Research Article

Statistics in Medicine

Received 5 July 2010, Accepted 9 August 2012 Published online 11 September 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5594

Diagnostics for repeated measurements in linear mixed effects models

Jungwon Mun[a*†] **and Mary J. Lindstrom**[b]

Most currently available methods for detecting discordant subjects and observations in linear mixed effects model fits adapt existing methods for single-level regression data. The most common methods are generalizations of deletion-based approaches, primarily Cook's distance. This article describes the limitations of modifications to Cook's distance and local influence, and suggests a new nondeletion subject-level method, studentized residual sum of squares (TRSS) plots. We also suggest a new observation-level deletion method that detects discordant observations as an application of TRSS plots. The proposed method provides greater information on repeated measurements by utilizing revised residuals and efficiently evaluating the effect of discordant subjects and observations on the estimation of parameters including variance components. We compare the performance of the proposed methods with current methods by using the orthodontic growth data: a longitudinal dataset with 27 subjects each observed four times. TRSS plots successfully identified discordant subjects that were missed by modified Cook's distance methods and the local influence approach. Extensions of TRSS plots are also described.
Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: Cook's distance; discordant subjects; influential observations; local influence; longitudinal data; outliers; TRSS plots; PTRSS plot

1. Introduction and motivation

Repeated measurements are commonly used in many fields such as medical research [1, 2], engineering research, linguistic research [3], and so forth. Repeated measurements have more than one experimental unit because multiple observations are taken from each subject over time (or space). Linear mixed effects (LME) models have been popular and effective models to fit repeated measurements because of their ability to model serial correlation. A great deal progress has been made on mixed effects models in the past two decades on estimation of parameters, inference of estimates, and residual analysis [4–7]. However, effective diagnostics for repeated measures are still needed.

Several diagnostic methods have been proposed for repeated measurements [8–12]. Most of these proposed methods are based on well-known tools designed for cross-sectional data; Cook's distance [13], AP_i [14], DFBETAS, DFFITS, and COVRATIO [15], each of which measures the influence of observations on estimators of coefficients emphasizing different aspects of each observation's impact. Cook's distance is one of the most popular diagnostic devices and has been modified in versatile ways in many papers. For example, [10] proposed a straightforward modification of Cook's distance at the population level by eliminating the entire data vector of one subject at a time: This approach is referred to as the subject-wise Cook's distance (SCD). [16] suggested a modification at the subject level by removing one observation at a time, which is referred to as observation-wise Cook's distance (OCD).

Modifications of Cook's distance have limited application to repeated measurements. First, as mentioned earlier, repeated measurements have an intrinsic hierarchical structure. LME models account for this structure with multiple levels of parameters: population parameters and subject-specific parameters.

[a] *Department of Mathematics and Statistics, California State Polytechnic University at Pomona, 3801 W. Temple ave., Pomona, CA 91768, U.S.A.*
[b] *Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, K6/446 Clinical Sciences Center, 600 Highland Ave. Madison, WI 53792, U.S.A.*
**Correspondence to: Jungwon Mun, Department of Mathematics and Statistics, California State Polytechnic University at Pomona 3801 W. Temple ave. Pomona, CA 91768, U.S.A.*
[†] *E-mail: stat.chris@gmail.com*

Copyright © 2012 John Wiley & Sons, Ltd.

The SCD only considers the changes in the population parameters; it does not take into account changes in the subject-specific parameters. In the subject-wise investigation, if all the data from Subject i is excluded from the model, it is unlikely that the effect of Subject i on its own subject-specific parameters will be seen. Second, the key difference between the data in single-level regression and repeated measurements is that the latter concerns the trajectories that measurements show over time. We propose that diagnostics should detect the subjects with different trajectory (discordant subjects) and the observations that contribute the discordant trajectory (discordant observations). None of the current methods, on the basis of Cook's distance, consider the discordance of trajectory properly and focus solely on the changes in coefficients. SCD often fails to detect discordant subjects, and OCD sometimes fails to detect discordant observations [12]. A discordant (or outlying) subject may have influence on the variance components, whereas it has little influence on the fixed effects. When a discordant subject affects one or more of variance components, it could make the inferential statistics less powerful. Again, the limitations of Cook's distance and its modifications come from the features of repeated measurements that cross-sectional data do not have.

Local influence is another method used to detect influential observations in single-level regression analysis [17] that has been extended to LME models [18, 19] and generalized LME models [9]. Subject-level local influence measures the changes in the log-likelihood function with changing weights for subjects. Local influence is beneficial to investigate the sources of deviation, but it has to be applied cautiously. As shown in an application to the orthodontic growth data, subject-level local influence fails to detect critical discordant subjects identified by our proposed method.

This paper introduces a new diagnostic tool that uses the sum of squares of revised residuals in LME models. Two types of deviations are considered, and they can be investigated simultaneously by the new method. In single-level regression data, influential outliers are often, although not always, far from the fitted mean. An analogous deviation for repeated measurements is defined to measure a subject-specific mean that is far from the population mean. A second deviation measures the distance of individual trajectories from one another and from the typical population trajectory. The former deviation will be referred to as 'L-type' deviation ('L' stands for **L**ocation) and the latter 'S-type' deviation ('S' for **S**hape), respectively. In addition, we describe an observation-wise application of this new diagnostic tool.

Simple visual investigation for repeated measurements is less informative than it is for single-level data and can often be misleading. For example, whereas scatter plots lose much information if observations within subject are not connected by lines, scatter plots with these connections may be too complicated to extract useful information. Trellis or lattice plots that display curves one by one or in groups in an array can be useful but will not always clearly show influential subjects. Our proposed method provides an alternative plot, called the studentized residual sum of squares (TRSS) plot, which displays the effects of subjects in an efficient and informative way.

The rest of the article is organized as follows. Section 2 provides the outline of LME models and their notations in brief and explains the challenges of the current methods (SCD, OCD, conditional Cook's distance, and local influence). Our new method and an application are introduced in Section 3 along with a simulation study and extensions of the new method. The application of TRSS and partial TRSS (PTRSS) plots to a real dataset is presented in Section 4 including a comparison with the modified Cook's distances and the local influence approach. Discordant subjects and observations can be influential on parameter estimates: not only on the coefficients but also on variance components. This issue is discussed in Sections 2.1, 3.3, and 4. Section 5 provides concluding remarks.

2. Linear mixed effects models

2.1. Model specification and Cook's distance modifications

In this article, assume that the data structure has one level of nesting, which is quite common in repeated measurements. For notational convenience, the population and the subject levels will be denoted as levels 0 and 1, respectively. An LME model for Subject i, containing n_i observations, will be represented as $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{X}_i$ is an $n_i \times p$ design matrix for fixed effects, $\boldsymbol{Z}_i$ is an $n_i \times q$ design matrix for random effects, $\boldsymbol{\beta}$ is a $p \times 1$ column vector of fixed effects, $\boldsymbol{b}_i$ is a $q \times 1$ vector of random effects following a normal distribution with zero mean and covariance matrix $\boldsymbol{D}$, $\boldsymbol{\varepsilon}_i$ is an $n_i \times 1$ vector of independent error terms following a normal distribution with zero mean and variance matrix $\sigma^2\boldsymbol{I}$, and $\boldsymbol{\varepsilon}_i$ is independent of $\boldsymbol{b}_i$. Popular estimators for fixed effects and random effects in LME models are $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{y}$ and $\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{D}}\boldsymbol{Z}_i^t\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})$, where $\boldsymbol{y} = (\boldsymbol{y}_1^t, \ldots, \boldsymbol{y}_M^t)^t$,

Copyright © 2012 John Wiley & Sons, Ltd.

$\boldsymbol{X} = (\boldsymbol{X}_1^t, \ldots, \boldsymbol{X}_M^t)^t$, $\boldsymbol{V}_i = \sigma^2\boldsymbol{I} + Z_i\boldsymbol{D}Z_i^t$, and $\boldsymbol{V} = \text{diag}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_M)$ [20]. $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator based on the marginal distribution of the data and best linear unbiased estimator having a normal distribution conditional on $\boldsymbol{D}$, whereas $\hat{\boldsymbol{b}}_i$ is the best linear unbiased predictor of $\boldsymbol{b}_i$. An efficient algorithm calculating these estimates for repeated measurements can be found in [21], which is implemented in many statistical packages.

