SORT 35 (2) July-December 2011, 125-144

Diagnostic measures for linear mixed measurement error models

Karim Zare and Abdolrahman Rasekh*

Abstract

In this paper, we present case deletion and mean shift outlier models for linear mixed measurement error models using the corrected likelihood of Nakamura (1990). We derive the corrected score test statistic for outliers detection based on mean shift outlier models. Furthermore, several case deletion diagnostics are constructed as a tool for influence diagnostics. It is found that they can be written in terms of studentized residuals of model, error contrast matrix and the inverse of the response variable covariance matrix. Our influence diagnostics are illustrated through a real data set.

MSC: 62J20 and 62J99

Keywords: Case deletion, corrected score method, influential points, linear mixed measurement error models, mean shift outlier model.

1. Introduction

Since all the observations in a data set do not play an equal role in determining estimators, tests and other statistics, it is important to consider influential points in data analysis. To identify anomalous observations, various approaches, including case deletion model (CDM) and mean shift outlier model (MSOM), have been proposed in the literature (Cook and Weisberg, 1982).

In linear mixed models, CDM, MSOM and related diagnostics are studied more widely by different authors including, Christensen et al. (1992), Banerjee and Frees (1997), Zhong and Wei (1999), Haslett and Dillane (2004), Zewotir and Galpin (2005) and Li et al. (2009).

* *Address for correspondence:* Department of Statistics, Shahid Chamran University, Ahvaz, Iran. Rasekh_a@scu.ac.ir, Tel: +98-611-3331043, Fax: +98-611-3331043.
Received: March 2011
Accepted: August 2011

Christensen et al. (1992) proposed case deletion diagnostics for both fixed effects and variance components. Banerjee and Frees (1997) proposed case deletion diagnostics for both fixed effects and random subject effects in linear longitudinal models. Zhong and Wei (1999) presented a unified diagnostic method for linear mixed models based upon the joint likelihood given by Robinson (1991). They showed that the estimates of parameters in CDM are equivalent to those in MSOM. Haslett and Dillane (2004) proved a 'delete = replace' identity in linear models and applied it to deletion diagnostics for estimators of variance components. Zewotir and Galpin (2005) provided routine diagnostic tools for fixed effects, random effects and variance components, which are computationally inexpensive. Li et al. (2009) considered subset deletion diagnostics for fixed effects, random effects and one variance component in varying coefficient mixed models.

As pointed out by Davidian and Giltinan (1995), independent variables in the models are often measured with non-negligible errors. Hence it is of great interest to study the measurement error models. On regression diagnostics for linear measurement error models, only some works has been done by Kelly (1984), Fuller (1987), Wellman and Gunst (1991), Zhong et al. (2000). Zhong et al. (2000) obtained CDM and MSOM for linear measurement error models. Also, they derived several diagnostics via CDM.

In linear mixed measurement error models the only work is due to Fung et al. (2003). However, in this paper, the corrected score function and the other relevant relations are not derived correctly. This problem also exists in diagnostic methods such as case deletion diagnostic on fixed effects. Furthermore, some of the relations in Fung et al. (2003) are somewhat different with Zhong et al. (2002).

Since there is no outstanding work in diagnostic methods for linear mixed measurement error models, in this paper, we concentrate on diagnostic methods for these models upon the corrected score function of Nakamura (1990). In Section 2, we present the model and the corrected score method for estimation of parameters. By using the corrected score method, Section 3 deals with two diagnostic models: CDM and MSOM. Besides, since MSOM is efficient to detect outliers, we construct a corrected score test for detecting outliers. In Section 4, we develop case deletion diagnostics for detecting influential points in linear mixed measurement error models. The given diagnostics are similar to diagnostics in linear mixed models and so are easy to compute. An influence analysis of a data set on hedonic housing-prices is given to illustrate the results in Section 5. Concluding remarks are given in Section 6.

2. Model definition and estimation

Consider the following linear mixed model with measurement errors in fixed effects:

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\beta} + \mathbf{U}\,\mathbf{b} + \boldsymbol{\varepsilon}, \\ \mathbf{X} &= \mathbf{Z} + \boldsymbol{\Delta}. \end{aligned} \tag{1}$$

In this model $\boldsymbol{\beta}$ is a $p \times 1$ vector of unobservable parameters, which are called fixed effects; $\mathbf{Z}$ and $\mathbf{U} = [\mathbf{U}_1 |\mathbf{U}_2| \ldots |\mathbf{U}_m]$ are $n \times p$ and $n \times q$ matrices of "regressors", respectively, where $\mathbf{U}_i$ is an $n \times q_i$ known design matrix of the random effect factor i; $\mathbf{b}^\mathsf{T} = \left(\mathbf{b}_1^\mathsf{T}, \mathbf{b}_2^\mathsf{T}, \ldots, \mathbf{b}_m^\mathsf{T}\right)$, where $\mathbf{b}_i$ is a $q_i \times 1$ vector of unobservable random effects from $N\left(\mathbf{0}, \sigma_i^2 \mathbf{I}\right)$, $i = 1, \ldots, m$; $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of unobservable random errors from $N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$. The variances σ^2 and σ_i^2, $i = 1, \ldots, m$ are called variance components. $\mathbf{X}$ is the observed value of $\mathbf{Z}$ with the measurement error $\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is an $n \times p$ random matrix from $N(\mathbf{0}, \mathbf{I} \otimes \boldsymbol{\Lambda})$. We assume that $\mathbf{b}_i, \boldsymbol{\varepsilon}$ and $\boldsymbol{\Delta}$ are mutually independent. One may also write $\mathbf{b} \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma}$ is a block diagonal matrix with the ith block being $\gamma_i \mathbf{I}$, for $\gamma_i = \sigma_i^2 / \sigma^2$, so that $\mathbf{y}$ has a multivariate normal distribution with $E(\mathbf{y}) = \mathbf{Z}\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}) = \sigma^2 \mathbf{V}$, in which $\mathbf{V} = \mathbf{I} + \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T} = \mathbf{I} + \sum_{i=1}^m \gamma_i \mathbf{U}_i \mathbf{U}_i^\mathsf{T}$. The conditional distribution of $\mathbf{b}|\mathbf{y}$ is $\mathbf{b}|\mathbf{y} \sim N\left(\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}), \sigma^2 \boldsymbol{\Sigma}\mathbf{T}\right)$ where $\mathbf{T} = \left(\mathbf{I} + \mathbf{U}^\mathsf{T}\mathbf{U}\boldsymbol{\Sigma}\right)^{-1}$. The log-likelihood of $\mathbf{y}$ is given by

$$l\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Upsilon}; \mathbf{Z}, \mathbf{y}\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log(|\mathbf{V}|) - \frac{1}{2\sigma^2}\left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})^\mathsf{T}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right],$$

where $\left(\sigma^2, \boldsymbol{\Upsilon}\right) = \left(\sigma^2, \gamma_1, \ldots, \gamma_m\right)$ belongs to $\Omega = \left\{\left(\sigma^2, \boldsymbol{\Upsilon}\right) : \sigma^2 > 0, \gamma_i \geq 0 \; (i = 1, \ldots, m)\right\}$. Also, the conditional log-likelihood of $\mathbf{b}|\mathbf{y}$ is given by

$$\begin{aligned} l_{\mathbf{b}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Upsilon}; \mathbf{Z}, \mathbf{y}\right) = & -\frac{q}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log(|\boldsymbol{\Sigma}\mathbf{T}|) \\ & - \frac{1}{2\sigma^2}\left\{\left[\mathbf{b} - \boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right]^\mathsf{T}(\boldsymbol{\Sigma}\mathbf{T})^{-1}\left[\mathbf{b} - \boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right]\right\}. \end{aligned}$$

Suppose that as in the model (1), the covariate $\mathbf{Z}$ is measured with error and the correlated structure arises from the random effects. If we simply replace $\mathbf{Z}$ by $\mathbf{X}$, then the estimates obtained from the score functions are not consistent in general. Various ways are proposed in dealing with measurement error models. In this paper, we use corrected score method proposed by Nakamura (1990) that is a common approach in measurement error models (see also Nakamura, 1992; Hanfelt and Liang, 1997; Gimenz and Bolfarine, 1997 and Zhong et al., 2000). In this method, we have to find a corrected score function whose expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true independent variables. For the model (1), Zhong et al. (2002) derived the corrected score estimates of fixed and random effects.

