Chapter 1

Linear Mixed Effects Models

While the method comparison problem is conventionally poised in the context of two methods of measurements, LME models allow for a straightforward analysis whereby several methods of measurement can be measured simulataneously. However simple models can only indicate agreement of lack thereof, and the presence of inter-method bias. To consider more complex questions, more complex LME models would be required.

Linear mixed effects (LME) models can facilitate greater understanding of the potential causes of bias and differences in precision between two sets of measurement. Due to computation complexity, LME models have not seen widespread use until many well known statistical software applications began facilitating them. Consequently LME approaches have seen increased use as a framework for method comparison studies in recent years (Lai & Shaio, Carstensen (2004); Carstensen et al. (2008) and Roy (2009) as examples), due to a higher profile in scientific literature, and the increased availability of capable software. Additionally development of a framework for LME model diagnostics has progressed, thanks to authors such as Schabenberger (2004), Christensen et al. (1992), Cook (1986) West et al. (2007), amongst others.

1.1 Linear Mixed Effects Models

A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects (also known as variance components). LME models are a generalization of the classical linear model, which contain fixed effects only. When the levels of factors are considered to be sampled from a population, and each level is not of particular interest, they are considered random quantities with associated variances.

The effects of the levels, as described, are known as random effects. Random effects are represented by unobservable normally distributed random variables. Conversely fixed effects are considered non-random and the levels of each factor are of specific interest.

Fisher (1918) introduced variance components models for use in genetical studies. Whereas an estimate for variance must take an non-negative value, an individual variance component, i.e. a component of the overall variance, may be negative.

The methodology has developed since, including contributions from Tippett (1931), who extend the use of variance components into linear models, and Eisenhart (1947), who introduced the 'mixed model' terminology and formally distinguished between mixed and random effects models. Henderson (1950) devised a methodology for deriving estimates for both the fixed effects and the random effects, using a set of equations that would become known as 'mixed model equations' or 'Henderson's equations'. LME methodology is further enhanced by Henderson's later works (Henderson, 1953; Henderson et al., 1959, 1963, 1973, 1984). The key features of Henderson's work provide the basis for the estimation techniques.

Hartley and Rao (1967) demonstrated that unique estimates of the variance components could be obtained using maximum likelihood methods. However these estimates are known to be biased 'downwards' (i.e. underestimated), because of the assumption that the fixed estimates are known, rather than being estimated from the data. **?** produced an alternative set of estimates, known as the restricted maximum likelihood (REML) estimates, that do not require the fixed effects to be known.

Thusly there is a distinction the REML estimates and the original estimates, now commonly referred to as ML estimates.

1.1.1 Laird-Ware Model

Laird and Ware (1982) provides a form of notation for notation for LME models that has since become the standard form, or the basis for more complex formulations. Due to computation complexity, linear mixed effects models have not seen widespread use until many well known statistical software applications began facilitating them. SAS Institute added PROC MIXED to its software suite in 1992 (Singer, 1998). Pinheiro and Bates (1994) described how to compute LME models in the `S-plus` environment.

Linear mixed effects models (LME) differs from the conventional linear model in that it has both fixed effects and random effects regressors, and coefficients thereof. The notation provided here is generic, and will be adapted to accord with complex formulations that will be encountered in due course. Using Laird-Ware form, the LME model is commonly described in matrix form,

$$Y = X\beta + Zb + \epsilon \tag{1.1}$$

Y is the $n \times 1$ response vector, where n is the number of observations. β is a $p \times 1$ vector of fixed p effects, with the first element being the population mean. X and Z are $n \times p$ and $n \times q$ "model matrices" for fixed effects and random effects respectively, comprising 0s or 1s, depending on the observation is question. The vector of residuals, v(e) has dimension $n \times 1$. The random effects are contained in the $q \times 1$ vector b.

1.1.2 LME Model Estimation

Estimation of LME models involve two complementary estimation issues'; estimating the vectors of the fixed and random effects estimates $\hat{\beta}$ and $\hat{b}$ and estimating the variance covariance matrices G and Σ. Inference about fixed effects have become known as 'estimates', while inferences about random effects have become known as 'predictions'.

