Supplementary Materials for
A Diagnostic of Influential Cases
Based on the Information Complexity Criteria
in Generalized Linear Mixed Models

Junfeng Shang *
Department of Mathematics and Statistics
Bowling Green State University, USA

Abstract: Modeling diagnostics assess models by means of a variety of criteria. Each criterion typically performs its evaluation upon a specific inferential objective. For instance, the well-known DFBETAS in linear regression models are a modeling diagnostic which is applied to discover the influential cases in fitting a model. To facilitate the evaluation of generalized linear mixed models (GLMM), we develop a diagnostic for detecting influential cases based on the Information Complexity Criteria (ICOMP) for detecting the influential cases which substantially affect the model selection criterion ICOMP. In a given model, the diagnostic compares the ICOMP between the full data set and a case-deleted data set. The computational formula of the ICOMP is evaluated using the Fisher information matrix. A simulation study is accomplished and a real data set of cancer cells is analyzed using the logistic linear mixed model for illustrating the effectiveness of the proposed diagnostic in detecting the influential cases.

Key words: GLMM, ICOMP, case-deletion, Fisher information matrix, logistic regression model.

3.3 The Presentation of the Application Results

To describe the cancer cell data and to avoid the presence of extra-binomial variation, the model is written as

$$\log \frac{\mu_{ij}}{n - \mu_{ij}} = \beta + \xi_{1i} + \xi_{2ij},$$ (0.1)

where $i = 1, \ldots, 9$, $j = 1, \ldots, 3$. The number of locations is 9, and $\xi_{1i}$ are the random effects originating from the location. The number of dishes for each location is 3, and $\xi_{2ij}$ are the random effects coming from the dish of the location. That is, the random effects are initiated from the location and the error term for each dish. Therefore the total observed $y_i$ is 27, i.e., $N = 27$ and $n_i = 400$ for the binomial distribution.

*Phone: (419) 372-7457. Fax: (419) 372-6092. e-mail: jshang@bgsu.edu. Department of Mathematics and Statistics, 450 Math Science Building, Bowling Green State University, Bowling Green, OH 43403.
For model (0.1), $\beta$ is the fixed effect, so $p = 1$, and the corresponding design matrix $X$ is a $27 \times 1$ vector consisting of all 1’s. The dimension of random effects is 2, so $r = 2$.

The $\delta_{ICOMP(i)}$ values are calculated and plotted versus the index of case numbers in Figure 4. The values of $\delta_{ICOMP(i)}$ for cases 3 and 8 are quite outstanding, they are therefore possibly justified as influential, and labeled by the symbol of “#”.

For this cancer cell data set, we may want to see the estimates of the two cases whose $\delta_{ICOMP(i)}$ values are outstanding are individually eliminated from the data set. Figure 5 features the plot of case-deleted parameter estimates $\beta$ versus $\sigma_1^2$ in (a) and the plot of $\beta$ versus $\sigma_2^2$ in (b). The dots for cases 3 and 8 are labeled. It is easy to see that these estimates are quite away from the others, indicating that when case 3 or 8 is eliminated from the data set, the parameter estimates are significantly changed. As a result, the corresponding $\delta_{ICOMP(i)}$ values are very large. The results therefore demonstrate that the diagnostic $\delta_{ICOMP(i)}$ can effectively detect the comprehensive change of the parameter estimates when case is removed from the data set and further can successfully evaluate the magnitude of the influence of each case.

Figure 6 features the $S.\delta_{ICOMP(i)}$ values, and it shows that only $S.\delta_{ICOMP(3)}$ is greater than 2, which is 2.01. Thus, case 3 is influential, however, $S.\delta_{ICOMP(8)}$ is only 1.3, indicating that case 3 is not influential by the criterion of our benchmark.