Figure 1 depicts three different simulation datasets with intended discordant subjects and/or observations. Twenty conforming subjects are generated on the basis of the model $\boldsymbol{y}_i = \beta_0 + \beta_1\boldsymbol{x}_i + b_i + \boldsymbol{\varepsilon}_i$ with $(\beta_0, \beta_1, \sigma_b^2, \sigma^2) = (5, 2.5, 2, 0.6)$, where $i = 1, 2, \ldots, 20$ and $\boldsymbol{x}_i = (1, 2, \ldots, 9)$ for all i. In Case 1, Subject 21 is added with a discordant observation at $x_{ij} = 5$. In Case 2, Subject 21 has a wiggly response pattern, and in Case 3, two discordant subjects are added that are similar to Subject 21 in Case 1. All three examples shown in Figure 1 have 'S-type' discordant subject(s) or observation(s). Case 1 is very similar to an example used in [12] where Cook's distance did not detect discordant observation (s) properly.

The top panels of Figure 1 are scatter plots with line connection for subjects with a discordant observation(s) or discordant pattern in each case. The bottom panels of Figure 1 present modified Cook's distances in each case. SCD_i indicates SCD values without Subject i, and OCD_{ij} implies OCD values without observation j of Subject i:

$$\text{SCD}_i = \frac{1}{p}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(i)})^t\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(i)}) \quad \text{and} \quad \text{OCD}_{ij} = \frac{1}{p}(\boldsymbol{\beta} - \boldsymbol{\beta}_{i(j)})^t\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_{i(j)})$$

where, $\boldsymbol{\beta}_{(i)}$ and $\boldsymbol{\beta}_{i(j)}$ are the coefficients estimated without the entire observations from Subject i and the the coefficients estimated without the jth observation from Subject i, respectively.

The SCD for Case 1 does not detect Subject 21 and OCD for Case 1 detects the ninth observation of Subject 21 as the most influential observation instead of the fifth, which is consistent with the conclusion of [12]. The fifth observation of Subject 21 in Case 1 is not detected by any other modified methods, not because it is uninfluential, but because the points at either end of the line have more influence in determining the slope. This is the reason that OCD identified the ninth observation as the most influential observation instead of the fifth. In addition, the fifth observation is an influential observation with respect to the variance components, but SCD and OCD only focus on the fixed effects.

SCD for Case 2 does not detect Subject 21 despite its evidently different trajectory from those of other subjects. This happens because its subject-specific mean is close to the population mean. and the elimination of Subject 21 does not cause a significant change on the estimates of the fixed effects in the

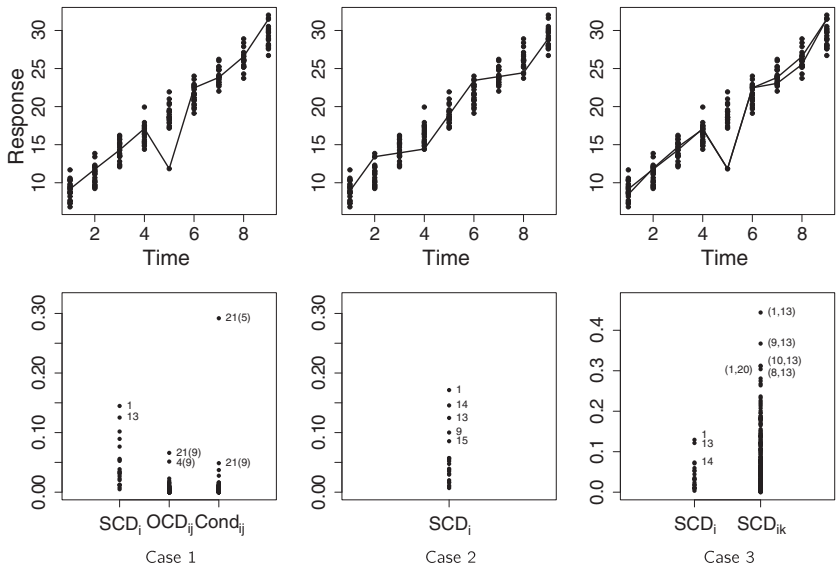

Figure 1. Plots for simulated longitudinal data with three different types of discordant subjects. The top panels show the scatter plots of each case emphasizing the intended outliers with line connection. The bottom panels illustrate modified Cook's distance in each case. SCD, OCD and the conditional Cook's distance for the Case 1, SCD for Case 2, and SCD with elimination of one subject (i) and two subjects (i and k), respectively.

Copyright © 2012 John Wiley & Sons, Ltd.

model. However, the inclusion of this discordant Subject 21 results in upwards biased estimates for the variance components. A similar phenomenon is observed with the orthodontic growth data in Section 4.

Case 3 contains two similar discordant subjects, so the elimination of one of them will not make any significant changes on parameter estimators. This phenomenon is referred to as "Masking effect" or "Joint effect" [25]. As shown in the SCD_i for Case 3 in Figure 1, SCD detects neither Subject 21 nor 22. The K-deletion (subject-wise or observation-wise) approach is often used where multiple outliers exist. The SCD_{ik} for Case 3 in Figure 1 is the SCD when two subjects i and k are omitted at the same time. However, this modification did not help in detecting Subject 21 and 22. The K-deletion approach sometimes causes the 'swamp' phenomenon, which happens in the situation where top choices of K outliers contain a common subset. The swamp phenomenon can be worse for repeated measurements with a large number of subjects or a large number of observations per subject.

[12] proposed an observation-wise investigation to take account of random effects when detecting influential observations.

$$\begin{aligned}\mathrm{Cond}_{i,j} &= \frac{1}{\sigma^2\{(M-1)q+p\}}\sum_{i=1}^{M}\left((\boldsymbol{X}_i\hat{\boldsymbol{\beta}}+\boldsymbol{Z}_i\hat{\boldsymbol{b}}_i)-(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{i(j)}+\boldsymbol{Z}_i\hat{\boldsymbol{b}}_{i(j)})\right)^t\\ &\quad\times\left((\boldsymbol{X}_i\hat{\boldsymbol{\beta}}+\boldsymbol{Z}_i\hat{\boldsymbol{b}}_i)-(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{i(j)}+\boldsymbol{Z}_i\hat{\boldsymbol{b}}_{i(j)})\right)\\ &= \frac{(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{i(j)})^t\boldsymbol{X}^t\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{i(j)})}{\sigma^2\{(M-1)q+p\}}+\frac{\sum_{i=1}^{M}(\hat{\boldsymbol{b}}_i-\hat{\boldsymbol{b}}_{i(j)})^t\boldsymbol{Z}_i^t\boldsymbol{Z}_i(\hat{\boldsymbol{b}}_i-\hat{\boldsymbol{b}}_{i(j)})}{\sigma^2\{(M-1)q+p\}}\\ &\quad+\frac{2(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{i(j)})^t\sum_{i=1}^{M}\boldsymbol{X}_i^t\boldsymbol{Z}_i(\hat{\boldsymbol{b}}_i-\hat{\boldsymbol{b}}_{i(j)})}{\sigma^2\{(M-1)q+p\}}\\ &= \mathrm{Cond1}_{i,j}+\mathrm{Cond2}_{i,j}+\mathrm{Cond3}_{i,j}\end{aligned}\tag{1}$$

where $\hat{\boldsymbol{b}}_{i(j)}$ is the predicted random effects without the jth observation of Subject i. $\mathrm{Cond}_{i,j}$ measures the influence of the jth observation of Subject i. The three terms in Equation (1), respectively, indicate a distance measure for the estimated fixed effects parameters, a distance measure for the change in the estimated subject-specific parameters (random effects), and a measure of covariation between a change in the average profile and a change in the position of the subject-specific profiles relative to the average profile. The conditional Cook's distance detects the fifth observation of Subject 21 as the most influential observation in Case 1.

2.2. *Local influence*

Local influence [17] is another diagnostic measure used to find outliers in single-level regression. It measures the influence of observations on the likelihood by assigning a weight to each observation. The author suggests investigating the local behavior of the likelihood displacement, a metric to measure local influence, by using the normal curvature. [18] extended local influence to LME models assuming independent random effects, and [19] further extended this measure in general LME context.

