

## 0.1 Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

## 0.2 Influence Analysis for LME Models

The linear mixed effects model is a useful methodology for fitting a wide range of models. However, linear mixed effects models are known to be sensitive to outliers. ? advises that identification of outliers is necessary before conclusions may be drawn from the fitted model.

Standard statistical packages concentrate on calculating and testing parameter estimates without considering the diagnostics of the model. The assessment of the effects of perturbations in data, on the outcome of the analysis, is known as statistical influence analysis. Influence analysis examines the robustness of the model. Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models. Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (?). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

# 1 Extension of techniques to LME Models

Model diagnostic techniques, well established for classical models, have since been adapted for use with linear mixed effects models. Diagnostic techniques for LME models are inevitably more difficult to implement, due to the increased complexity.

Beckman, Nachtsheim and Cook (1987) ? applied the local influence method of Cook (1986) to the analysis of the linear mixed model.

While the concept of influence analysis is straightforward, implementation in mixed models is more complex. Update formulae for fixed effects models are available only when the covariance parameters are assumed to be known.

If the global measure suggests that the points in  $U$  are influential, the nature of that influence should be determined. In particular, the points in  $U$  can affect the following

- the estimates of fixed effects,
- the estimates of the precision of the fixed effects,
- the estimates of the covariance parameters,
- the estimates of the precision of the covariance parameters,
- fitted and predicted values.

## 1.1 Residuals diagnostics in mixed models

The marginal and conditional means in the linear mixed model are  $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$  and  $E[\mathbf{Y}|\mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ , respectively.

A residual is the difference between an observed quantity and its estimated or predicted value. In the mixed model you can distinguish marginal residuals  $r_m$  and conditional residuals  $r_c$ .

## 1.2 Marginal and Conditional Residuals

A marginal residual is the difference between the observed data and the estimated (marginal) mean,  $r_{mi} = y_i - x_i' \hat{\mathbf{b}}$  A conditional residual is the difference between the observed data and the predicted value of the observation,  $r_{ci} = y_i - x_i' \hat{\mathbf{b}} - z_i' \hat{\boldsymbol{\gamma}}$

In linear mixed effects models, diagnostic techniques may consider ‘conditional’ residuals. A conditional residual is the difference between an observed value  $y_i$  and the conditional predicted value  $\hat{y}_i$ .

$$\epsilon_{\hat{y}_i} = y_i - \hat{y}_i = y_i - (X_i \hat{\boldsymbol{\beta}} + Z_i \hat{\boldsymbol{\gamma}})$$

However, using conditional residuals for diagnostics presents difficulties, as they tend to be correlated and their variances may be different for different subgroups, which can lead to erroneous conclusions.

$$r_{mi} = x_i^T \hat{\beta} \tag{1}$$

### 1.3 Marginal Residuals

$$\begin{aligned} \hat{\beta} &= (X^T R^{-1} X)^{-1} X^T R^{-1} Y \\ &= BY \end{aligned}$$

## 2 Standardized and studentized residuals

To alleviate the problem caused by inconstant variance, the residuals are scaled (i.e. divided) by their standard deviations. This results in a ‘standardized residual’. Because true standard deviations are frequently unknown, one can instead divide a residual by the estimated standard deviation to obtain the ‘studentized residual’.

### 2.1 Standardization

A random variable is said to be standardized if the difference from its mean is scaled by its standard deviation. The residuals above have mean zero but their variance is unknown, it depends on the true values of  $\theta$ . Standardization is thus not possible in practice.

### 2.2 Studentization

Instead, you can compute studentized residuals by dividing a residual by an estimate of its standard deviation.

### 2.3 Internal and External Studentization

If that estimate is independent of the  $i$ -th observation, the process is termed ‘external studentization’. This is usually accomplished by excluding the  $i$ -th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be internally studentized.

Externally studentized residual require iterative influence analysis or a profiled residuals variance.

