{11}$	0.9992	0.0257
$\hat{\beta}_{12}$	1.0164	0.0100
$\hat{\beta}_{20}$	0.9922	0.0448
$\hat{\beta}_{21}$	0.8018	0.0326
$\hat{\beta}_{22}$	0.9998	0.0110

The estimated **G** matrix and its associated correlation matrix, **G** are:

$$\overline{\hat{\mathbf{G}}} = \begin{array}{c} \\ b_1 \\ b_2 \end{array} \begin{pmatrix} \overset{b_1}{0.8996} & \overset{b_2}{0.1925} \\ 0.1925 & 1.3020 \end{pmatrix} \qquad \overline{\hat{\mathcal{G}}} = \begin{array}{c} \\ b_1 \\ b_2 \end{array} \begin{pmatrix} \overset{b_1}{1.0000} & \overset{b_2}{0.1778} \\ 0.1778 & 1.0000 \end{pmatrix}$$

We can see the estimated $\overline{\hat{\rho}}_G = 0.1778$ is close to the true value $\rho_G = 0.20$. The model fits well for the 1000 simulations.

The estimated $\boldsymbol{\Sigma}$ matrix averaged over the 1000 simulations and the associated correlation matrix are:

$$\overline{\hat{\boldsymbol{\Sigma}}} = \begin{array}{c} \\ y_1 \\ y_2 \end{array} \begin{pmatrix} \overset{y_1}{1.2171} & \overset{y_2}{0.4976} \\ 0.4976 & 0.9962 \end{pmatrix} \qquad \overline{\hat{\boldsymbol{\Sigma}}} = \begin{array}{c} \\ y_1 \\ y_2 \end{array} \begin{pmatrix} \overset{y_1}{1.0000} & \overset{y_2}{0.4519} \\ 0.4519 & 1.0000 \end{pmatrix}$$

We can see the estimated $\hat{\rho}_R = 0.4519$, the true value $\rho_R = 0.50$. The model fits well.

Our multivariate conditional Cook's distance successfully detected the "known" influential observation in 925 of the 1000 datasets. The original Cook's distance only detected the "known" influential observation in 262 of the 1000 datasets. In Table 2.2, a contingency table for the multivariate conditional Cook's distance and the original Cook's distance summarizes the result from the 1000 simulations.

Table 2.2: Number of detections for conditional and original Cook's distance

	Original Cook's D		
Conditional Cook's D	No	Yes	Total
No	75	0	75
Yes	663	262	925
Total	738	262	1000

2.5 APPLICATION

We applied our method to glaucoma clinical data. In this study, patients' eyes were repeatedly measured for multiple responses. The purpose of our investigation is to jointly model the thicknesses of retina nerve fiber layer (RNFL) and retinal ganglion cells complex (GCC), and find out if there are some abnormal measurements. The dataset is from UPMC eye center.

The following summarizes some information about the data:

- Total 487 eyes from 256 patients;
- Eyes were divided into three diagnostic groups: healthy(H), glaucoma suspect(GS) and glaucoma(G);
- There were 97 healthy eyes, 279 glaucoma suspect eyes and 111 glaucoma eyes;
- Patients' baseline ages (in years) vary from 40.5 to 81.9;
- The follow-up duration (in years) varies from 1.322 to 6.398;
- The outcomes were retina nerve fiber layer (RNFL) thickness and retinal ganglion cells complex (GCC) thickness;
- There were a total of $5,994$ observations among the 256 patients.

The simultaneous outcomes of RNFL and GCC are typically correlated, and these outcomes were longitudinally measured.

Note that in this model, all eyes were assumed to be independent. This assumption is NOT correct, because two eyes from one patient are typically correlated. But here we just use the dataset to demonstrate the method. In the next chapter we will show a multilevel multivariate mixed effect model which takes into account the correlation between both eyes for each patient.

2.5.1 The Model

We fitted the following bivariate linear mixed effect model:

$$Y_{RNFL} = (\beta_{10}N + \beta_{11}GS + \beta_{12}G + b_{10})$$
$$+ (\beta_{13}N + \beta_{14}GS + \beta_{15}G + b_{11}) Fu + \beta_{16}Age + \epsilon_{RNFL}$$
$$Y_{GCC} = (\beta_{20}N + \beta_{21}GS + \beta_{22}G + b_{20})$$
$$+ (\beta_{23}N + \beta_{24}GS + \beta_{25}G + b_{21}) Fu + \beta_{26}Age + \epsilon_{GCC}$$

where Age indicates baseline age; Fu indicates Follow-up (in years). We assume that:

- $\boldsymbol{b} = [b_{10}, b_{20}, b_{11}, b_{21}]^T \sim N(\boldsymbol{0}, \mathbf{G})$,

- $\boldsymbol{\epsilon} = [\epsilon_{RNFL}, \epsilon_{GCC}]^T \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{2\times2})$

Table 3.1 shows the estimated parameters of fixed effects:

Table 2.3: The estimated parameters of fixed effects

β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}
111.70	105.38	94.391	-0.0606	-0.676	-0.753	-0.186

β_{20}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}
107.70	102.48	94.968	-0.972	-0.592	-0.475	-0.221

The estimated variance-covriance matrices of the random effect ($\mathbf{G}$) is:

$$
\hat{\mathbf{G}} = \begin{array}{c} \\ b_{10} \\ b_{20} \\ b_{11} \\ b_{21} \end{array} \begin{array}{cccc} b_{10} & b_{20} & b_{11} & b_{21} \\ \begin{pmatrix} 94.507 & 51.264 & -0.965 & -0.255 \\ 51.264 & 51.848 & -0.0523 & -0.258 \\ -0.965 & -0.0523 & 0.904 & 0.0957 \\ -0.255 & -0.258 & 0.0957 & 0.744 \end{pmatrix} \end{array}
$$

The correlation matrix of $\mathbf{G}$ is:

$$
\hat{\mathcal{G}} = \begin{array}{c} \\ b_{10} \\ b_{20} \\ b_{11} \\ b_{21} \end{array} \begin{array}{cccc} b_{10} & b_{20} & b_{11} & b_{21} \\ \begin{pmatrix} 1.000 & 0.732 & -0.104 & -0.0305 \\ 0.732 & 1.000 & -0.00764 & -0.0415 \\ -0.104 & -0.00764 & 1.000 & 0.117 \\ -0.0305 & -0.0415 & 0.117 & 1.000 \end{pmatrix} \end{array}
$$

The estimated residual variance-covariance matrix is:

$$
\hat{\boldsymbol{\Sigma}} = \begin{array}{c} \\ Y_{RNFL} \\ Y_{GCC} \end{array} \begin{array}{cc} Y_{RNFL} & Y_{GCC} \\ \begin{pmatrix} 13.387 & 0.292 \\ 0.292 & 8.458 \end{pmatrix} \end{array}
$$

Its associated correlation matrix is:

$$
\hat{\boldsymbol{\Sigma}} = \begin{array}{c} \\ Y_{RNFL} \\ Y_{GCC} \end{array} \begin{array}{cc} Y_{RNFL} & Y_{GCC} \\ \begin{pmatrix} 1.0000 & 0.0274 \\ 0.0274 & 1.0000 \end{pmatrix} \end{array}
$$

2.5.2 The influential observations (observation level)

Using our method, we calculated the observation level conditional Cook's distance and the decomposed C_{A1}, C_{A2} and C_{A3}. "Observation level" means that the subset A to be removed is the whole observation of the ith subject at j^{th} time point. That is, A contains both RNFL and GCC values measured at the jth time point for the ith subject.

Figure 3.3 is the illustration of the 10 observations in 10 eyes. and the Table 2.4 contains a list of the 10 observations with largest value of conditional Cook's distance. Note that the blue circles and lines indicate the observed RNFL values and individual fitted regression lines for RNFL. The blue dotted lines indicate the patient-level fitted regression lines for RNFL. The light blue lines indicate the marginal fitted regression lines for RNFL. Similarly, the red cirles, lines and pink lines are for GCC as well.

Figure 2.5: 10 observations with largest conditional cook's distance in 10 eyes

Table 2.4: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance

Eye ID	Follow-up (in years)	Follow-up (in days)	Diagonstic group	Condional Cook SD ($\times 10^{-3}$)	C_{A1} ($\times 10^{-3}$)	C_{A2} ($\times 10^{-3}$)	C_{A3} ($\times 10^{-3}$)
2:204:OS	0.5448	199	2	3.597	0.022	3.577	−0.002
2:102:OD	2.0205	738	2	2.645	0.036	2.615	−0.006
2:102:OS	2.0205	738	2	2.321	0.033	2.294	−0.006
3:170:OS	2.9103	1063	3	2.292	0.101	2.240	−0.048
4:174:OS	1.5387	562	3	2.151	0.059	2.102	−0.010
2:191:OS	0.9391	343	3	1.932	0.035	1.905	−0.009
4:91:OS	0.8487	310	2	1.921	0.013	1.908	0.000
4:91:OD	4.8953	1788	3	1.848	0.027	1.811	0.010
2:72:OD	5.5633	2032	2	1.822	0.008	1.807	0.007
2:191:OD	0.9391	343	3	1.767	0.031	1.743	−0.007
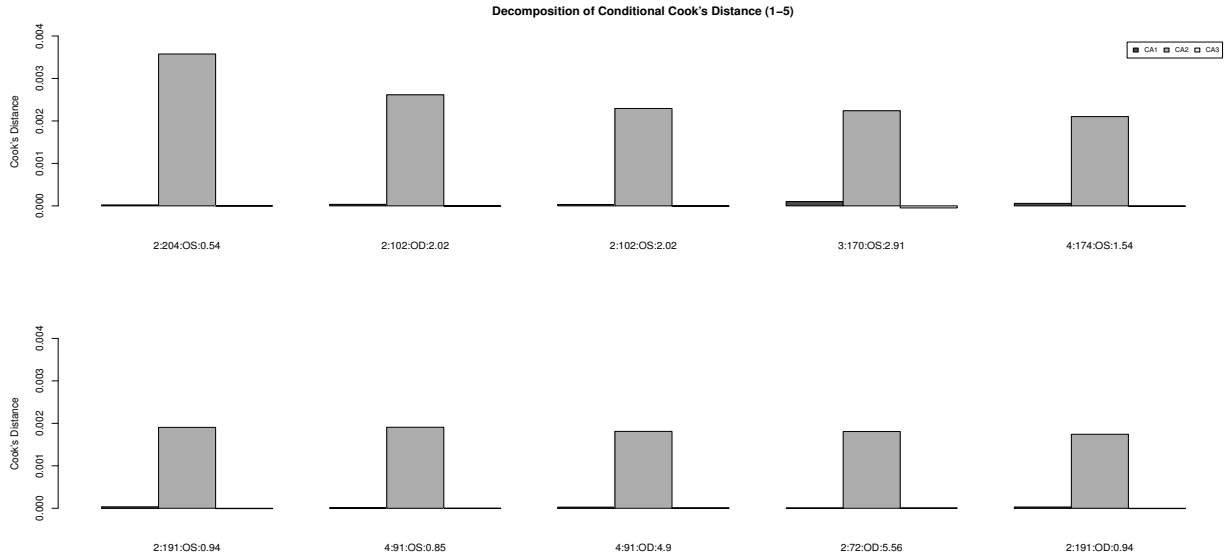
Figure 2.6: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance

Figure 2.6 shows the values of the three decomposition terms of the multivariate conditional Cook's distance. We notice that for most observations, the distance measurement of the random effects are much greater than the distance measurement of the fixed effects, that is, $C_{A2} \gg C_{A1}$. This is obvious because the subject-specific effects are much more sensitive to the influential observation than the marginal effects. Also, the covariance between the distance measurements of fixed and random random effects, that is, C_{A3}, is very small. This is similar to Tan's [20] conclusion in the univariate case.

2.5.3 The influential observations (component level)

Using our method, we also calculated the component level conditional Cook's distance and the decomposed C_{A1}, C_{A2} and C_{A3}. "Component level" means that the subset A to be removed is only one component of the whole observation of the ith subject at j^{th} time point. That is, A contains only one of the two components, either RNFL or GCC value measured at the jth time point for the ith subject.

Figure 2.7 illustrates the 10 components in 10 eyes. and the Table 2.5 contains a list of the 10 components with largest value of conditional Cook's distance.
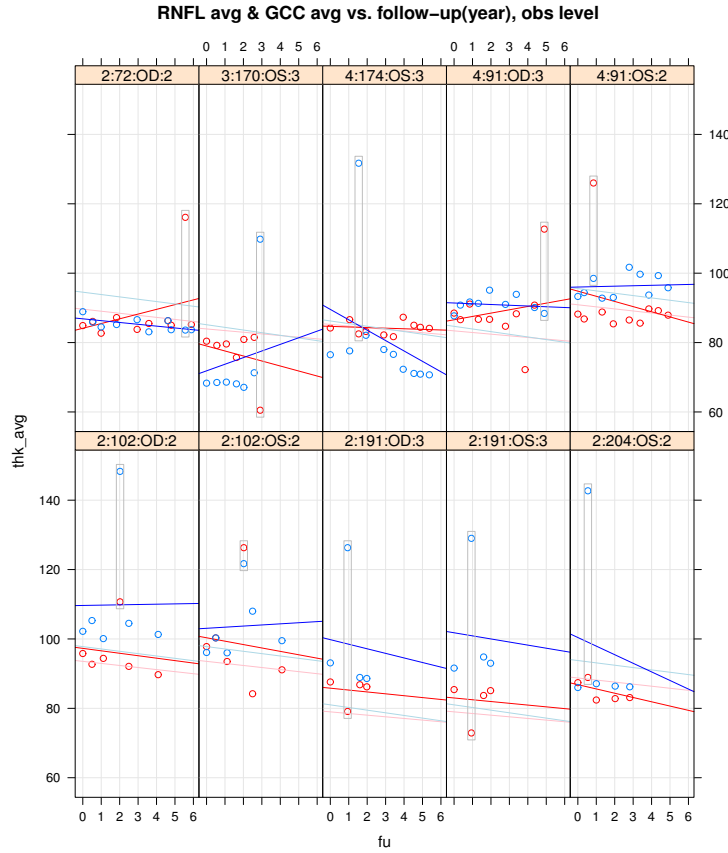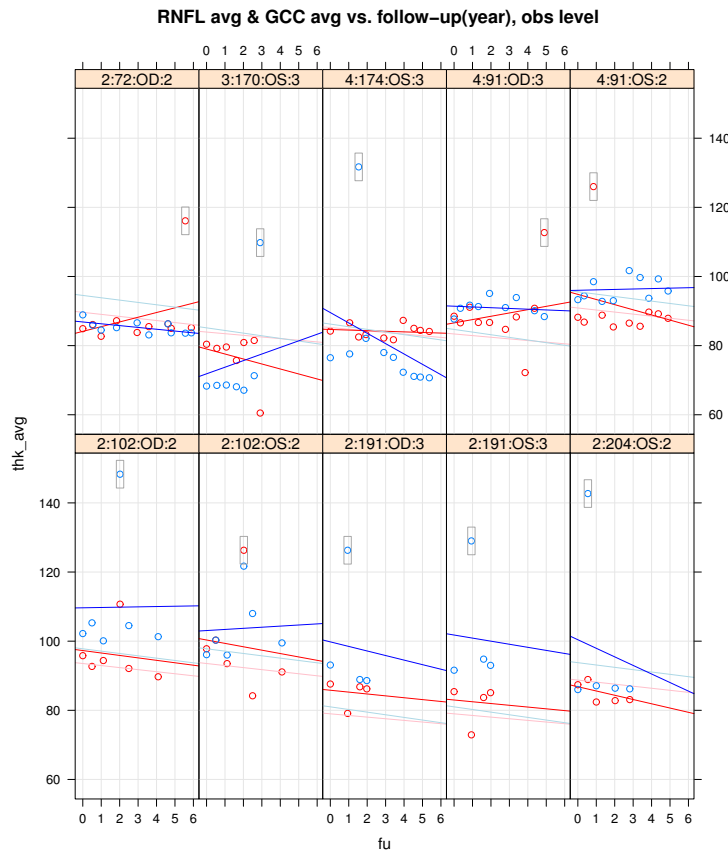

Figure 2.7: 10 components with largest conditional cook's distance in 10 eyes

Table 2.5: Decomposition of the Conditional Cook's Distance for the 10 components with largest conditional cook's distance

| Eye ID	Type	Follow-up (years)	Follow-up (days)	Diag group	Conditional Cook SD ($\times 10^{-3}$)	C_{A1} ($\times 10^{-3}$)	C_{A2} ($\times 10^{-3}$)	C_{A3} ($\times 10^{-3}$)
2:204:OS	RNFL	0.5448	199	2	3.491	0.0220	3.471	−0.00235
4:174:OD	RNFL	1.5387	562	3	2.167	0.0586	2.118	−0.00981
2:102:OD	RNFL	2.0205	738	2	2.079	0.0297	2.054	−0.00488
4:91:OS	GCC	0.8487	310	2	1.888	0.0129	1.875	0.00570
4:91:OD	GCC	4.8953	1788	3	1.848	0.0270	1.811	0.00959
2:72:OD	GCC	5.5633	2032	2	1.823	0.0077	1.808	0.00750
2:191:OS	RNFL	0.9391	343	3	1.836	0.0314	1.812	−0.00787
2:102:OS	GCC	2.0205	738	2	1.808	0.0270	1.786	−0.00544
3:170:OS	RNFL	2.9103	1063	3	1.821	0.0838	1.777	−0.00394
2:191:OD	RNFL	0.9391	343	3	1.770	0.0292	1.748	−0.00682
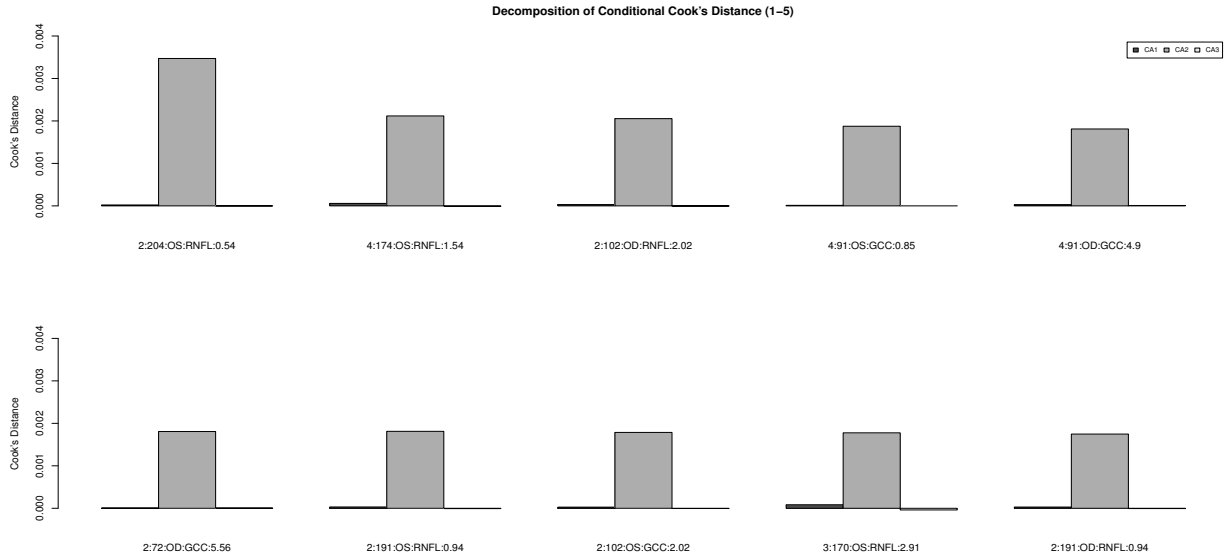
Figure 2.8: Decomposition of the Conditional Cook's Distance for the 10 components with largest conditional cook's distance

Figure 2.8 shows the values of the three decomposition terms of the multivariate conditional Cook's distance. Similar to the observation level, for most components, the distance measurement of the random effects are much greater than the distance measurement of the fixed effects, that is, $C_{A2} \gg C_{A1}$. This is obvious because the subject-specific effects are much more sensitive to the influential observation than the marginal effects. Also, the covariance between the distance measurements of fixed and random random effects, that is, C_{A3}, is very small. This is similar to Tan's [20] conclusion in the univariate case.