The authors in [19] derived the normal curvature of Subject i in the direction of the $M\times 1$ vector containing zeros everywhere except on the ith position: this measure is called 'total local influence' and denoted by C_i. As per the authors, the subjects with large values of C_i are identified as influential subjects, but C_i does not provide information on the types of deviations causing large C_i. To remedy this, the measure was decomposed into two interpretable parts, $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{D},\sigma)$: the former measures the local influence of Subject i on the fixed-effect parameters, and the latter measures the local influence on the variance components. This decomposition helps to examine the source of the deviation. The authors further decomposed the two parts to identify the essential components contributing to large values of $C_i(\boldsymbol{\beta})$ and/or $C_i(\boldsymbol{D},\sigma)$. The second-stage influence measures are

$$||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^t||,\ ||\boldsymbol{\mathcal{R}}_i||^2,\ ||\boldsymbol{\mathcal{Z}}_i\boldsymbol{\mathcal{Z}}_i^t||^2,\ ||I-\boldsymbol{\mathcal{R}}_i\boldsymbol{\mathcal{R}}_i^t||^2,\ \text{and}\ ||\boldsymbol{V}_i^{-1}||^2\tag{2}$$

where $\boldsymbol{\mathcal{X}}_i=\boldsymbol{V}_i^{-1/2}\boldsymbol{X}_i, \boldsymbol{\mathcal{Z}}_i=\boldsymbol{V}_i^{-1/2}\boldsymbol{Z}_i, \boldsymbol{\mathcal{R}}_i=\boldsymbol{V}_i^{-1/2}\boldsymbol{r}_i$, and $\boldsymbol{r}_i=\boldsymbol{y}_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}$. The decomposition of $C_i(\boldsymbol{\beta})$ only involves $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^t||$ and $||\boldsymbol{\mathcal{R}}_i||^2$, whereas that of $C_i(\boldsymbol{D},\sigma)$ only involves $||\boldsymbol{\mathcal{Z}}_i\boldsymbol{\mathcal{Z}}_i^t||^2$, $||I-\boldsymbol{\mathcal{R}}_i\boldsymbol{\mathcal{R}}_i^t||^2$, and $||\boldsymbol{V}_i^{-1}||^2$.

The authors in [19] set up a two-stage investigation of local influence with the examination of several index plots: (i) identify influential subjects on the basis of the total local influence C_i, $C_i(\boldsymbol{\beta})$ and

Copyright © 2012 John Wiley & Sons, Ltd.

$C_i(\boldsymbol{D},\sigma)$. If the values of C_i, $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{D},\sigma)$ are greater than $2\sum C_i/M$, $2\sum C_i(\boldsymbol{\beta})/M$ and $2\sum C_i(\boldsymbol{D},\sigma)/M$, respectively, the corresponding Subject i is considered as an influential subject; and (ii) investigate index plots of the five second-stage measures in (2) for the subjects identified in the first step. This allows identification of the source of deviation. Local influence does not adjust for imbalance in the data structure such as different number of observations per subjects. As the sizes of interpretable components are directly related to the n_i, the authors advised to examine the index plot of n_i in this step. This two-stage approach is proposed by the authors on the grounds that large values of the second-stage measures do not necessarily indicate the large value of the total local influence. The authors emphasize the need to examine the index plots of the second-stage measures only for the subjects identified in the first step. A counter example is shown in the application with the prostate data example in their paper: two subjects have similarly large values of $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^t||^2$, $||\boldsymbol{\mathcal{R}}_i||^2$, and $||I-\boldsymbol{\mathcal{R}}_i\boldsymbol{\mathcal{R}}_i^t||^2$, but one of them is identified as an influential subject with a large value of C_i, but the other, having a small value of C_i, is not. However, it seems that this two-stage approach can be misleading. First, [17] mentioned that the local influence might not reflect global effect even for single-level regression data. In the application of local influence for linear models to the orthodontic growth data, this approach failed to detect an influential subject. After the inspection of several simulation data cases, it seems more effective to consider all index plots together, not in a hierarchical manner. The authors in [9] extended local influence to generalized LME models. A comparison of local influence is provided with the real example in Section 4.

3. Proposed methods: TRSS plots

3.1. Nondeletion subject-wise approach

The conventional definition of residuals at the population (subject) level is defined as the difference between the response and the population (subject-specific) mean. In this article, we use revised residuals that decompose conventional residuals at level 0 as

$$\underbrace{(\text{response}-\text{subject specific mean})}_{\boldsymbol{e}_{i,1}}+\underbrace{(\text{subject specific mean}-\text{population mean})}_{\boldsymbol{e}_{i,0}}.$$

As described in Section 2.1, the response vector for Subject i is modeled as $\boldsymbol{y}_i=\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{b}_i+\boldsymbol{\varepsilon}_i$, and the population mean and subject specific mean are $\boldsymbol{X}_i\boldsymbol{\beta}$ and $\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{b}_i$, respectively. If $\boldsymbol{\eta}_i=[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^t,\boldsymbol{b}_i^t,\boldsymbol{\varepsilon}_i^t]^t$ and $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of the fixed effects, then $\boldsymbol{\eta}_i$ follows a multivariate normal distribution with zero mean on the basis of the model assumptions and the properties of $\hat{\boldsymbol{\beta}}$. This allows us to reexpress $\boldsymbol{e}_{i,1}$ and $\boldsymbol{e}_{i,0}$ as linear forms of the normally distributed vector $\boldsymbol{\eta}_i$.

$$\boldsymbol{e}_{i,1}=(\boldsymbol{I}_i-\boldsymbol{K}_i)\,[-\boldsymbol{X}_i\ \ \boldsymbol{Z}_i\ \ \boldsymbol{I}_i]\,\boldsymbol{\eta}_i \text{ and } \boldsymbol{e}_{i,0}=\boldsymbol{K}_i\,[-\boldsymbol{X}_i\ \ \boldsymbol{Z}_i\ \ \boldsymbol{I}_i]\,\boldsymbol{\eta}_i$$

where $\boldsymbol{K}_i=\boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i^t\boldsymbol{V}_i^{-1}$.

The variance–covariance matrix of $\boldsymbol{\eta}_i$ is also obtained from the model assumptions and the properties of the parameter estimators. If $\boldsymbol{H}_i$ is defined as $(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_i^t\boldsymbol{V}_i^{-1}$, the covariance matrix of $\boldsymbol{\eta}_i$ is

$$\text{Cov}(\boldsymbol{\eta}_i)=\begin{pmatrix}\text{Cov}(\hat{\boldsymbol{\beta}}) & \boldsymbol{H}_i\boldsymbol{Z}_i\boldsymbol{D} & \sigma^2\boldsymbol{H}_i\\ \boldsymbol{D}\boldsymbol{Z}_i^t\boldsymbol{H}_i^t & \boldsymbol{D} & \boldsymbol{0}\\ \sigma^2\boldsymbol{H}_i^t & \boldsymbol{0} & \sigma^2\boldsymbol{I}_i\end{pmatrix}\equiv\boldsymbol{T}_i \tag{3}$$

For unknown σ^2 and $\boldsymbol{D}$, we use their estimators in $\boldsymbol{T}_i$. The residual sum of squares (RSS) at levels 0 and 1 are defined as $\text{RSS}_{i,0}=\boldsymbol{e}_{i,0}^t\boldsymbol{e}_{i,0}$ and $\text{RSS}_{i,1}=\boldsymbol{e}_{i,1}^t\boldsymbol{e}_{i,1}$, respectively. Intuitively, $\text{RSS}_{i,0}$'s ($\text{RSS}_{i,1}$'s) contain the information on 'L-type' ('S-type') deviation. $\text{RSS}_{i,0}$ and $\text{RSS}_{i,1}$ still depend on the unit of measurement and the number of measurements per subject. This must be accounted for in cases where the data are not balanced in terms of the number of measurements per individual.

Note that $\text{RSS}_{i,1}$ and $\text{RSS}_{i,0}$ are quadratic forms of a normally distributed vector $\boldsymbol{\eta}_i$. Define $\boldsymbol{e}_{i,1}=\boldsymbol{Q}_{i,1}\boldsymbol{\eta}_i$ and $\boldsymbol{e}_{i,0}=\boldsymbol{Q}_{i,0}\boldsymbol{\eta}_i$, where $\boldsymbol{Q}_{i,1}=(\boldsymbol{I}-\boldsymbol{K}_i)[-\boldsymbol{X}_i\ \ \boldsymbol{Z}_i\ \ \boldsymbol{I}]$ and $\boldsymbol{Q}_{i,0}=\boldsymbol{K}_i[-\boldsymbol{X}_i\ \ \boldsymbol{Z}_i\ \ \boldsymbol{I}_i]$. Because $\boldsymbol{\eta}_i\overset{\sim}{}N(\boldsymbol{0},\boldsymbol{T}_i)$, the expectation and variance of $\text{RSS}_{i,1}$ and $\text{RSS}_{i,0}$ can be derived by properties of quadratic forms: $E(\boldsymbol{\eta}_i^t\boldsymbol{A}\boldsymbol{\eta}_i)=tr(\boldsymbol{A}\boldsymbol{T}_i)$ and $\text{Var}(\boldsymbol{\eta}_i^t\boldsymbol{A}\boldsymbol{\eta}_i)=2\text{tr}(\boldsymbol{A}\boldsymbol{T}_i\boldsymbol{A}\boldsymbol{T}_i)$, where $\boldsymbol{A}=\boldsymbol{Q}_{i,1}^t\boldsymbol{Q}_{i,1}$