The most common approach to obtain estimators are Best Linear Unbiased Estimator (BLUE) and Best Linear Unbiased Predictor (BLUP). For an LME model given by (1.1), the BLUE of $\hat{\beta}$ is given by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y, \tag{1.2}$$

whereas the BLUP of $\hat{b}$ is given by

$$\hat{b} = DZ'V^{-1}(y - X\hat{\beta}). \tag{1.3}$$

Henderson's Equations

Because of the dimensionality of V (i.e. $n \times n$) computing the inverse of V can be difficult. As a way around the this problem Henderson (1953); Henderson et al. (1959, 1963, 1973, 1984) offered a more simpler approach of jointly estimating $\hat{\beta}$ and $\hat{b}$. Henderson (1950) made the (ad-hoc) distributional assumptions $y|b \sim \mathrm{N}(X\beta + Zb, \Sigma)$ and $b \sim \mathrm{N}(0, D)$, and proceeded to maximize the joint density of y and b

$$\left|\begin{matrix} G & 0 \\ 0 & \Sigma \end{matrix}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix}'\begin{pmatrix} G & 0 \\ 0 & \Sigma \end{pmatrix}^{-1}\begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix}\right\}, \tag{1.4}$$

with respect to β and b, which ultimately requires minimizing the criterion

$$(y - X\beta - Zb)'\Sigma^{-1}(y - X\beta - Zb) + b'D^{-1}b. \tag{1.5}$$

This leads to the mixed model equations

$$\begin{pmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & X'\Sigma^{-1}X + G^{-1} \end{pmatrix}\begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{pmatrix}. \tag{1.6}$$

Using these equations, obtaining the estimates requires the inversion of a matrix of dimension $p + q \times p + q$, considerably smaller in size than V. Henderson et al. (1963) shows that these mixed model equations do not depend on normality and that $\hat{\beta}$ and $\hat{b}$ are the BLUE and BLUP under general conditions, provided G and Σ are known.

Robinson (1991) points out that although Henderson (1950) initially referred to the estimates $\hat{\beta}$ and $\hat{b}$ from (1.6) as "joint maximum likelihood estimates", Henderson (1973) later advised that these estimates should not be referred to as "maximum likelihood" as the function being maximized in (1.5) is a joint density rather than a likelihood function. Lee et al. (2006) remarks that it is clear that Henderson used joint estimation for computational purposes, without recognizing the theoretical implications.

Estimation of the Fixed Parameters

The vector y has marginal density $y \sim \mathrm{N}(X\beta, V)$, where $V = \Sigma + ZDZ'$ is specified through the variance component parameters θ. The log-likelihood of the fixed parameters (β, θ) is

$$\ell(\beta, \theta|y) = -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta), \tag{1.7}$$

and for fixed θ the estimate $\hat{\beta}$ of β is obtained as the solution of

$$(X'V^{-1}X)\beta = X'V^{-1}y. \tag{1.8}$$

Substituting $\hat{\beta}$ from (1.8) into $\ell(\beta, \theta|y)$ from (1.7) returns the *profile* log-likelihood

$$\begin{aligned} \ell_P(\theta \mid y) &= \ell(\hat{\beta}, \theta \mid y) \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) \end{aligned}$$

of the variance parameter θ. Estimates of the parameters θ specifying V can be found by maximizing $\ell_P(\theta \mid y)$ over θ. These are the ML estimates. For REML estimation the *restricted* log-likelihood is defined as

$$\ell_R(\theta \mid y) = \ell_P(\theta \mid y) - \frac{1}{2}\log|X'VX|.$$

The REML approach does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function derived from a data set, transformed to remove the irrelevant influences (Dodge, 2003). Restricted maximum likelihood is often preferred to maximum likelihood because REML estimation reduces the bias in the variance component by taking into account the loss of degrees of freedom that results from estimating the fixed effects in β. Restricted maximum likelihood also handles high correlations more effectively, and is less sensitive to outliers than maximum likelihood. The problem with REML for model building is that the likelihoods obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML should be used instead.

Estimation of the Random Effects

The established approach for estimating the random effects is to use the best linear predictor of b from y, which for a given β equals $GZ'V^{-1}(y-X\beta)$. In practice β is replaced by an estimator such as $\hat{\beta}$ from (1.8) so that $\hat{b} = DZ'V^{-1}(y - X\hat{\beta})$. Pre-multiplying by the appropriate matrices it is straightforward to show that these estimates $\hat{\beta}$ and $\hat{b}$ satisfy the equations in (1.6).