Let E^* denotes the conditional mean with respect to $\mathbf{X}$ given $\mathbf{y}$. The corrected log-likelihood $l^*\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Upsilon}; \mathbf{X}, \mathbf{y}\right)$ for our model should satisfy

$$\begin{aligned} E^*\left[\partial l^*\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Upsilon}; \mathbf{X}, \mathbf{y}\right) / \partial \boldsymbol{\beta}\right] &= \partial l\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Upsilon}; \mathbf{Z}, \mathbf{y}\right) / \partial \boldsymbol{\beta}, \\ E^*\left[\partial l_1^*\left(\sigma^2, \boldsymbol{\Upsilon}; \mathbf{X}, \mathbf{y}\right) / \partial \sigma^2\right] &= \partial l_1\left(\sigma^2, \boldsymbol{\Upsilon}; \mathbf{Z}, \mathbf{y}\right) / \partial \sigma^2 \end{aligned}$$

and

$$E^*\left[\partial l_1^*\left(\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\gamma_i\right]=\partial l_1\left(\sigma^2,\boldsymbol{\Upsilon};\mathbf{Z},\mathbf{y}\right)/\partial\gamma_i,\quad i=1,\ldots,m,$$

where $l_1\left(\sigma^2,\boldsymbol{\Upsilon};\mathbf{Z},\mathbf{y}\right)=l\left(\hat{\boldsymbol{\beta}}\left(\boldsymbol{\Upsilon}\right),\sigma^2,\boldsymbol{\Upsilon};\mathbf{Z},\mathbf{y}\right)$, in which $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}\left(\boldsymbol{\Upsilon}\right)$ is maximum likelihood estimate of $\boldsymbol{\beta}$ and $l_1^*\left(\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)=l^*\left(\hat{\boldsymbol{\beta}}\left(\boldsymbol{\Upsilon}\right),\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)$, in which $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}\left(\boldsymbol{\Upsilon}\right)$ is the solution of the equation $\partial l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}=\mathbf{0}$. Also, the conditional corrected log-likelihood $l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)$ should satisfy

$$E^*\left[\partial l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\right]=\partial l_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{Z},\mathbf{y}\right)/\partial\mathbf{b}.$$

The following equation is useful to find such l^* and $l_{\mathbf{b}}^*$,

$$E^*\left(\mathbf{X}^{\mathsf{T}}\mathbf{A}\mathbf{X}\right)=\mathbf{Z}^{\mathsf{T}}\mathbf{A}\mathbf{Z}+\operatorname{tr}\left(\mathbf{A}\right)\boldsymbol{\Lambda}.$$

Given $\boldsymbol{\Lambda}$, l^* and $l_{\mathbf{b}}^*$ are obtained as

$$\begin{aligned}l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)=&-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\mathbf{V}|\right)\\&-\frac{1}{2\sigma^2}\left\{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathsf{T}}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}\end{aligned}$$

and

$$\begin{aligned}l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)=&-\frac{q}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\boldsymbol{\Sigma}\mathbf{T}|\right)\\&-\frac{1}{2\sigma^2}\left\{\left[\mathbf{b}-\boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]^{\mathsf{T}}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\left[\mathbf{b}-\boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\right.\\&\left.-\operatorname{tr}\left(\mathbf{I}-\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}.\end{aligned}$$

If the γ_i's (and hence $\mathbf{V}$) are known, by solving the equations $\partial l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}=\mathbf{0}$, $\partial l_1^*\left(\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\sigma^2=0$ and $\partial l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Upsilon};\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}=\mathbf{0}$, the corrected score estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$, respectively, are given by (See Zhong et al., 2002 and Zare et al., 2011 for more details)

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=\left(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{y},\\\hat{\sigma}^2&=\frac{1}{n}\left[\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^{\mathsf{T}}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}\right],\\\tilde{\mathbf{b}}&=\boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right).\end{aligned}$$

If the γ_i's are unknown, the corrected score estimates are substituted back into $\mathbf{\Sigma}$ to obtain $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\tilde{\mathbf{b}}$. For the estimation of γ_i's, we can use the corrected score estimates of $\sigma_1^2,\ldots,\sigma_m^2$ that are given by (Zare et al., 2011)

$$\hat{\sigma}_i^2 = \frac{1}{q_i - \operatorname{tr}(\mathbf{T}_{ii})}\left[\tilde{\mathbf{b}}_i^{\mathsf{T}}\tilde{\mathbf{b}}_i - \operatorname{tr}\left(\hat{\mathbf{D}}_i^{\mathsf{T}}\hat{\mathbf{D}}_i\right)\hat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{\Lambda}\hat{\boldsymbol{\beta}}\right], \quad i = 1,\ldots,m,$$

where $\mathbf{T}_{ij}$ is ijth block of matrix $\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \cdots & \mathbf{T}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{T}_{m1} & \cdots & \mathbf{T}_{mm} \end{bmatrix}$, $\hat{\mathbf{D}}_i = \hat{\gamma}_i\mathbf{U}_i^{\mathsf{T}}\mathbf{V}^{-1} = \left(\hat{\sigma}_i^2/\hat{\sigma}^2\right)$ $\mathbf{U}_i^{\mathsf{T}}\mathbf{V}^{-1}$ and $\tilde{\mathbf{b}}_i = \hat{\mathbf{D}}_i\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$.

The above results show that we must use an iterative numerical procedure to obtain the corrected score estimates of parameters. We use the iterative algorithm given in Zare et al. (2011). Also, Zare et al. (2011) showed the corrected score estimates of γ_i's are consistent. In continuing, we assume that the γ_i's are known.

For notational simplicity, $\mathbf{A}_{(i)}$ denotes an $n\times m$ matrix $\mathbf{A}$ with ith row removed, $\mathbf{A}_{[i]}$ denotes a matrix $\mathbf{A}$ with the ith row and column removed, $\mathbf{a}_i^{\mathsf{T}}$ denotes the ith row of $\mathbf{A}$ and a_{ij} denotes the ijth element of $\mathbf{A}$. Similarly, $\mathbf{a}_{(i)}$ denotes vector $\mathbf{a}$ with the ith element removed and a_i denotes the ith element of $\mathbf{a}$. Without loss of generality, we partition the matrices as if the ith deleted case is the first row; i.e. $i = 1$. Then

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_i^{\mathsf{T}} \\ \mathbf{X}_{(i)} \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{z}_i^{\mathsf{T}} \\ \mathbf{Z}_{(i)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_i \\ \mathbf{y}_{(i)} \end{bmatrix} \text{ and } \mathbf{C} = \mathbf{V}^{-1} = \begin{bmatrix} c_{ii} & \mathbf{c}_{i(i)}^{\mathsf{T}} \\ \mathbf{c}_{i(i)} & \mathbf{V}_{[i]}^{-1} + \mathbf{c}_{i(i)}\mathbf{c}_{i(i)}^{\mathsf{T}}/c_{ii} \end{bmatrix}.$$

3. Mean shift outlier and case deletion model

In regression diagnostics, there are two commonly used models: CDM amd MSOM (Cook and Weisberg, 1982). Each of models has its own advantage in practice. CDM's are used to obtain case deletion diagnostics for detecting influential observations. MSOM's are used for detecting outlier observations. It is well known that in linear (mixed) models maximum likelihood estimates of parameters in CDM and MSOM are equal. In linear measurement error models, the estimates are approximately equal.