3.0 MULTILEVEL MULTIVARIATE LONGITUDINAL EXTENSION OF COOK'S DISTANCE

The multilevel multivariate mixed effect model is an extension of models 2.2, 2.3 and 2.4. The extension is obtained for cases when repeated multivariate measurements on subjects are further clustered into larger groups. Such designs arise in many applications. An example is the ophthalmology clinical data. Multivariate measurements (e.g. retina thickness and mean deviation) were repeatedly measured on each eye (cluster), and each patient (group) has two eyes (clusters). Another example is in educational studies. Students are repeatedly taking exams of multiple courses (reading, math, science, etc), with students belonging to schools, which are in turn clustered within school districts, and so forth.

The extended model will be motivated by the analyses of ophthalmology clinical data. In ophthalmic data, most of the patients contribute two eyes, and for each eye a set of m characteristics are repeatedly measured for n_i times. Obviously, two eyes from one patient are dependent, and they are typically highly correlated.

3.1 MULTILEVEL MULTIVARIATE MIXED EFFECT MODEL

The following describes our two-level multivariate mixed effect model. The model for the observations at j^{th} time point of hth cluster in i^{th} group is:

$$\underbrace{\boldsymbol{y}_{(ih)j}}_{m \times 1} = \underbrace{\mathbf{X}_{ij}}_{m \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\mathbf{Z}_{ij}}_{m \times q} \underbrace{\boldsymbol{b}_i}_{q \times 1} + \underbrace{\mathbf{W}_{(ih)j}}_{m \times r} \underbrace{\boldsymbol{c}_{ih}}_{r \times 1} + \underbrace{\boldsymbol{\epsilon}_{(ih)j}}_{m \times 1} \tag{3.1}$$

The model for h^{th} cluster in i^{th} group for all variables and time points is given as:

$$\underbrace{\boldsymbol{y}_{(ih)}}_{mn_i \times 1} = \underbrace{\mathbf{X}_i}_{mn_i \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\mathbf{Z}_i}_{mn_i \times q} \underbrace{\boldsymbol{b}_i}_{q \times 1} + \underbrace{\mathbf{W}_{(ih)}}_{mn_i \times r} \underbrace{\boldsymbol{c}_{ih}}_{r \times 1} + \underbrace{\boldsymbol{\epsilon}_{(ih)}}_{mn_i \times 1} \tag{3.2}$$

where:

- $h = 1, 2, ..., s$, indicates the clusters in each group. There are a total of s clusters in each group. For example, in the ophthalmic data, $h = 1, 2$, $s = 2$, indicating two clusters (eyes) in each group (patient).

- $\boldsymbol{b}_i$ is the group level random effect for the ith group, $\boldsymbol{b}_i \sim N_q(\mathbf{0}, \mathbf{G}_1)$, $\mathbf{Z}_i$ is the design matrix of random effects for the ith group. In the ophthalmic data example, $\boldsymbol{b}_i$ is the patient level random effect for the ith patient.

- $\boldsymbol{c}_{ih}$ is the cluster level random effect for the hth cluster in the ith group, $\boldsymbol{c}_{ih} \sim N_r(\mathbf{0}, \mathbf{G}_2)$, $\mathbf{W}_{ih}$ is the design matrix of random effects for the hth cluster in the ith group. In the ophthalmic data example, $\boldsymbol{c}_{ih}$ is the eye level random effect for the hth eye of the ith patient. We assume that $\boldsymbol{c}_{ih}$ are independently distributed given $\boldsymbol{b}_i$.

The model for the ith group (patient) is:

$$
\begin{aligned}
\boldsymbol{y}_i &= \begin{bmatrix} \boldsymbol{y}_{i1} \\ \boldsymbol{y}_{i2} \\ \vdots \\ \boldsymbol{y}_{is} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \mathbf{W}_{i1}\boldsymbol{c}_{i1} + \boldsymbol{\epsilon}_{i1} \\ \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \mathbf{W}_{i2}\boldsymbol{c}_{i2} + \boldsymbol{\epsilon}_{i2} \\ \vdots \\ \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \mathbf{W}_{is}\boldsymbol{c}_{is} + \boldsymbol{\epsilon}_{is} \end{bmatrix} \\[2ex]
&= \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_i \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_i \\ \mathbf{Z}_i \\ \vdots \\ \mathbf{Z}_i \end{bmatrix}\boldsymbol{b}_i + \begin{bmatrix} \mathbf{W}_{i1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{i2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{W}_{is} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_{i1} \\ \boldsymbol{c}_{i2} \\ \vdots \\ \boldsymbol{c}_{is} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{i1} \\ \boldsymbol{\epsilon}_{i2} \\ \vdots \\ \boldsymbol{\epsilon}_{is} \end{bmatrix} \\[2ex]
&= (\mathbf{1}_s \otimes \mathbf{X}_i)\boldsymbol{\beta} + (\mathbf{1}_s \otimes \mathbf{Z}_i)\boldsymbol{b}_i + \mathbf{W}_i\boldsymbol{c}_i + \boldsymbol{\epsilon}_i \\
&= \mathbf{X}_i^*\boldsymbol{\beta} + \mathbf{Z}_i^*\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_i
\end{aligned}
\tag{3.3}
$$

where:

- $\mathbf{W}_i = \text{diag}(\mathbf{W}_{i1}, \dots, \mathbf{W}_{is})$
- $\mathbf{X}_i^* = \mathbf{1}_s \otimes \mathbf{X}_i$; $\mathbf{Z}_i^* = [\mathbf{1}_s \otimes \mathbf{Z}_i, \ \mathbf{W}_i]$.

- $\boldsymbol{b}_i^* = \left[\boldsymbol{b}_i^T, \boldsymbol{c}_{i1}^T, \boldsymbol{c}_{i2}^T, \ldots, \boldsymbol{c}_{is}^T\right]^T$; $\boldsymbol{c}_i^* = \left[\boldsymbol{c}_{i1}^T, \boldsymbol{c}_{i2}^T, \ldots, \boldsymbol{c}_{is}^T\right]^T$.

- $\boldsymbol{b}_i^* \sim N(\boldsymbol{0}, \mathbf{G}^*)$. Note that $\boldsymbol{c}_{ih}$ are independently distributed given $\boldsymbol{b}_i$. So, we have $\mathbf{G}^* = \mathrm{diag}(\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_2)$.

- $\mathbf{J}_s$ is the $s \times s$ matrix with all entries 1. $\mathbf{I}_s$ is the $s \times s$ identity matrix.

The variance of $\boldsymbol{y}_i$ is:

$$\mathbf{V}_i = \mathrm{var}(\boldsymbol{y}_i) = \mathbf{Z}_i^* \mathbf{G}^* \mathbf{Z}_i^{*T} + \mathbf{R}_i^*$$

$$= \begin{bmatrix} \mathbf{Z}_i & \mathbf{W}_{i1} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_i & \mathbf{0} & \ldots & \mathbf{W}_{is} \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_i^T & \ldots & \mathbf{Z}_i^T \\ \mathbf{W}_{i1}^T & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{W}_{is}^T \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T + \mathbf{W}_{i1} \mathbf{G}_2 \mathbf{W}_{i1}^T & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T & \ldots & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T \\ \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T + \mathbf{W}_{i2} \mathbf{G}_2 \mathbf{W}_{i2}^T & \ldots & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T & \ldots & \mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T + \mathbf{W}_{is} \mathbf{G}_2 \mathbf{W}_{is}^T \end{bmatrix}$$

$$+ \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})$$

$$= \mathbf{J}_s \otimes (\mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T) + \begin{bmatrix} \mathbf{W}_{i1} \mathbf{G}_2 \mathbf{W}_{i1}^T & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{W}_{is} \mathbf{G}_2 \mathbf{W}_{is}^T \end{bmatrix} + \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})$$

$$= \mathbf{J}_s \otimes (\mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T) + \begin{bmatrix} \mathbf{W}_{i1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{W}_{is} \end{bmatrix} \begin{bmatrix} \mathbf{G}_2 & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{W}_{i1}^T & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{W}_{is}^T \end{bmatrix}$$

$$+ \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})$$

$$= \mathbf{J}_s \otimes (\mathbf{Z}_i \mathbf{G}_1 \mathbf{Z}_i^T) + \mathbf{W}_i(\mathbf{I}_s \otimes \mathbf{G}_2)\mathbf{W}_i^T + \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})$$

Without conditioning on each group, cluster and characteristic, the marginal model of $\boldsymbol{y}$ is:

$$\boldsymbol{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \quad \mathbf{J}_s \otimes (\mathbf{Z}_i\mathbf{G}_1\mathbf{Z}_i^T) + \mathbf{W}_i(\mathbf{I}_s \otimes \mathbf{G}_2)\mathbf{W}_i^T + \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}))$$

.

we have following covariances:

$$\text{cov}(\boldsymbol{y}_i, \boldsymbol{b}_i) = \text{cov}\left((\mathbf{1}_s \otimes \mathbf{X}_i)\boldsymbol{\beta} + (\mathbf{1}_s \otimes \mathbf{Z}_i)\boldsymbol{b}_i + \mathbf{W}_i\boldsymbol{c}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{b}_i\right) = (\mathbf{1}_s \otimes \mathbf{Z}_i)\text{cov}(\boldsymbol{b}_i, \boldsymbol{b}_i)$$

$$= (\mathbf{1}_s \otimes \mathbf{Z}_i)\mathbf{G}_1$$

$$\text{cov}(\boldsymbol{y}_i, \boldsymbol{c}_i) = \text{cov}\left((\mathbf{1}_s \otimes \mathbf{X}_i)\boldsymbol{\beta} + (\mathbf{1}_s \otimes \mathbf{Z}_i)\boldsymbol{b}_i + \mathbf{W}_i\boldsymbol{c}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{c}_i\right) = \mathbf{W}_i\text{cov}(\boldsymbol{c}_i, \boldsymbol{c}_i) = \mathbf{W}_i\mathbf{G}_2$$

Under the assumption of normality, the joint distribution of $\boldsymbol{y}_i$, $\boldsymbol{b}_i$, $\boldsymbol{c}_i$, and $\boldsymbol{\epsilon}_i$ is multivariate normal:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \\ \boldsymbol{c}_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Psi}), \quad \text{where: } \boldsymbol{\mu} = \begin{bmatrix} (\mathbf{1}_s \otimes \mathbf{X}_i)\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \text{ and}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{J}_s \otimes (\mathbf{Z}_i\mathbf{G}_1\mathbf{Z}_i^T) + \mathbf{W}_i(\mathbf{I}_s \otimes \mathbf{G}_2)\mathbf{W}_i^T + \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}) & \mathbf{G}_1(\mathbf{1}_s \otimes \mathbf{Z}_i)^T & \mathbf{G}_2\mathbf{W}_i^T & \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}) \\ (\mathbf{1}_s \otimes \mathbf{Z}_i)\mathbf{G}_1^T & \mathbf{G}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_i\mathbf{G}_2^T & \mathbf{0} & \mathbf{G}_2 & \mathbf{0} \\ \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}) & \mathbf{0} & \mathbf{0} & \mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}) \end{bmatrix}$$

3.2 INFLUENCE FUNCTIONS

As was emphasized earlier, influence functions are functions for assessing the effect (or "influence") of removing an observation (or a subset of observations) on the value of a statistic without having to re-compute that statistic. Here, we derive the functions to calculate the estimation with removal of an observation (or a subset of observation), $\boldsymbol{\beta}_{(A)}$, $\boldsymbol{b}_{(A)}$, $\boldsymbol{c}_{(A)}$ from the complete data estimation, $\boldsymbol{\beta}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.

From 3.3 we have the model for the complete dataset:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon}_i \tag{3.4}$$

where: $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^* \\ \vdots \\ \mathbf{X}_N^* \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1^* & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{Z}_N^* \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1^* \\ \vdots \\ \boldsymbol{b}_N^* \end{bmatrix}$, $\boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix}$,

$$\boldsymbol{b} \sim N(\mathbf{0}, \mathbf{G}), \quad \mathbf{G} = \text{diag}_N(\mathbf{G}^*, \ldots, \mathbf{G}^*)$$

For simplicity of notation, in this section, we use i to denote the ith observation in the complete dataset, and k to denote the number of observations in the subset of interest. Hence, the subset A to be removed can be identified as beginning at the ith observation, having a total of k observations. Thus, A has a cardinality of k.

Without loss of generality, we partition the matrices as if the subset A of observations to be removed are the beginning k observations; That is, from the 1st to the kth of the total N observations. Using a similar notation as in [25], we have:

$$\underset{N \times 1}{\boldsymbol{y}} = \begin{bmatrix} \underset{k \times 1}{\boldsymbol{y}_i^T} \\ \underset{(N-k) \times 1}{\boldsymbol{y}_{(i)}} \end{bmatrix}, \quad \underset{N \times p}{\mathbf{X}} = \begin{bmatrix} \underset{k \times p}{\mathbf{X}_i^T} \\ \underset{(N-k) \times p}{\mathbf{X}_{(i)}} \end{bmatrix}, \quad \underset{N \times q}{\mathbf{Z}} = \begin{bmatrix} \underset{k \times q}{\mathbf{Z}_i^T} \\ \underset{(N-k) \times q}{\mathbf{Z}_{(i)}} \end{bmatrix}, \qquad (3.5)$$

$$\text{and} \quad \underset{N \times N}{\mathbf{V}} = \begin{bmatrix} \underset{k \times k}{\mathbf{V}_{ii}} & \underset{k \times (N-k)}{\mathbf{V}_i^T} \\ \underset{(N-k) \times k}{\mathbf{V}_i} & \underset{(N-k) \times (N-k)}{\mathbf{V}_{[i]}} \end{bmatrix}$$

We first need to derive $\mathbf{A}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{B}_{(i)}$. If we know $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ can be partitioned in a same way as $\mathbf{X}$ and $\mathbf{Z}$ above, then

$$\underset{N \times a}{\mathbf{A}} = \begin{bmatrix} \underset{k \times a}{\mathbf{A}_i^T} \\ \underset{(N-k) \times a}{\mathbf{A}_{(i)}} \end{bmatrix}, \quad \underset{N \times b}{\mathbf{B}} = \begin{bmatrix} \underset{k \times b}{\mathbf{B}_i^T} \\ \underset{(N-k) \times b}{\mathbf{B}_{(i)}} \end{bmatrix}$$

From the inverse of partitioned matrix, we have:

$$\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{V}^{ii} & \mathbf{\Lambda}_i^T \\ \mathbf{\Lambda}_i & \mathbf{\Lambda}_{[i]} \end{bmatrix} = \begin{bmatrix} (\mathbf{V}_{ii} - \mathbf{V}_i^T \mathbf{V}_{[i]}^{-1} \mathbf{V}_i)^{-1} & -\mathbf{V}_{ii}^{-1} \mathbf{V}_i^T (\mathbf{V}_{[i]} - \mathbf{V}_i \mathbf{V}_{ii}^{-1} \mathbf{V}_i^T)^{-1} \\ -\mathbf{V}_{[i]}^{-1} \mathbf{V}_i (\mathbf{V}_{ii} - \mathbf{V}_i^T \mathbf{V}_{[i]}^{-1} \mathbf{V}_i)^{-1} & (\mathbf{V}_{[i]} - \mathbf{V}_i \mathbf{V}_{ii}^{-1} \mathbf{V}_i^T)^{-1} \end{bmatrix}$$

$$(3.6)$$

$$\mathbf{\Lambda}_i^T = -\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T(\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)^{-1}$$

$$= -(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T(\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)^{-1}$$

$$= -(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}(\mathbf{V}_i^T - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)(\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)^{-1}$$

$$= -(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}(\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)(\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)^{-1}$$

$$= -(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}$$

This confirms that $\mathbf{V}$ is symmetric.