Algorithms for Likelihood Function Optimization

Iterative numerical techniques are used to optimize the log-likelihood function and estimate the covariance parameters θ. The procedure is subject to the constraint that R and G are both positive definite. The most common iterative algorithms for optimizing the likelihood function are the Newton-Raphson method, which is the preferred method, the expectation maximization (EM) algorithm and the Fisher scoring methods.

The EM algorithm, introduced by Dempster et al. (1977), is an iterative technique for maximizing complicated likelihood functions. The algorithm alternates between performing an expectation (E) step and the maximization (M) step. The 'E' step computes the expectation of the log-likelihood evaluated using the current estimate for the variables. In the 'M' step, parameters that maximize the expected log-likelihood, found on the previous 'E' step, are computed. These parameter estimates are then used to determine the distribution of the variables in the next 'E' step. The algorithm alternatives between these two steps until convergence is reached.

The main drawback of the EM algorithm is its slow rate of convergence. Consequently the EM algorithm is rarely used entirely in LME estimation, instead providing an initial set of values that can be passed to other optimization techniques.

The Newton-Raphson (NR) method is the most common, and recommended technique for ML and REML estimation. The NR algorithm minimizes an objective function defines as -2 times the log likelihood for the covariance parameters θ. At every iteration the NR algorithm requires the calculation of a vector of partial derivatives, known as the gradient, and the second derivative matrix with respect to the covariance parameters. This is known as the observed Hessian matrix. Due to the Hessian matrix, the NR algorithm is more time-consuming, but convergence is reached with fewer iterations compared to the EM algorithm. The Fisher scoring algorithm is an variant of the NR algorithm that is more numerically stable and likely to converge, but not recommended to obtain final estimates.

The Extended Likelihood

The desire to have an entirely likelihood-based justification for estimates of random effects, in contrast to Henderson's equation, has motivated Pawitan (2001, page 429) to define the *extended likelihood.* He remarks "In mixed effects modelling the extended likelihood has been called *h-likelihood* (for hierarchical likelihood) by Lee and Nelder (1996), while in smoothing literature it is known as the *penalized likelihood* (e.g. Green and Silverman 1994)." The extended likelihood can be written $L(\beta, \theta, b|y) = p(y|b;\beta,\theta)p(b;\theta)$ and adopting the same distributional assumptions used by Henderson (1950) yields the log-likelihood function

$$\begin{aligned}\ell_h(\beta,\theta,b|y) &= -\frac{1}{2}\left\{\log|\Sigma| + (y - X\beta - Zb)'\Sigma^{-1}(y - X\beta - Zb)\right.\\ &\qquad\qquad \left. + \log|D| + b'D^{-1}b\right\}.\end{aligned}$$

Given θ, differentiating with respect to β and b returns Henderson's equations in (1.6).

The LME model as a general linear model

Henderson's equations in (1.6) can be rewritten $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

$$\delta = \begin{pmatrix} \beta \\ b \end{pmatrix}, \ y_a = \begin{pmatrix} y \\ \psi \end{pmatrix}, \ T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}, \text{ and } W = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix},$$

where Lee et al. (2006) describe $\psi = 0$ as quasi-data with mean $\mathrm{E}(\psi) = b$. Their formulation suggests that the joint estimation of the coefficients β and b of the linear mixed effects model can be derived via a classical augmented general linear model $y_a = T\delta + \varepsilon$ where $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = W$, with *both* β and b appearing as fixed parameters. The usefulness of this reformulation of an LME as a general linear model will be revisited.

1.2 Repeated Measurements in LME models

In many statistical analyzes, the need to determine parameter estimates where multiple measurements are available on each of a set of variables often arises. Further to Lam et al. (1999), Hamlett et al. (2004) performs an analysis of the correlation of replicate measurements, for two variables of interest, using LME models.

Let y_{Aij} and y_{Bij} be the jth repeated observations of the variables of interest A and B taken on the ith subject. The number of repeated measurements for each variable may differ for each individual. Both variables are measured on each time points. Let n_i be the number of observations for each variable, hence $2 \times n_i$ observations in total.