3.1. Mean shift outlier model

A commonly used diagnostic model is MSOM (Cook and Weisberg, 1982). MSOM can be represented as

$$\begin{aligned} &y_j = \mathbf{z}_j^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{u}_j^{\mathsf{T}}\mathbf{b} + \varepsilon_j \quad \text{for} \quad j \neq i,\ j = 1,\ldots,n,\ y_i = \mathbf{z}_i^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{u}_i^{\mathsf{T}}\mathbf{b} + \tau + \varepsilon_i, \\ &\mathbf{x}_k^{\mathsf{T}} = \mathbf{z}_k^{\mathsf{T}} + \boldsymbol{\delta}_k^{\mathsf{T}} \quad \text{for} \quad k = 1,\ldots,n, \end{aligned} \tag{2}$$

where τ is an extra parameter to indicate the presence of an outlier (Cook and Weisberg, 1982). Obviously, if value of τ is nonzero, then it no longer comes from the original model, and so ith case may be an outlier. An outlier test can be formulated as a test of the null hypothesis that $\tau = 0$. The corrected likelihood estimates of $\boldsymbol{\beta}$, σ^2, τ, and $\mathbf{b}$ in (2) are denoted by $\hat{\boldsymbol{\beta}}_{mi}$, $\hat{\sigma}^2_{mi}$, $\hat{\tau}_{mi}$ and $\tilde{\mathbf{b}}_{mi}$, respectively.

Theorem 1 *For model (2), we have*

$$\hat{\boldsymbol{\beta}}_{mi} = \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - tr\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

$$\hat{\tau}_{mi} = \frac{\hat{v}_i}{r_{ii}}, \quad \hat{\sigma}^2_{mi} = \frac{n - t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n}\hat{\sigma}^2 \quad \text{and} \quad \tilde{\mathbf{b}}_{mi} = \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\frac{\hat{v}_i}{r_{ii}},$$

where $\mathbf{c}_i^{\mathsf{T}}$ and $\mathbf{r}_i^{\mathsf{T}}$ are ith rows of $\mathbf{V}^{-1}$ and $\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - tr\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}$, respectively, c_{ii} and r_{ii} are the ith diagonal elements of $\mathbf{V}^{-1}$ and $\mathbf{R}$, $\hat{v}_i = y_i - \mathbf{x}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}} - \mathbf{u}_i^{\mathsf{T}}\tilde{\mathbf{b}}$ is ith residual of model and $t_i = \hat{v}_i/\left(\hat{\sigma}_v\sqrt{r_{ii}}\right)$ is ith studentized residual of model, in which $\hat{\sigma}_v^2 = \hat{\sigma}^2 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}$.

Theorem 2 *For MSOM, the score test statistic for the hypothesis $H_0 : \tau = 0$ is given by*

$$SC_i = \frac{\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}} = t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right).$$

This theorem shows that score statistic SC_i is a multiple of the square of studentized residual of model that is an adequate diagnostic statistic as often used in linear regression diagnostics.

3.2. Case deletion model

As mentioned, CDM is the fundamental basis for constructing effective diagnostics. CDM can be represented as

$$y_j = \mathbf{z}_j^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{u}_j^{\mathsf{T}}\mathbf{b} + \varepsilon_j, \ \mathbf{x}_j^{\mathsf{T}} = \mathbf{z}_j^{\mathsf{T}} + \boldsymbol{\delta}_j^{\mathsf{T}} \quad \text{for} \quad j \neq i, \ j = 1, \ldots, n.$$

Let $\hat{\boldsymbol{\beta}}_{(i)}$, $\hat{\sigma}^2_{(i)}$ and $\tilde{\mathbf{b}}_{(i)}$ denote the estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$ when the ith case is deleted, respectively.

Theorem 3 *For model (3), we have*

$$\hat{\boldsymbol{\beta}}_{(i)} \approx \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

$$\hat{\sigma}^2_{(i)} \approx \frac{n}{n-1}\hat{\sigma}^2 - \frac{\hat{v}_i^2}{(n-1)\,r_{ii}} = \frac{n - t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\hat{\sigma}^2,$$

$$\tilde{\mathbf{b}}_{(i)} \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\frac{\hat{v}_i}{r_{ii}}.$$

Comparing results of the theorems 1 and 3, it is obvious that the estimates of the parameters are approximately the same. In the following section we derive different diagnostic measures based on CDM.

4. Influence diagnostics

It is well known that results from an analysis can be substantially influenced by one or a few observations; that is, all the observations have not equal effect in statistical models. Case deletion diagnostics are the usual methods to measure the influence of individual observations in the statistical models with dropping the observation from data set and computing a convenient norm of the change in the parameters. Let the corrected Fisher information matrix of $\mathbf{y}$ for $\boldsymbol{\beta}$ be $\mathbf{I}^*(\boldsymbol{\beta})$, then

$$\mathbf{I}^*(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right].$$

Also, the corrected Fisher information matrix of $\mathbf{y}$ for $\mathbf{b}$ is

$$\mathbf{I}^*(\mathbf{b}) = \frac{1}{\sigma^2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U} + \boldsymbol{\Sigma}^{-1}\right).$$

4.1. Analogue of generalized Cook's distance

4.1.1. Analogue of generalized Cook's distance for fixed effects

The generalized Cook (1977) distance is the norm of $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ with respect to certain weight matrix $\mathbf{M} > \mathbf{0}$, i.e.

$$CD_i(\boldsymbol{\beta}) = \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\mathbf{M}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right).$$

Choosing $\mathbf{M} = \hat{\mathbf{I}}^*(\boldsymbol{\beta}) = \hat{\sigma}^{-2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]$, where $\hat{\mathbf{I}}^*(\boldsymbol{\beta})$ is estimate of $\mathbf{I}^*(\boldsymbol{\beta})$, yields

$$CD_i(\boldsymbol{\beta}) = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}^2}.$$

Since

$$\hat{\boldsymbol{\beta}}_{(i)} \approx \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

we can get, approximately,

$$CD_i(\boldsymbol{\beta}) = \frac{(c_{ii} - r_{ii})\,\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}^2} = \frac{c_{ii} - r_{ii}}{r_{ii}} t_i^2 \left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right).$$