From Woodbury's formula, we have

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}$$

and thus

$$\mathbf{\Lambda}_{[i]} = (\mathbf{V}_{[i]} - \mathbf{V}_i\mathbf{V}_{ii}^{-1}\mathbf{V}_i^T)^{-1}$$

$$= (\mathbf{V}_{[i]} + \mathbf{V}_i(-\mathbf{V}_{ii}^{-1})\mathbf{V}_i^T)^{-1}$$

$$= \mathbf{V}_{[i]}^{-1} - \mathbf{V}_{[i]}^{-1}\mathbf{V}_i(-\mathbf{V}_{ii} + \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}$$

$$= \mathbf{V}_{[i]}^{-1} + \mathbf{V}_{[i]}^{-1}\mathbf{V}_i(\mathbf{V}_{ii} - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}$$

Letting $\underset{k\times k}{\mathbf{S}_i} = \underset{k\times k}{\mathbf{V}_{ii}} - \underset{k\times(N-k)}{\mathbf{V}_i^T}\underset{(N-k)\times(N-k)}{\mathbf{V}_{[i]}^{-1}}\underset{(N-k)\times k}{\mathbf{V}_i}$, we have:

$$\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{S}_i^{-1} & -\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1} \\ -\mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1} & \mathbf{V}_{[i]}^{-1} + \mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1} \end{bmatrix}$$

Hence,

$$\mathbf{A}^T\mathbf{V}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{A}_i & \mathbf{A}_{(i)}^T \end{bmatrix} \begin{bmatrix} \mathbf{S}_i^{-1} & -\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1} \\ -\mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1} & \mathbf{V}_{[i]}^{-1} + \mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}_i^T \\ \mathbf{B}_{(i)} \end{bmatrix}$$

$$= \mathbf{A}_i\mathbf{S}_i^{-1}\mathbf{B}_i^T - \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1}\mathbf{B}_i^T - \mathbf{A}_i\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)} + \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)}$$

$$+ \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i\mathbf{S}_i^{-1}\mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)}$$

$$= \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)} + (\mathbf{A}_i - \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)\mathbf{S}_i^{-1}(\mathbf{B}_i^T - \mathbf{V}_i^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)})$$

$$= \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{B}_{(i)} + (\mathbf{A}_i - \mathbf{A}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)\mathbf{S}_i^{-1}(\mathbf{B}_i - \mathbf{B}_{(i)}^T\mathbf{V}_{[i]}^{-1}\mathbf{V}_i)^T$$

Letting $\widetilde{\mathbf{A}}_i = \mathbf{A}_i - \mathbf{A}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{V}_i$ and $\widetilde{\mathbf{B}}_i = \mathbf{B}_i - \mathbf{B}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{V}_i$,

$$\mathbf{A}^T \mathbf{V}^{-1} \mathbf{B} = \mathbf{A}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{B}_{(i)} + \widetilde{\mathbf{A}}_i \mathbf{S}_i^{-1} \widetilde{\mathbf{B}}_i^T$$

Thus, we have:

$$\mathbf{A}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{B}_{(i)} = \mathbf{A}^T \mathbf{V}^{-1} \mathbf{B} - \widetilde{\mathbf{A}}_i \mathbf{S}_i^{-1} \widetilde{\mathbf{B}}_i^T \tag{3.7}$$

Letting $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{V}_i$ and $\widetilde{\boldsymbol{y}}_i = \boldsymbol{b}_i - \boldsymbol{y}_{(i)} \mathbf{V}_{[i]}^{-1} \mathbf{V}_i$, from 3.7,

$$
\begin{aligned}
\mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)} &= \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} - \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\mathbf{X}}_i^T \\
\mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \boldsymbol{y}_{(i)} &= \mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{y} - \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i
\end{aligned}
\tag{3.8}
$$

Define $\mathbf{H}_{ii} = \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i$, from 3.8 and applying Woodbury's Formula,

$$
\begin{aligned}
(\mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)})^{-1} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} - \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\mathbf{X}}_i^T)^{-1} \\
&= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\
&\quad + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i [\mathbf{S}_i - \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i]^{-1} \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\
&= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}
\end{aligned}
$$

For complete dataset, the estimated fixed effect parameter vector and random effect parameter are

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{y} \\
\hat{\boldsymbol{b}} &= \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1} (\boldsymbol{y} - \mathbf{X} \hat{\boldsymbol{\beta}})
\end{aligned}
$$

Thus, if we remove a subset A from the dataset, the estimated fixed effect parameter vector $\boldsymbol{\beta}$ is:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{(A)} &= (\mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{V}_{[i]}^{-1} \boldsymbol{y}_{(i)} \\
&= \left[(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \right] \\
&\quad \left[\mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{y} - \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \right] \\
&= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{y} - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&\quad + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \boldsymbol{y} \\
&\quad - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&= \hat{\boldsymbol{\beta}} - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&\quad + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}} \\
&\quad - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \mathbf{H}_{ii} \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&= \hat{\boldsymbol{\beta}} + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}} \\
&\quad - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} (\mathbf{S}_i - \mathbf{H}_{ii}) \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&\quad - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \mathbf{H}_{ii} \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \\
&= \hat{\boldsymbol{\beta}} + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}} \\
&\quad - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \left[(\mathbf{S}_i - \mathbf{H}_{ii}) \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i + \mathbf{H}_{ii} \mathbf{S}_i^{-1} \widetilde{\boldsymbol{y}}_i \right] \\
&= \hat{\boldsymbol{\beta}} + (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}} - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} \widetilde{\boldsymbol{y}}_i \\
&= \hat{\boldsymbol{\beta}} - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \widetilde{\mathbf{X}}_i (\mathbf{S}_i - \mathbf{H}_{ii})^{-1} (\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}})
\end{aligned}
$$

And the estimated random effect parameter vector $\boldsymbol{b}$ is:

$$\hat{\boldsymbol{b}}_{(A)} = \mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}(\boldsymbol{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}_{(A)})$$

$$=\mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\left[\boldsymbol{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}} + \mathbf{X}_{(i)}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})\right]$$

$$=\mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\boldsymbol{y}_{(i)} - \mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}$$

$$+ \mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\mathbf{G}(\mathbf{Z}^T\mathbf{V}^{-1}\boldsymbol{y} - \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\boldsymbol{y}}_i) - \mathbf{G}(\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X} - \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\mathbf{X}}_i^T)\hat{\boldsymbol{\beta}}$$

$$+ \mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\boldsymbol{y} - \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{G}\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\boldsymbol{y}}_i + \mathbf{G}\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}}$$

$$+ \mathbf{G}\mathbf{Z}_{(i)}^T \mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\hat{\boldsymbol{b}} - \mathbf{G}\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$+ \mathbf{G}(\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X} - \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\mathbf{X}}_i^T)(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\hat{\boldsymbol{b}} - \mathbf{G}\left[\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1} - (\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X} - \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\widetilde{\mathbf{X}}_i^T)(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}\right](\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\hat{\boldsymbol{b}} - \mathbf{G}\left[\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1} - \mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1} + \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\mathbf{H}_{ii}(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}\right]$$

$$(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\hat{\boldsymbol{b}} - \mathbf{G}\left[\widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}(\mathbf{S}_i - \mathbf{H}_{ii}) - \mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i + \widetilde{\mathbf{Z}}_i\mathbf{S}_i^{-1}\mathbf{H}_{ii}\right]$$

$$(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$=\hat{\boldsymbol{b}} - \mathbf{G}\left[\widetilde{\mathbf{Z}}_i - \mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i\right](\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

It is also essential that $\mathbf{V}_{[i]}^{-1}$ be easily computed. A simpler computational formula for $\mathbf{V}_{[i]}^{-1}$ can be derived. Since $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}$, from 3.5 and 3.6 we have:

$$\begin{bmatrix} \mathbf{V}_{ii} & \mathbf{V}_i^T \\ \mathbf{V}_i & \mathbf{V}_{[i]} \end{bmatrix} \begin{bmatrix} \mathbf{V}^{ii} & \boldsymbol{\Lambda}_i^T \\ \boldsymbol{\Lambda}_i & \boldsymbol{\Lambda}_{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

Then,

$$\mathbf{V}_i\boldsymbol{\Lambda}_i^T + \mathbf{V}_{[i]}\boldsymbol{\Lambda}_{[i]} = \mathbf{I}$$

$$\mathbf{V}_i\mathbf{V}^{ii} + \mathbf{V}_{[i]}\boldsymbol{\Lambda}_i = \mathbf{0}$$

42

From the second of the above two equations,

$$\mathbf{V}_i = -\mathbf{V}_{[i]}\mathbf{\Lambda}_i(\mathbf{V}^{ii})^{-1}$$

Substituting into the first of the two equations,

$$\mathbf{V}_{[i]}\left[\mathbf{\Lambda}_{[i]} - \mathbf{\Lambda}_i(\mathbf{V}^{ii})^{-1}\mathbf{\Lambda}_i^T\right] = \mathbf{I}$$

and

$$\mathbf{V}_{[i]}^{-1} = \mathbf{\Lambda}_{[i]} - \mathbf{\Lambda}_i(\mathbf{V}^{ii})^{-1}\mathbf{\Lambda}_i^T \tag{3.9}$$

In Equation 3.9, we still need to compute the inverse matrix of $\underset{k \times k}{\mathbf{V}^{ii}}$. But, usually the number of removed observations is much less than the total number of observations, that is, $k \ll N$. That means computing $\underset{k \times k}{(\mathbf{V}^{ii})^{-1}}$ is much more efficient than directly computing $\underset{(N-k)\times(N-k)}{\mathbf{V}_{[i]}^{-1}}$.

So, the influence functions for $\hat{\boldsymbol{\beta}}_{(A)}$ and $\hat{\boldsymbol{b}}_{(A)}$ are:

$$\hat{\boldsymbol{\beta}}_{(A)} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i(\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$\hat{\boldsymbol{b}}_{(A)} = \hat{\boldsymbol{b}} - \mathbf{G}\left[\widetilde{\mathbf{Z}}_i - \mathbf{Z}^T\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\widetilde{\mathbf{X}}_i\right](\mathbf{S}_i - \mathbf{H}_{ii})^{-1}(\widetilde{\boldsymbol{y}}_i - \widetilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}})$$

$$\mathbf{V}_{[A]}^{-1} = \mathbf{\Lambda}_{[i]} - \mathbf{\Lambda}_i(\mathbf{V}^{ii})^{-1}\mathbf{\Lambda}_i^T$$

We can estimate the parameters $\boldsymbol{\beta}_{(A)}$ and $\boldsymbol{b}_{(A)}$ after removing subset A without having to re-fit the multivariate mixed effect model. Here $\boldsymbol{b}_{(A)} = [\boldsymbol{b}_{1(A)}^{*T} \ldots \boldsymbol{b}_{N(A)}^{*T}]^T$, and $\boldsymbol{b}_i^* = [\boldsymbol{b}_i^T, \boldsymbol{c}_{i1}^T, \boldsymbol{c}_{i2}^T, \ldots, \boldsymbol{c}_{is}^T]^T$.

3.3 MULTILEVEL MULTIVARIATE LONGITUDINAL EXTENSION OF COOK'S DISTANCE

Similar to the single-level Multivariate conditional Cook's distance we proposed in previous chapter, we here propose a multilevel multivariate longitudinal extension. Conditioning on all of the individuals and each characterisitc of the individuals, we have the following likelihood:

$$L(\boldsymbol{\phi}) = |\mathbf{S}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})^T \mathbf{S}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})\right]$$

and its corresponding log-likelihood:

$$l(\boldsymbol{\phi}) = -\frac{1}{2}\log|\mathbf{S}| - \frac{1}{2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})^T \mathbf{S}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})$$

where $\mathbf{S} = \text{diag}\left[\mathbf{I}_s \otimes (\mathbf{I}_{n_1} \otimes \boldsymbol{\Sigma}), \ldots, \mathbf{I}_s \otimes (\mathbf{I}_{n_N} \otimes \boldsymbol{\Sigma})\right]$.

Here we use $\boldsymbol{\phi}$ to denote the vector containing all the fixed and random effects parameters to be estimated, that is, $\boldsymbol{\phi} = (\boldsymbol{\beta}^T, \boldsymbol{b}^T, \boldsymbol{c}^T)^T$. Using model defined by 3.1, 3.2 and from 2.5, we have the score function:

$$\frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{\beta}} \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{b}} \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \boldsymbol{c}} \end{bmatrix} = \begin{bmatrix} -(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})\boldsymbol{\beta} + (\boldsymbol{y} - \mathbf{Z}\boldsymbol{b} - \mathbf{W}\boldsymbol{c})^T\mathbf{S}^{-1}\mathbf{X} \\ -(\mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z})\boldsymbol{b} + (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\boldsymbol{c})^T\mathbf{S}^{-1}\mathbf{Z} \\ -(\mathbf{W}^T\mathbf{S}^{-1}\mathbf{W})\boldsymbol{b} + (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^T\mathbf{S}^{-1}\mathbf{W} \end{bmatrix}$$

and Fisher Information:

$$\mathbf{B} = I(\boldsymbol{\phi}) = -\frac{\partial^2}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^T}l(\boldsymbol{\phi}) = \begin{bmatrix} \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{W} \\ \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{W} \\ \mathbf{W}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{W}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{W}^T\mathbf{S}^{-1}\mathbf{W} \end{bmatrix}$$

then, the conditional Cook's distance can be written as

$$CD(A) = \frac{(\hat{\phi}_{(A)} - \hat{\phi})^T \mathbf{B}(\hat{\phi}_{(A)} - \hat{\phi})}{c}$$

$$= \frac{1}{c} \begin{bmatrix} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T & (\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T & (\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})^T \end{bmatrix} \begin{bmatrix} \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{X}^T\mathbf{S}^{-1}\mathbf{W} \\ \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{W} \\ \mathbf{W}^T\mathbf{S}^{-1}\mathbf{X} & \mathbf{W}^T\mathbf{S}^{-1}\mathbf{Z} & \mathbf{W}^T\mathbf{S}^{-1}\mathbf{W} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}} \end{bmatrix}$$

$$= \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c} + \frac{(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c} + \frac{(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})^T \mathbf{W}^T\mathbf{S}^{-1}\mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c}$$

$$+ \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c} + \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c} + \frac{2(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \mathbf{Z}^T\mathbf{S}^{-1}\mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c}$$

$$= C_{A1} + C_{A2} + C_{A3} + C_{A4} + C_{A5} + C_{A6}$$

(3.10)

where

$$c = (Nsm - 1)q + p$$

Here, p is the total number of fixed effect predictors, that is, $p = \sum_{k=1}^{m} p_k$; q is the total number of random effect predictors, $q = \sum_{k=1}^{m} q_k$; m is the number of characteristics for each individual; and A indicates a subset of the observations. Other quantities include $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{c}}$ and $\hat{\boldsymbol{\beta}}_{(A)}$, $\hat{\boldsymbol{b}}_{(A)}$, $\hat{\boldsymbol{c}}_{(A)}$ which are the fitted values of $\boldsymbol{\beta}$, $\boldsymbol{b}$, $\boldsymbol{c}$ using the samples with and without the subset of observations, respectively. The subset, A, could be one data point, all observations in one individual, or some other specified subset.

From equation (3.2), we see that $CD(A)$ can be decomposed into six parts: C_{A1}, C_{A2}, C_{A3}, C_{A4}, C_{A5} and C_{A6}, which are as follows:

$$C_{A1} = \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1} \mathbf{X}_i (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \sum_{h=1}^{s} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{X}_i (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= \sum_{i=1}^{N} \frac{s(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{X}_i (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

$$= s \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_{ij} (\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})}{c}$$

45

C_{A1} is the total distance measurement for the fixed (marginal) effect between the complete dataset and the data with subset A removed. The term $\frac{(\hat{\boldsymbol{\beta}}_{(A)}-\hat{\boldsymbol{\beta}})^T\mathbf{X}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_{ij}(\hat{\boldsymbol{\beta}}_{(A)}-\hat{\boldsymbol{\beta}})}{c}$ is the overall marginal Cook's distance for the hth eye of the ith patient at the jth time point. It is the total distance measurement of m characteristics, but only normalizing the residual variance-covariance matrix, without normalizing the random variance-covariance matrices[20]. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2,\ldots,\sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^m \frac{(\hat{\boldsymbol{\beta}}_{(A)}-\hat{\boldsymbol{\beta}})^T\mathbf{X}_{ij}^T\mathbf{X}_{ij}(\hat{\boldsymbol{\beta}}_{(A)}-\hat{\boldsymbol{\beta}})}{c\sigma_k^2}$, which is the simple summation of the distance measurements for all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the fixed (marginal) effect also takes into account the correlations among all the m characteristics.

Continuing with C_{A2}, we have:

$$C_{A2} = \frac{(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T\mathbf{Z}^T\mathbf{S}^{-1}\mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c}$$
$$= \sum_{i=1}^N \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T\mathbf{Z}_i^T[\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1}\mathbf{Z}_i(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$
$$= \sum_{i=1}^N \sum_{h=1}^s \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T\mathbf{Z}_i^T[\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}]\mathbf{Z}_i(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$
$$= \sum_{i=1}^N \frac{s(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T\mathbf{Z}_i^T[\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}]\mathbf{Z}_i(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$
$$= s\sum_{i=1}^N \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T\mathbf{Z}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}_{ij}(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

C_{A2} is the total distance measurement for the first level (individual level) random effect parameters between the complete dataset and the data with subset A removed. The term $\frac{(\hat{\boldsymbol{b}}_{i(A)}-\hat{\boldsymbol{b}}_i)^T\mathbf{Z}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}_{ij}(\hat{\boldsymbol{b}}_{i(A)}-\hat{\boldsymbol{b}}_i)}{c}$ is actually the overall distance measurement of the individual level random effects for both eyes of the ith patient at the jth time point. It is the total distance measurement of m characteristics, normalizing the residual variance-covariance matrix. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2,\ldots,\sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^m \frac{(\hat{\boldsymbol{b}}_{i(A)}-\hat{\boldsymbol{b}}_i)^T\mathbf{Z}_{ij}^T\mathbf{Z}_{ij}(\hat{\boldsymbol{b}}_{i(A)}-\hat{\boldsymbol{b}}_i)}{c\sigma_k^2}$, which is a simple summation of the distance measurements for all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the individual level random effect also takes into account the correlations among all the m characteristics.

C_{A3} is the total distance measurement for the second level (subject level) random effect parameters between the complete dataset and the data with subset A removed.

$$C_{A3} = \frac{(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})^T \mathbf{W}^T \mathbf{S}^{-1} \mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c}$$

$$= \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{c}}_{i(A)} - \hat{\boldsymbol{c}}_i)^T \mathbf{W}_i^T [\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1} \mathbf{W}_i (\hat{\boldsymbol{c}}_{i(A)} - \hat{\boldsymbol{c}}_i)}{c}$$

$$= \sum_{i=1}^{N} \sum_{h=1}^{s} \frac{(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})^T \mathbf{W}_{ih}^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{W}_{ih} (\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$$

$$= \sum_{i=1}^{N} \sum_{h=1}^{s} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})^T \mathbf{W}_{(ih)j}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j} (\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$$

The term $\frac{(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})^T \mathbf{W}_{(ih)j}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j} (\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$ is actually the overall distance measurement of the subject level random effects for the hth subject of ith individual at j^{th} time point. It is the total distance measurement of m characteristics, normalizing the residual variance-covariance matrix. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = $ diag$[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})^T \mathbf{W}_{(ih)j}^T \mathbf{W}_{(ih)j} (\hat{\boldsymbol{b}}_{ih(A)} - \hat{\boldsymbol{b}}_{ih})}{c \sigma_k^2}$, which is the simple summation of the distance measurements for all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the subject level random effect also takes into account the correlations among all the m characteristics.

C_{A4} is the distance measure of covariation between the change in the population average profile and the change in the first level (individual level) subject-specific profile relative to the population average profile.

$$C_{A4} = \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{S}^{-1} \mathbf{Z}(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})}{c}$$

$$= 2 \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1} \mathbf{Z}_i (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

$$= 2 \sum_{i=1}^{N} \frac{s(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{Z}_i (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

$$= 2s \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$$

The term $\frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}_{ij} (\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c}$ is actually the overall distance measurement of the covariation between the change in the population average profile and the change in the

first level (individual level) subject-specific profile relative to the population average profile for both eyes of the ith patient at the jth time point. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \mathbf{Z}_{ij}(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)}{c\sigma_k^2}$, which is a simple summation of the distance measurements for the covariance of all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the covariance also takes into account the correlations among all the m characteristics.

C_{A5} is the distance measure of covariation between the change in the population average profile and the change in the second level (subject level) subject-specific profile relative to the first level subject-specific profile.

$$
\begin{aligned}
C_{A5} &= \frac{2(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{S}^{-1} \mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c} \\
&= 2 \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1} \mathbf{W}_i(\hat{\boldsymbol{c}}_{i(A)} - \hat{\boldsymbol{c}}_i)}{c} \\
&= 2 \sum_{i=1}^{N} \sum_{h=1}^{s} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{W}_{ih}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c} \\
&= 2 \sum_{i=1}^{N} \sum_{h=1}^{s} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}
\end{aligned}
$$

The term $\frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$ is actually the overall distance measurement of the covariation between the change in the population average profile and the change in the second level (subject level) subject-specific profile relative to the first level subject-specific profile for the hth eyes of the ith patient at the jth time point. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{\beta}}_{(A)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}_{ij}^T \mathbf{W}_{(ih)j}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c\sigma_k^2}$, which is a simple summation of the distance measurements for the covariance of all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the covariance also takes into account the correlations among all the m characteristics.