Copyright © 2012 John Wiley & Sons, Ltd.

for $\mathrm{RSS}_{i,1}$ and $\boldsymbol{A} = \boldsymbol{Q}_{i,0}^t \boldsymbol{Q}_{i,0}$ for $\mathrm{RSS}_{i,0}$. Studentized measures of $\mathrm{RSS}_{i,1}$ and $\mathrm{RSS}_{i,0}$, $\mathrm{TRSS}_{i,1}^*$ and $\mathrm{TRSS}_{i,0}^*$, are defined as

$$\mathrm{TRSS}_{i,1}^* = \frac{\mathrm{RSS}_{i,1} - \mathrm{E}(\mathrm{RSS}_{i,1})}{\sqrt{\mathrm{Var}(\mathrm{RSS}_{i,1})}} \quad \text{and} \quad \mathrm{TRSS}_{i,0}^* = \frac{\mathrm{RSS}_{i,0} - \mathrm{E}(\mathrm{RSS}_{i,0})}{\sqrt{\mathrm{Var}(\mathrm{RSS}_{i,0})}}$$

This procedure of studentization makes $\mathrm{TRSS}_{i,1}^*$ and $\mathrm{TRSS}_{i,0}^*$ unit free and handles the situation of unequal n_i's automatically. The deviations in positive direction are of greater concern than those in negative direction. Residual sum of squares smaller than its expectation simply means that the given model fits better for that subject than others or the corresponding subject is less noisy than others. Hence, only large positive values of $\mathrm{TRSS}_{i,1}^*$ and $\mathrm{TRSS}_{i,0}^*$ are of interest, so $\mathrm{TRSS}_{i,1}$ and $\mathrm{TRSS}_{i,0}$ are truncated at zero: $\mathrm{TRSS}_{i,1} = \max\{0, \mathrm{TRSS}_{i,1}^*\}$ and $\mathrm{TRSS}_{i,0} = \max\{0, \mathrm{TRSS}_{i,0}^*\}$. The TRSS plot is the scatter plot of $\mathrm{TRSS}_{i,1}$ and $\mathrm{TRSS}_{i,0}$ and shows discordant subject(s) and their type(s) of deviation. Large $\mathrm{TRSS}_{i,0}$ compared with small $\mathrm{TRSS}_{i,1}$ indicates that Subject i is far from the marginal mean ('L-type'). In contrast, large $\mathrm{TRSS}_{i,1}$ compared with small $\mathrm{TRSS}_{i,0}$ suggests that the corresponding subject may have different trajectory from others ('S-type') or have a different correlation structure from others.

The TRSS plots show both types of deviations simultaneously and allow subject-wise inspection of the data without deleting the entire observation vector of each subject. In other words, TRSS is a non-deletion subject-wise deviation measure, whereas most other proposed statistics are deletion approach methods. If there is more than one discordant subject, TRSS plots enable us to detect any clustering of them visually. Even when one prefers to use other statistics based on the deletion approach, TRSS plots can be used to suggest the proper size of K in the K-deletion method.

3.2. *Reference lines*

Discordant subjects and their types of deviations are determined by their distance from the origin and their direction in TRSS plots. It is beneficial to have a reference line to judge if a subject can be considered as discordant or not. We established three plausible reference lines on the basis of the assumption that the TRSS points form a truncated bivariate normal. We use the highest probability density (HPD), the local quantile (LQ), and the rotated quantile regression (RQR) and compare their performances in simulation studies (detailed simulation results not shown). As shown in Figure 2, the HPD finds a cutoff value and an ellipsoid on the first quadrant to meet the nominal coverage probability. LQ sets up an angle from the origin and finds the quantiles of the data at the nominal level in term of the distance from the origin. RQR rotates the truncated bivariate points by 45° and attains a nonparametric quantile regression line at the given nominal level. This line of quantiles is rotated back and forms the RQR reference line. These three reference lines emphasize different aspects in direction and strength of the discordance.

Among these three reference lines, simulation shows that RQR approach achieves the nominal coverage probability more often than the other two. As shown in plot (a) of Figure 2, HPD approach generates a convex type line and gives equally likely significance to points on the plane and ones at the axes. The

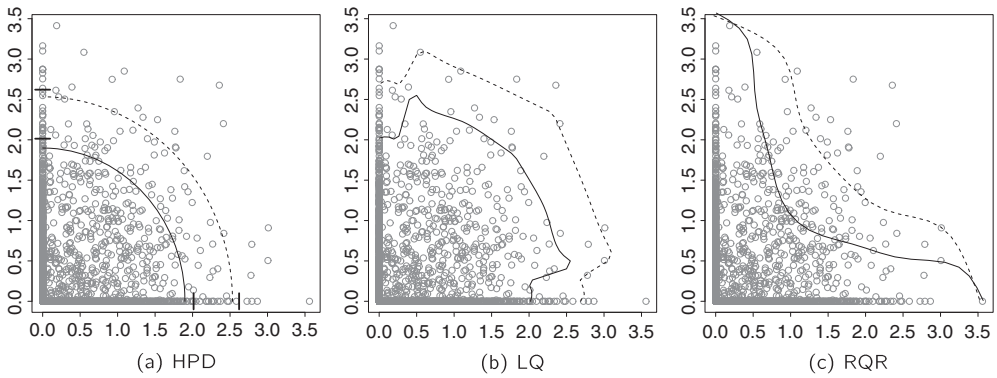

Figure 2. Reference lines: three different reference lines are presented. In each plot, a solid (dotted) line is a suggested reference line at 99% (95%) level. (a) is the reference line based on the highest probability density (HPD) approach. (b) and (c) present the reference lines based on local quantile (LQ) and rotated quantile regression (RQR) approaches, respectively.

Copyright © 2012 John Wiley & Sons, Ltd.

LQ approach in Figure 2(b) also builds a convex type line giving greater scrutiny to the points at the axes. The RQR approach in Figure 2(c) produces a concave type line and assigns more attention to the subjects that deviate moderately in both L-direction and S-direction (points on the interior of the quadrant) than those that deviate in one direction (points at the axes).

Reference values for diagnostic tools are not always provided even for single-level regression models. For example, the original Cook's distance does not have a clear cutoff criteria, neither do the modified Cook's distance. A few researchers proposed cutoff values [22], but there are no dominant choices among statisticians. Statistical practitioners have paid attention to observations with relatively large values of diagnostics measures such as Cook's distance, AP_i, DFBETA, and DFFITS. Similarly, the local influence has crude cutoff values for C_i, $C_i(\boldsymbol{\beta})$, and $C_i(\boldsymbol{D}, \sigma)$ to determine influential subjects. These reference values are obtained by approximation for large M, and there are no clear reference values for the second-stage measures and is only advised to examine the subjects with relatively large values of the second-stage measures. The three reference lines for TRSS points are ad hoc approach for a practitioner's convenience. The judgment based on a reference line is not definite. We advise users to choose a proper reference line or value in the context of the data and take a close look at subjects with isolated TRSS values. From our experience, it is worthwhile to pay more attention to isolated subjects in TRSS plots and subjects with $TRSS_{i,1}$ and/or $TRSS_{i,0}$ greater than $2 \sim 3$. The RQR reference lines will be used in the examples in the remainder of this manuscript (R package for TRSS plot provides the option for reference lines for those who prefer the HPD or the LQ approaches.) The 95% (99%) RQR reference lines are marked with sold (dotted) lines in the TRSS plots presented in the subsequent sections.

3.3. Discordant subjects in simulation datasets and interpretation

This section examines the three simulation cases defined in Section 2.1 for comparison of the modified Cook's distances and the TRSS method. Table I lists the changes of the coefficients ($\hat{\beta}_0$ and $\hat{\beta}_1$) and the variance components ($\hat{\sigma}^2$ and $\hat{\sigma}_b^2$) in Cases 1–3. The second column shows the estimated parameters and variances of error terms and random intercept with the concordant 20 subjects, that is, without Subject 21 (and 22). The remaining columns show the fixed and variance parameters of the model with intended discordant subject(s). As shown in Table I, fixed effect parameters β_0 and β_1 hardly change with and without intended discordant subject(s), whereas variance components σ and σ_b change in a much wider range, especially σ. This explains the reason SCD failed to detect Subject 21 (and 22) in all three cases.

All TRSS points will lie inside the reference line if there are no significant outliers (discordant subjects or discordant observations). This is observed in Figure 3(a) that presents the TRSS plot only with 20 conforming subjects. As set up in the simulation, 20 random subjects generated from the given model have small $TRSS_{i,1}$. There are a few subjects with large $TRSS_{i,0}$ compared with others, which result from the large values of their random effects. Random effects can be associated with both L-type and S-type deviation, and the discordant subjects that affect the population parameters are often associated with S-type.