Let $\mathbf{d}_k$ be a p-vector with 1 at the kth position and zero elsewhere, then $\hat{\beta}_k = \mathbf{d}_k^{\mathsf{T}}\hat{\boldsymbol{\beta}}$ has the standard error s_k and t-value $t_k = \hat{\beta}_k/s_k$, where $s_k^2 = \hat{\sigma}^2\mathbf{d}_k^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{d}_k$. The joint $100\,(1-\alpha)\,\%$ confidence region for parameter $\boldsymbol{\beta}$ is

$$\left\{\boldsymbol{\beta} : \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \le p\hat{\sigma}^2 F\left(p, n-p, \alpha\right)\right\},$$

where $F\left(p, n-p, \alpha\right)$ denotes the upper α percentile of the Fisher's distribution with p and $n-p$ degrees of freedom. Suppose that $CD_i(\boldsymbol{\beta}) \simeq pF\left(p, n-p, \alpha\right)$, then the removal of the ith case moves corrected score estimate to the edge of the $100\,(1-\alpha)\,\%$ confidence region. Such a situation may be cause for concern and so more attention should be paid to that case. Usually, one would like each $\hat{\boldsymbol{\beta}}_{(i)}$ to stay well within a 90%, say, confidence region. Then case i can be considered a highly influential point if

$$CD_i(\boldsymbol{\beta}) > pF\left(p, n-p, 0.1\right).$$

Based on Cook (1977), we have $CD_i(\beta_k) = t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right) G_i^2\left(\mathbf{d}_k^{\mathsf{T}}\right)$, where, for any $q' \times p$ matrix $\mathbf{A}$ of rank q', $G_i(\mathbf{A})$ is defined as

$$G_i(\mathbf{A}) = \frac{1}{\sqrt{r_{ii}}}\left[\mathbf{A}\left(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{A}^{\mathsf{T}}\right]^{-1/2}\mathbf{A}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i.$$

As similar, case i can be considered a highly influential point if $CD_i(\beta_k) > F\left(1, n-1, 0.1\right)$, since this case, if deleted, would move the estimate of $\hat{\beta}_k$ to the edge of the 90% confidenc region. Rio (1988) argued that $G_i^2\left(\mathbf{u}^{\mathsf{T}}\right)$ can be used to measure the influence of case i on the precision of the estimation of $\mathbf{u}^{\mathsf{T}}\boldsymbol{\beta}$. Therefore, based on Rio (1988), ith case is said to have high influential on the estimate of $\hat{\beta}_k$ if $G_i^2\left(\mathbf{d}_k^{\mathsf{T}}\right)$ is sufficiently large.

Let $\mathbf{A}$ denote a $q' \times p$ rank q' matrix and let $\Psi = \mathbf{A}\boldsymbol{\beta}$ denote the combinations of interest. A generalized measure of the importance of the ith case is defined as

$$CD_i(\Psi) = \frac{\left(\hat{\Psi}_{(i)} - \hat{\Psi}\right)^{\mathsf{T}}\left[\mathbf{A}\left(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{A}^{\mathsf{T}}\right]^{-1}\left(\hat{\Psi}_{(i)} - \hat{\Psi}\right)}{q'\hat{\sigma}^2},$$

where $\hat{\boldsymbol{\Psi}}_{(i)} = \mathbf{A}\hat{\boldsymbol{\beta}}_{(i)}$ and $\hat{\boldsymbol{\Psi}} = \mathbf{A}\hat{\boldsymbol{\beta}}$. Since

$$\hat{\boldsymbol{\Psi}}_{(i)} - \hat{\boldsymbol{\Psi}} \approx \mathbf{A}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

we can get, approximately,

$$CD_i(\Psi) = \frac{t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{q'}G_i^{\mathsf{T}}(\mathbf{A})\,G_i(\mathbf{A}).$$

To obtain the levels of significance the values of this generalized measure should be compared to the probability points of the central Fisher distribution with q' and $n-q'$ degrees of freedom.

4.1.2. Analogue of generalized Cook's distance for random effects

The proposed diagnostic measure examines the squared distance from the complete data predictor of the random effects to ith case deleted predictor of the random effects, relative to $\mathbf{M} = \hat{\mathbf{I}}^*(\mathbf{b}) = \hat{\sigma}^{-2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right)$. This is the generalized Cook distance and can be written as

$$CD_i(\mathbf{b}) = \left(\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}}\mathbf{M}\left(\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_{(i)}\right) = \frac{\left(\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right)\left(\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_{(i)}\right)}{\hat{\sigma}^2}.$$

Since $\tilde{\mathbf{b}}_{(i)} \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\,\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\dfrac{\hat{v}_i}{r_{ii}}$, we can get, approximately,

$$\begin{aligned}CD_i(\mathbf{b}) &= \mathbf{r}_i^{\mathsf{T}}(\mathbf{V}-\mathbf{I})\,\mathbf{V}\mathbf{r}_i\frac{\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}^2}\\ &= \mathbf{r}_i^{\mathsf{T}}(\mathbf{V}-\mathbf{I})\,\mathbf{V}\mathbf{r}_i\frac{t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{r_{ii}}. \end{aligned} \tag{4}$$

Also, from (4) we have

$$CD_i(\mathbf{b}_j) = \mathbf{r}_i^{\mathsf{T}}\mathbf{U}_j\mathbf{T}_{jj}^{-1}\mathbf{U}_j^{\mathsf{T}}\mathbf{r}_i\frac{\gamma_j t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{r_{ii}}, \quad \text{for} \quad j=1,\ldots,m.$$

4.2. Analogue of Welsch's distance

Welsch (1982) has suggested using Welsch's distance as a diagnostic tool and, for $n > 15$, using $3\sqrt{p}$ as a cutoff point for linear models. Welsch's distance gives more emphasize to high leverage points. It has similar rationale as Cook's distance Essential difference between these two methods is in the choice of scale. (Chatterjee and Hadi, 1986). For the fixed effects it is given as

$$W_i(\boldsymbol{\beta}) = \left[(n-1) \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}} \left(\mathbf{X}_{(i)}^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)} - \operatorname{tr}\left(\mathbf{V}_{[i]}^{-1}\right) \boldsymbol{\Lambda}\right) \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}_{(i)}^2} \right]^{1/2}$$
$$\approx \left[(n-1) \frac{c_{ii} - r_{ii}}{c_{ii} r_{ii} \hat{\sigma}_{(i)}^2} \right]^{1/2} |\hat{v}_i| .$$

Welsch (1982) suggested using W_i as a diagnostic tool. The analogue of Welsch's distance for random effects, is

$$W_i(\mathbf{b}) = \left[(n-1) \frac{\left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}} \left(\mathbf{U}_{(i)}^{\mathsf{T}} \mathbf{U}_{(i)} + \boldsymbol{\Sigma}^{-1}\right) \left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)}{\hat{\sigma}_{(i)}^2} \right]^{1/2}$$
$$\approx \left[(n-1) \frac{\mathbf{r}_i^{\mathsf{T}} (\mathbf{V} - \mathbf{I}) \mathbf{V} \mathbf{r}_i - \left(r_{ii} - \mathbf{c}_i^{\mathsf{T}} \mathbf{r}_i\right)^2}{r_{ii}^2 \hat{\sigma}_{(i)}^2} \right]^{1/2} |\hat{v}_i| .$$

4.3. Analogue of the likelihood distance

Another popular measure to assess the influence of the ith case on corrected score estimate is the likelihood distance (Cook and Weisberg, 1982). Let $l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ and $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ be the corrected log-likelihood evaluated at $\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\right)$ and $\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2\right)$, respectively. A measure of the influence of the ith case on $\hat{\boldsymbol{\beta}}$ can be derived based on the distance between $l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ and $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$. The likelihood distance is defined as

$$LD_i(\boldsymbol{\beta}) = 2\left[l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right) - l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)\right].$$

Taylor expansion of $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ at $\hat{\boldsymbol{\beta}}$ gives

$$LD_i(\boldsymbol{\beta}) = 2\left\{ \left[\left. \frac{\partial l^*\left(\boldsymbol{\beta}, \sigma^2; \mathbf{X}, \mathbf{y}\right)}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \sigma^2 = \hat{\sigma}^2} \right]^{\mathsf{T}} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right) \right.$$

$$+\frac{1}{2}\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[-\frac{\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\sigma^2=\hat{\sigma}^2}\right]\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)\Bigg\}$$

$$=\frac{\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}^2}.$$

This result is exact because the third derivative is zero. As seen, we have $LD_i(\boldsymbol{\beta}) = CD_i(\boldsymbol{\beta})$ As before, it can be shown that $LD_i(\mathbf{b}) = CD_i(\mathbf{b})$.