C_{A6} is the distance measure of covariation between the change in the first level (individual level) subject-specific profile relative to the population average profile and the change in the

second level (subject level) subject-specific profile relative to the first level subject-specific profile.

$$C_{A6} = \frac{2(\hat{\boldsymbol{b}}_{(A)} - \hat{\boldsymbol{b}})^T \mathbf{Z}^T \mathbf{S}^{-1} \mathbf{W}(\hat{\boldsymbol{c}}_{(A)} - \hat{\boldsymbol{c}})}{c}$$
$$= 2 \sum_{i=1}^{N} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_i^T [\mathbf{I}_s \otimes (\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma})]^{-1} \mathbf{W}_i (\hat{\boldsymbol{c}}_{i(A)} - \hat{\boldsymbol{c}}_i)}{c}$$
$$= 2 \sum_{i=1}^{N} \sum_{h=1}^{s} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_i^T [\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}^{-1}] \mathbf{W}_{ih} (\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$$
$$= 2 \sum_{i=1}^{N} \sum_{h=1}^{s} \sum_{j=1}^{n_i} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j} (\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$$

The term $\frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{W}_{(ih)j}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c}$ is actually the overall distance measurement of the covariation between the change in the first level (individual level) subject-specific profile relative to the population average profile for both eyes of the ith patient at the jth time point and the change in the second level (subject level) subject-specific profile relative to the first level subject-specific profile for the hth eye of the ith patient at the jth time point. If we assume the residual covariance matrix is diagonal, that is, $\boldsymbol{\Sigma} = \text{diag}\,[\sigma_1^2, \ldots, \sigma_m^2]$, then the term can be reduced to $\sum_{k=1}^{m} \frac{(\hat{\boldsymbol{b}}_{i(A)} - \hat{\boldsymbol{b}}_i)^T \mathbf{Z}_{ij}^T \mathbf{W}_{(ih)j}(\hat{\boldsymbol{c}}_{ih(A)} - \hat{\boldsymbol{c}}_{ih})}{c\sigma_k^2}$, which is a simple summation of the distance measurements for the covariance of all the characteristics. When $\boldsymbol{\Sigma}$ is NOT diagonal, the total distance measurement for the covariance also takes into account the correlations among all the m characteristics.

3.4 APPLICATION

In the previous chapter, we applied the single level multivariate Cook's distance to the glaucoma clinical data. Note that for the single level multivariate Cook's distance, all eyes were assumed to be independent. This assumption is obviously NOT correct because two eyes from one patient are typically correlated. We were just demonstrating the method.

Now we apply the multi-level multivariate Cook's distance to the glaucoma clinical data we have used before. The multi-level multivariate Cook's distance will take into account the correlations between two eyes within each patient.

3.4.1 The Model

We fitted the following two-level bivariate linear mixed effect model:

$$Y_{RNFL} = \left[\beta_{10}N + \beta_{11}GS + \beta_{12}G + b_{10(pt)} + c_{10(eye)} \right]$$
$$+ \left[\beta_{13}N + \beta_{14}GS + \beta_{15}G + b_{11(pt)} + c_{11(eye)} \right] Fu + \beta_{16}\text{Age} + \epsilon_{RNFL}$$
$$Y_{GCC} = \left[\beta_{20}N + \beta_{21}GS + \beta_{22}G + b_{20(pt)} + c_{20(eye)} \right]$$
$$+ \left[\beta_{23}N + \beta_{24}GS + \beta_{25}G + b_{21(pt)} + c_{21(eye)} \right] Fu + \beta_{26}\text{Age} + \epsilon_{GCC}$$

where Age indicates baseline age; Fu indicates Follow-up (in years), Subscripts (pt) and (eye) indicate patient-level and eye-level random effects.

We assume that:

$$\boldsymbol{b} = [b_{10(pt)}, b_{20(pt)}, b_{11(pt)}, b_{21(pt)}]^T \sim N(\mathbf{0}, \mathbf{G}_1),$$
$$\boldsymbol{c} = [c_{10(eye)}, c_{20(eye)}, c_{11(eye)}, c_{21(eye)}]^T \sim N(\mathbf{0}, \mathbf{G}_2),$$
$$\boldsymbol{\epsilon} = [\epsilon_{RNFL}, \epsilon_{GCC}]^T \sim N(\mathbf{0}, \mathbf{\Sigma}_{2\times2})$$

Although, in general, the error terms in this model are correlated due to correlations between different characteristics at the same time point, in specific applications, the correlation effect can be dominated by the combination of random effects. For computational reasons, in this example, we restricted the error terms to be uncorrelated. That is, $\mathbf{\Sigma}_{2\times2}$ is assumed to be diagonal, $\mathbf{\Sigma}_{2\times2} = \begin{bmatrix} \sigma^2_{RNFL} & 0 \\ 0 & \sigma^2_{GCC} \end{bmatrix}$ Table 3.1 shows the estimated parameters of the fixed effects:

Table 3.1: The estimated parameters of fixed effects

Parameters	Estimated value	Standard Deviation	p value
$\hat{\beta}_{10}$	111.7124	3.7328	< 0.0001
$\hat{\beta}_{11}$	104.3342	3.8464	< 0.0001
$\hat{\beta}_{12}$	95.8305	3.9444	< 0.0001
$\hat{\beta}_{13}$	-0.08998	0.1859	0.0049
$\hat{\beta}_{14}$	-0.7453	0.1051	< 0.0001
$\hat{\beta}_{15}$	-0.6221	0.1518	< 0.0001
$\hat{\beta}_{16}$	-0.1811	0.06190	< 0.0001
$\hat{\beta}_{20}$	107.7753	2.7568	< 0.0001
$\hat{\beta}_{21}$	101.9587	2.8392	< 0.0001
$\hat{\beta}_{22}$	95.8144	2.9178	< 0.0001
$\hat{\beta}_{23}$	-0.9624	0.1506	< 0.0001
$\hat{\beta}_{24}$	-0.5836	0.08332	< 0.0001
$\hat{\beta}_{25}$	-0.4858	0.1267	< 0.0001
$\hat{\beta}_{26}$	-0.2180	0.04572	< 0.0001

The estimated variance-covriance matrix of the patient level random effect ($\mathbf{G}_1$) and its correlation matrix are:

$$
\hat{\mathbf{G}}_1 = \begin{array}{c} \\ b_{10(pt)} \\ b_{20(pt)} \\ b_{11(pt)} \\ b_{21(pt)} \end{array}
\begin{array}{c} \begin{array}{cccc} b_{10(pt)} & b_{20(pt)} & b_{11(pt)} & b_{21(pt)} \end{array} \\
\left(\begin{array}{cccc}
73.629 & 39.668 & -0.664 & 0.0729 \\
39.668 & 38.489 & 0.308 & -0.0802 \\
-0.664 & 0.308 & 0.931 & 0.0239 \\
0.0729 & -0.0802 & 0.0239 & 0.356
\end{array} \right) \end{array}
$$

$$
\hat{\mathcal{G}}_1 = \begin{array}{c} \\ b_{10(pt)} \\ b_{20(pt)} \\ b_{11(pt)} \\ b_{21(pt)} \end{array}
\begin{array}{c} \begin{array}{cccc} b_{10(pt)} & b_{20(pt)} & b_{11(pt)} & b_{21(pt)} \end{array} \\
\left(\begin{array}{cccc}
1.000 & 0.745 & -0.0801 & 0.0142 \\
0.745 & 1.000 & 0.0515 & -0.0217 \\
-0.0801 & 0.0515 & 1.000 & 0.0416 \\
0.0142 & -0.0217 & 0.0416 & 1.000
\end{array} \right) \end{array}
$$

The estimated variance-covriance matrices of the eye level random effect $(\mathbf{G}_2)$ and its correlation matrix are:

$$
\hat{\mathbf{G}}_2 = \begin{array}{c} \\ c_{10(eye)} \\ c_{20(eye)} \\ c_{11(eye)} \\ c_{21(eye)} \end{array}
\begin{array}{cccc} c_{10(pt)} & c_{20(pt)} & c_{11(pt)} & c_{21(pt)} \\ \end{array}
\left(\begin{array}{cccc}
24.713 & 12.338 & -1.060 & -0.233 \\
12.338 & 12.977 & -0.487 & 0.117 \\
-1.060 & -0.487 & 0.233 & 0.131 \\
-0.233 & 0.117 & 0.131 & 0.184
\end{array} \right)
$$

$$
\hat{\mathcal{G}}_2 = \begin{array}{c} \\ c_{10(eye)} \\ c_{20(eye)} \\ c_{11(eye)} \\ c_{21(eye)} \end{array}
\begin{array}{cccc} c_{10(pt)} & c_{20(pt)} & c_{11(pt)} & c_{21(pt)} \\ \end{array}
\left(\begin{array}{cccc}
1.000 & 0.689 & -0.442 & -0.109 \\
0.689 & 1.000 & -0.280 & 0.0760 \\
-0.442 & -0.280 & 1.000 & 0.635 \\
-0.109 & 0.0760 & 0.635 & 1.000
\end{array} \right)
$$

The estimated Residual Variance-Covariance Matrix is:

$$
\hat{\mathbf{\Sigma}} = \begin{array}{c} \\ Y_{RNFL} \\ Y_{GCC} \end{array}
\begin{array}{cc} Y_{RNFL} & Y_{GCC} \\ \end{array}
\left(\begin{array}{cc}
13.8980 & 0 \\
0 & 8.4852
\end{array} \right)
$$

Table 3.2 shows the estimated residual variances.

Table 3.2: The estimated residual variances

Parameters	Estimated value	Standard Deviation	p value
$\hat{\sigma}^2_{RNFL}$	13.8980	0.4324	$< .0001$
$\hat{\sigma}^2_{GCC}$	8.4852	0.2625	$< .0001$

3.4.2   The influential observations

Using our method, we calculated the conditional Cook's distance and the decomposed terms C_{A1}, C_{A2}, C_{A3}, C_{A4}, C_{A5} and C_{A6} for each observation (paired components). We picked out ten observations with the largest conditional Cook's distances in ten eyes. Figure 3.1 illustrates the ten observations in the ten eyes and Table 3.3 lists the ten observations with the largest values of the conditional Cook's distance. Note that the blue circles and lines indicate the observed RNFL values and individual fitted regression lines for RNFL. The blue dotted lines indicate the patient-level fitted regression lines for RNFL. The light blue lines indicate the marginal fitted regression lines for RNFL. Similarly, the red cirles, lines and pink lines are for GCC as well. The grey rectangles indicate influential observations (pairs of components).
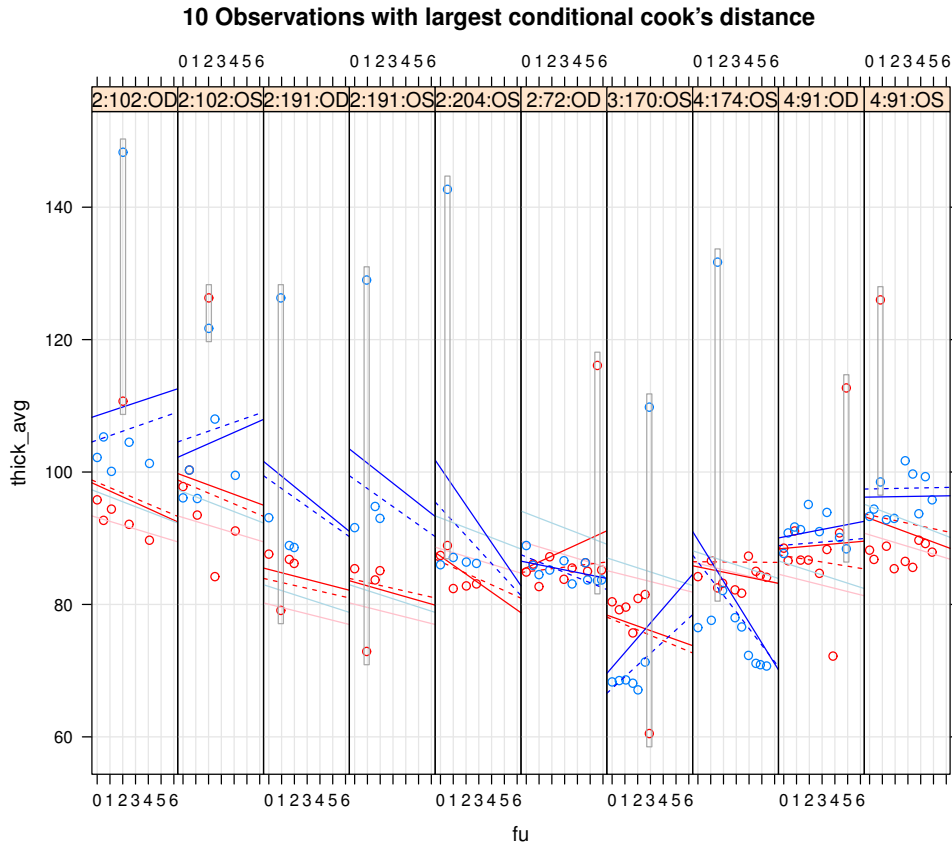

Figure 3.1: 10 observations with largest conditional cook's distance in 10 eyes.

Table 3.3: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance.

| Eye ID	Follow-up (in years)	Follow-up (in days)	Diag grp	Cond Cooksd	C_{A1}	C_{A2}	C_{A3}	C_{A4}	C_{A5}	C_{A6}
2:204:OS	0.5448	199	2	0.02972	0.000163	0.01113	0.01371	-0.000059	0.0000151	0.00477
4:174:OS	1.5387	562	3	0.02522	0.000167	0.01231	0.00881	0.000127	0.0000128	0.00379
3:170:OS	2.9103	1063	3	0.02070	0.000606	0.02060	0.01008	-0.001201	0.0000059	-0.00939
2:102:OD	2.0205	738	2	0.01519	0.000066	0.00739	0.00741	-0.000042	-0.0000018	0.00037
2:191:OS	0.9391	343	3	0.01261	0.000093	0.00492	0.00593	0.000017	-0.0000100	0.00165
4:174:OS	0.0000	0	3	0.01230	0.000073	0.00463	0.00694	-0.000031	0.0000068	0.00068
2:191:OD	0.9391	343	3	0.01206	0.000114	0.00492	0.00586	-0.000014	-0.0000045	0.00118
4:91:OS	0.8487	310	2	0.01068	0.000208	0.00705	0.00381	-0.000211	0.0000073	-0.00018
2:204:OS	0.0000	0	2	0.00770	0.000052	0.00235	0.00486	-0.000055	0.0000135	0.00049
2:102:OS	2.0205	738	2	0.00638	0.000118	0.00224	0.00310	-0.000066	-0.0000077	0.00099
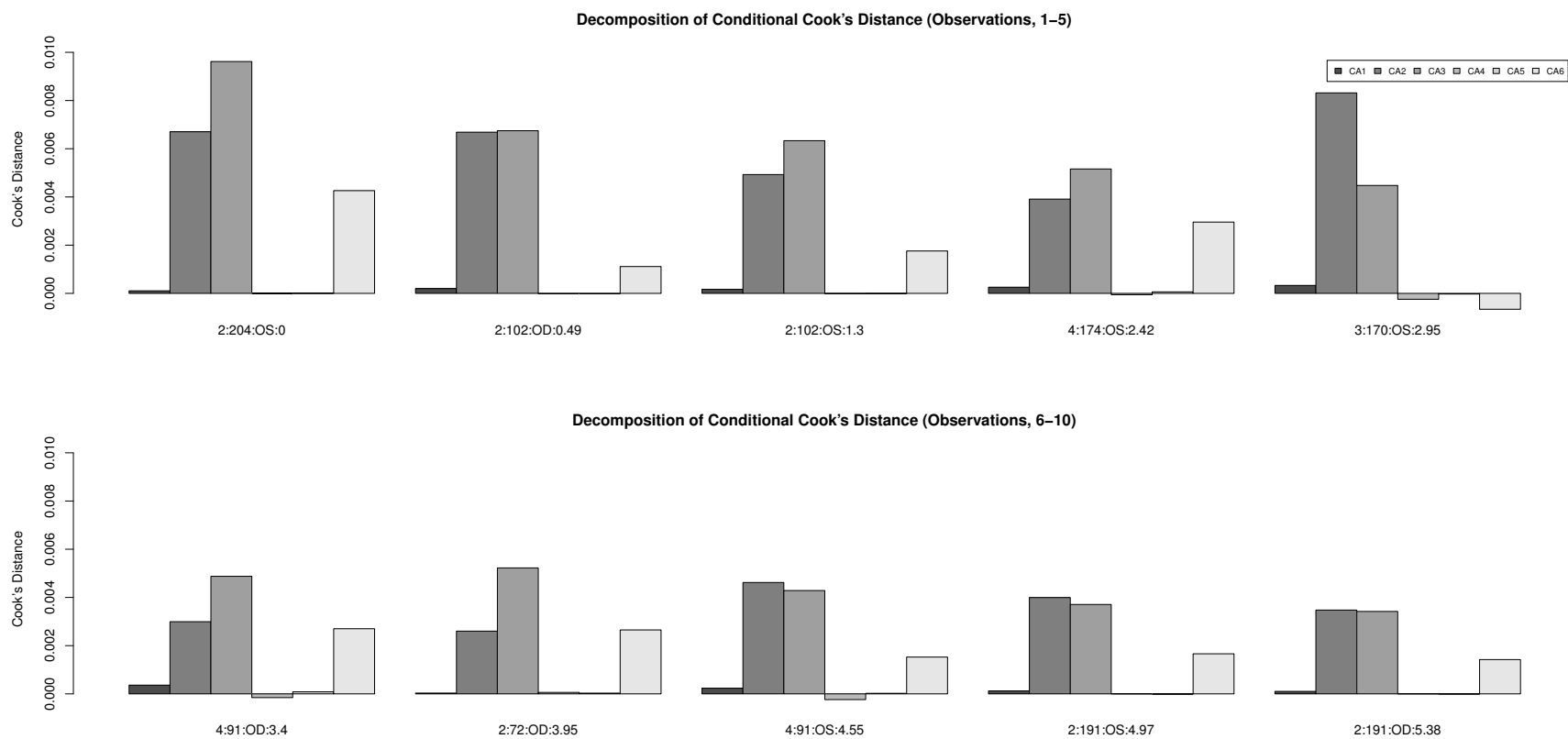
Figure 3.2: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance.

3.4.3 The influential components

Using our method, we calculated the conditional Cook's distance and the decomposed parts C_{A1}, C_{A2}, C_{A3}, C_{A4}, C_{A5} and C_{A6} for each component. We also picked out ten components with the largest conditional Cook's distances in eight eyes. Figure 3.3 illustrates the ten components in the eight eyes and Table 3.4 lists the ten observations with the largest values of the conditional Cook's distance.