In Figure 3(b)–(d), TRSS plots accurately identify intended discordant subjects in all cases implying the source of deviation, and the TRSS plot for Case 3 indicates that Subjects 21 and 22 behave similarly as they form a cluster in the plot.

Table I. Changes of the parameters with and without a designated discordant subject(s).

Parameter	With 20 subjects	Case 1	Case 2	Case 3
β_0	6.4122	6.3714	6.4439	6.3346
β_1	2.5039	2.5101	2.4976	2.5136
σ^2	0.6103	0.9291	0.6766	1.2253
σ_b^2	1.0832	0.9923	1.0183	0.9154

The second column lists the estimates of the respective fixed effects and variance components of the linear mixed effects model with 20 ordinary subjects. The remaining columns show the fixed effects and variance components of the model including the intended discordant subject(s).

Copyright © 2012 John Wiley & Sons, Ltd.

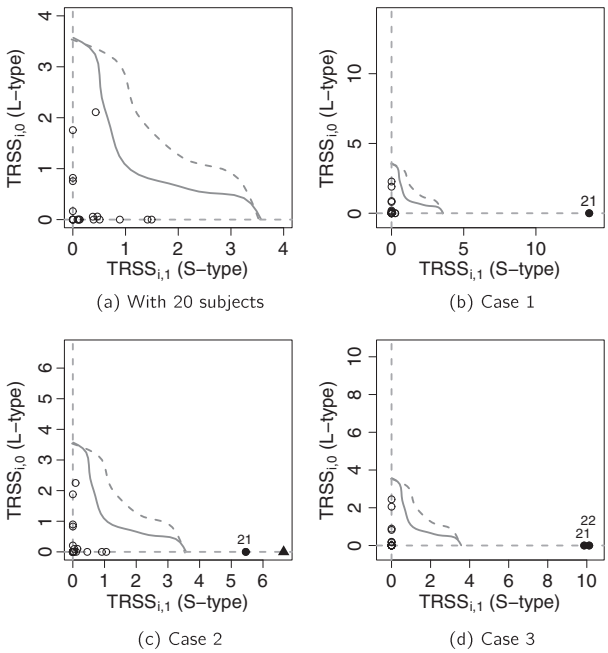

Figure 3. Studentized residual sum of squares (TRSS) plots for simulated longitudinal data in three different situations. (a) illustrates the TRSS plots with the conforming 20 subjects. Plot (b)–(d) depict TRSS plots for the simulation data of Cases 1–3, respectively. The intended discordant subjects are marked as black circles, and they all are outside the reference lines in their TRSS plot.

In conclusion, both modified methods of Cook's distance and TRSS plots are effective to detect influential subjects (observations) that have influences on fixed effects. However, modified Cook's methods have limitation on detecting subjects (observations) that affect random effects and variance components.

The degree of discordance that would be detected in TRSS plots has no straightforward answer because it would be affected by many factors, such as the number of subjects, the number of observations per subject, and the degree of variation in trajectory across subjects. Case 1 is an exaggerated example. It is presented to compare the performance of conditional Cook's distance as set up similarly to the example in [12]. Because the deviation of the fifth observation is obvious, it can be easily seen. As stated previously, simple visual detection becomes much less informative for repeated measurements, and Case 2 is presented to demonstrate the limitation. If the level of deviation in Case 1 is weaker, say Case 1B, the discordant subject (observation) is similar to that in Case 1 in terms of its pattern, but it is not easily detected by visual detection with or without line connection (see Figure 4). Subject 21 in Case 1B is still detected as a discordant subject, but its S-type deviation is less evident than the deviation of Subject 21 in Case 1 (Figure 4(a)) . With fewer observations per subject, the degree of deviation for Subject 21 is different in the TRSS plot (Figure 4(b)), even though the discordant drop at $x = 5$ remains the same as in Case 1.

3.4. Extensions of studentized residual sum of squares approach

This section further investigates alleged discordant subject(s) from TRSS plots. The effect of single observation from a suspicious subject can be assessed by observing the changes in the $\text{TRSS}_{i,1}$, $\text{TRSS}_{i,0}$ plots with and without the observation under consideration. The TRSS points without one or more observations are called the PTRSS points, and the PTRSS points without the kth observation of Subject i, x_{ik}, is denoted as $\text{TRSS}_{i(k),1}$ and $\text{TRSS}_{i(k),0}$. These PTRSS values are helpful to probe the effect of each measurement in a particular discordant subject. An observation can be thought of as an 'S-type' discordant observation, if its $\text{TRSS}_{i(k),1}$ is considerably less than $\text{TRSS}_{i,1}$. Examining the value of $(\text{TRSS}_{i(k),1} - \text{TRSS}_{i,1})$ detects the fifth observation of Subject 21 in Case 1 as the most discordant observation.

Independent error terms and one level of nesting group have been assumed so far in the development of the TRSS plot. Without much difficulty, TRSS values are extended for multilevel models and correlated

Copyright © 2012 John Wiley & Sons, Ltd.

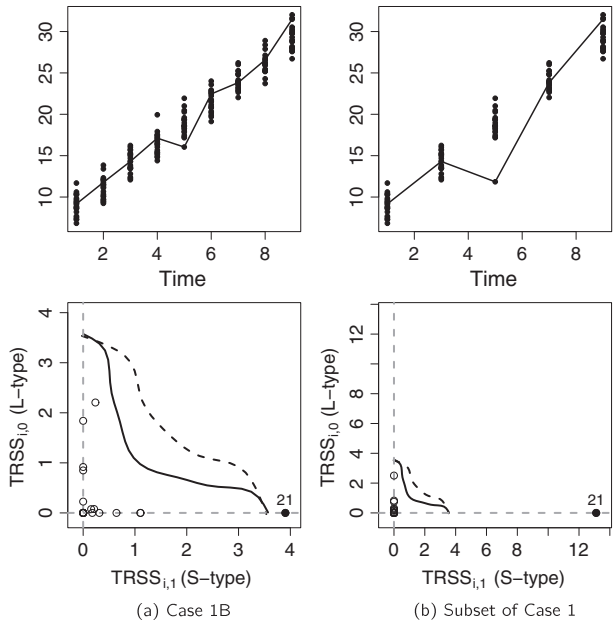

Figure 4. Case 1B illustrates a similar case to Case 1 with a weaker discordant pattern. The left panels show the scatter plot and studentized residual sum of squares (TRSS) plot of Case 1B. The right panels show the scatter plot and TRSS plot for a subset of Case 1 using only five of the original nine observations (observations at $x = 1, 3, 5, 7,$ and 9).

error terms including heteroskedasticity. For a multilevel model, say a model with three levels, $\text{TRSS}_{i,3}$, $\text{TRSS}_{i,2}$, $\text{TRSS}_{i,1}$, and $\text{TRSS}_{i,0}$ (TRSS values at each level), can be derived in a similar fashion. It is, of course, not easy to inspect all four components at the same time, but they can be examined pairwise, and similar interpretation can be made with caution. For correlated error terms, say $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Lambda}_i$, TRSS plots are derived by replacing the elements of $\sigma^2\boldsymbol{H}_i$ and $\sigma^2\boldsymbol{I}_i$ in Equation (3) with $\boldsymbol{H}_i\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Lambda}_i$, respectively.

It is noticed that a discordant subject(s) may affect $\text{Var}(\hat{\boldsymbol{\beta}})$ or $\text{Var}(\boldsymbol{\varepsilon}_i)$, which are used in the derivation of TRSS values. This influences the expectation and variance of $\text{RSS}_{i,1}$ and $\text{RSS}_{i,0}$. It is similar to the issue between the internally studentized residuals and externally studentized residuals in single-level regression. The suggested TRSS measures in this paper are a type of internally studentized RSS. An analogous externally studentized RSS (ETRSS) can be designed for repeated measurements. A naive approach for generating the ETRSS values for Subject i, $\text{ETRSS}_{i,1}$ and $\text{ETRSS}_{i,0}$, is to use the relationships

$$\text{ETRSS}^*_{i,1} = \frac{\text{RSS}_{i,1} - \text{MERSS}^{(i)}_1}{\sqrt{\text{MVRSS}^{(i)}_1}} \quad \text{and} \quad \text{ETRSS}^*_{i,0} = \frac{\text{RSS}_{i,0} - \text{MERSS}^{(i)}_0}{\sqrt{\text{MVRSS}^{(i)}_0}}$$

$$\text{ETRSS}_{i,1} = \max\{0, \text{ETRSS}^*_{i,1}\} \quad \text{and} \quad \text{ETRSS}_{i,0} = \max\{0, \text{ETRSS}^*_{i,0}\}$$