### 2.4 Computation

The computation of internally studentized residuals relies on the diagonal entries of  $V(\hat{\theta}) - Q(\hat{\theta})$ , where  $Q(\hat{\theta})$  is computed as

$$Q(\hat{\theta}) = X(X'Q(\hat{\theta})^{-1}X)X^{-1}$$

### 2.5 Pearson Residual

Another possible scaled residual is the ‘Pearson residual’, whereby a residual is divided by the standard deviation of the dependent variable. The Pearson residual can be used when the variability of  $\hat{\beta}$  is disregarded in the underlying assumptions.

### 3 Covariance Parameters

The unknown variance elements are referred to as the covariance parameters and collected in the vector  $\theta$ .

#### 3.1 Methods and Measures

The key to making deletion diagnostics useable is the development of efficient computational formulas, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model.

? lists several established methods of analyzing influence in LME models. These methods include

- Cook's distance for LME models,
- likelihood distance,
- the variance (information) ration,
- the Cook-Weisberg statistic,
- the Andrews-Prebignon statistic.

## 4 Influence analysis

Likelihood based estimation methods, such as ML and REML, are sensitive to unusual observations. Influence diagnostics are formal techniques that assess the influence of observations on parameter estimates for  $\beta$  and  $\theta$ . A common technique is to refit the model with an observation or group of observations omitted.

? examines a group of methods that examine various aspects of influence diagnostics for LME models. For overall influence, the most common approaches are the ‘likelihood distance’ and the ‘restricted likelihood distance’.

### 4.1 Cook’s 1986 paper on Local Influence

Cook 1986 introduced methods for local influence assessment. These methods provide a powerful tool for examining perturbations in the assumption of a model, particularly the effects of local perturbations of parameters of observations.

The local-influence approach to influence assessment is quite different from the case deletion approach, comparisons are of interest.

### 4.2 Overall Influence

An overall influence statistic measures the change in the objective function being minimized. For example, in OLS regression, the residual sums of squares serves that purpose. In linear mixed models fit by maximum likelihood (ML) or restricted maximum likelihood (REML), an overall influence measure is the likelihood distance [Cook and Weisberg ].

## 5 Iterative and non-iterative influence analysis

? highlights some of the issue regarding implementing mixed model diagnostics.

A measure of total influence requires updates of all model parameters.

however, this doesnt increase the procedures execution time by the same degree.

### 5.1 Iterative Influence Analysis

For linear models, the implementation of influence analysis is straightforward. However, for LME models, the process is more complex. Update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. A measure of total influence requires updates of all model parameters. This can only be achieved in general is by omitting observations, then refitting the model.

? describes the choice between iterative influence analysis and non-iterative influence analysis.

## 6 Computation and Notation

with  $\mathbf{V}$  unknown, a standard practice for estimating  $\mathbf{X}\boldsymbol{\beta}$  is to estimate the variance components  $\sigma_j^2$ , compute an estimate for  $\mathbf{V}$  and then compute the projector matrix  $\mathbf{A}$ ,  $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$ .

? remarks that  $\mathbf{D}$  is a block diagonal with the  $i$ -th block being  $u\mathbf{I}$



## 7 Measures 2

### 7.1 Cook's Distance

- For variance components  $\gamma$

Diagnostic tool for variance components

$$C_{\theta i} = (\hat{(\theta)}_{[i]} - \hat{(\theta)})^T \text{cov}(\hat{(\theta)})^{-1} (\hat{(\theta)}_{[i]} - \hat{(\theta)})$$

## 8 Haslett's Analysis

For fixed effect linear models with correlated error structure Haslett (1999) showed that the effects on the fixed effects estimate of deleting each observation in turn could be cheaply computed from the fixed effects model predicted residuals.