Similar to Figure 3.2, in Figure 3.3, blue indicates RNFL, while red indicates GCC. Circles indicates the components. Solid lines indicate the eye-level fitted regression lines, while dotted lines indicate the patient-level fitted regression lines. Light solid lines indicate the marginal fitted regression lines.
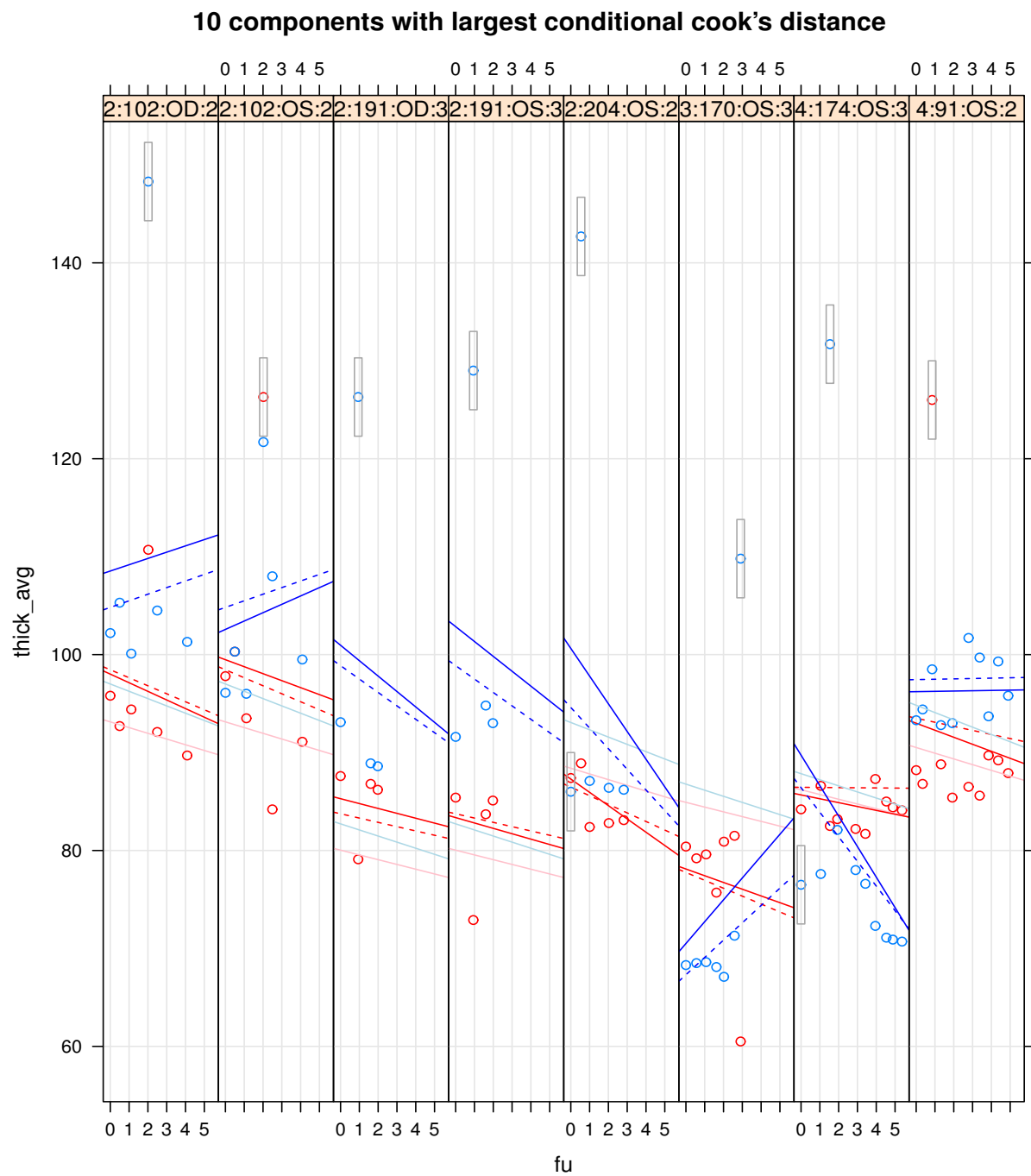
Figure 3.3: 10 components with largest conditional cook's distance in 8 eyes.

Table 3.4: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance.

| Eye ID	Outcome Type	Follow-up (in years)	Follow-up (in days)	Diag grp	Cond Cooksd	C_{A1}	C_{A2}	C_{A3}	C_{A4}	C_{A5}	C_{A6}
2:204:OS	RNFL	0.5448	199	2	0.02972	0.000163	0.01113	0.01371	-0.000059	0.0000151	0.00477
4:174:OS	RNFL	1.5387	562	3	0.02522	0.000167	0.01231	0.00881	0.000127	0.0000128	0.00379
3:170:OS	RNFL	2.9103	1063	3	0.02070	0.000606	0.02060	0.01008	-0.001201	0.0000059	-0.00939
2:102:OD	RNFL	2.0205	738	2	0.01519	0.000066	0.00739	0.00741	-0.000042	-0.0000018	0.00037
2:191:OS	RNFL	0.9391	343	3	0.01261	0.000093	0.00492	0.00593	0.000017	-0.0000100	0.00165
4:174:OS	RNFL	0.0000	0	3	0.01230	0.000073	0.00463	0.00694	-0.000031	0.0000068	0.00068
2:191:OD	RNFL	0.9391	343	3	0.01206	0.000114	0.00492	0.00586	-0.000014	-0.0000045	0.00118
4:91:OS	GCC	0.8487	310	2	0.01068	0.000208	0.00705	0.00381	-0.000211	0.0000073	-0.00018
2:204:OS	RNFL	0.0000	0	2	0.00770	0.000052	0.00235	0.00486	-0.000055	0.0000135	0.00049
2:102:OS	GCC	2.0205	738	2	0.00638	0.000118	0.00224	0.00310	-0.000066	-0.0000077	0.00099
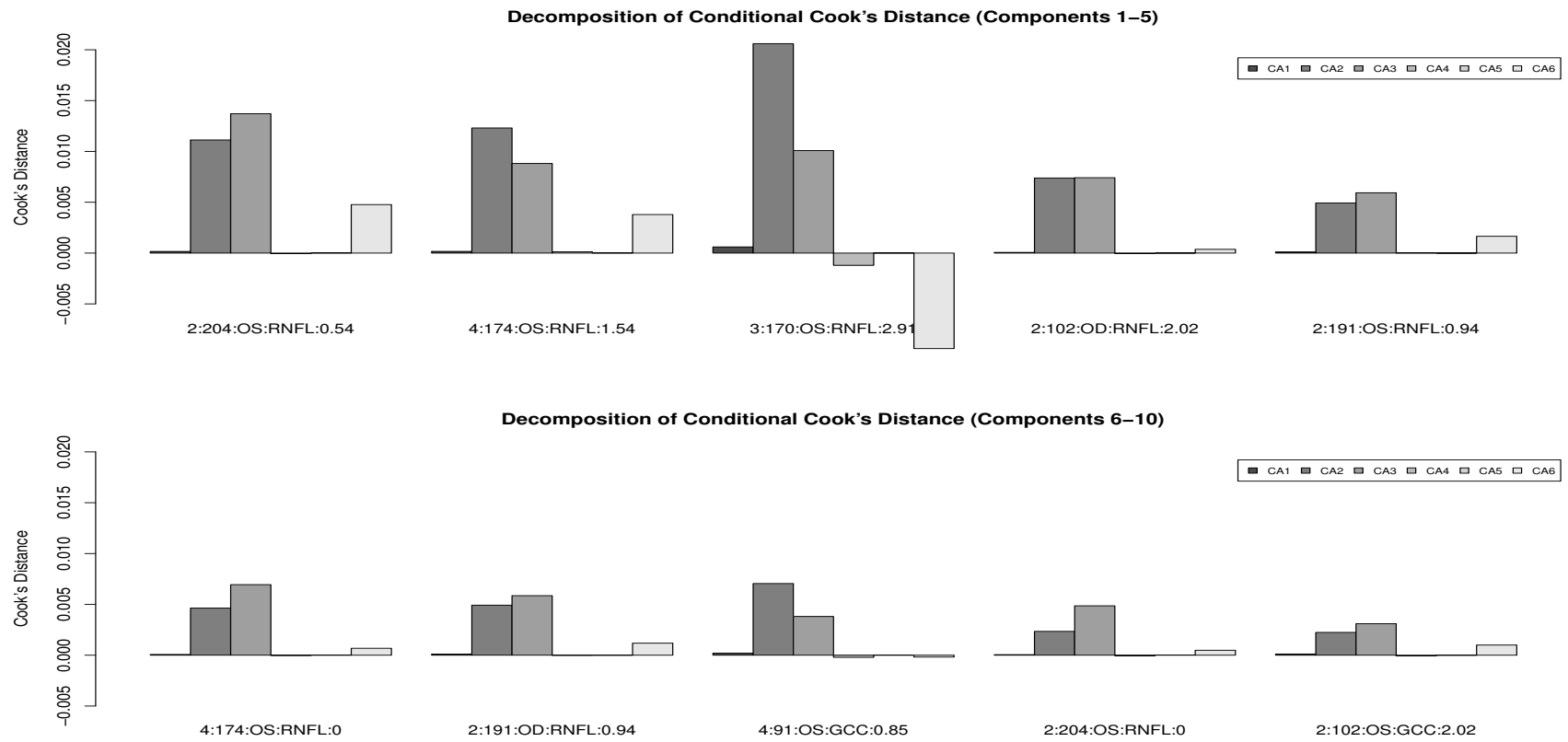
Figure 3.4: Decomposition of the Conditional Cook's Distance for the 10 observations with largest conditional cook's distance

3.4.4 Decomposition of Conditional Cook's Distance

Figure 3.4 shows the values of the six decomposition terms of the multivariate conditional Cook's distance. We notice that for most observations, the distance measurements of the random effects are much greater than the distance measurement of the fixed effects, that is, $C_{A2}, C_{A3} \gg C_{A1}$. As was pointed out earlier, this is obvious because the subject-specific effects are much more sensitive to the influential observation than the marginal effects. It is also noticable that for some influential observations, such as the three most influential observations, the covariance between the distance measurements of first and second level random effects, that is, C_{A6}, is relatively large. Also, the covariance C_{A6} can be either positive (the two most influential observations) or negative (the third most influential observation). We will further explain the six decomposed effects in the conditional Cook's distance by considering two influential observations.

3.4.4.1 Influential Component Example 1 Consider the third most influential observation in patient number 3:170 (See Figure 3.3), left eye (OS), RNFL (1st characteristic), at 2.91 years of follow-up, in diagnostic group 3 (G). There is a large negative value of C_{A6}, which indicates a large negative covariance between the distance measurements of first and second level random effects.

Table 3.5 shows the changes of estimated parameter values after removing the influential observation patient number 3:170, left eye (OS), RNFL (1st characteristic), at 2.91 years of follow-up, including changes of fixed intercepts and slopes, first level subject-specific intercepts and slopes, second level subject-specific intercepts and slopes, for both characteristics (RNFL and GCC). Figure 3.5 shows the 3:170:OS and 3:170:OD observations and predicted trend lines based on complete dataset and after the influential observation was removed.

Table 3.5: Parameter estimation based on complete dataset and removal of influential observation in 3:170:OS.

|	Parameters	Est based on complete data	Est w influential point removed	Change
F.E.	$\hat{\beta}_{12}$	95.8305	95.6579	-0.1726
(pt ID 3:170)	$\hat{\beta}_{15}$	-0.6221	-0.6630	-0.04090
	$\hat{\beta}_{16}$	-0.1811	-0.1761	0.004960
	$\hat{\beta}_{22}$	95.8144	95.7490	-0.06541
	$\hat{\beta}_{25}$	-0.4858	-0.4872	-0.001351
	$\hat{\beta}_{26}$	-0.2180	-0.2164	0.001561
1st level R.E.	$\hat{b}_{10(pt)}$	-19.4635	-19.5020	-0.03848
(pt ID 3:170)	$\hat{b}_{11(pt)}$	2.3902	0.1599	-2.2302
	$\hat{b}_{20(pt)}$	-7.1572	-7.6951	-0.5379
	$\hat{b}_{21(pt)}$	-0.3212	-0.3086	0.01264
2nd level R.E.	$\hat{c}_{10(eye)}$	3.1883	2.1130	-1.0753
(OS, left eye)	$\hat{c}_{11(eye)}$	0.4625	-0.2265	-0.6890
	$\hat{c}_{20(eye)}$	0.3516	1.0986	0.7471
	$\hat{c}_{21(eye)}$	0.1193	-0.2272	-0.3465
2nd level R.E.	$\hat{c}_{10(eye)}$	-11.5232	-8.8544	2.6688
(OD, right eye)	$\hat{c}_{11(eye)}$	0.2208	0.4803	0.2595
	$\hat{c}_{20(eye)}$	-3.9869	-3.4371	0.5498
	$\hat{c}_{21(eye)}$	0.08816	0.1873	0.09913
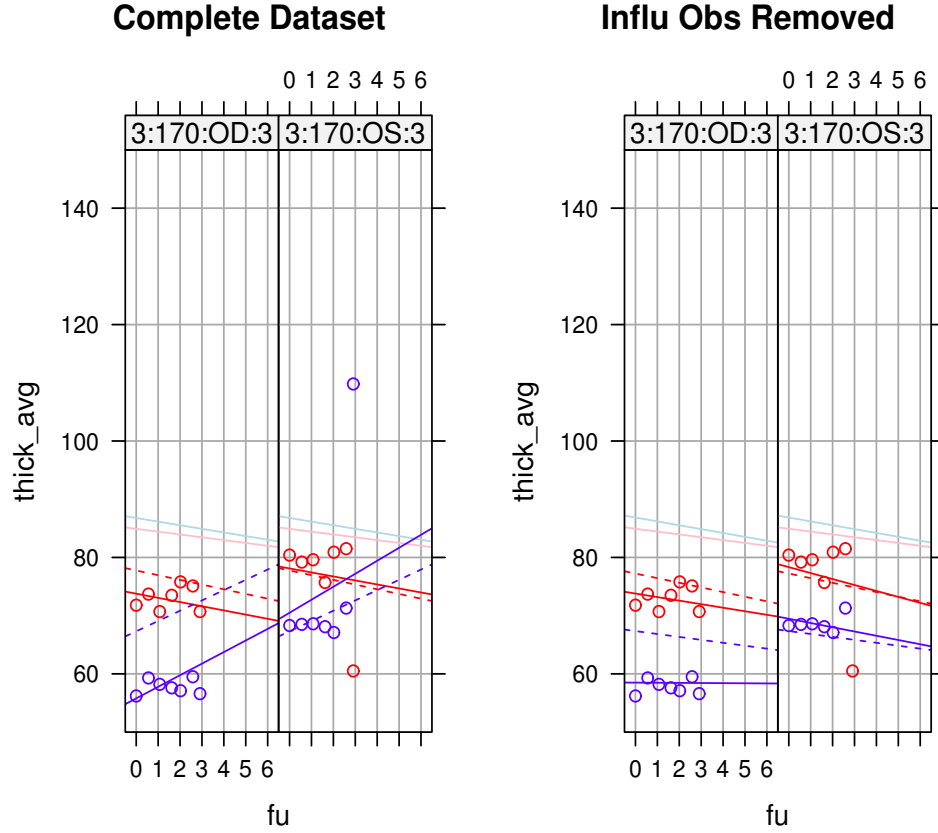
Figure 3.5: 3:170:OS and 3:170:OD

The marginal effects (fixed effect parameters) only have very slight changes. This is consistent with the decomposition barplot shown in Figure 3.4.

Consider the patient level (1st level) subject-specific effects. The intercept of RNFL (1st characteristic) slightly decreased by 0.03848, but the slope had a large decrease of 2.230218. Since the patient level (1st level) subject-specific effects are the offset between the marginal effects and the patient's 1st level actual intercept and slope values, and the marginal effects almost had no change, the 1st level subject-specific slope and intercept changes are close to the changes of the 1st level predicted slope and intercept.

Next, we look at the eye level (2nd level) subject-specific effects. For OS (the left eye), the eye level random intercept of RNFL (1st characteristic) decreased by 1.07529, and the slope decreased by 0.689004. Note that the eye level (2nd level) subject-specific effects are

the offset between the 1st level subject-specific effects and each eye's predicted intercepts and slopes values. Hence, they are quite small compare to the actual changes. For OD (the right eye), we notice that the intercept of RNFL (1st characteristic) increased by 2.66884, which is large compared to other estimated parameter value changes.

We now calcuate the 1st and 2nd level random effects. In this example, the eyes were measured at follow-up times (in years) of: 0, 0.5585216, 1.073238, 1.620808, 2.023272, 2.584531, 2.910335. Thus, the matrices $\mathbf{Z}$, $\mathbf{W}$ and $\boldsymbol{\Sigma}^{-1}$ are:

$$
\mathbf{Z} = \mathbf{W} = \begin{bmatrix}
1 & 0.000000 & 0 & 0 \\
0 & 0 & 1 & 0.000000 \\
1 & 0.558522 & 0 & 0 \\
0 & 0 & 1 & 0.558522 \\
1 & 1.073238 & 0 & 0 \\
0 & 0 & 1 & 1.073238 \\
1 & 1.620808 & 0 & 0 \\
0 & 0 & 1 & 1.620808 \\
1 & 2.023272 & 0 & 0 \\
0 & 0 & 1 & 2.023272 \\
1 & 2.584531 & 0 & 0 \\
0 & 0 & 1 & 2.584531 \\
1 & 2.910335 & 0 & 0 \\
0 & 0 & 1 & 2.910335
\end{bmatrix}
\qquad
\boldsymbol{\Sigma}^{-1} = \begin{bmatrix}
0.072264788 & 0 \\
0 & 0.118237494
\end{bmatrix}
$$

Calculating the influence of the 1st level random effects, and noting that $(\boldsymbol{b}_{(i)} - \boldsymbol{b}) = (-0.03848, -2.230218, -0.537915, 0.012644)^T$, we have:

$$
f_1 = (\boldsymbol{b}_{(i)} - \boldsymbol{b})^T \, \mathbf{Z}^T \, [\mathbf{I}_7 \otimes \boldsymbol{\Sigma}^{-1}] \, \mathbf{Z} \, (\boldsymbol{b}_{(i)} - \boldsymbol{b}) = 8.744124
$$

Then we calculate the influence of the 2nd level random effects. For OS (left eye), we know that $(\boldsymbol{c}_{(i)} - \boldsymbol{c}) = (-1.07529, -0.689004, 0.747056, -0.346515)^T$, then we have:

$$
f_{2,\,OS} = (\boldsymbol{c}_{(i)} - \boldsymbol{c})^T \, \mathbf{W}^T \, [\mathbf{I}_7 \otimes \boldsymbol{\Sigma}^{-1}] \, \mathbf{W} \, (\boldsymbol{c}_{(i)} - \boldsymbol{c}) = 2.672573
$$

For OD (right eye), we know that $(\boldsymbol{c}_{(i)} - \boldsymbol{c}) = (2.66884, 0.259482, 0.549785, 0.0991345)^T$, and so we have:

$$
f_{2,\,OD} = (\boldsymbol{c}_{(i)} - \boldsymbol{c})^T \, \mathbf{W}^T \, [\mathbf{I}_7 \otimes \boldsymbol{\Sigma}^{-1}] \, \mathbf{W} \, (\boldsymbol{c}_{(i)} - \boldsymbol{c}) = 5.210717
$$

Thus, the total influence of the 2nd level random effects is:

$$
f_2 = f_{2,\,OS} + f_{2,\,OD} = 2.672573 + 5.210717 = 7.88329
$$

The ratio of the influence of the 1st level random effect and the 2nd level random effect is:

$$f_1/f_2 = 8.744124/7.88329 = 1.109197$$

Consequently, both the 1st level and 2nd level random effects had a strong effect due to the removal of the influential observation, and the ratio of the two effects is 1.11, which means, they are quite close. Thus, C_{A2} and C_{A3} should be close to each other. But in Figure 3.4, the barplot shows C_{A2} (0.02060) is almost double the value of C_{A3} (0.01008), which is not reasonable, because the influence on 1st level random effects is not that high.