4.4. Analogue of the corrected Fisher information ratio

4.4.1. Analogue of corrected Fisher information ratio for fixed effects

As suggested by Belsley et al. (1980), the influence of the ith case on corrected Fisher information matrix for $\boldsymbol{\beta}$ can be measured by comparing the ratio of $\left|\hat{\mathbf{I}}^*(\boldsymbol{\beta})\right|$ to $\left|\hat{\mathbf{I}}^*_{ci}(\boldsymbol{\beta})\right|$; that is,

$$CFIR1_i(\boldsymbol{\beta})=\frac{\left|-\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2}}{\left|-\partial^2 l^*_{ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2_{(i)}}}$$

$$=\frac{\left|\hat{\sigma}^{-2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\right|}{\left|\hat{\sigma}^{-2}_{(i)}\left[\mathbf{X}^{\mathsf{T}}_{(i)}\mathbf{V}^{-1}_{[i]}\mathbf{X}_{(i)}-\operatorname{tr}\left(\mathbf{V}^{-1}_{[i]}\right)\boldsymbol{\Lambda}\right]\right|}$$

$$=\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^p\frac{\left|\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right|}{\left|\mathbf{X}^{\mathsf{T}}_{(i)}\mathbf{V}^{-1}_{[i]}\mathbf{X}_{(i)}-\operatorname{tr}\left(\mathbf{V}^{-1}_{[i]}\right)\boldsymbol{\Lambda}\right|}.$$

We can get, approximately,

$$CFIR1_i(\boldsymbol{\beta})=\left(\frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\right)^p\frac{c_{ii}}{r_{ii}}.$$

As this is close to 1 if the point is not influential, it seems sensible to use the relative measure $|CFIR1_i(\boldsymbol{\beta})-1|$ as a criterion for assessing the influence of the ith case on $\mathbf{I}^*(\boldsymbol{\beta})$. The larger the statistic $|CFIR1_i(\boldsymbol{\beta})-1|$, the higher the influence of the ith case.

If one uses the trace instead of the determinant, the corrected fisher information ratio becomes

$$
\begin{aligned}
CFIR2_i(\boldsymbol{\beta}) = \operatorname{tr}\Bigg\{ & \left[-\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right]\Big|_{\sigma^2=\hat{\sigma}^2} \\
& \left[-\partial^2 l^*_{ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right]^{-1}\Big|_{\sigma^2=\hat{\sigma}^2_{(i)}} \Bigg\} \\
& \approx \frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\left(\frac{c_{ii}}{r_{ii}}+p-1\right).
\end{aligned}
$$

If removing the ith case does not change the trace, $CFIR2_i(\boldsymbol{\beta})$ will be close to p and so we could use the relative measure $|CFIR2_i(\boldsymbol{\beta})-p|$ as a criterion for assessing the influence of the ith case on the corrected Fisher information for fixed effects.

4.4.2. Analogue of corrected Fisher information ratio for random effects

As similar, the influence of the ith case on corrected Fisher information matrix for $\mathbf{b}$ can be measured by comparing the ratio of $\left|\hat{\mathbf{I}}^*(\mathbf{b})\right|$ to $\left|\hat{\mathbf{I}}^*_{ci}(\mathbf{b})\right|$; that is,

$$
\begin{aligned}
CFIR1_i(\mathbf{b}) &= \frac{\left|-\partial^2 l^*_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2}}{\left|-\partial^2 l^*_{\mathbf{b}_{ci}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2_{(i)}}} \\
&= \frac{\left|\hat{\sigma}^{-2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right)\right|}{\left|\hat{\sigma}^{-2}_{(i)}\left[\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right]\right|} = \left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^q \frac{\left|\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right|}{\left|\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right|}.
\end{aligned}
$$

We can get, approximately,

$$
CFIR1_i(\mathbf{b}) = \left(\frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\right)^q \frac{1}{c_{ii}}.
$$

Also, if one uses the trace instead of the determinant, the corrected fisher information ratio becomes

$$
\begin{aligned}
CFIR2_i(\mathbf{b}) = \operatorname{tr}\Bigg\{ & \left[-\partial^2 l^*_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right]\Big|_{\sigma^2=\hat{\sigma}^2} \\
& \left[-\partial^2 l^*_{\mathbf{b}_{ci}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right]^{-1}\Big|_{\sigma^2=\hat{\sigma}^2_{(i)}} \Bigg\} \\
& \approx \frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\left(c_{ii}^{-1}+q-1\right).
\end{aligned}
$$

If removing the ith case does not change the trace, $CFIR2_i(\mathbf{b})$ will be close to q. Hence, we could use the relative measure $|CFIR2_i(\mathbf{b})-q|$ as a criterion for assessing the influence of the ith case on the corrected Fisher information for random effects The ith observation is influential observation if $|CFIR2_i(\mathbf{b})-q|$ is sufficiently large.

5. Example

Diagnostic measures developed in the previous sections are applied to analyse a set of real data which is known as the Boston Housing data set. This data set was the basis for a 1978 paper by Harrison and Rubinfeld, which discussed approaches for using housing market data to estimate the willingness to pay for clean air. The authors employed a hedonic price model, based on the premise that the price of the property is determined by structural attributes (such as size, age, condition) as well as neighborhood attributes (such as crime rate, accessibility, environmental factors). This type of approach is often used to quantify the effects of environmental factors that affect the price of a property. A description of this data set can be found in Harrison and Rubinfeld (1978) and Belsley et al. (1980).

Zhong et al. (2002) considered this data set and used the data of $n=132$ census tracts within the 15 districts of the Boston city (as a part of 506 observations on census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970). They followed the regression model of Harrison and Rubinfeld (1978). However, the census tracts within districts are taken as repeated measurements. All independent variables can be measured precisely except the pollution variable NOXSQ which is taken to have measurement errors. Therefore, a linear mixed measurement error model was employed.

Now, we consider the same data set and derive different diagnostic measures for linear mixed measurement error model given in previous section. Figures 1-3 give the index plot of the diagnostic measures for fixed effects and Figures 4-6 give the index plot of the diagnostic measures for random effects, respectively. Based on generalized Cook's distance, a glance at Figures 1 and 4 shows that cases 9 and 15 have more influence on fixed effect and cases 9, 16 and 36 have more influence on random effects. The plots for W_i in Figures 2 and 5, respectively, for fixed and random effects have almost identical behavior as CD_i

Table 1 gives the parameter estimates from corrected score method with the full data and with only case 9 deleted. As seen, after deleting case 9, the NOXSQ variable does not have any significant effect. The RM, AGE and CHAS variables, in each case, do not have any significant effects but after deleting case 9 their signs have been changed. Figure 3 show that case 36 is the most influential point on corrected Fisher information for fixed effects, while Figure 6 indicate that this case is the most influential point on corrected Fisher information for random effects.