To further examine the effectiveness of the proposed diagnostic, we then artificially changed the values for cases 3 and 8. We change case 3 because $\delta_{ICOMP(3)}$ is large. The Values for cases 3 and 8 are originally 66, 75, 80 and 88, 76, 90 respectively. For Version 1, we changed them to 104, 105, 116 and 120, 110, 117. Then the $\delta_{ICOMP(i)}$ values are featured in Figure 7. Since the two cases are changed to be close to the most of the other data, the $\delta_{ICOMP(i)}$ values for cases 3 and 8 are not significantly larger any more, and they become the normal sizes compared to the other $\delta_{ICOMP(i)}$ values. The two $\delta_{ICOMP(i)}$ values are marked by “#” in Figure 7. The $S.\delta_{ICOMP(i)}$ values are shown in Figure 8, none of them is greater than 2, so there is no influential case this time.

For Version 2, we changed them to 120, 111, 130 and 123, 134, 125 respectively, which are closer to the other data compared to those in Version 1. Then the $\delta_{ICOMP(i)}$ values are featured in Figure 9. The $\delta_{ICOMP(i)}$ values generally become smaller than those for Version 1. The two $\delta_{ICOMP(i)}$ values for cases 3 and 8 are specially marked by “#” in Figure 9 and the $S.\delta_{ICOMP(i)}$ values are featured in Figure 10. It is easy to report that they are not influential either this time.

From the previous application results, it is exhibited that the proposed diagnostic $\delta_{ICOMP(i)}$ performs well in detecting an influential case in the GLMMs.
Figure 1: $\delta_{ICOMP(i)}$ vs. Case Index $i$ for the Simulated Data

Figure 2: Parameter Estimates for the Simulated Data
Figure 3: Standardized $\delta_{ICOMP(i)}$ vs. Case Index $i$ for the Simulated Data

Table 1: Parameter Estimates for Cancer Cell Data

<table>
<thead>
<tr>
<th>Estimates</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$\sigma^2_j$</th>
<th>$\sigma^2_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full data</td>
<td>-0.7579</td>
<td>1.1294</td>
<td>0.1958</td>
<td>0.0057</td>
</tr>
<tr>
<td>Case 1 deleted</td>
<td>-0.8437</td>
<td>1.1165</td>
<td>0.1516</td>
<td>0.0038</td>
</tr>
<tr>
<td>Case 2 deleted</td>
<td>-0.7330</td>
<td>1.0416</td>
<td>0.2151</td>
<td>0.0109</td>
</tr>
<tr>
<td>Case 3 deleted</td>
<td>-0.6648</td>
<td>1.0543</td>
<td>0.1454</td>
<td>0.0093</td>
</tr>
<tr>
<td>Case 4 deleted</td>
<td>-0.7498</td>
<td>1.1792</td>
<td>0.2230</td>
<td>0.0062</td>
</tr>
<tr>
<td>Case 5 deleted</td>
<td>-0.7773</td>
<td>1.0097</td>
<td>0.2156</td>
<td>0.0038</td>
</tr>
<tr>
<td>Case 6 deleted</td>
<td>-0.7495</td>
<td>1.1211</td>
<td>0.2228</td>
<td>0.0076</td>
</tr>
<tr>
<td>Case 7 deleted</td>
<td>-0.8316</td>
<td>1.0950</td>
<td>0.1696</td>
<td>0.0069</td>
</tr>
<tr>
<td>Case 8 deleted</td>
<td>-0.6882</td>
<td>1.1500</td>
<td>0.1771</td>
<td>0.0062</td>
</tr>
<tr>
<td>Case 9 deleted</td>
<td>-0.7570</td>
<td>1.1347</td>
<td>0.2193</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

Appendix: The Fisher Information Matrix in Logistic Linear Models with Random Effects

In what follows, we will derive the Fisher Information matrix in logistic linear models with random effects. Note that in this derivation, we do not consider the extra-binomial variation because it is the scale parameter which can not affect the calculation of $C(Q)$. Although this scale parameter does affect the log-likelihood function in the ICOMP, for
Figure 4: $\delta_{ICOMP(i)}$ vs. Case Index $i$ for Cancer Cell Data

Figure 5: Parameter Estimates for Cancer Cell Data

the logistic regression model, realistically we try to model the data without extra-binomial variation. So ignoring the scale parameter $\sigma^2$ in the application part is quite practical.