It is assumed that the pair y_{Aij} and y_{Bij} follow a bivariate normal distribution.

$$\begin{pmatrix} y_{Aij} \\ y_{Bij} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\mu} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$$

The matrix $\boldsymbol{\Sigma}$ represents the variance component matrix between response variables at a given time point j.

$$\mathbf{\Sigma} = \left(\begin{array}{cc} \sigma^2_A & \sigma_{AB} \\ \sigma_{AB} & \sigma^2_B \end{array} \right)$$

σ^2_A is the variance of variable A, σ^2_B is the variance of variable B and σ_{AB} is the covariance of the two variable. It is assumed that $\mathbf{\Sigma}$ does not depend on a particular time point, and is the same over all time points.

1.3 The Variance Covariance Matrix

The LME model can be written

$$\mathrm{y_i = X_i\beta + Z_ib_i + \epsilon_i},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is a vector of fixed effects, and $\boldsymbol{X}_i$ is a corresponding $2n_i \times 3$ design matrix for the fixed effects. The random effects are expressed in the vector $\boldsymbol{b} = (b_1, b_2)'$, with $\boldsymbol{Z}_i$ the corresponding $2n_i \times 2$ design matrix. The vector $\boldsymbol{\epsilon}_i$ is a $2n_i \times 1$ vector of residual terms. Random effects and residuals are assumed to be independent of each other. The variance matrix of $\mathbf{Y}$, denoted $\mathbf{V}$, is an $n \times n$ matrix that can be expressed as

$$\mathbf{V} = \mathrm{Var}(\mathbf{Xb} + \mathbf{Zb} + \mathbf{e}) \tag{1.9}$$

$$\mathbf{V} = \mathrm{Var}(\mathbf{Xb}) + \mathrm{Var}(\mathbf{Zb}) + \mathrm{Var}(\mathbf{e}), \tag{1.10}$$

and $\mathrm{Var}(\mathbf{Xb})$ is known to be zero. The variance of the random effects $\mathrm{Var}(\mathbf{Zu})$ can be written as $Z\mathrm{Var}(\mathbf{b})Z'$.

$$\mathrm{var}\left(\begin{array}{c} b \\ \epsilon \end{array} \right) = \left(\begin{array}{cc} D & 0 \\ 0 & \Sigma \end{array} \right)$$

where G and $\mathbf{\Sigma}$ are positive definite matrices parameterized by an unknown variance component parameter vector θ. The variance-covariance matrix for the vector of observations y is given by $V = ZGZ' + \Sigma$. This implies $y \sim (X\beta, V) = (X\beta, \boldsymbol{ZGZ}' + \Sigma)$.

$\boldsymbol{R}_i$ is the variance covariance matrix for the residuals, i.e. the within-item sources of variation between both methods. Computational analysis of linear mixed effects models allow for the explicit analysis of both $\boldsymbol{G}$ and $\boldsymbol{R_i}$. The above terms can be used to express the variance covariance matrix $\boldsymbol{\Omega}_i$ for the responses on item i ,

$$\boldsymbol{\Omega}_i = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i' + \boldsymbol{R}_i.$$

It is assumed that $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$, $\boldsymbol{\epsilon}_i$ is a matrix of random errors distributed as $N(0, \boldsymbol{R}_i)$ and that the random effects and residuals are independent of each other. Assumptions made on the structures of $\boldsymbol{G}$ and $\boldsymbol{R}_i$ will be discussed in due course.

The random effects are assumed to be distributed as $\boldsymbol{b}_i \sim \mathcal{N}_2(0, \boldsymbol{G})$. The between-item variance covariance matrix $\boldsymbol{G}$ is constructed as follows:

$$\boldsymbol{G} = \text{Var}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{pmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{pmatrix}$$

The distribution of the random effects is described as $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$. Similarly random errors are distributed as $\boldsymbol{\epsilon}_i \sim N(0, \boldsymbol{R}_i)$. The random effects and residuals are assumed to be independent. The variance-covariance matrix for the vector of observations y is given by $V = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \Sigma$. This implies $y \sim (X\beta, V) = (X\beta, ZGZ' + \Sigma)$.