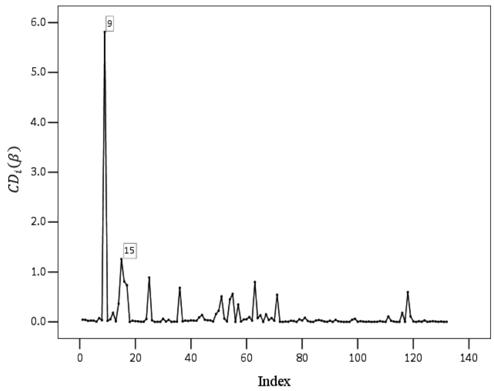

Figure 1: *Index plot of* $CD_i(\boldsymbol{\beta})$.

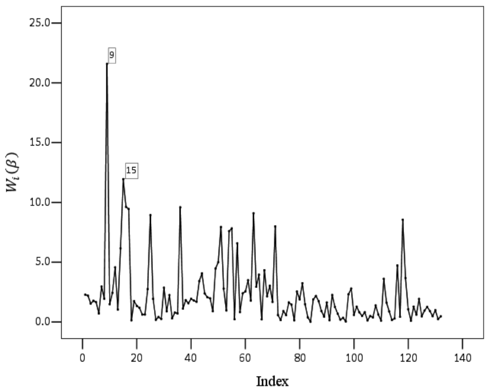

Figure 2: *Index plot of* $W_i(\boldsymbol{\beta})$.

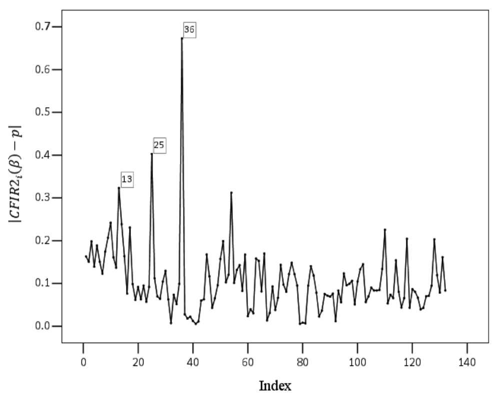

Figure 3: *Index plot of* $|CFIR2_i(\boldsymbol{\beta})-p|$.

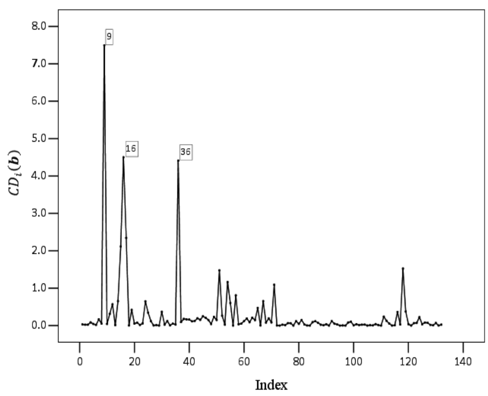

Figure 4: *Index plot of* $CD_i(\boldsymbol{b})$.

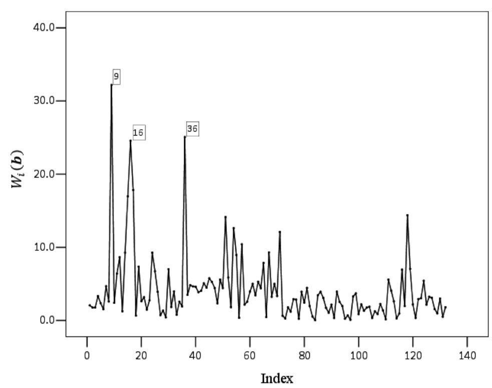

Figure 5: *Index plot of* $W_i(\boldsymbol{b})$.

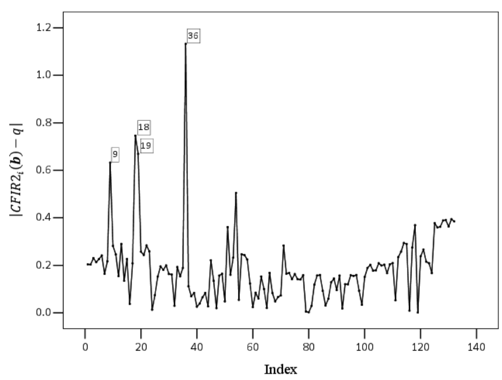

Figure 6: *Index plot of* $|CFIR2_i(\boldsymbol{b})-q|$.

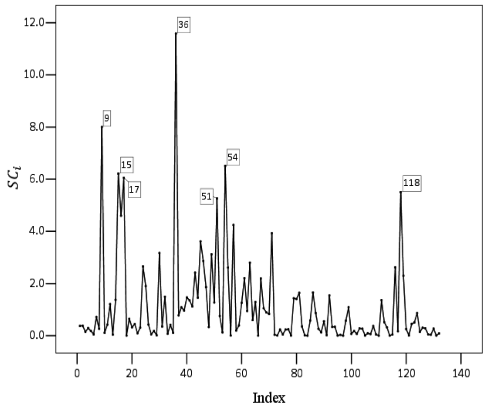

Figure 7: *Index plot of SC_i.*

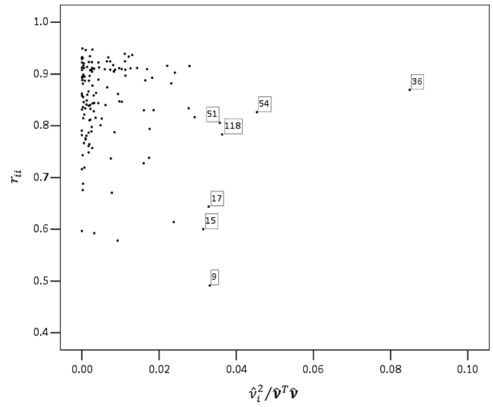

Figure 8: *Scatter plot of r_{ii} versus $\hat{\mathbf{v}}_i^2/\hat{\mathbf{v}}^\top\hat{\mathbf{v}}$.*

Table 1: *Corrected score estimates for the hedonic housing price data of Boston city. The t-ratios are in parentheses.*

Variable	Full data	Case 9 deleted	% change
Intercept	9.07 (28.72)	8.90 (28.11)	1.9
RM	-1.4×10^{-3} (−0.57)	3.6×10^{-3} (1.26)	352.2
AGE	7.6×10^{-4} (0.4)	-3.7×10^{-4} (−0.2)	150.0
DIS	8.8×10^{-2} (0.59)	1.8×10^{-1} (1.16)	101.2
B	4.6×10^{-1} (2.95)	5.1×10^{-1} (3.43)	12.2
LSTAT	-5.3×10^{-1} (−8.65)	-4.8×10^{-1} (−7.77)	10.5
CRIM	-7.3×10^{-3} (−5.36)	-6.6×10^{-3} (−5.00)	9.7
CHAS	-3.0×10^{-2} (−0.33)	4.6×10^{-3} (0.05)	115.5
NOXSQ	-1.0×10^{-2} (−2.34)	-7.8×10^{-3} (−1.74)	24.1
σ_1^2	4.8×10^{-3}	6.8×10^{-3}	41.7
σ^2	2.8×10^{-3}	2.5×10^{-3}	10.7

Table 2: *Corrected score estimates for the hedonic housing price data of Boston city after deleting pair cases {9, 15} and {9, 16}. The t-ratios are in parentheses.*