We also notice that C_{A6} is a negative value (-0.00939). If we add up C_{A2} and C_{A6}, we get 0.01121, which is close to C_{A3} (0.01008). Moreover, the ratio of C_{A2} plus C_{A6} and C_{A3} is $0.01121/0.01008 = 1.112103$, which is roughly equal to the ratio of the influence of the 1st level random effect and the 2nd level random effect.

In summary, the C_{A6} seems to compensate for the overestimation of the patient level effect (C_{A2}). Hence, the 1st level (patient level) subject-specific effects is actually a combination of the effects of the two 2nd level (eye level) subject-specific effects.

3.4.4.2 Influential Component Example 2 We now consider the most influential observation, patient number 2:204, left eye (OS), RNFL (1st characteristic), at 0.54 years of follow-up, in diagnostic group 2 (GS). We observe a moderately large positive value of C_{A6}, which indicates a large positive covariance between the distance measurements of the first and second level random effects.

Table 3.6 shows the changes of estimated parameter values after we removed the most influential observation, including changes of fixed intercepts and slopes, first level subject-specific intercepts and slopes, second level subject-specific intercepts and slopes, for both characteristics (RNFL and GCC). Figure 3.6 shows the 2:204:OS and 2:204:OD observations and predicted trend lines based on the complete dataset and after the influential observation was removed.

Table 3.6: Parameter estimation based on complete dataset and removal of influential observation in 2:204:OS.

|	Parameters	Est based on complete data	Est influential point removed	Change
F.E. (pt ID 2:204)	$\hat{\beta}_{11}$	104.3342	104.3013	-0.03294
	$\hat{\beta}_{14}$	-0.7453	-0.7243	0.02100
	$\hat{\beta}_{16}$	-0.1811	-0.1816	-0.00047
	$\hat{\beta}_{21}$	101.9587	102.0001	0.04140
	$\hat{\beta}_{24}$	-0.5836	-0.5842	-0.00058
	$\hat{\beta}_{26}$	-0.2180	-0.2187	-0.00069
1st level R.E. (pt ID 2:204)	$\hat{b}_{10(pt)}$	1.5315	-5.6106	-7.1421
	$\hat{b}_{11(pt)}$	-1.3594	0.3528	1.7122
	$\hat{b}_{20(pt)}$	-1.9316	-1.8514	0.08019
	$\hat{b}_{21(pt)}$	-0.2932	-0.2805	0.01266
2nd level R.E. (OS, left eye)	$\hat{c}_{10(eye)}$	6.0322	-1.2572	-7.2895
	$\hat{c}_{11(eye)}$	-0.7231	0.009808	0.7329
	$\hat{c}_{20(eye)}$	0.8239	-0.5073	-1.3312
	$\hat{c}_{21(eye)}$	-0.4799	-0.06592	0.4140
2nd level R.E. (OD, right eye)	$\hat{c}_{10(eye)}$	-4.0796	-0.2700	3.8095
	$\hat{c}_{11(eye)}$	0.2872	0.02951	-0.2577
	$\hat{c}_{20(eye)}$	-0.8624	-0.3181	0.5443
	$\hat{c}_{21(eye)}$	0.1303	-0.04079	-0.1711
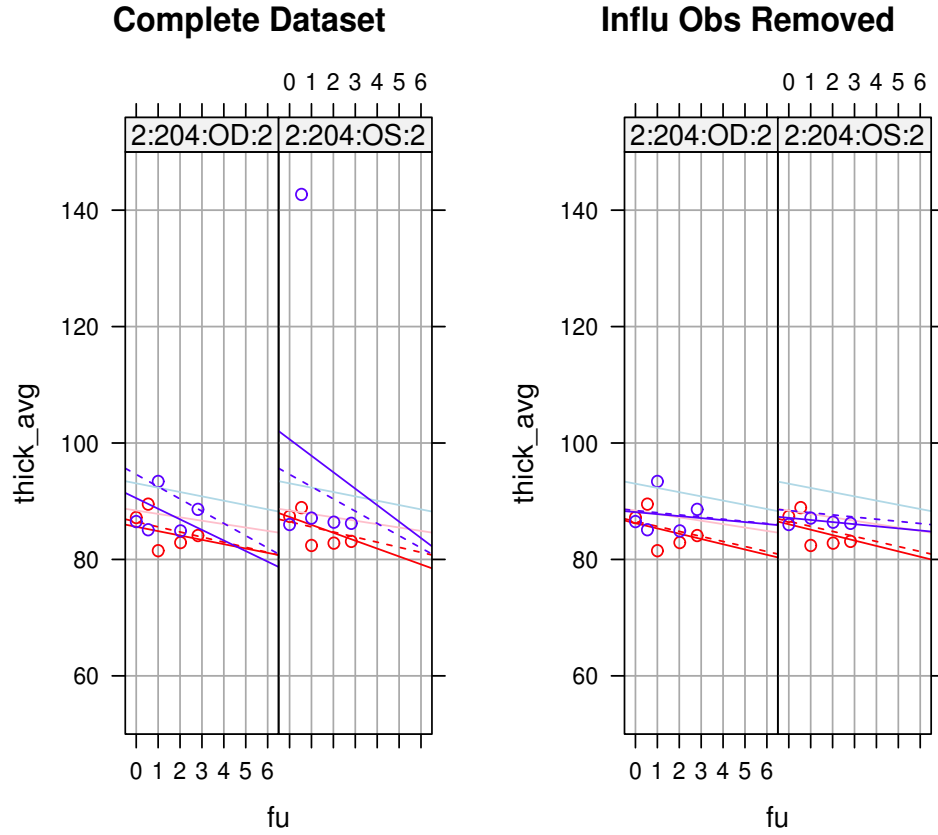
Figure 3.6: 2:204:OS and 2:204:OD.

Like the first example, the marginal effects (fixed effect parameters) only have very slight changes. This is consistent with the decomposition barplot shown in Figure 3.4.

Examining the patient level (1st level) subject-specific effects, we see that the intercept of RNFL (1st characteristic) had a large decrease of 6.142142, and the slope had a large increase of 1.712221 (the trend line became much "flatter"). Since the patient level (1st level) subject-specific effects are the offset between the marginal effects and the patient's 1st level actual intercepts and slopes values, and the marginal effects almost had no change, the 1st level subject-specific slope and intercept changes are close to the changes of the 1st level actual slopes and intercepts.

Next, considering the eye level (2nd level) subject-specific effects, we see that for OS (the left eye), the eye level random intercept of RNFL (1st characteristic) decreased by 7.289487,

and the eye level random slope increased by 0.7328908. Note that the eye level (2nd level) subject-specific effects are the offset between the 1st level subject-specific effects and each eye's predicted intercepts and slopes values. From the plot we can see before we remove the influential observation, the OS intercept is much larger than the 1st level intercept. After removal, the OS intercept is slightly less than the 1st level intercept. So, it makes sense that the 2nd level random intercept decreased a lot. Also, before removal, the OS trend line is much "steeper" than the 1st level trend line, which means the slope of the OS trend line is much less ("more negative") than the 1st level trend line. After removal, the two trend lines are almost parallel. Hence, it makes sense that the 2nd level random slope increased substantially. For OD (the right eye), the changes in the 2nd level subject-specific effects were similiar to OS, but slightly smaller in magnitude. That follows because the two eyes are highly correlated.

For both characteristics, it is noticable that the 1st level random effects (both intercepts and slopes) had much smaller changes than the 2nd level random effects. After removal of the influential observation, the 2nd level random effects became much smaller and the trend lines were closer to the 1st level trend lines.

We now calcuate the 1st and 2nd level random effects. In this example, the eye was measured at follow-up time (in years) are: 0, 0.5448323, 1.0020534, 2.0232717, 2.8172485. Thus, the matrices $\mathbf{Z}$, $\mathbf{W}$ and $\mathbf{\Sigma}^{-1}$ are:

$$
\mathbf{Z} = \mathbf{W} = \begin{bmatrix}
1 & 0.000000 & 0 & 0 \\
0 & 0 & 1 & 0.000000 \\
1 & 0.5448323 & 0 & 0 \\
0 & 0 & 1 & 0.5448323 \\
1 & 1.0020534 & 0 & 0 \\
0 & 0 & 1 & 1.0020534 \\
1 & 2.0232717 & 0 & 0 \\
0 & 0 & 1 & 2.0232717 \\
1 & 2.8172485 & 0 & 0 \\
0 & 0 & 1 & 2.8172485
\end{bmatrix}
\quad
\mathbf{\Sigma}^{-1} = \begin{bmatrix}
0.072264788 & 0 \\
0 & 0.118237494
\end{bmatrix}
$$

Again, calculating the influence of the 1st level random effects, and nothing that $(\boldsymbol{b}_{(i)} - \boldsymbol{b}) = (-7.142142, 1.712221, 0.080185, 0.0126628)^T$, then we have:

$$
f_1 = (\boldsymbol{b}_{(i)} - \boldsymbol{b})^T \, \mathbf{Z}^T \, [\mathbf{I}_7 \otimes \mathbf{\Sigma}^{-1}] \, \mathbf{Z} \, (\boldsymbol{b}_{(i)} - \boldsymbol{b}) = 9.971814
$$

As before, we next calculate the influence of the 2nd level random effects. For OS (left eye), and so we know that $(\boldsymbol{c}_{(i)} - \boldsymbol{c}) = (-7.289487, 0.7328908, -1.331182, 0.4139648)^T$, then we have

$$f_{2,\ OS} = (\boldsymbol{c}_{(i)} - \boldsymbol{c})^T \mathbf{W}^T \left[\mathbf{I}_7 \otimes \boldsymbol{\Sigma}^{-1}\right] \mathbf{W} (\boldsymbol{c}_{(i)} - \boldsymbol{c}) = 15.27044$$

For OD (right eye), we know that $(\boldsymbol{c}_{(i)} - \boldsymbol{c}) = (3.809533, -0.2577002, 0.54433, -0.17111)^T$, then we have

$$f_{2,\ OD} = (\boldsymbol{c}_{(i)} - \boldsymbol{c})^T \mathbf{W}^T \left[\mathbf{I}_7 \otimes \boldsymbol{\Sigma}^{-1}\right] \mathbf{W} (\boldsymbol{c}_{(i)} - \boldsymbol{c}) = 4.482049$$

The total influence of the 2nd level random effects is:

$$f_2 = f_{2,\ OS} + f_{2,\ OD} = 15.27044 + 4.482049 = 19.75249$$

The ratio of the influence of the 1st level random effect and the 2nd level random effect is:

$$f_1/f_2 = 9.971814/19.75249 = 0.5048383$$

Thus, in our example, the 2nd level random effects exhibited a much stronger influence due to the removal of the influential observation than the 1st level random effects. That is, $C_{A3} \gg C_{A2}$. But in Figure 3.4, the barplot shows C_{A3} (0.01371) is only slightly higher than C_{A2} (0.01113).

We note that C_{A6} is a positive value (0.00477). C_{A3} and C_{A6} sum up to 0.01848, which is much higher than C_{A2} (0.01113). And the ratio of C_{A2} and C_{A3} plus C_{A6} is $0.01113/0.01848 = 0.6036139$, which is roughly close to the ratio of the influence of the 1st level random effect and the 2nd level random effect. That follows from the effect of the covariance between the distance measurements of 1st and 2nd level of random effects, C_{A6}. The reason is, the 1st level (patient level) subject-specific effects is actually a "marginal" effect of the two 2nd level (eye level) subject-specific effects.

4.0   CONCLUSIONS AND FUTURE WORK

4.1   CONCLUSIONS

In the first part of this dissertation, we derived the multivariate conditional Cook's distance, which takes into account the three kinds of correlations in multivariate longitudinal data. Our results show that the extended multivariate conditional Cook's distance is superior to the unconditional Cook's distance in a mixed effect multivariate longitudinal data analysis. In 1000 simulations, our method successfully detected the influential vector component 925 times but the unconditional Cook's distance only detected that component for 262 times. We successfully extended Tan's conditional Cook's distance [20] to a multivariate context, where two characteristics are measured at each time point, and also extended Tan's conditional Cook's distance [20] from one component of observations to a subset of observations.

In the second part of this dissertation, we derived a multilevel extension of the multivariate conditional Cook's distance, which takes into account the correlations between the subjects within a cluster, in addition to the three kinds of correlations in single level multivariate longitudinal data. We use a two level random effect model to handle the subject level and cluster level correlations among different time points, and a residual covariance matrix to handle correlations among different responses. We also explored the six parts of our two-level multivariate conditional Cook's distance, and found that the covariation between the subject level and cluster level random effects has a relatively large impact on the Cook's distance measurement. We demonstrated our method in a real data example from a glaucoma study. A multilevel multivariate extension of the influence function was derived. A set of SAS and R programs were developed to implement this method.

4.2 FUTURE WORK

In this dissertation, we have computed the Cook's distance for multiple characteristics of each observation vector. Figure 4.1 shows the 20 eye-characteristics (all RNFL) with largest conditional cook's distances, and Table 4.1 shows the decomposition of the 20 Cook's distances. Figure 4.1 and Table 4.1 clearly show that the 20 eye-characteristics have the largest numbers of observations, but do not necessarily contain influential observations. Consequently, deleting subsets with different numbers of observations introduces different degrees of perturbation to the current model fitted to the data and the magnitude of Cook's distance associated with the degree of the perturbation. Thus, the multivariate conditional Cook's distance for subsets with different degrees of perturbation are not directly comparable. Accordingly, Zhu and Ibrahim [37] have developed a scaled version of Cook's distance to address this fundamental issue. Accordingly, our further research will consider extension to a scaled version of our present multivariate conditional Cook's distance.
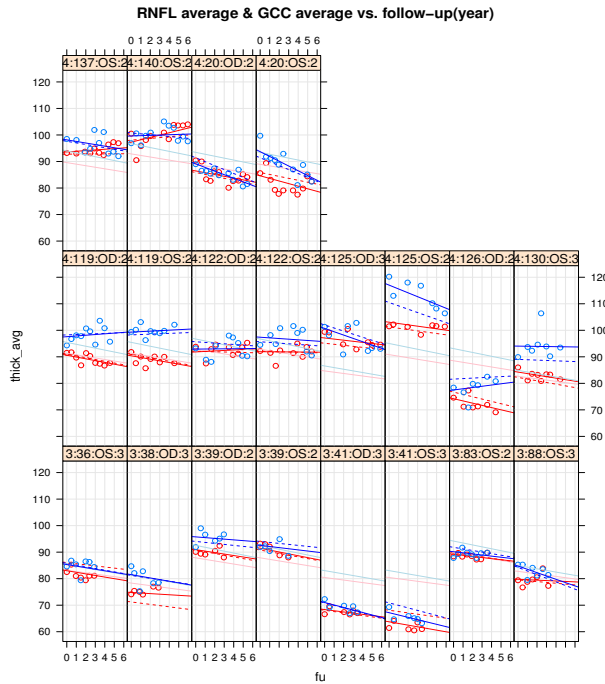

Figure 4.1: 20 eye-characteristics with largest conditional cook's distance

Table 4.1: Conditional Cook's Distance for the 20 eye-characteristics with largest conditional cook's distance.

Eye ID	Outcome Type	Diag grp	Number of Observations	Cond Cooksd	C_{A1}	C_{A2}	C_{A3}	C_{A4}	C_{A5}	C_{A6}
4:20:OD	RNFL	2	11	0.2114	0.0071	1.1768	1.5365	-0.0137	0.0008	-2.4961
4:20:OS	RNFL	2	11	0.1595	0.0071	1.1001	1.2793	-0.0164	0.0033	-2.2138
4:125:OS	RNFL	2	7	0.1566	0.0676	1.5757	1.3166	-0.1331	-0.0089	-2.6613
4:122:OS	RNFL	2	9	0.1479	0.0225	1.8638	1.8448	-0.0487	0.0104	-3.5450
4:137:OS	RNFL	2	10	0.1342	0.0199	2.3371	2.2882	-0.0416	0.0031	-4.4725
4:119:OD	RNFL	2	10	0.1166	0.0082	1.9603	2.0206	-0.0120	-0.0036	-3.8570
3:88:OS	RNFL	3	7	0.1099	0.0242	3.2979	3.0727	-0.0659	0.0154	-6.2344
3:83:OS	RNFL	2	8	0.1067	0.0104	2.7103	2.5830	-0.0151	-0.0031	-5.1786
3:39:OS	RNFL	2	5	0.1065	0.0087	0.6925	0.6342	-0.0195	0.0000	-1.2093
4:130:OS	RNFL	3	8	0.0951	0.0113	1.0902	1.0278	-0.0247	0.0036	-2.0131
4:119:OS	RNFL	2	9	0.0883	0.0069	1.0283	1.0772	-0.0098	-0.0046	-2.0096
4:126:OD	RNFL	2	7	0.0843	0.0145	1.9138	1.7834	-0.0337	0.0040	-3.5977
4:125:OD	RNFL	3	8	0.0835	0.0041	1.1295	1.1621	-0.0072	0.0004	-2.2053
3:41:OS	RNFL	3	6	0.0830	0.0319	1.8221	1.7933	-0.0592	-0.0010	-3.5041
3:41:OD	RNFL	3	6	0.0828	0.0305	2.0089	1.9079	-0.0715	0.0164	-3.8094
4:140:OS	RNFL	2	11	0.0734	0.0052	1.6561	1.5290	-0.0100	-0.0011	-3.1058
3:39:OD	RNFL	2	6	0.0657	0.0069	1.5981	1.5458	-0.0149	0.0012	-3.0713
3:38:OD	RNFL	3	6	0.0626	0.0229	1.8019	1.5861	-0.0549	0.0078	-3.3012
4:122:OD	RNFL	2	9	0.0558	0.0069	1.3763	1.3979	-0.0083	-0.0031	-2.7138
3:36:OS	RNFL	3	7	0.0409	0.0034	0.4339	0.3432	-0.0060	-0.0013	-0.7324

Second, for multivariate longitudinal data, an associated problem with joint modeling is: as the number of characteristics goes up, a convergence issue becomes more and more severe. To resolve this computational difficulty, a pairwise fitting approach originally proposed by Fieuws and Verbeke [13] can be adopted into our method to estimated the random effects and correlations between different characteristics.

Third, currently, we are applying our method to linear multivariate mixed effect model. In the future, an extension to multivariate generalized linear mixed models can be developed, where nonlinear models are considered.