By letting $\text{var}(b) = D$ (i.e $\mathbf{b}$ $N(0, \mathbf{G})$), this becomes $\boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$. This specifies the covariance due to random effects. The residual covariance matrix $\text{var}(e)$ is denoted as R, ($\mathbf{e}$ $N(0, \mathbf{R})$). Residual are uncorrelated, hence $\mathbf{R}$ is equivalent to $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The variance matrix $\mathbf{V}$ can therefore be written as:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}. \tag{1.11}$$

The methodology proposed by Roy (2009) is largely based on Hamlett et al. (2004), which in turn follows on from Lam et al. (1999). Hamlett re-analyses the data of Lam et al. (1999) to generalize their model to cover other settings not covered by the Lam method. In many cases, repeated observation are collected from each subject in sequence and/or longitudinally.

1.4 Hamlett's Model

Hamlett et al. (2004) demonstrated how the between-subject and within subject variabilities can be expressed in terms of correlation terms with D and Σ specified as follows:

$$G = \left(\begin{array}{cc} \sigma^2_A\rho_A & \sigma_A\sigma_b\rho_{AB}\delta \\ \sigma_A\sigma_b\rho_{AB}\delta & \sigma^2_B\rho_B \end{array} \right) \qquad l\Sigma = \left(\begin{array}{cc} \sigma^2_A(1-\rho_A) & \sigma_{AB}(1-\delta) \\ \sigma_{AB}(1-\delta) & \sigma^2_B(1-\rho_B) \end{array} \right).$$

ρ_A describe the correlations of measurements made by the method A at different times. Similarly ρ_B describe the correlation of measurements made by the method B at different times. Correlations among repeated measures within the same method are known as intra-class correlation coefficients. The correlation of measurements taken at the same same time by both methods is denoted ρ_{AB}. The coefficient δ is added for when the measurements are taken at different times, and is a constant of less than 1 for linked replicates. This is based on the assumption that linked replicates measurements taken at the same time would have greater correlation than those taken at different times. For unlinked replicates δ is simply 1. Hamlett et al. (2004) provides a useful graphical depiction of the role of each correlation coefficients.

Lam et al. (1999) used ML estimation to estimate the true correlation between the variables when the measurements are linked over time. The methodology relies on the assumption that the two variables with repeated measures follow a multivariate normal distribution. The methodology currently does not extend to any more than two cases. The MLE of the correlation takes into account the dependency among repeated measures.

The true correlation ρ_{xy} is repeated measurements can be considered as having two components: between subject and within-subject correlation. The usefulness of estimating repeated measure correlation coefficients is the calculation of between-method and within-method variabilities are produced as by-products.

Chapter 2

Model Diagnostics for Method Comparison

Linear Mixed Effects models are a useful framework for fitting a wide range of models. However, they are known to be sensitive to outliers, specifically the likelihood based estimation techniques, such as ML and REML.

A full and comprehensive analysis that comprises residual analysis and influence analysis for testing model assumptions, should be carried out. Therefore, a suite of diagnostic procedures should be specified for method comparison in mind. However it has been noted by several papers (Christensen et al., 1992; Schabenberger, 2004) that model diagnostics do not often accompany LME model analyses.

Further to the analysis of residuals, Schabenberger (2004) recommends the examination of the following questions:

- Does the model-data agreement support the model assumptions?
- Should model components be refined, and if so, which components?
- Are individual data points or groups of cases particularly influential on the analysis?

The last of these three questions, regarding influential points, is of particular interest in the context of method comparison. After fitting an LME model, it is important to carry put model diagnostics to check whether distributional assumptions for the residuals as satisfied and whether the fit the model is sensitive to unusual assumptions.

Model diagnostics techniques determine whether or not the distributional assumptions are satisfied, and to assess the influence of unusual observations. For classical linear models model diagnostics are now considered a required part of any statistical analysis, and are commonly available in statistical packages and standard textbooks on applied regression.

Model diagnostic techniques, well established for classical models, have since been adapted for use with linear mixed effects models. Schabenberger (2004) remarks that the concept of critiquing the model-data agreement applies in LME models in the same way as in classical linear models. West et al. (2007) argues that model and data diagnostics are even more important because of the more complex model structure.

Diagnostic techniques for LME models are inevitably more difficult to implement, due to the increased complexity. Christensen et al. (1992) advises that outlier identification is necessary before any conclusions may be drawn from the fitted model, with the leverage of an observation a further consideration. Schabenberger (2004) discusses the state of LME diagnostics tools, providing a useful summary of established measures. Prominent in literature is the taxonomy of residuals for LME Models, distinguishing between condition residuals, marginal residuals and EBLUPS, including ?Schabenberger (2004); West et al. (2007); ?.