As mentioned previously, for the logistic regression model, the link function can be written as

$$g(\mu) = \eta = \log\left(\frac{\pi}{1 - \pi}\right) = \log\left(\frac{\mu}{n - \mu}\right) = X\beta + U\xi.$$

By first-order Taylor expansion, the response variable has shifted to $Z$ from the original one $Y$ in the model. In the logistic regression model with random effects, the covariance matrix of the response variable $Z$ is $\Sigma = W + \sum_{i=1}^{r} U_i U_i' \sigma_i^2$, and note that the $W$ is a function of $\beta$ in the model, because of this, it is not trivial to derive the Fisher Information matrix
for the model. Based on the logit link function, the \( W \) is then written as

\[
W = \text{diag}\left\{ \frac{n_i}{\mu_i(n_i - \mu_i)} \right\}, \quad i = 1, \ldots, N.
\]

We suppose that the design matrix \( X \) is an \( N \times p \) matrix with each element \( \{x_{ij}\} \), \( i = 1, \ldots, N, \quad j = 1, \ldots, p \). This notation will be used in the following derivation.

First, without considering the constant \( 2\pi \), we have the log-likelihood function

\[
\log L(\beta, \sigma_1^2, \ldots, \sigma_r^2 | Z) = -\frac{1}{2} \log |\Sigma| - (Z - X\beta)'\Sigma^{-1}(Z - X\beta)/2. \tag{A.1}
\]

Now, we need to find the Fisher Information matrix \( F \) in this scenario, i.e., we need to derive the formula of the matrix \( F = -E_\theta \left\{ \frac{\partial^2 \log L(\beta, \sigma_1^2, \ldots, \sigma_r^2 | Z)}{\partial \theta \partial \theta'} \right\} \).
To find the first derivatives of the log-likelihood function with respect to the parameters, we have

\[
\frac{\partial \log L(\cdot)}{\partial \beta} = X'\Sigma^{-1} (Z - X\beta) - \frac{1}{2} \left[ \begin{array}{c} \text{tr}(\Sigma^{-1}\Sigma_{\beta_1}) \\ \vdots \\ \text{tr}(\Sigma^{-1}\Sigma_{\beta_p}) \end{array} \right] + \frac{1}{2} \left[ \begin{array}{c} (Z - X\beta)'\Sigma^{-1}\Sigma_{\beta_1}\Sigma^{-1}(Z - X\beta) \\ \vdots \\ (Z - X\beta)'\Sigma^{-1}\Sigma_{\beta_p}\Sigma^{-1}(Z - X\beta) \end{array} \right]
\]

and

\[
\frac{\partial \log L(\cdot)}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr}(\Sigma^{-1}U_iU_i') + \frac{1}{2}(Z - X\beta)'\Sigma^{-1}U_iU_i'\Sigma^{-1}(Z - X\beta), \ i = 1, \cdots, r,
\]
Figure 10: Standardized $\delta_{ICOMP(i)}$ vs. Case Index $i$ for Changed Cell Data Version 2

where

$$
\Sigma_{\beta_j} = \begin{bmatrix}
X_1j \left( \frac{\mu_1}{n_1(n_1-\mu_1)} - \frac{n_1-\mu_1}{n_1\mu_1} \right) & 0 & \cdots & 0 \\
0 & X_2j \left( \frac{\mu_2}{n_2(n_2-\mu_2)} - \frac{n_2-\mu_2}{n_2\mu_2} \right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & X_{Nj} \left( \frac{\mu_N}{n_N(n_N-\mu_N)} - \frac{n_N-\mu_N}{n_N\mu_N} \right)
\end{bmatrix},
$$

$j = 1, \ldots, p$, and $\mu_i = n_i\pi_i$ for $i = 1, \ldots, N$. It is noted that $n_i$ and $\pi_i$ are the parameters for the binomial response variable $y_i$.