APPENDIX A

SIMULATION

```
# The function generating bivariate normal distribution data:
rbivariate <- function(mean.x=0, sd.x=1, mean.y=0, sd.y=1, rho=.50, iter=100) {
    z1 <- rnorm(iter)
    z2 <- rnorm(iter)
    x <- sqrt(1-rho^2)*sd.x*z1 + rho*sd.x*z2 + mean.x
    y <- sd.y*z2 + mean.y
    return(list(x,y))
}


# Generate yi dataset:
ryi <- function(n=50, rho.y=0.8, rho.b=0.2, rho.e=0.5, minni=1, maxni=9) {

    # Generate ui's:
    data.u <- rbivariate(mean.x=0, mean.y=0, sd.x=1, sd.y=1, rho=rho.y, iter=n)
    var(matrix(c(data.u[[1]], data.u[[2]]), nrow=50, ncol=2))

    # Generate matrix B (bi's):
    data.b <- rbivariate(mean.x=0, mean.y=0, sd.x=1, sd.y=1, rho=rho.b, iter=n)
    B <- matrix(c(data.b[[1]], data.b[[2]]), nrow=50, ncol=2)
    var(B)

    # Generate ni's:
    ni <- as.integer(runif(n, 1, 10))
    ni[n] <- 9
    total_length <- sum(ni)

    # Generate matrix E (epsilon's):
    data.e <- rbivariate(mean.x=0, mean.y=0, sd.x=1, sd.y=1, rho=rho.e, iter=total_length)
    E <- matrix(c(data.e[[1]], data.e[[2]]), nrow=total_length, ncol=2)
    var(E)
    e1i <- data.e[[1]]
    e2i <- data.e[[2]]

    # Generate ID's:
    for(i in 1:n) {
        if(i==1) id <- rep(i,ni[i])
        else id <- c(id, rep(i,ni[i]))
    }

    # Generate tij's (tt and ttc):
    # The first individual has 1 measurement
    # and the last individual has 5 measurements
    for(i in 1:n) {
        if(i==1) { tt <- log(1:ni[i]); ttc <- 1:ni[i] }
        else {tt <- c(tt, log(1:ni[i])); ttc <- c(ttc, 1:ni[i]) }
    }

    # Generate u1i,u2i, b1i,b2i, e1i,e2i:
    for(i in 1:n) {
        if(i==1) {
            u1i <- rep(data.u[[1]][i], ni[i])
            u2i <- rep(data.u[[2]][i], ni[i])
            b1i <- rep(data.b[[1]][i], ni[i])
            b2i <- rep(data.b[[2]][i], ni[i])
```

```
        }
        else {
            u1i <- c(u1i, rep(data.u[[1]][i], ni[i]))
            u2i <- c(u2i, rep(data.u[[2]][i], ni[i]))
            b1i <- c(b1i, rep(data.b[[1]][i], ni[i]))
            b2i <- c(b2i, rep(data.b[[2]][i], ni[i]))
        }
    }

    # Make the "Outliers":
    b1i[total_length-4] <- 7
    # b2i[total_length-4] <- 6

    # Generate the data frame:
    y1 <- y2 <- rep(0, total_length)
    simu.data <- data.frame(id, y1, y2, tt, ttc, u1i,u2i, b1i,b2i, e1i,e2i)

    names(simu.data)[ 6] <- "u1"
    names(simu.data)[ 7] <- "u2"
    names(simu.data)[ 8] <- "b1"
    names(simu.data)[ 9] <- "b2"
    names(simu.data)[10] <- "e1"
    names(simu.data)[11] <- "e2"

    beta10 <- beta11 <- beta12 <- 1
    beta20 <- beta21 <- beta22 <- 1

    simu.data$y1 <- with(simu.data, beta10 + beta11*u1 + beta12*tt + b1 + e1)
    simu.data$y2 <- with(simu.data, beta20 + beta21*u2 + beta22*tt + b2 + e2) # only use u1, not u2 ?
    simu.data$int.id <- 1

    return(simu.data)
}

# Compute the Cook's distances:
COOKSD <- function(full.model, simu.data) {
    simu.data.u <- simu.data

    Sigma_inv <- solve(full.model$sigma)
    # V = Sigma + ZGZ' = Sigma + G (In this model, Z = I)
    V <- full.model$sigma + full.model$psi
    V_inv <- solve(V)

    # First, calcualte the yhat based on the complete dataset:
    beta10hat <- full.model$beta[1,1]
    beta11hat <- full.model$beta[2,1]
    beta12hat <- full.model$beta[3,1]
    beta20hat <- full.model$beta[1,2]
    beta21hat <- full.model$beta[2,2]
    beta22hat <- full.model$beta[3,2]

    eb <- data.frame(t(full.model$eb))
    names(eb)[1] <- "b1hat"
    names(eb)[2] <- "b2hat"
    eb$id <- rownames(eb)

    # For conditional:
    simu.data <- merge(x=simu.data, y=eb, by.x="id", by.y="id")
    simu.data$y1hat <- with(simu.data,  beta10hat + beta11hat*u1 + beta12hat*tt + b1hat)
    simu.data$y2hat <- with(simu.data,  beta20hat + beta21hat*u1 + beta22hat*tt + b2hat)

    simu.data$cooksd1 <- 0
    simu.data$cooksd2 <- 0
    simu.data$y1hat_i <- 0
    simu.data$y2hat_i <- 0

    simu.data$d_beta10hat_y1 <- 0
    simu.data$d_beta11hat_y1 <- 0
    simu.data$d_beta12hat_y1 <- 0
    simu.data$d_beta20hat_y1 <- 0
    simu.data$d_beta21hat_y1 <- 0
    simu.data$d_beta22hat_y1 <- 0

    simu.data$d_beta10hat_y2 <- 0
    simu.data$d_beta11hat_y2 <- 0
    simu.data$d_beta12hat_y2 <- 0
    simu.data$d_beta20hat_y2 <- 0
    simu.data$d_beta21hat_y2 <- 0
    simu.data$d_beta22hat_y2 <- 0
```

```
    simu.data$d_b1_50_hat_y1 <- 0
    simu.data$d_b2_50_hat_y1 <- 0

    simu.data$d_b1_50_hat_y2 <- 0
    simu.data$d_b2_50_hat_y2 <- 0

    # For unconditional:
    simu.data.u$y1hat <- with(simu.data,  beta10hat + beta11hat*u1 + beta12hat*tt)
    simu.data.u$y2hat <- with(simu.data,  beta20hat + beta21hat*u1 + beta22hat*tt)

    simu.data.u$cooksd1 <- 0
    simu.data.u$cooksd2 <- 0
    simu.data.u$y1hat_i <- 0
    simu.data.u$y2hat_i <- 0

    # For y1's:
    for(i in 1:dim(simu.data)[1]) {
        y_i <- y
        y_i[i,1] <- NA
        y_i[i,2] <- NA # set the other component to NA (missing)
        xcol <- 1:3
        zcol <- 1
        ith.model <- mlmmm.em(y_i, subj, pred, xcol, zcol, maxits=200, eps=0.0001)
        beta10hat_i <- ith.model$beta[1,1]
        beta11hat_i <- ith.model$beta[2,1]
        beta12hat_i <- ith.model$beta[3,1]
        beta20hat_i <- ith.model$beta[1,2]
        beta21hat_i <- ith.model$beta[2,2]
        beta22hat_i <- ith.model$beta[3,2]

        eb_i <- data.frame(t(ith.model$eb))
        names(eb_i)[1] <- "b1hat_i"
        names(eb_i)[2] <- "b2hat_i"
        eb_i$id <- rownames(eb_i)

        simu.data <- merge(x=simu.data, y=eb_i, by.x="id", by.y="id")
        simu.data$y1hat_i <- with(simu.data,  beta10hat_i + beta11hat_i*u1 + beta12hat_i*tt + b1hat_i)
        simu.data$y2hat_i <- with(simu.data,  beta20hat_i + beta21hat_i*u1 + beta22hat_i*tt + b2hat_i)

        cooksd <- 0
        for(j in 1:dim(y)[1]) {
            dy_i <- with(simu.data, matrix(c(y1hat_i[j]-y1hat[j], y2hat_i[j]-y2hat[j]), nrow=2) )
            cooksd <- cooksd + t(dy_i) %*% Sigma_inv %*% dy_i
        }
        simu.data$cooksd1[i] <- cooksd

        simu.data$d_beta10hat_y1[i] <- beta10hat_i - beta10hat
        simu.data$d_beta11hat_y1[i] <- beta11hat_i - beta11hat
        simu.data$d_beta12hat_y1[i] <- beta12hat_i - beta12hat
        simu.data$d_beta20hat_y1[i] <- beta20hat_i - beta20hat
        simu.data$d_beta21hat_y1[i] <- beta21hat_i - beta21hat
        simu.data$d_beta22hat_y1[i] <- beta22hat_i - beta22hat

        simu.data$d_b1_50_hat_y1[i] <- eb_i$b1hat_i[50] - eb$b1hat[50]
        simu.data$d_b2_50_hat_y1[i] <- eb_i$b2hat_i[50] - eb$b2hat[50]

        simu.data <- simu.data[, -c(37,38)]

        # Unconditional:
        simu.data.u$y1hat_i <- with(simu.data.u,  beta10hat_i + beta11hat_i*u1 + beta12hat_i*tt)
        simu.data.u$y2hat_i <- with(simu.data.u,  beta20hat_i + beta21hat_i*u1 + beta22hat_i*tt)

        cooksd <- 0
        for(j in 1:dim(y)[1]) {
            dy_i <- with(simu.data.u, matrix(c(y1hat_i[j]-y1hat[j], y2hat_i[j]-y2hat[j]), nrow=2) )
            cooksd <- cooksd + t(dy_i) %*% V_inv %*% dy_i
        }
        simu.data.u$cooksd1[i] <- cooksd
    }
    return(list(simu.data, simu.data.u))
}

# Program starts here:
# -------------------
library(mlmmm)

n.rep <- 1000

list.beta  <- vector("list", n.rep)
list.psi   <- vector("list", n.rep)
```

```
list.sigma <- vector("list", n.rep)

results.ccd <- rep(0, n.rep)
results.ucd <- rep(0, n.rep)

# Start the clock!
ptm <- proc.time()

for(i in 1:n.rep) {
    # Generate data:
    simu.data <- ryi()

    # Fit the full model (with individual effects):
    y <- with(simu.data, cbind(y1,y2))
    subj <- simu.data$id
    pred <- with(simu.data, cbind(int.id, u1, tt))
    xcol <- 1:3
    zcol <- 1
    fit.mlmmm.full <- mlmmm.em(y, subj, pred, xcol, zcol, maxits=200, eps=0.0001)
    list.beta[[i]]  <- matrix(fit.mlmmm.full$beta, nrow=6) # convert to vector
    list.psi[[i]]   <- fit.mlmmm.full$psi
    list.sigma[[i]] <- fit.mlmmm.full$sigma

    # Conditional and Naive Cook's Distances:
    D2 <- COOKSD(full.model=fit.mlmmm.full, simu.data=simu.data)
    simu.data.ccd <- D2[[1]]
    simu.data.ucd <- D2[[2]]
    simu.data.ccd.ordered <- simu.data.ccd[order(simu.data.ccd$cooksd1, decreasing=TRUE), ]
    simu.data.ucd.ordered <- simu.data.ucd[order(simu.data.ucd$cooksd1, decreasing=TRUE), ]

    if(as.integer(rownames(simu.data.ccd.ordered)[1])==dim(simu.data.ccd.ordered)[1]-4) {results.ccd[i] <- 1}
    if(as.integer(rownames(simu.data.ucd.ordered)[1])==dim(simu.data.ucd.ordered)[1]-4) {results.ucd[i] <- 1}
    print(i)
    print(results.ccd[i])
    print(results.ucd[i])
}

# Stop the clock
proc.time() - ptm

# Calculate the average estimated value:
all.beta <- s.beta <- list.beta[[1]]
for(i in 2:n.rep) {
s.beta <- s.beta + list.beta[[i]]
all.beta <- cbind(all.beta, list.beta[[i]])
}
all.beta <- t(all.beta)
var(all.beta)
beta <- s.beta/n.rep
beta

s.psi <- list.psi[[1]]
for(i in 2:n.rep) s.psi <- s.psi + list.psi[[i]]
psi <- s.psi/n.rep
psi
cov2cor(psi)

s.sigma <- list.sigma[[1]]
for(i in 2:n.rep) s.sigma <- s.sigma + list.sigma[[i]]
sigma <- s.sigma/n.rep
sigma
cov2cor(sigma)

results.ccd
sum(results.ccd)/n.rep

results.ucd
sum(results.ucd)/n.rep

table(results.ccd, results.ucd)
```

APPENDIX B

APPLICATION - MULTILEVEL MULTIVARIATE COOK'S DISTANCE

B.1 TWO-LEVEL MULTIVARIATE MIXED EFFECT MODEL

```
data long3sas;
infile 'D:\LINGYUN\dropbox\Dropbox\Yun Ling\Model_04\Data\long3sas2.csv' dsd delimiter='09'x;
input id $ ideye $ thick_avg type $ fu dx bsage type_num fu_rnfl fu_gcc bsage_rnfl bsage_gcc int_rnfl int_gcc ideyetype $ fu_order;
run;

proc print data=long3sas;
title 'Multilevel Multivaraite Mixed Effect Models';
run;

** 1-level;
proc mixed data=long3sas covtest noclprint;
class ideye dx fu_order type_num;
model thick_avg = int_rnfl*dx int_gcc*dx fu_rnfl*dx fu_gcc*dx bsage_rnfl bsage_gcc /noint solution outpm=resid_1;
random int_rnfl int_gcc fu_rnfl fu_gcc /subject=ideye type=un g gcorr;
repeated type_num/subject=ideye*fu_order type=un r rcorr;
** ods output covparms=cov_1 solutionF=Fixed_1;
run;

** 2-level (Nested Model), but residuals are independent;
** This model works well !!!;
** 2013/12/01...;
proc mixed data=long3sas covtest noclprint;
class id ideye dx fu_order type_num;
model thick_avg = int_rnfl*dx int_gcc*dx fu_rnfl*dx fu_gcc*dx bsage_rnfl bsage_gcc /noint solution outpm=resid_1;
random int_rnfl int_gcc fu_rnfl fu_gcc /subject=id type=un g gcorr s;
random int_rnfl int_gcc fu_rnfl fu_gcc /subject=ideye(id) type=un g gcorr s;
** repeated type_num/subject=ideye*fu_order type=vc r rcorr;
repeated /type=vc group=type_num subject=ideye(id) r rcorr;
where type_num=1 or type_num=2;
ods output covparms=cov_1 solutionF=Fixed_1 solutionR=Random_1;
run;

proc print data=Fixed_1;
format estimate 10.6;
run;

PROC EXPORT DATA= WORK.Random_1
            OUTFILE= "D:\LINGYUN\dropbox\Dropbox\Yun Ling\Thesis_Model_6
\SAS\SAS_Output_20131201_random.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;
```