2.1 Residual Analysis

Analysis of residuals, the differences between observed values and fitted values, is a widely used model validation technique used to examine model assumptions, as well as to detect outliers and potentially influential data points.

As with classical models, there are two key graphical techniques: a residual plot and the normal probability plot. The rationale is that, if the model is properly fitted to the model, then the residuals would approximate the random errors that one should expect. If the residuals behave randomly, with no discernible trend, the model has fitted the data reasonably well.

The underlying assumptions for LME models are similar to those of classical linear models. However, for LME models the matter of residuals are more complex, both from a theoretical point of view and from the practicalities of implementing a comprehensive analysis using statistical software.

Statistical software environments, such as the `R` programming language, provides a suite of tests and graphical procedures for appraising a fitted LME model, with several of these procedures analysing the model residuals. Texts such as Pinheiro and Bates (1994); West et al. (2007); Gałecki and Burzykowski (2013) describe what can be implemented for LME residual analyses with statistical software, such as `R` and `SAS`.

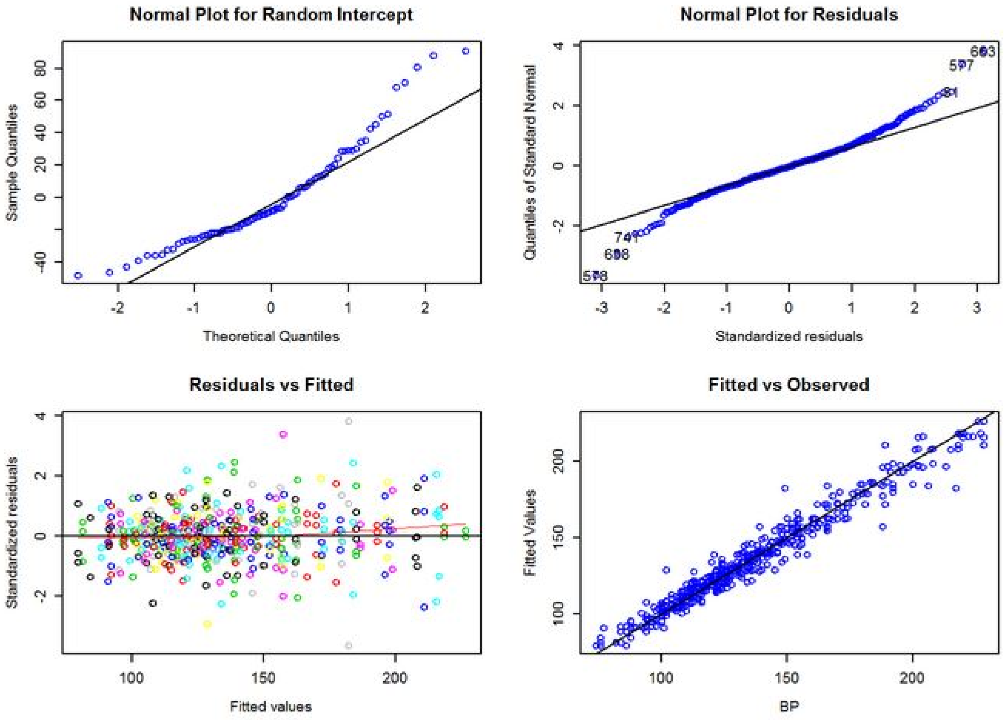
Residual diagnostics are typically implemented as a plot of the observed residuals and the predicted values. A visual inspection for the presence of trends inform the analyst on the validity of distributional assumptions, and to detect outliers and influential observations.

Analysis of the residuals could determine if the methods of measurement disagree systematically, or whether or not erroneous measurements associated with a subset of the cases are the cause of disagreement.

In the context of method comparison, a residual analysis would be carried out just as any other LME model would, testing normality. Pinheiro and Bates (1994) provide some insight into how to compute and interpret model diagnostic plots for LME models. Unfortunately this aspect of LME theory is not as expansive as the corresponding body of work for linear models.

Residual analysis must be carried out on a method-by-method basis, for a model fitted by Roy's approach. When considering the whole data, erroneous conclusions will be drawn.