For the purpose of simpler notation for the preceding first derivatives, we have

$$
L_1 = \begin{bmatrix}
\text{tr}(\Sigma^{-1}\Sigma_{\beta_1}) \\
\vdots \\
\text{tr}(\Sigma^{-1}\Sigma_{\beta_p})
\end{bmatrix}
$$

and

$$
L_2 = \begin{bmatrix}
(Z - X\beta)'\Sigma^{-1}\Sigma_{\beta_1}\Sigma^{-1}(Z - X\beta) \\
\vdots \\
(Z - X\beta)'\Sigma^{-1}\Sigma_{\beta_p}\Sigma^{-1}(Z - X\beta)
\end{bmatrix}.
$$

Based on the first derivatives, we can then obtain the second derivatives

$$
\frac{\partial^2 X'\Sigma^{-1}(Z - X\beta)}{\partial \beta \partial \beta'} = -X'\Sigma^{-1}X,
$$

and

$$
\frac{\partial^2 L_1}{\partial \beta \partial \beta'} = \begin{bmatrix}
\text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_p}) \\
\text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_p}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_p})
\end{bmatrix},
$$

$$
\frac{\partial^2 L_2}{\partial \beta \partial \beta'} = \begin{bmatrix}
\text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_1}\Sigma_{\beta_p}) \\
\text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_2}\Sigma_{\beta_p}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_1}) & \text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_2}) & \cdots & \text{tr}(\Sigma^{-1}\Sigma_{\beta_p}\Sigma_{\beta_p})
\end{bmatrix}.
$$

where

$$
\Sigma_{\beta_i\beta_j} = \begin{bmatrix}
X_1jX_1j \left( \frac{\mu_1}{n_1(n_1-\mu_1)} + \frac{n_1-\mu_1}{n_1\mu_1} \right) & 0 & \cdots & 0 \\
0 & X_2jX_2j \left( \frac{\mu_2}{n_2(n_2-\mu_2)} + \frac{n_2-\mu_2}{n_2\mu_2} \right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & X_{Nj}X_{Nj} \left( \frac{\mu_N}{n_N(n_N-\mu_N)} + \frac{n_N-\mu_N}{n_N\mu_N} \right)
\end{bmatrix},
$$
we therefore have the second derivative of $L_2$ as
\[
\frac{\partial^2 L_2}{\partial \beta \partial \beta'} = -2I_1 - 2I_2 + I_3.
\]

Further, we have the second derivative of the interaction term for the Fisher Information matrix $F$ as
\[
\frac{\partial \log L(\cdot)}{\partial \beta \partial \sigma^2_i} = -X'\Sigma^{-1}U_iU_i'\Sigma^{-1}(Z - X\beta) + \frac{1}{2} \begin{bmatrix}
\text{tr}(\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_1}) \\
\vdots \\
\text{tr}(\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_j}) \\
\vdots \\
\text{tr}(\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_p})
\end{bmatrix}
\]
\[
- \begin{bmatrix}
(Z - X\beta)'\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_1}\Sigma^{-1}(Z - X\beta) \\
\vdots \\
(Z - X\beta)'\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_j}\Sigma^{-1}(Z - X\beta) \\
\vdots \\
(Z - X\beta)'\Sigma^{-1}U_iU_i'\Sigma^{-1}\Sigma_{\beta_p}\Sigma^{-1}(Z - X\beta)
\end{bmatrix}.
\]

For the first derivative of the log-likelihood function in (A.1) for model (??) with respect to the variance $\sigma^2_i$, we have
\[
\frac{\partial \log L(\cdot)}{\partial \sigma^2_i} = -\frac{1}{2} \text{tr}(\Sigma^{-1}U_iU_i') + \frac{1}{2}(Z - X\beta)'\Sigma^{-1}U_iU_i'\Sigma^{-1}(Z - X\beta),
\]
\[i = 1, \ldots, r.\]
Relying on the previous equation, we can have the second derivatives with respect to the variance term \( \sigma_i^2 \), we can therefore have

\[
\frac{\partial \log L(\cdot)}{\partial \sigma_i^2} = \frac{1}{2} \text{tr}(\Sigma_i^{-1}U_i'U_i\Sigma_i^{-1}U_i')
\]

\[
-(Z - X\beta)'\Sigma_i^{-1}U_i'U_i\Sigma_i^{-1}U_i'\Sigma_i^{-1}(Z - X\beta),
\]

\(i = 1, \ldots, r\).