Appendices

## 9 The Hat Matrix

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (2)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (3)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned} \text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2 \end{aligned} \quad (4)$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 10 Sherman Morrison Woodbury Formula

The ‘Sherman Morrison Woodbury’ Formula is a well known result in linear algebra;

$$(A + a^T B)^{-1} = A^{-1} - A^{-1} a^T (I - b A^{-1} a^T)^{-1} b A^{-1} \quad (5)$$

This result is highly useful for analyzing regression diagnostics, and for matrices inverses in general. Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T) (X^T X)^{-1}}{1 - x_i^T (X^T X)^{-1} x_i} \quad (6)$$

The projection matrix  $H$  (also known as the hat matrix), is a well known identity that maps the fitted values  $\hat{Y}$  to the observed values  $Y$ , i.e.  $\hat{Y} = HY$ .

$$H = X(X^T X)^{-1} X^T \quad (7)$$

$H$  describes the influence each observed value has on each fitted value. The diagonal elements of the  $H$  are the ‘leverages’, which describe the influence each observed value has on the fitted value for that same observation. The residuals ( $R$ ) are related to the observed values by the following formula:

$$R = (I - H)Y \quad (8)$$

The variances of  $Y$  and  $R$  can be expressed as:

$$\begin{aligned}\text{var}(Y) &= H\sigma^2 \\ \text{var}(R) &= (I - H)\sigma^2\end{aligned}\tag{9}$$

Updating techniques allow an economic approach to recalculating the projection matrix,  $H$ , by removing the necessity to refit the model each time it is updated. However this approach is known for numerical instability in the case of down-dating.

## 10.1 Hat Values for MCS regression

With  $A$  as the averages and  $D$  as the casewise differences.

`fit = lm(D~A)`

$$H = A (A^\top A)^{-1} A^\top,$$

## 11 Cross Validation

Cross validation techniques for linear regression employ the use ‘leave one out’ recalculations. In such procedures the regression coefficients are estimated for  $n - 1$  covariates, with the  $Q^{th}$  observation omitted.

Let  $\hat{\beta}$  denote the least square estimate of  $\beta$  based upon the full set of observations, and let  $\hat{\beta}^{-Q}$  denote the estimate with the  $Q^{th}$  case excluded.

In leave-one-out cross validation, each observation is omitted in turn, and a regression model is fitted on the rest of the data. Cross validation is used to estimate the generalization error of a given model. alternatively it can be used for model selection by determining the candidate model that has the smallest generalization error.

Evidently leave-one-out cross validation has similarities with ‘jackknifing’, a well known statistical technique. However cross validation is used to estimate generalization error, whereas the jackknife technique is used to estimate bias.

### 11.1 Cross Validation: Updating standard deviation

The variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (10)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1} \quad (11)$$

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n - 2} \quad (12)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n - 2} \quad (13)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n - 2} \quad (14)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (15)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n - 1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n - 1}, \quad (16)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n - 1}. \quad (17)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (18)$$

## 12 Updating Estimates

### 12.1 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row.

### 12.2 Updating Standard deviation

A simple, but useful, example of updating is the updating of the standard deviation when an observation is omitted, as practised in statistical process control analyzes. From first principles, the variance of a data set can be calculated using the following formula.

$$S^2 = \frac{\sum_{i=1}^n (x_i^2) - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1} \quad (19)$$

While using bivariate data, the notation  $Sxx$  and  $Syy$  shall apply hither to the variance of  $x$  and of  $y$  respectively. The covariance term  $Sxy$  is given by

$$Sxy = \frac{\sum_{i=1}^n (x_i y_i) - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{n - 1}. \quad (20)$$

### 12.3 Updating of Regression Estimates

Updating techniques are used in regression analysis to add or delete rows from a model, allowing the analyst the effect of the observation associated with that row. In time series problems, there will be scientific interest in the changing relationship between variables. In cases where there a single row is to be added or deleted, the procedure used is equivalent to a geometric rotation of a plane.