B.2 INFLUENCE FUNTION AND MULTILEVEL MULTIVARIATE COOK'S DISTANCE

```
# Random Effects:
# ---------------
re <- read.table("/media/yun/DATA/LINGYUN/dropbox/Dropbox/Yun Ling/Model_04/SAS/SAS_Output_20131209_random.csv", header=TRUE, sep=',')
# re <- read.table("D:/LINGYUN/dropbox/Dropbox/Yun Ling/Model_04/SAS/SAS_Output_20131209_random.csv", header=TRUE, sep=',')

# Remove those incorrect records......
re[re$Estimate==0, ]
factor(re[re$Estimate==0, ]$ideye) # 26 levels...
levels(factor(re[re$Estimate==0, ]$ideye))
re <- re[re$Estimate!=0, ]
re <- re[, -c(5:8)]
re <- re[order(re$id, re$ideye), ]
re$ideye <- factor(re$ideye)
save.image()

# Fixed Effects:
# --------------
beta.hat.rnfl <- c(111.66, 104.27, 95.8123, -0.08193, -0.7430, -0.6128, -0.1803)
beta.hat.gcc  <- c(107.77, 101.96, 95.7668, -0.95520, -0.5873, -0.4479, -0.2179)
beta.hat <- c(beta.hat.rnfl, beta.hat.gcc)
beta.hat <- matrix(beta.hat, ncol=1)

# Variance-covariance matrices:
# -----------------------------
library(Matrix)
library(bigmemory)

# Note: b10 b20 b11 b21
G1 <- matrix( c(73.2727, 39.6156, -0.4922,  0.1135,
39.6156, 38.5626,  0.3086, -0.1125,
-0.4922,  0.3086,  0.8820, 0.01798,
 0.1135, -0.1125, 0.01798,  0.3775), nrow=4)
# Note: c10 c20 c11 c21
G2 <- matrix( c(23.7693, 12.4620, -0.7705, -0.4184,
12.4620, 13.8827, -0.4175, -0.1597,
-0.7705, -0.4175, 0.04496,  0.1472,
-0.4184, -0.1597,  0.1472,  0.3505), nrow=4)
# GG1 <- bdiag(G1,G2,G2)
# GG2 <- bdiag(G1,G2) # Some subject missing one eye...
Sigma <- matrix( c(13.8980, 0, 0, 8.4852), nrow=2)
save.image()

# Import the dataset:
# -------------------
long <- read.table("/media/yun/DATA/LINGYUN/dropbox/Dropbox/Yun Ling/Model_04/Data/long3sas.csv", header=TRUE, sep='\t')
long$type <- with(long, factor(type, c("RNFL", "GCC")))
long <- long[order(long$id, long$id.eye, long$type, long$fu), ]
long <- long[, -c(8:16)]
names(long)
long$dx <- factor(long$dx)

# (1) Create X matrix:
long2 <- long
long2$beta10 <- with(long2, ifelse(dx=="1" & type=="RNFL", 1, 0))
long2$beta11 <- with(long2, ifelse(dx=="2" & type=="RNFL", 1, 0))
long2$beta12 <- with(long2, ifelse(dx=="3" & type=="RNFL", 1, 0))
long2$beta13 <- with(long2, ifelse(dx=="1" & type=="RNFL", fu, 0))
long2$beta14 <- with(long2, ifelse(dx=="2" & type=="RNFL", fu, 0))
long2$beta15 <- with(long2, ifelse(dx=="3" & type=="RNFL", fu, 0))
long2$beta16 <- with(long2, ifelse(type=="RNFL", baseline.age, 0))
long2$beta20 <- with(long2, ifelse(dx=="1" & type=="GCC", 1, 0))
long2$beta21 <- with(long2, ifelse(dx=="2" & type=="GCC", 1, 0))
long2$beta22 <- with(long2, ifelse(dx=="3" & type=="GCC", 1, 0))
long2$beta23 <- with(long2, ifelse(dx=="1" & type=="GCC", fu, 0))
long2$beta24 <- with(long2, ifelse(dx=="2" & type=="GCC", fu, 0))
long2$beta25 <- with(long2, ifelse(dx=="3" & type=="GCC", fu, 0))
long2$beta26 <- with(long2, ifelse(type=="GCC", baseline.age, 0))
X.data <- long2[, 8:21]
X <- as.matrix(x=X.data)

# (2) Create Z (Z and W) and G (GG) matrix:
id <- with(long, levels(factor(id)))
idx <- 1
for(i in id)
```

```
{
re.id <- re[re$id==i, ]
long.id <- long[long$id==i, ]
ideye <- levels(factor(re.id$ideye))
idx.j <- 1
for(j in ideye)
{
# Create G* (Gstar) and Zi:
if(j=="")
Gstar <- G1
else if(idx.j==2) {
Gstar <- bdiag(Gstar, G2)

# Create Zi, Wi1, Wi2:
fu1 <- long.id[long.id$type=="RNFL" & long.id$id.eye==j, ]$fu
l1 <- length(fu1)
Zij1 <- cbind(rep(1,l1), rep(0,l1), fu1, rep(0,l1))

fu2 <- long.id[long.id$type=="GCC"  & long.id$id.eye==j, ]$fu
l2 <- length(fu1)
Zij2 <- cbind(rep(0,l1), rep(1,l1), rep(0,l1), fu2)

Wij <- Zij <- rbind(Zij1, Zij2)
Zi1star <- cbind(Zij, Wij)

Zistar <- Zi1star # The Z* matrix
ZZi <- Zij # The Z matrix
WWi <- Wij # The W matrix
}
else if(idx.j==3) {
Gstar <- bdiag(Gstar, G2)

# Create Zi, Wi1, Wi2:
fu1 <- long.id[long.id$type=="RNFL" & long.id$id.eye==j, ]$fu
l1 <- length(fu1)
Zij1 <- cbind(rep(1,l1), rep(0,l1), fu1, rep(0,l1))

fu2 <- long.id[long.id$type=="GCC"  & long.id$id.eye==j, ]$fu
l2 <- length(fu1)
Zij2 <- cbind(rep(0,l1), rep(1,l1), rep(0,l1), fu2)

Wij <- Zij <- rbind(Zij1, Zij2)
Zi2star <- cbind(Zij, matrix(0, nrow=dim(Zij)[1], ncol=dim(Zij)[2]), Wij)

Zistar <- cbind(Zistar, matrix(0, nrow=dim(Zistar)[1], ncol=dim(Wij)[2]))
Zistar <- rbind(Zistar, Zi2star)
ZZi <- rbind(ZZi, Zij) # The Z matrix
WWi <- bdiag(WWi, Wij) # The W matrix
}

idx.j <- idx.j + 1
}
if(idx==1) {
Z <- Zistar
ZZ <- ZZi
WW <- WWi
GG <- Gstar
}
else {
Z <- bdiag(Z, Zistar)
ZZ <- bdiag(ZZ, ZZi)
WW <- bdiag(WW, WWi)
GG <- bdiag(GG, Gstar)
}
idx <- idx + 1
}

Z <- as.big.matrix(as.matrix(Z))
ZZ <- as.big.matrix(as.matrix(ZZ))
WW <- as.big.matrix(as.matrix(WW))
GG <- as.big.matrix(as.matrix(GG))

# Create the R matrix (assume R is diagonal):
sigma.rnfl <- 13.8980
sigma.gcc  <-  8.4852

RRinv <- RR <- diag(dim(long)[1])
for(k in 1:dim(RR)[1]) RR[k,k] <- ifelse(long$type[k]=="RNFL", sigma.rnfl, sigma.gcc)
for(k in 1:dim(RRinv)[1]) RRinv[k,k] <- ifelse(long$type[k]=="RNFL", 1/sigma.rnfl, 1/sigma.gcc)
```

```
RR <- as.big.matrix(RR)
RRinv <- as.big.matrix(RRinv)

b.hat <- re$Estimate
b.hat <- as.big.matrix(matrix(b.hat, ncol=1))

y.hat <- X %*% beta.hat + Z %*% b.hat

bb.hat <- re[re$ideye=="",]$Estimate
cc.hat <- re[re$ideye!="",]$Estimate
bb.hat <- as.big.matrix(matrix(bb.hat, ncol=1))
cc.hat <- as.big.matrix(matrix(cc.hat, ncol=1))

y.hat.2 <- X %*% beta.hat + ZZ %*% bb.hat + WW %*% cc.hat

# Create the V matrix:
V <- Z %*% GG %*% t(Z) + RR

long9 <- long
long9$y.hat <- y.hat
long9$y.hat.2 <- y.hat.2
long9$epsilon <- long9$thick_avg - long9$y.hat
long9$diff <- with(long9, y.hat - y.hat.2)

# Create other quantities:
Vinv <- solve(V)
Xt.Vinv.X_inv <- solve(t(X) %*% Vinv %*% X)
Zt.Vinv.X_Xt.Vinv.X_inv <- t(Z) %*% Vinv %*% X %*% Xt.Vinv.X_inv


y <- long$thick_avg
y <- matrix(y, ncol=1)

# Using the dataset "long":
long7 <- long
long7$cooksd <- 0

m <- 2 # number of characteristics...
N <- 487 # number of eyes...
p <- 14 # number of fixed effects...
q <- 2 # number of random effects...
dfr <- (N*m-1)*2+p


# Calculate the decomposed Cook's distance for the largest 10 observations:
# ----------------------------------------------------------------------
cooksd_comp_10$ca1 <- 0
cooksd_comp_10$ca2 <- 0
cooksd_comp_10$ca3 <- 0
cooksd_comp_10$ca4 <- 0
cooksd_comp_10$ca5 <- 0
cooksd_comp_10$ca6 <- 0
cooksd_comp_10$cooksd2 <- 0

# Start the clock!
ptm <- proc.time()
for(i in 1:dim(cooksd_comp_10)[1])
{
# ii contains the index of the observations to be removed:
ii <- long8[long8$id.eye==cooksd_comp_10[i,]$id.eye & long8$fu==cooksd_comp_10[i,]$fu & long8$type==cooksd_comp_10[i,]$type, ]$idx

# yi, yit, y.i (y(i))
yit <- matrix(y[ii], ncol=1)
yi <- t(yit)
y.i <- matrix(y[-ii], ncol=1)

# Xi, Xit, X.i (X(i)):
Xit <- X[ii, ]
if(length(ii)==1) Xit <- matrix(Xit, nrow=1)
Xi <- t(Xit)
X.i <- X[-ii, ]

# Zi, Zit, Z.i (Z(i)):
Zit <- Z[ii, ]
if(length(ii)==1) Zit <- matrix(Zit, nrow=1)
Zi <- t(Zit)
Z.i <- Z[-ii, ]

# Vii, Vit, Vi, V.i (V[i]):
Vii <- V[ ii, ii]
```

```
V.i  <- V[-ii,-ii]
Vit <- V[ ii,-ii]
if(length(ii)==1) Vit <- matrix(Vit, nrow=1)
Vi  <- V[-ii, ii]

# V[i]^-1 * Vi:
V.i.inv <- solve(V.i)
V.i.inv.Vi <- V.i.inv %*% Vi


yi.tilda <- yi - t(y.i) %*% V.i.inv.Vi
Xi.tilda <- Xi - t(X.i) %*% V.i.inv.Vi
Zi.tilda <- Zi - t(Z.i) %*% V.i.inv.Vi

# Si and Hii:
Si <- Vii - Vit %*% V.i.inv %*% Vi
Hii <- t(Xi.tilda) %*% Xt.Vinv.X_inv %*% Xi.tilda

# beta.hat.A and b.hat.A:
# -----------------------
beta.hat.A <- beta.hat  - Xt.Vinv.X_inv %*% Xi.tilda %*% solve(Si-Hii) %*% (t(yi.tilda) - t(Xi.tilda)%*%beta.hat)

b.hat.A <- b.hat  -  GG %*% (Zi.tilda - Zt.Vinv.X_Xt.Vinv.X_inv%*%Xi.tilda) %*% solve(Si-Hii) %*% (t(yi.tilda) - t(Xi.tilda)%*%beta.hat)
b.hat.A <- as.matrix(b.hat.A)

y.hat.A <- X %*% beta.hat.A + Z %*% b.hat.A
y.hat.A <- as.matrix(y.hat.A) # convert to normal matrix...


# bb.hat.A and cc.hat.A:
# ----------------------
re$b.hat.A <- b.hat.A
bb.hat.A <- re[re$ideye=="",]$b.hat.A
cc.hat.A <- re[re$ideye!="",]$b.hat.A
bb.hat.A <- matrix(bb.hat.A, ncol=1)
cc.hat.A <- matrix(cc.hat.A, ncol=1)

# Remove some matrices:
rm(Vi); rm(Vit); rm(Vii); rm(V.i); rm(V.i.inv); rm(V.i.inv.Vi)
rm(Zi); rm(Zit); rm(Z.i)

# Compute the Conditional Cook's Distance (decomposed):
CA1 <- t(beta.hat.A - beta.hat) %*% t(X) %*% RRinv %*% X %*% (beta.hat.A - beta.hat) / dfr
CA2 <- t(bb.hat.A - bb.hat) %*% t(ZZ) %*% RRinv %*% ZZ %*% (bb.hat.A - bb.hat) / dfr
CA3 <- t(cc.hat.A - cc.hat) %*% t(WW) %*% RRinv %*% WW %*% (cc.hat.A - cc.hat) / dfr
CA4 <- 2*t(beta.hat.A - beta.hat) %*% t(X) %*% RRinv %*% ZZ %*% (bb.hat.A - bb.hat) / dfr
CA5 <- 2*t(beta.hat.A - beta.hat) %*% t(X) %*% RRinv %*% WW %*% (cc.hat.A - cc.hat) / dfr
CA6 <- 2*t(bb.hat.A - bb.hat) %*% t(ZZ) %*% RRinv %*% WW %*% (cc.hat.A - cc.hat) / dfr
cooksd_comp_10[i,]$ca1 <- CA1[1,1]
cooksd_comp_10[i,]$ca2 <- CA2[1,1]
cooksd_comp_10[i,]$ca3 <- CA3[1,1]
cooksd_comp_10[i,]$ca4 <- CA4[1,1]
cooksd_comp_10[i,]$ca5 <- CA5[1,1]
cooksd_comp_10[i,]$ca6 <- CA6[1,1]
cooksd_comp_10[i,]$cooksd2 <- cooksd_comp_10[i,]$ca1 + cooksd_comp_10[i,]$ca2 + cooksd_comp_10[i,]$ca3 + cooksd_comp_10[i,]$ca4 + cooksd_comp_10[i,]$ca5 + cooksd_comp_10[

# Compute the Conditional Cook's Distance:
# cooksd_comp_10[i,]$cooksd <- t(y.hat.A - y.hat) %*% RRinv %*% (y.hat.A - y.hat) / dfr

print(i)
}


# Stop the clock
proc.time() - ptm
```

BIBLIOGRAPHY

[1] J. T. Newsom, "A multilevel structural equation model for dyadic data", *Structural Equation Modeling*, Vol.9, pp. 431-447, 2002.

[2] R. Thiebaut, H. Jacqmin-gadda, G. Cheng, C. Leport, D. Commenges, "Bivariate linear mixed models using SAS proc MIXED", *Computer Methods, Programs in Biomedicine*, vol.69, pp. 249-256, 2002

[3] A. B. Chapman, L. M. Guay-Woodford, J. J. Grantham, V. E. Torres, K. T. Bae, D. A. Baumgarten, P. J. Kenney, B. F. King, J. F. Glockner, L. H. Wetzel, M. E. Brummer, W. C. O'Neill, M. L. Robbin, W. M. Bennett, S. Klahr, G. H. Hirschman, P. L. Kimmel, P. A. Thompson, J. P. Miller, "Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): the Consortium for Radiology Imaging Study of Polycystic Kidney Disease (CRISP) cohort", *Kidney International*, vol.64, pp. 1035-1045, 2003

[4] F. Gao, P. Thompson, C. Xiong, J. P. Miller, "Analyzing multivariate longitudinal data using SAS", *Proceedings of the Thirty-first Annual SAS Users Group International Conference*, pp. 187-31, 2006

[5] J. N. M. Ouwens, F. E. S. Tan, M. P. F. Berger, "Local influence for repeated measures generalized linear mixed models", *Proceedings 14th International Workshop Statistical Modeling*, 1999

[6] S. Xu, M. Styner, J. Gilmore, G. Gerig, "Multivariate longitudinal statistics for neonatal-pediatric brain tissue development", *Proceedings of SPIE*, pp. 6914, 69140C, 2008

[7] G. Reinsel, "Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure", *Journal of the American Statistical Association*, vol.77, pp. 190-95, 1982

[8] G. Reinsel, "Estimation, prediction in a multivariate random effect generalized linear model", *Journal of the American Statistical Association*, vol.79, pp. 406-14, 1984

[9] A. Shah, N. M. Laird, D. Schoenfeld, "A random effects model with multiple characteristics with possibly missing data", *Journal of the American Statistical Association*, vol.92, pp. 775-79, 1997

[10] C. H. Morrell, L. J. Brant, S. Sheng, S. Najjar, "Using multivariate mixed effect model to predict hypertension", *Joint Statistical Meetings, Biometrics Section*, 2003,

[11] J. Reeves, G. MacKenzie, "A bivariate regression model with serial correlation", *The statistician*, vol.47, pp. 607-15, 1998

[12] B. A. Coull, J. Staudenmayer, "Self-modeling regression for multivariate curve data, *Statistica Sinica*, vol.14, pp. 695-711, 2004

[13] S. Fieuws, G. Verbeke, "Pairwise fitting of mixed models for the joint modeling of multivariate longtudinal profiles: pitfalls of the random effect approach", *Biometrics*, vol.62, pp. 424-31, 2006

[14] N. W. Shock, R. C. Greulich, R. Andres, "Normal human aging: the Baltimore Longitudinal Study of Aging", *U.S. Government Printing Office, Washington, DC (NIH Publication)*, vol.1984, pp. 84-2450,

[15] J. G. Ibrahim, M. H. Chen, D. Sinha, "Bayesian methods for joint modeling of longitudinal, survival data with applications to cancer vaccine studies", *Statistica Sinica*, vol.14, pp. 847-867, 2004

[16] J. Xu, S. L. Zeger, "Joint analysis of longitudinal data comprising repeated measures, times to events", *Applied Statistics*, vol.50, pp. 375-387, 2001

[17] J. Xu, S. L. Zeger, "The evaluation of multiple surrogate endpoints", *Biometrics*, vol.57, pp. 81-87, 2001

[18] X. Song, M. Davidian, A. A. Tsiatis, "An estimator for the proportional hazards model with multiple longitudinal covariates measured with errors", *Biostatistics*, vol.3, pp. 511-528, 2001

[19] A. A. Tsiatis, M. Davidian, "Joint modeling of longitudinal, time-to-event data: An overview", *Statistica Sinica*, vol.14, pp. 793-818, 2004

[20] Frans E. S. Tan, Mario J. N. Ouwens, Martijn P. F. Berger, "Detection of influential observation in longitudinal mixed effects regression data", *Journal of the Royal Statistical Society*, series. D, vol.50, No.3, pp. 271-284, 2001

[21] R. D. Cook, "Detection of influential observation in linear regression, *Econometrics*, vol.19, pp. 15-18, 1977

[22] R. D. Cook, "Assessment of local influence (with discussion)", *Journal of the Royal Statistical Society, Series B*, vol.48, pp. 133-169, 1986

[23] R. J. Beckman, C. J. Nachtsheim, R. D. Cook, "Diagnostics for mixed model analysis of variance", *Technometrics*, vol.29, pp. 413-426, 1987

[24] B. E. Barrett, R. F. Ling, "General classes of influence measures for multivariate regression", *Journal of American Statistic Association*, vol.87, pp. 187-191, 1992

[25] R. Christensen, L. M. Pearson, W. Johnson, "Case-deletion diagnostics for mixed models", *Technometrics*, vol.34, pp. 38-45, 1992

[26] B. Altunkaynak, M. Ekni, "Detection of influential observation vectors for multivariate linear regression", *Hacettepe Journal of Mathematics, Statistics*, 2002, vol.31, pp. 139-151, 2002

[27] F. R. Hampel, "The influence curve, its role in robust estimation", *Journal of American Statistical Association*, vol.69, pp. 383-393, 1974

[28] R. E. Weiss, "Residuals, outliers in repeated measures random effects models", *Technical Report*, Department of Biostatistics, UCLA School of Public Health, 1995

[29] Andrea Cerioli, "Multivariate outlier detection with high-breakdown estimators", *Journal of American Statistic Association*, vol.105:489, pp. 147-156, 2010

[30] J. A. Diaz-Garcia, G. Gonzalez-Farias, "A note on the Cook's Distance", *Journal of Statistical Planning, Inference*, vol.120, pp. 119-136, 2004

[31] A. Hossain, D. N. Naik, "Detection of influential observation in multivariate regression", *Journal of Applied Statistics*, vol.16, pp. 25-37, 1989

[32] D. N. Naik, "Diagnostic methods for univariate, multivariate normal data", *Handbook of Statistics*, vol.22, pp. 957-993, 2003

[33] M. S. Srivastava, D. von Rosen, "Outliers in multivariate regression model", *Journal of Multivariate Analysis*, vol.65, pp. 195-208, 1998

[34] M. Banerjee, "Cook's distance in linear longitudinal models", *Communs Stat. Theory Meth*, vol.27, pp. 2973-2983, 1998

[35] M. Banerjee, E. W. Frees, "Influence diagnostics for linear longitudinal models", *Journal of American Statistics Association*, vol.92, pp. 999-1005, 1997

[36] H. Zhu, J. G. Ibrahim, Y-Y. Chi, N. Tang, "Bayesian influence measures for joint models for longitudinal survival Data, *Biometrics*, vol.68, series.3, pp. 954-964, 2012

[37] H. Zhu, J. G. Ibrahim, H. Cho, "Perturbation, scaled Cooks distance", *The Annals of Statistics*, vol.40, series.2, pp. 785-811, 2012

[38] Y-Y. Chi, J. G. Ibrahim, "Joint models for multivariate longitudinal, multivariate survival data", *Biometrics*, vol.62, pp. 432-445, 2006

[39] A. S. Hadi, J. S. Simonoff, "Procedures for the identification of multiple outliers in linear models", *Journal of American Statistics Association*, vol.88, pp. 1264-1272, 1993

[40] R. I. Jennrich, M. D. Schluchter, "Unbalanced repeated measures models with structured covariance matrices", *Biometrics*, vol.38, pp. 967-974, 1986

[41] N. M. Laird, N. Lange, D. Stram, "Maximum likelihood computations with repeated measures application of the EM algorithm", *Journal of the American Statistical Association*, vol.82, pp. 97-105, 1987

[42] N. M. Laird, J. H. Ware, "Random effects models for longitudinal data", *Biometrics*, vol.38, pp. 963-974, 1982

[43] M. J. Lindstrom, D. M. Bates, "Newton Raphson, EM algorithms for linear mixed effects models for repeated measures data", *Journal of the American Statistical Association*, vol.83, pp. 1014-1022, 1988

[44] J. L. Schafer, R. M. Yucel, "Computational strategies for multivariate linear mixed effects models with missing values", *Journal of Computational, Graphical Statistics*, vol.11, pp. 437-457, 2002

[45] D. A. Belsley, E. Kuh, R. E. Welsch, *Regression Diagnosis: Identifying Influential Data, Source of Collinearity*, Wiley-Interscience, New York, 1980

[46] R. C. Littell, G. A. Milliken, W. W. Stroup, R. D. Wolfinger, *SAS System for Mixed Models*, SAS Institute Inc, 1996

[47] J. D. Singer, J. B. Willett, *Applied Longitudinal Data Analysis: Modeling Change, Event Occurrence*, Oxford University Press, New York, 2002

[48] G. M. Fitzmaurice, N. M. Laird, J. H. Ware, *Applied Longitudinal Data Analysis*, Wiley-Interscience, New Jersey, 2004

[49] P. J. Diggle, P. Heagerty, K-Y. Liang, S. L. Zeger, *Analysis of Longitudinal Data, 2nd Edition*, Oxford University Press, New York, 2002

[50] V. Barnett, T. Lewis, *Outliers in Statistical Data, 3rd Edition*, Wiley Series in Probability, Mathematical Statistics, 1994

[51] A. J. Lawrance, *Directions in Robust Statistics and Diagnostics, Part I*, vol.29, pp. 141-157, Springer Verlag, 1990

[52] R. M. Yucel, *R mlmmm package: fitting multivariate linear mixed-effects models with missing values*, 2007