Next, we need to obtain the expectations of the second derivatives and then to derive the matrix \( F \). We first have the expectation regarding the \( \beta \) term as

\[
E_\theta \left[ \frac{\partial^2 \log L(\cdot)}{\partial \beta \partial \beta'} \right] = -X'\Sigma^{-1}X - \frac{1}{2} E_\theta \left[ \frac{\partial^2 L_1}{\partial \beta \partial \beta'} \right] - 2E(I_1) - 2E(I_2) + E(I_3). \tag{A.3}
\]

To obtain the expectations of all parts in equation (A.3), we have

\[
E(I_1) = 0,
\]

\[
E(I_2) = \begin{bmatrix}
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) \\
\text{tr}(\Sigma_{\beta_2}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_2}\Sigma_{\beta_1}^{-1}) \\
\vdots & \ddots & \vdots \\
\text{tr}(\Sigma_{\beta_p}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_p}\Sigma_{\beta_1}^{-1})
\end{bmatrix}, \tag{A.4}
\]

\[
E(I_3) = \begin{bmatrix}
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) \\
\text{tr}(\Sigma_{\beta_2}\Sigma_{\beta_2}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_2}\Sigma_{\beta_2}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_2}\Sigma_{\beta_2}^{-1}) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\text{tr}(\Sigma_{\beta_p}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_p}\Sigma_{\beta_1}^{-1}) & \cdots & \text{tr}(\Sigma_{\beta_p}\Sigma_{\beta_1}^{-1})
\end{bmatrix}, \tag{A.5}
\]

\[
E_\theta \left[ \frac{\partial^2 \log L(\cdot)}{\partial \beta \partial \sigma_i^2} \right] = \frac{1}{2} \begin{bmatrix}
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) \\
\vdots \\
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) \\
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1}) \\
\vdots \\
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1})
\end{bmatrix} - \begin{bmatrix}
U_iU_i'\Sigma_{\beta_1}^{-1}
\vdots
U_iU_i'\Sigma_{\beta_1}^{-1}
\vdots
U_iU_i'\Sigma_{\beta_1}^{-1}
\end{bmatrix}, \text{ and} \tag{A.6}
\]

\[
E_\theta \left[ \frac{\partial^2 \log L(\cdot)}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{2} \begin{bmatrix}
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1})
\vdots
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1})
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1})
\vdots
\text{tr}(\Sigma_{\beta_1}\Sigma_{\beta_1}^{-1})
\end{bmatrix}, \tag{A.7}
\]

\(i = 1, \ldots, r\).

Thus, we can arrive at

\[
F = -E_\theta \left( \frac{\partial^2 \log L(\beta, \sigma_1^2, \ldots, \sigma_r^2 | Z)}{\partial \beta \partial \theta'} \right) = \begin{bmatrix} F_1(p \times p) & F_2(p \times r) \\ F_3(r \times p) & F_4(r \times r) \end{bmatrix},
\]

\[10\]
where

\[
F_1 = -E_\theta \left[ \frac{\partial^2 \log L(\cdot)}{\partial \beta \partial \beta'} \right] \quad \text{and} \\
F_2 = -E_\theta \left[ \frac{\partial^2 \log L(\cdot)}{\partial \beta \partial \sigma_i^2} \right],
\]

and the element of \( F_3 \) is \( \frac{1}{2} \text{tr}(\Sigma^{-1}U_iU_i^T) \) with diagonal elements for \( i = j \) and non-diagonal elements for \( i \neq j \). Note that \( F_1, F_2, \) and \( F_3 \) are determined by (A.3), (A.4), (A.5), (A.6), and (A.7).

Hence, the Fisher information matrix \( F \) for the logistic regression model is established. The estimate of \( F \) can be calculated by simply plugging in the estimates of parameters.