where $\text{MERSS}^{(i)}_L$ and $\text{MVRSS}^{(i)}_L$ are the averages of $\text{E}(\text{RSS}^{(i)}_{k,L})$ and $\text{Var}(\text{RSS}^{(i)}_{k,L})$, respectively, for $L = 0, 1$, $k = 1, \ldots, M$ and $k \neq i$. $\text{RSS}^{(i)}_{k,L}$ is $\text{RSS}_{k,L}$ for Subject k from the model without Subject i; hence, $\text{MERSS}^{(i)}_L = \frac{1}{(M-1)}\sum^{M}_{k\neq i}\text{E}(\text{RSS}^{(i)}_{k,L})$. If Subject i is an 'S-type' ('L-type') discordant subject, this results in smaller $\text{MERSS}^{(i)}_1$ and $\text{MVRSS}^{(i)}_1$ ($\text{MERSS}^{(i)}_0$ and $\text{MVRSS}^{(i)}_0$) than $\text{E}(\text{RSS}_{i,1})$ and $\text{Var}(\text{RSS}_{i,1})$ ($\text{E}(\text{RSS}_{i,0})$ and $\text{Var}(\text{RSS}_{i,0})$), respectively. As a consequence, it would elucidate the distinction between Subject i and the others in the TRSS plot.

This naive ETRSS point ($\text{ETRSS}_{i,1}$, $\text{ETRSS}_{i,0}$) for Subject 21 in Case 2 is presented as a triangle in the TRSS plots of Figure 3. It is noticeable that the ETRSS point for Subject 21 is a bit farther from the original TRSS and illustrates a stronger 'S-type' deviation. Therefore, if the level of deviation is

Copyright © 2012 John Wiley & Sons, Ltd.

moderate, a type of external studentization would perform better. The suggested crude approach is computationally expensive compared with the internal approach. A more formal derivation of the external studentization needs to be investigated further.

Furthermore, depending on the context of the research, different modifications can be considered for $\mathrm{TRSS}^*_{i,0}$ and $\mathrm{TRSS}^*_{i,1}$. In this paper, we do not consider a subject with less noise as discordant, and the truncation at zero for both directions is taken to obtain TRSS values. If researchers consider such subjects as discordant and want to identify them, TRSS values can be taken as the absolute value of $\mathrm{TRSS}^*_{i,0}$ and/or $\mathrm{TRSS}^*_{i,1}$. To comply with the purpose of the study, a different version of TRSS plots can be designed with a proper modification (including using the original values of $\mathrm{TRSS}^*_{i,0}$ and/or $\mathrm{TRSS}^*_{i,1}$). In these cases, the interpretation will differ from those presented in this manuscript and must be made with care.

4. Orthodontic growth data

In this section, we investigate the orthodontic growth data reported in [23] and available in the nlme R package. The distance between the pituitary and the pterygomaxillary fissure was measured every 2 years from 27 children (16 males and 11 females) from ages 8 to 14. Those two points are easily identified on X-ray exposures of the side of the head. A model

$$\mathrm{Distance}_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})(\mathrm{Age}_{ij} - 11) + \beta_2 \mathrm{Sex}_i + \beta_3 (\mathrm{Age}_{ij} - 11) \times \mathrm{Sex}_i \qquad (4)$$

is fitted as suggested in [24]. Figure 5 shows select subjects in need of careful attention. The population mean and subject-specific means are illustrated with dotted and solid lines, respectively.

Figure 6 displays the performances of SCD, OCD, and conditional Cook's distance for the orthodontic growth dataset. The first two clusters are SCD and OCD, and the third cluster is the conditional Cook's distance of each observation. The three clusters on the right show the components of the conditional Cook's distance for Subject i: the sum of these three clusters becomes the conditional Cook's distance as suggested in (1). For the observation-wise investigation, the corresponding ages are marked in parenthesis.

The pattern of four observations in M09 deviates severely from the linear alignment compared with the other subjects, and the deviation should be identified. However, similar to Subject 21 in Case 2, the subject-specific mean for Subject M09 appears close to the population mean for males, and this makes

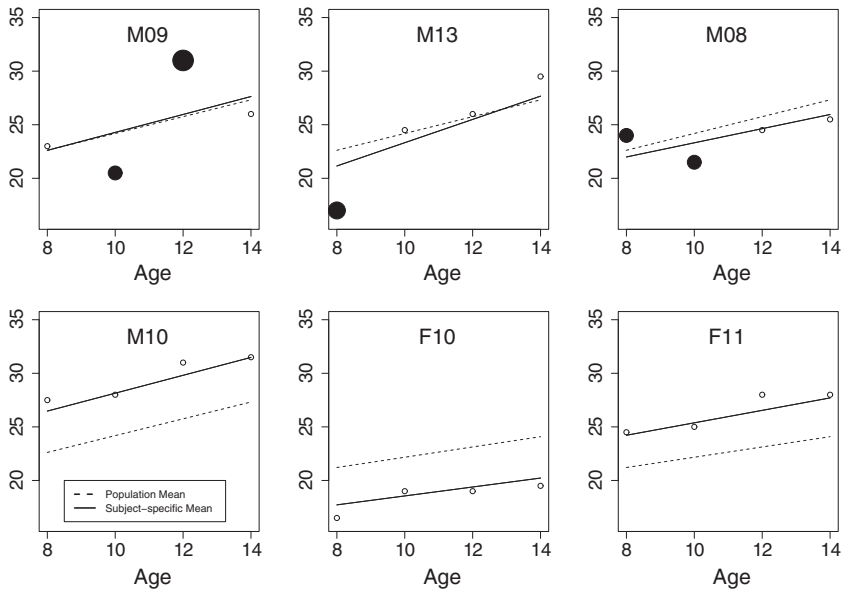

Figure 5. Subset of the orthodontic growth data: the six select subjects that are identified by SCD, OCD, conditional Cook's distance, and/or TRSS plots. The filled circles are the discordant observations identified by the PTRSS plot and their sizes correspond to the sizes of the effects on their $\mathrm{TRSS}_{i,1}$'s.

Copyright © 2012 John Wiley & Sons, Ltd.

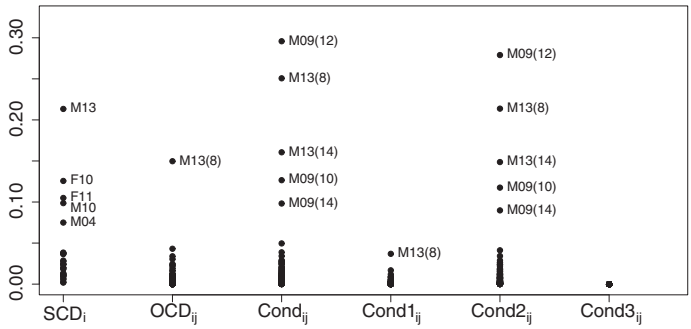

Figure 6. Modified Cook's distance methods for the orthodontic growth data: the first two clusters are SCD and OCD and the third cluster is the conditional Cook's distance of each observation. The three clusters on the right show the components of conditional Cook's distance for Subject i. Labels indicate corresponding subject IDs and ages within parenthesis.

SCD fail to identify M09. It is because the absence of M09 does not cause a serious change on the fixed effects and SCD is not designed to consider the patterns of individual subjects. Subject M09 may not have a significant effect on the fixed effects, but it affects the variance components because of its discordant trajectory. For instance, the estimate of the standard deviation of error terms (σ) is 1.31 and 0.98 with and without M09, respectively. Instead of M09, SCD detects M13 as the most influential subject, as its discordant trajectory due to a relatively small observation at age 8 results in a significant change on the slope of the subject-specific mean. Subjects F10, F11, and M10 form a second cluster, next to Subject M13. The responses of these three subjects seem to be linear, but their subject-specific means are a bit far from the population mean. OCD only identifies the measurement of M13 at age 8 as the most influential observation. Unfortunately, any measurements of M09 are not identified as an unusual observations by OCD.

Conditional Cook's distance for LME models improves on the diagnostic performance of OCD. For the orthodontic growth data, it indicates the observation of M09 at age 12, M13 at age 8 and 14, and M09 at age 10 as discordant measurements in the order. From Figure 6, it seems that all observations detected by conditional Cook's distance are mainly based on the changes of random effects, the second component of the measure. Conditional Cook's distance did not detect any points of F10, F11, or M10 as unusual because their deviation does not result from a single observation but from the overall effect of their trajectories.