Variable	Pair case {9, 15} deleted	% change	Pair case {9, 16} deleted	% change
Intercept	8.76 (28.05)	3.3	8.86 (28.97)	2.3
RM	3.7×10^{-3} (1.34)	358.1	2.3×10^{-3} (0.81)	264.3
AGE	-1.0×10^{-4} (−0.06)	113.7	-3.4×10^{-4} (−0.19)	144.7
DIS	2.1×10^{-1} (1.4)	144.6	1.4×10^{-1} (0.98)	59.1
B	4.9×10^{-1} (3.41)	7.6	4.6×10^{-1} (3.12)	0.0
LSTAT	-5.2×10^{-1} (−8.35)	2.3	-5.1×10^{-1} (−8.56)	3.8
CRIM	-6.5×10^{-3} (−5.11)	11.6	-6.9×10^{-3} (−5.20)	5.5
CHAS	5.1×10^{-2} (0.60)	270.5	1.1×10^{-1} (1.14)	466.7
NOXSQ	-7.5×10^{-3} (−1.7)	27.3	-6.8×10^{-3} (−1.59)	32.0
σ_1^2	8.4×10^{-3}	75.0	4.3×10^{-3}	10.4
σ^2	2.3×10^{-3}	17.9	2.6×10^{-3}	7.1

Table 2 gives the corrected score estimates after deleting pairs of cases $\{9, 15\}$and $\{9, 16\}$from data set. Deleting these pairs have almost the same effect with deleting case 9 on parameters of model. The only difference is that deleting cases $\{9, 15\}$has more influence on CHAS variable and σ_1^2 while deleting cases $\{9, 16\}$has more influence on CHAS variable. Table 3 indicates the maximum percentage of changes in determinant of corrected Fisher information after deleting case 36. Finally, Figures 7 and 8 indicate that case 36 is also an outlier observation (see Zewotir and Galpin, 2007 for details about plot of Figure 8).

Table 3: *The determinant of the corrected Fisher information (DCFI) for the hedonic housing price data of Boston city.*

DCFI	Full data	Case 36 deleted	% change
Fixed effects	$6.08 \times 10^{+31}$	$1.17 \times 10^{+32}$	91.7
Random effects	$4.52 \times 10^{+36}$	$1.49 \times 10^{+37}$	229.8

6. Concluding remarks

We have presented case deletion and mean shift outlier models for linear mixed measurement error models that appear to be useful and can play important role in data analysis. Also, based on the corrected likelihood, we obtained case deletion diagnostics for detecting influential observations in linear mixed measurement error models. All the diagnostic measures are similar to diagnostics in linear mixed models. They are functions of studentized residuals of model, error contrast matrix $(\mathbf{R})$ and the inverse of the response variable covariance matrix $(\mathbf{C})$. Although no formal cutoff points are presented for these measures, it appears that relative comparisons such as ranking or simple index plots are a promising and practical approach to pinpoint influential observations. Here, the results obtained with the assumption that the γ_i's are known. In practice, we do not know the γ_i's. So, the corrected score estimates of the γ_i's are used and the results are useful as an approximation. In this paper, we fitted a linear mixed model with measurement error in fixed effects (and not in random effects) by specifying the covariance structure of $\mathbf{b}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Delta}$. Here we have assumed that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are known and $\boldsymbol{\Sigma}$ has diagonal structure with the ith block being $\gamma_i \mathbf{I}$. However, if random effects are also measured with errors and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are unknown, extending our diagnostics is an area of future research.

Acknowledgments

We wish to thank the Editor and the referee for their valuable comments which enhanced the presentation of this paper.

Appendix

Proof of Theorem 1:

It follows from (2) that the corrected log-likelihood of $\mathbf{y}$ and the conditional corrected log-likelihood of $\mathbf{b}|\mathbf{y}$ for MSOM, respectively, are given by

$$
\begin{aligned}
l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right) = & -\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\mathbf{V}|\right)-\frac{1}{2\sigma^2}\Bigg\{\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\Big(\mathbf{V}^{-1}_{[i]} \\
& +\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\Big)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+c_{ii}\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)^2 \\
& +2\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\Bigg\},
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
l^*_{\mathbf{b}_{mi}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right) = & -\frac{q}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\boldsymbol{\Sigma}\mathbf{T}|\right)-\frac{1}{2\sigma^2}\Bigg\{\mathbf{b}^{\mathsf{T}}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\mathbf{b} \\
& -2\mathbf{b}^{\mathsf{T}}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\boldsymbol{\Sigma}\Big[c_{ii}\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{u}_i+\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{c}_{i(i)} \\
& +\mathbf{u}_i\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+\mathbf{U}^{\mathsf{T}}_{(i)}\left(\mathbf{V}^{-1}_{[i]}+\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\right)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\
& +F\left(\boldsymbol{\beta},\tau\right)\Big]\Bigg\}
\end{aligned}
\tag{6}
$$

where

$$
\begin{aligned}
F\left(\boldsymbol{\beta},\tau\right) = & \left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\left[\mathbf{I}-\mathbf{V}^{-1}_{[i]}-\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\right]\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\
& +\left(1-c_{ii}\right)\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)^2-2\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\
& -\operatorname{tr}\left(\mathbf{I}-\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}
\end{aligned}
$$

The corrected likelihood estimates of $\hat{\boldsymbol{\beta}}_{mi}$, $\hat{\sigma}^2_{mi}$, $\hat{\tau}_{mi}$ and $\tilde{\mathbf{b}}_{mi}$, are derived with differentiating (5) with respect to $\boldsymbol{\beta}$, σ^2 and τ and (6) with respect to $\mathbf{b}$.

Proof of Theorem 2:

Since corrected score estimate is asymptotically normal, the score test can be used (Cox and Hinkley, 1974). Let the corrected Fisher information matrix of $\mathbf{y}$ for $\boldsymbol{\beta}$ and τ be $\mathbf{J}\left(\boldsymbol{\beta},\tau\right)$, then the score statistic under $H_0:\tau=0$ is

$$
SC_i = \left[\frac{\partial l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\tau}\right]^{\mathsf{T}} J^{\tau\tau}\left[\frac{\partial l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\tau}\right]\Bigg|_{\left(\hat{\boldsymbol{\beta}},\hat{\sigma}^2\right)},
$$

where $J^{\tau\tau}$ is the lower right corner of $\mathbf{J}^{-1}\left(\boldsymbol{\beta},\tau\right)$. It is easily seen that under $H_0:\tau=0$

$$\frac{\partial l_{mi}^{*}\left(\boldsymbol{\beta},\sigma^{2};\mathbf{X},\mathbf{y}\right)}{\partial\tau}=\frac{1}{\sigma^{2}}\mathbf{c}_{i}^{\mathsf{T}}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right),$$

$$\mathbf{J}\left(\boldsymbol{\beta},\tau\right)=\frac{1}{\sigma^{2}}\begin{bmatrix}\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda} & \mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\\ \mathbf{c}_{i}^{\mathsf{T}}\mathbf{X} & c_{ii}\end{bmatrix},$$

and $J^{\tau\tau}=\dfrac{\sigma^{2}}{r_{ii}}$ then under $H_0:\tau=0$

$$SC_{i}=\left.\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathsf{T}}\mathbf{c}_{i}\mathbf{c}_{i}^{\mathsf{T}}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{\sigma^{2}r_{ii}}\right|_{\left(\hat{\boldsymbol{\beta}},\hat{\sigma}^{2}\right)}=\frac{\hat{v}_{i}^{2}}{\hat{\sigma}^{2}r_{ii}}=t_{i}^{2}\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^{2}\right).$$

Proof of Theorem 3:

It follows from (5) that the corrected log-likelihood of $\mathbf{y}$ and the conditional corrected log-likelihood of $\mathbf{b}|\mathbf{y}$ for CDM, respectively, are given by

$$\begin{aligned}l_{ci}^{*}\left(\boldsymbol{\beta},\sigma^{2};\mathbf{X},\mathbf{y}\right)&=-\frac{n-1}{2}\log\left(2\pi\sigma^{2}\right)-\frac{1}{2}\log\left(\left|\mathbf{V}_{[i]}\right|\right)-\\&-\frac{1}{2\sigma^{2}}\left\{\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\operatorname{tr}\left(\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}\end{aligned}\qquad(7)$$

$$\begin{aligned}l_{\mathbf{b}_{ci}}^{*}\left(\boldsymbol{\beta},\sigma^{2};\mathbf{X},\mathbf{y}\right)&=-\frac{q}{2}\log\left(2\pi\sigma^{2}\right)-\frac{1}{2}\left[\log\left(\left|\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right|\right)\right]-\frac{1}{2\sigma^{2}}\left\{\mathbf{b}^{\mathsf{T}}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}\right.\right.\\&\left.\left.+\ \boldsymbol{\Sigma}^{-1}\right]\mathbf{b}-2\mathbf{b}^{\mathsf{T}}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\Sigma}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+F'\left(\boldsymbol{\beta}\right)\right]\right\},\end{aligned}\qquad(8)$$

where

$$F'\left(\boldsymbol{\beta}\right)=\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\left(\mathbf{I}-\mathbf{V}_{[i]}^{-1}\right)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\operatorname{tr}\left(\mathbf{I}-\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}.$$

The corrected score estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$ will be obtained with differentiating (7) with respect to $\boldsymbol{\beta}$ and σ^2 and (8) with respect to $\mathbf{b}$. Then we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i)}&=\left[\mathbf{X}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}-\operatorname{tr}\left(\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\mathbf{y}_{(i)}\\&=\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\mathbf{c}_{i}^{\mathsf{T}}\mathbf{X}/c_{ii}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}+c_{ii}\boldsymbol{\Lambda}\right]^{-1}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{y}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\mathbf{c}_{i}^{\mathsf{T}}\mathbf{y}/c_{ii}\right]\\&=\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\mathbf{c}_{i}^{\mathsf{T}}\mathbf{X}/c_{ii}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{y}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\mathbf{c}_{i}^{\mathsf{T}}\mathbf{y}/c_{ii}\right]\\&+\mathbf{O}_{p}\left(n^{-1}\right)\approx\hat{\boldsymbol{\beta}}-\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_{i}\frac{\hat{v}_{i}}{r_{ii}},\end{aligned}$$

$$
\begin{aligned}
(n-1)\,\hat{\sigma}^2_{(i)} &= \left[\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}_{(i)}\right]^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \left[\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}_{(i)}\right] - \operatorname{tr}\left(\mathbf{V}_{[i]}^{-1}\right) \hat{\boldsymbol{\beta}}_{(i)}^{\mathsf{T}} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}_{(i)} \\
&= \mathbf{y}_{(i)}^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \mathbf{y}_{(i)} - \hat{\boldsymbol{\beta}}_{(i)}^{\mathsf{T}} \mathbf{X}_{(i)}^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \mathbf{y}_{(i)} \\
&= \mathbf{y}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y} - \mathbf{y}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y}/c_{ii} \\
&- \left[\hat{\boldsymbol{\beta}} - \left(\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right)^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \frac{\hat{v}_i}{r_{ii}} + \mathbf{O}_p\left(n^{-1}\right)\right]^{\mathsf{T}} \left(\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y}\right. \\
&\left. - \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y}/c_{ii}\right) \\
&= n\hat{\sigma}^2 - \mathbf{y}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y}/c_{ii} + \mathbf{c}_i^{\mathsf{T}} \mathbf{X} \left[\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right]^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y} \frac{\hat{v}_i}{r_{ii}} \\
&- \mathbf{c}_i^{\mathsf{T}} \mathbf{X} \left[\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right]^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} \frac{\hat{v}_i}{c_{ii} r_{ii}} + \hat{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y}/c_{ii} \\
&+ O_p(1) = n\hat{\sigma}^2 - \frac{\hat{v}_i^2}{r_{ii}} + O_p(1),
\end{aligned}
$$

and hence, $\hat{\sigma}^2_{(i)} \approx \dfrac{n - t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1} \hat{\sigma}^2,$

$$
\tilde{\mathbf{b}}_{(i)} = \boldsymbol{\Sigma} \mathbf{U}_{(i)}^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \left(\mathbf{y}_{(i)} - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)}\right) \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma} \mathbf{U}^{\mathsf{T}} \mathbf{r}_i \frac{\hat{v}_i}{r_{ii}}.
$$

References

Banerjee, M. and Frees, E. W. (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association*, 92, 999–1005.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

Chatterjee, S. and Hadi, A. S. (1986). Influential observation, high leverage points and outliers in linear regression (with discussion). *Statistical Science*, 1, 379–416.

Christensen, R., Pearson, L. M. and Johnson, W. (1992). Case deletion diagnostics for mixed models. *Technometrics*, 34, 38–45.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, London.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

Fung, W. K., Zhong, X. P. and Wei, B. C. (2003). On estimation and influence diagnostics in linear mixed measurement errors models. *American Journal of Mathematical and Management Sciences*, 23, 37–59.

Gimenz, P. and Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models. *The Australian Journal of Statistics*, 39, 325–344.

Hanfelt, J. J. and Liang, K. Y. (1997). Approximate likelihood for generalized linear errors-in-variables models. *Journal of the Royal Statistical Society. Series B*, 59, 627–637.

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102.

Haslett, J. and Dillane, D. (2004). Application of 'delete=replace' to deletion diagnostics for variance component estimation in linear mixed model. *Journal of the Royal Statistical Society. Series B*, 66, 131–143.

Kelly, G. E. (1984). The influence function in the errors in variables problems, *The Annals of Statistics*, 12, 87–100.

Li, Z., Xu, W., and Zhu, L. (2009). Influence diagnostics and outlier tests for varying coefficient mixed models. *Journal of Multivariate Analysis*, 100, 2002–2017.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77, 127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48, 829–838.

Rio, M. (1988). On the potential in the estimation of linear functions in regression. *Communications in Statistics. Theory and Methods*, 17, 729–738.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science*, 6, 15–51.

Wellman, J. M. and Gunst, R. F. (1991). Influence diagnostics for linear measurement errors models. *Biometrika*, 78, 373–380.

Welsch, R. E. (1982). Influence functions and regression diagnostics. In *Modern Data Analysis* (R.L. Launer and A.F. Siegel, eds.), Academic, New York.

Zare, K., Rasekh, A. and Rasekhi, A. (2011). Estimation of variance components in linear mixed measurement error models. *Statistical Papers*, DOI 10.1007/s00362-011-0387-0.

Zewotir, T. and Galpin, J. S. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science*, 3, 153–177.

Zewotir, T. and Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed model, *Test*, 16, 58–75.

Zhong, X. P. and Wei, B. C. (1999). Influence analysis on linear models with random effects. *Applied Mathematics. A Journal of Chinese Universities. Ser. B.*, 14, 169–176.

Zhong, X. P., Fung, W. K. and Wei, B. C. (2002). Estimation in linear models with random effects and errors-in-variables. *Annals of the Institute of Statistical Mathematics*, 54, 595–606.

Zhong, X. P., Wei, B. C. and Fung, W. K. (2000). Influence analysis for linear measurement error models. *Annals of the Institute of Statistical Mathematics*, 52, 367–379.