Consider a  $p \times p$  matrix  $X$ , from which a row  $x_i^T$  is to be added or deleted. ? sets  $A = X^T X$ ,  $a = -x_i^T$  and  $b = x_i^T$ , and writes the above equation as

$$(X^T X \pm x_i x_i^T)^{-1} = (X^T X)^{-1} \mp \frac{(X^T X)^{-1} (x_i x_i^T (X^T X)^{-1})}{1 - x_i^T (X^T X)^{-1} x_i} \quad (21)$$

### 12.4 Updating Regression Estimates

Let the observation  $j$  be omitted from the data set. The estimates for the variance identities can be updating using minor adjustments to the full sample estimates. Where  $(j)$  denotes that the  $j$ th has been omitted, these identities are

$$Sxx^{(j)} = \frac{\sum_{i=1}^n (x_i^2) - (x_j)^2 - \frac{((\sum_{i=1}^n x_i) - x_j)^2}{n-1}}{n-2} \quad (22)$$

$$Syy^{(j)} = \frac{\sum_{i=1}^n (y_i^2) - (y_j)^2 - \frac{((\sum_{i=1}^n y_i) - y_j)^2}{n-1}}{n-2} \quad (23)$$

$$Sxy^{(j)} = \frac{\sum_{i=1}^n (x_i y_i) - (y_j x_j) - \frac{((\sum_{i=1}^n x_i) - x_j)((\sum_{i=1}^n y_i) - y_j)}{n-1}}{n-2} \quad (24)$$

The updated estimate for the slope is therefore

$$\hat{\beta}_1^{(j)} = \frac{Sxy^{(j)}}{Sxx^{(j)}} \quad (25)$$

It is necessary to determine the mean for  $x$  and  $y$  of the remaining  $n-1$  terms

$$\bar{x}^{(j)} = \frac{(\sum_{i=1}^n x_i) - (x_j)}{n-1}, \quad (26)$$

$$\bar{y}^{(j)} = \frac{(\sum_{i=1}^n y_i) - (y_j)}{n-1}. \quad (27)$$

The updated intercept estimate is therefore

$$\hat{\beta}_0^{(j)} = \bar{y}^{(j)} - \hat{\beta}_1^{(j)} \bar{x}^{(j)}. \quad (28)$$

## 12.5 Inference on intercept and slope

$$\hat{\beta}_1 \pm t_{(\alpha, n-2)} \sqrt{\frac{S^2}{(n-1)S_x^2}} \quad (29)$$

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \quad (30)$$

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_0)} \quad (31)$$

### 12.5.1 Inference on correlation coefficient

This test of the slope is coincidentally the equivalent of a test of the correlation of the  $n$  observations of  $X$  and  $Y$ .

$$\begin{aligned} H_0 : \rho_{XY} &= 0 \\ H_A : \rho_{XY} &\neq 0 \end{aligned} \quad (32)$$



## 13 Lesaffre's paper.

Lesaffre considers the case-weight perturbation approach.

Cook's 86 describes a local approach wherein each case is given a weight  $w_i$  and the effect on the parameter estimation is measured by perturbing these weights. Choosing weights close to zero or one corresponds to the global case-deletion approach.

Lesaffre describes the displacement in log-likelihood as a useful metric to evaluate local influence

Lesaffre describes a framework to detect outlying observations that matter in an LME model. Detection should be carried out by evaluating diagnostics  $C_i$ ,  $C_i(\alpha)$  and  $C_i(D, \sigma^2)$ .

Lesaffre defines the total local influence of individual  $i$  as

$$C_i = 2|\Delta_i L^{-1} \Delta_i|. \quad (33)$$

The influence function of the MLEs evaluated at the  $i$ th point  $IF_i$ , given by

$$IF_i = -L^{-1} \Delta_i \quad (34)$$

can indicate how  $\hat{\theta}$  changes as the weight of the  $i$ th subject changes.

The manner by which influential observations distort the estimation process can be determined by inspecting the interpretable components in the decomposition of the above measures of local influence.

Lesaffre comments that there is no clear way of interpreting the information contained in the angles, but that this doesn't mean the information should be ignored.