Figure 7 displays local influence diagnostics with several select index plots. Because the orthodontic growth data are balanced, $||\boldsymbol{\mathcal{X}}_i \boldsymbol{\mathcal{X}}_i^t||^2$, $||\boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^t||^2$, $||\mathrm{V}_i^{-1}||$, and n_i are the same for all subjects and their index plots are not shown. According to the guideline in [19], M10, F10, and F11 are identified as influential subjects from the index plots of the total local influence with the values over the suggested criterion, $2\sum_i^M C_i/M$. In Figure 7, the subjects identified from TRSS plots are marked with filled circles (L-type) and diamonds (S-type). The size of points is proportional to the level of deviation examined from TRSS plots. Interestingly, M09 is not detected by the index plots of C_i, $C_i(\boldsymbol{\beta})$, or $C_i(\boldsymbol{D}, \sigma)$, but M09 has large values of both $||\mathcal{R}||^2$ and $||I - \mathcal{R}\mathcal{R}^t||^2$. Nevertheless, it is hard to say M09 is either a discordant or influential subjects on the basis of the investigation procedure [19] suggested. Following a similar argument in [19], M09 becomes an example for the case where large values of the second-stage measures do not necessarily indicate large values of the total local influence, C_i. Subject M13 is identified from SCD as the most influence subject, and it is identified by $C_i(\boldsymbol{\beta})$ (not by the total local influence, C_i). Note that the equation used to obtain the local influence measure is the same as (4) without the centering.

The TRSS plots for the orthodontic growth data are displayed in Figure 8. Figure 8(a) is the TRSS plots for the given model using all data. Because of the large $\mathrm{TRSS}_{i,1}$ value of M09, TRSS values for other subjects are squeezed making it difficult to see in detail. A TRSS plot from the same LME model without Subject M09 is provided in Figure 8(b). As mentioned earlier, the trajectory of M09 is quite different from the others, but its subject-specific mean is close to the population mean. This results in the small $\mathrm{TRSS}_{i,0}$ value and the large $\mathrm{TRSS}_{i,1}$ value for the subject in Figure 8(a). Also, because of the features of Subjects M10 and F10 described earlier, they have a small $\mathrm{TRSS}_{i,1}$ value and a large $\mathrm{TRSS}_{i,0}$ value. For Subject M13, the observation at age 8 appears to deviate from the linear pattern causing its $\mathrm{TRSS}_{i,1}$ to be large but its level of 'S-type' deviation is weaker than that of Subject M09.

Copyright © 2012 John Wiley & Sons, Ltd.

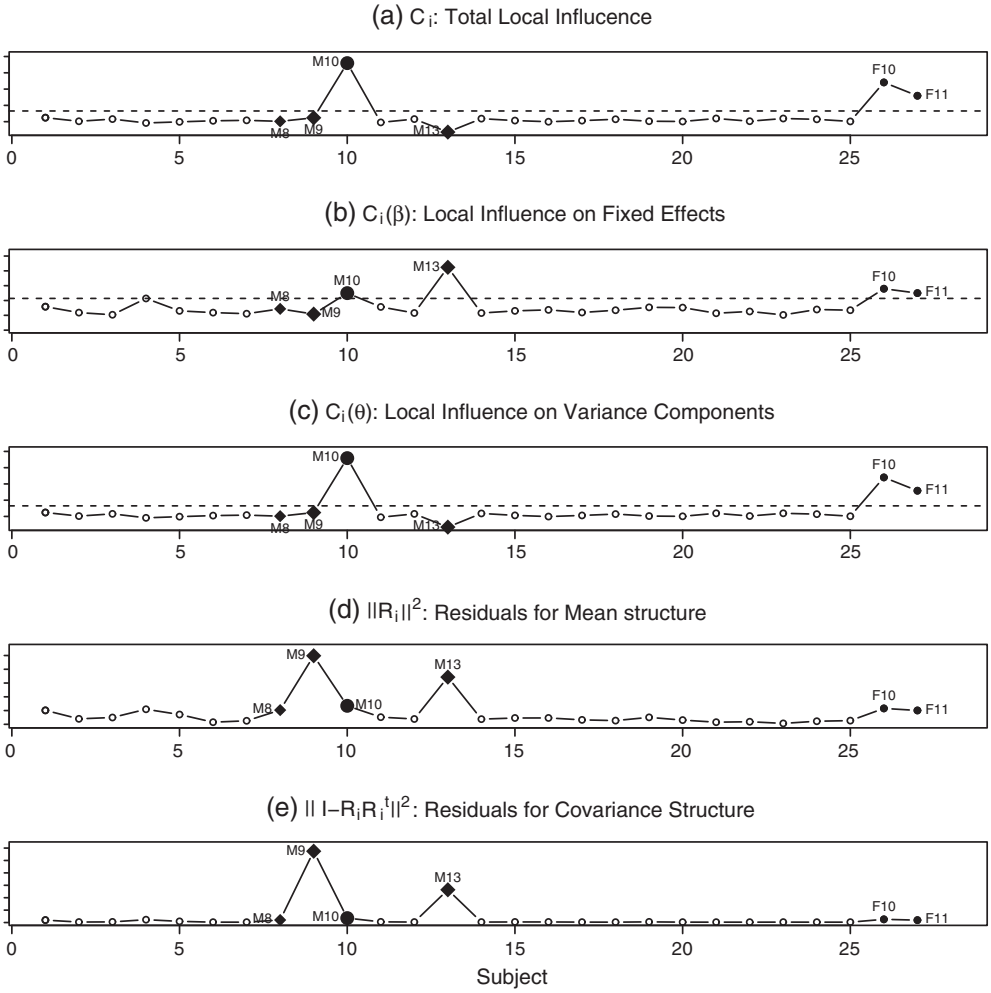

Figure 7. Local influence methods for the orthodontic growth data: five index plots of (a) the total local influence, C_i , (b) the local influence on the fixed effects, $C_i(\boldsymbol{\beta})$, (c) the local influence on the variance components, $C_i(\boldsymbol{V})$, (d) $||\boldsymbol{\mathcal{R}}_i||^2$, and (e) $||I - \boldsymbol{\mathcal{R}}_i\boldsymbol{\mathcal{R}}_i^t||^2$. The subjects identified from TRSS plots are marked with filled circles (L-type) and diamonds (S-type). The size of points are proportional to the level of deviation measured from TRSS plots.

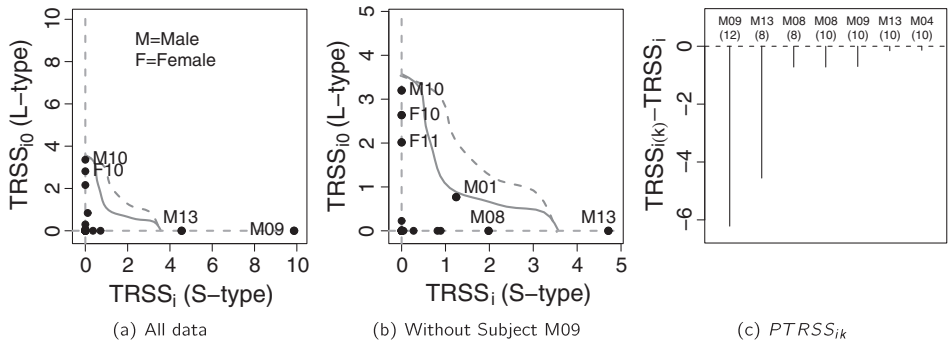

Figure 8. TRSS and PTRSS plots for the orthodontic growth data: plots (a) and (b) show the TRSS plot for the entire dataset and the data without the subject M09, respectively. Plot (c) is the PTRSS plot with the top seven differences between $\text{TRSS}_{i(k),1}$ and $\text{TRSS}_{i,1}$.

The measurement of Subject M08 at age 8 also contributes to its relatively large $\text{TRSS}_{i,1}$ in Figure 8(b). The TRSS plot without subject M09 indicates that Subject M08 has relatively large $\text{TRSS}_{i,1}$. The TRSS plot detects Subjects M09 and M13 as 'S-type' discordant subjects and M10, F10, and F11 as possible 'L-type' discordant subjects. Subject M08 and M01 may need further attention as well.

Copyright © 2012 John Wiley & Sons, Ltd.

Table II. Subjects identified from subject-wise investigation and observations identified from observation-wise investigation.

(a) Subject-wise investigation				(b) Observation-wise investigation		
Subject	SCD	Local influence	TRSS	OCD	Conditional Cook's	PTRSS
M10	**	***	** (L-type)	M13(8)	M09(12)	M09(12)
F10	**	**	** (L-type)		M13(8)	M13(8)
F11	**	**	** (L-type)		M13(14)	M08(8)
M09			*** (S-type)		M09(10)	M08(10)
M13	***	*	*** (S-type)		M09(14)	M09(10)
M08			* (S-type)			
M04	**					

(a) lists the subjects identified from subject-wise investigation, and the number of asterisks indicate the degree of deviation: three asterisk marks (***) for severe degree of deviation and one (*) for relatively weak deviation. (b) lists the observations identified from observation-wise investigations: the observations with isolated values in each measure. The discordant observations are listed in the order that they are detected by each method. SCD, subject-wise Cook's distance; TRSS, studentized residual sum of squares; OCD, observation-wise Cook's distance; PTRSS, partial studentized residual sum of squares.

Figure 8(c) is a PTRSS plot of the top 7 discordant observations as measured by the changes in $\text{TRSS}_{i,1}$ value after deletion. The y-axis is the values of $(\text{TRSS}_{i(k),1} - \text{TRSS}_{i,1})$, and it visualizes the effect of the specific observation on the model. The observation of Subject M09 at age 12 has the biggest effect on $\text{TRSS}_{i,1}$ values and on the fitted model. The observations with large changes in the PTRSS plot are indicated by filled circles in Figure 5.

Table II(a) displays a comparison of the discordant or influential subjects and observations identified by each subject-wise and observation-wise methods. All three subject-wise methods (SCD, local influence, and TRSS) detect Subject M10, F10, and F11 because of their large effect on the predicted random intercepts. The absolute value of random intercepts for these three subject is the three largest, and this is revealed in the TRSS plots as relatively large deviations in L-type direction. Subject M13 is also identified by all three methods, but subject-wise local influence barely identified it when following the two-stage approach. TRSS is the only diagnostic that distinguishes the discordant pattern of Subject M09. For observation-wise investigation presented in Table II(b), all three approaches are able to identify the observation of M13 at age 8: the conditional Cook's distance and PTRSS find the observation of M13 at age 8 as the second next to the observation of M09 at age 12. It is noticed that PTRSS is the only measure that identifies discordant observations in M08. In addition, among the top 5 choices of the conditional Cook's distance, three observations are from M09 and two from M13. Conditional Cook's distance and PTRSS agree with seriously discordant observations, but they produce somewhat different results for second tiers in their choices. This is no wonder as they measure different aspect of changes: the changes of TRSS values with and without the observation under consideration (PTRSS) and the influence on the fixed and random effects (conditional Cook's distance).

5. Conclusion

The TRSS plot is proposed to detect discordant subjects and observations for repeated measurements fit with LME models. The adoption of TRSS diagnostic tools does not alter the methods used for single-level regression analysis, and it naturally incorporates the features of repeated measurements. Hence, it overcomes or does not have the limitations and drawbacks of modified diagnostic approaches from single-level regression models. TRSS points are obtained by the reexpression of revised residuals into linear combinations of normal vectors and the studentization of their sum of squares. The new method derives two quantities that are unit free and contain two different aspects of information relevant for repeated measures models. The scatter plot of those two quantities enables us to examine the data in two dimensions at once and promise to reveal more information about the the shape and location of longitudinal observations of a subject. Discordant subjects either in the S-direction or L-direction are influential by affecting fixed parameters and/or variance components. As TRSS points indirectly measure the influence of subjects on fixed parameters and variance components, we recommend further examination of

Copyright © 2012 John Wiley & Sons, Ltd.

subjects identified in TRSS plots. For instance, the effect of a discordant subject (or observation) on the model parameters can be precisely measured by fitting a proposed model with and without it.

In addition to a subject-wise investigation, this article also suggests how to explore the data in an observation-wise manner. It is recommended to investigate the data subject-wise first and then observation-wise emphasizing the measurements of selected discordant subjects [16]. Unlike most diagnostic methods, the TRSS plot is a nondeletion subject-wise approach. Detection of clusters is simple, and the swamp phenomenon becomes less of a concern. PTRSS is a deletion approach to the observation-wise investigation.

The SCD and OCD are conceptually straightforward but often fail to recognize critical subjects and observations as they were designed for single-level regression models and do not consider the features of repeated measurements. Conditional Cook's distance is an improvement over OCD, but the concept is not suitable for extension to subject-wise investigation as it is still a deletion approach. For LME models, [16] suggested that it is desirable to examine the effects of subjects on the variance components before studying the effects on the coefficients. TRSS plots make this examination easy and effective by identifying subjects affecting variance components as well as ones affecting coefficients. The subject-wise local influence approach examines the effects of a subject on the likelihood function and needs to inspect five to nine index plots of different terms as mentioned in Section 2.2. It is sometimes confusing to statistical practitioners if they do not understand the theory and features of the method clearly. The interpretation of index plots for local influence should be delivered with much caution. In addition, many examples including the orthodontic growth data have indicated the need of improvement on current guideline for local influence in LME models.

As illustrated in simulated and real examples, TRSS plots are effective and efficient in detecting discordant subjects and also provide information for goodness-of-fit; if $\mathrm{TRSS}_{i,1}$ and $\mathrm{TRSS}_{i,0}$ display strong positive correlation, we may suspect a violation of model assumptions. Even in case of the absence of discordant subjects, TRSS plots may provide insights. For example, if TRSS points form a few clusters, this could indicate the possibility of missing important factor(s) in models. These will be areas of future study, and there is still room for improvement in reference lines for TRSS plots as well. The R package for TRSS plots is available at readers' requests.

References

1. Høyer AP, Jørgensen T, Grandjean P, Hartvig HB. Repeated measurements of organochlorine exposure and breast cancer risk (Denmark). *Cancer Causes and Control* 1999; **11**:177–184.
2. Choa MR, Wang CJ, Wu MT, Pan CH, Kuo CY, Yang HJ, Chang LW, Hu CW. Repeated measurements of urinary methylated/oxidative dna lesions, acute toxicity, and mutagenicity in coke oven workers. *Cancer Epidemiology Biomarkers & Prevention* 2008; **17**:3381–3389.
3. Egea B, Mouridsena SE. Linguistic development in a boy with the landau-kleffner syndrome: a five year follow-up study. *Journal of Neurolinguistics* 1998; **11**:321–328. DOI: 10.1016/S0911-6044(97)00013-4.
4. Hart JD, Wehrly TE. Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association* 1986; **81**:1080–1088.
5. Keselmana HJ, Alginab J, Kowalchuka RK, Wolfingerc RD. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Simulation and Computation* 1998; **27**:591 –604.
6. Lindsey JK, Lambert P. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* 1998; **17**:447–469.
7. Huang JZ, Wu CO, Zhou L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 2002; **89**:111–128.
8. Pan J, Fang K. Influential observation in the growth curve model with unstructured covariance matrix. *Computational Statistics & Data Analysis* 1996; **22**:71–87.
9. Ouwens MJNM, Tan FES, Berger MPF. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics* 2001; **57**:1166–1172.
10. Banerjee M, Frees EW. Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* 1997; **92**:999–1005.
11. Banerjee M. Cook's distance in linear longitudinal models. *Communications in Statistics: Theory and Methods* 1998; **27**:2973–2983.
12. Tan FES, Ouwens MJNM, Berger MPF. Detection of influential observations in longitudinal mixed effects regression models. *The Statistician* 2001; **50**:271–284.
13. Cook RD. Detection of influential observation in linear regression. *Technometrics* 1977; **19**:15–18.
14. Andrews DF, Pregibon D. Finding outliers that matter. *Journal of the Royal Statistical Society, Series B: Methodological* 1978; **40**:85–93.
15. Belsley DA, Kuh E, Welsch RE. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons: New York, NY, 1980.

Copyright © 2012 John Wiley & Sons, Ltd.

16. Christensen R, Pearson LM, Johnson W. Case-deletion diagnostics for mixed models. *Technometrics* 1992; **34**:38–45.
17. Cook DR. Assessment of local influence. *Journal of the Royal Statistical Society. Series B* 1986; **48**:133–169.
18. Beckman RJ, Cook RD. Diagnostics for mixed-model analysis of variance. *Technometrics* 1987; **29**:413–426.
19. Lesaffre E, Verbeke G. Local influence in linear mixed models. *Biometrics* 1998; **54**:570–582.
20. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
21. Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* 1988; **83**:1014–1022.
22. Kim C, Storer BE. Reference values for cook's distance. *Communications in Statistics: Simulation and Computation* 1996; **25**:691–708.
23. Potthoff RF, Roy SN. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 1964; **51**:313–326.
24. Pinheiro JC, Bates DM. *Mixed-effects Models in S and S-Plus*. Springer-Verlag Inc: New York, NY, 2000.
25. Xiang L, Tse S-K, Lee AH. Influence diagnostics for generalized linear mixed models: applications to clustered data. *Computational Statistics & Data Analysis* 2002; **40**:759–774.

Copyright © 2012 John Wiley & Sons, Ltd.