Figure 2.1 depicts residual plots for Roy's model for the systolic blood pressure data used in Bland and Altman (1999). Inspection of the normality probability plot would lead to the conclusion that the key assumption of residual normality being invalid. However, this plots depicts two difference populations of residuals, an incorrect use of the normal probability plot, and therefore it is not possible to tell if the assumption of normality for residuals is invalid.



Normal Plot for Random Intercept

Normal Plot for Residuals

Residuals vs Fitted

Fitted vs Observed

Figure 2.1:

For method comparison studies, one should create plots specific to each method, useful in determining which methods disagree with the rest.

Figure 2.2 depicts residual plot for the Systolic Blood Pressure example, panelled by the various measurement methods, confirming agreement between methods J and R. Lack of agreement between those two methods and method S is also indicated. However, little insight can be gained as to what actually causes this lack of agreement.

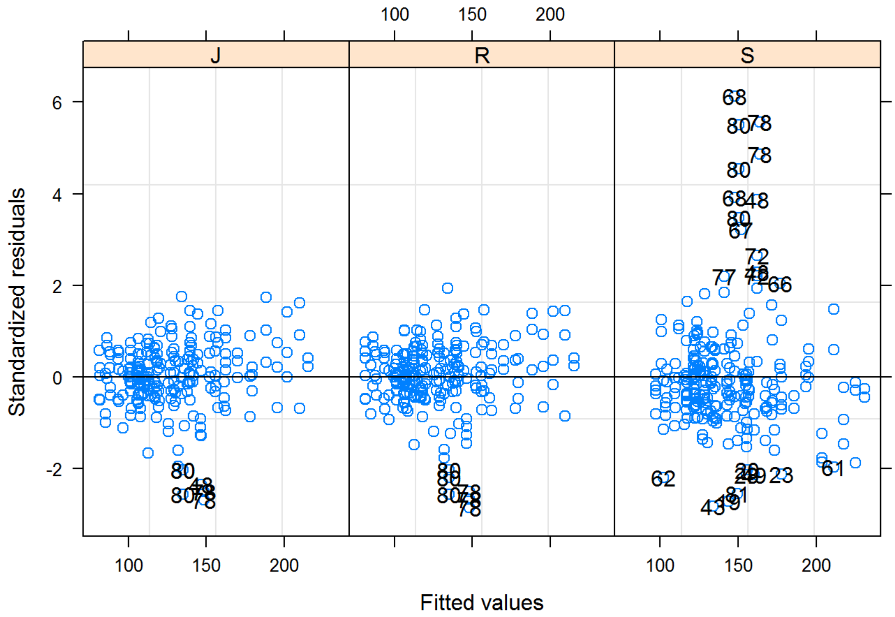


Figure 2.2: LME Residuals by Method (Blood Pressure Data)

Taxonomy of LME Residuals

Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis. To properly assess the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.

Pinheiro and Bates (1994) describes three types of residual that describe the variabilities present in LME models, marginal residuals which predict marginal errors, conditional residual, which predict conditional errors, and the BLUP, $Z\hat{b}$, that predicts random effects. Each type of residual is useful to evaluates some assumption of the model.

The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean. The definitions of both marginal residuals (r_m) and conditional residuals (r_c) follow from the definitions of marginal and conditional means in the LME model $E[Y] = X\beta$ and $E[Y|u] = X\beta + Zu$, respectively.

A marginal residual is the difference between the observed data and the estimated marginal mean, i.e. $r_{Mar} = y - X\hat{\beta}$. A conditional residual is the difference between the observed data and the predicted value of the observation. Conditional residuals include contributions from both fixed and random effects, whereas marginal residuals include contribution from only fixed effects. Marginal residuals should have mean of zero, but may show grouping structure. Also they may not be homoscedastic. In a model without random effects, both sets of residuals coincide.

Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components.

According to **?**, a residual is considered pure for a specfic type of error if it depends only on the fixed components and on the error that it is supposed to predict. Residuals that depend on other types of error are known as 'confounded errors'.

2.2 Influence Diagnostics

Model diagnostic techniques can determine whether or not the distributional assumptions are satisfied, but also to assess the influence of unusual observations. Following model specification and estimation, it is of interest to explore the model-data agreement by raising pertinent questions.

West et al. (2007) remarks that influence diagnostics play an important role in the interpretation of results, because influential data can negatively influence the statistical model and generalizability of the model.

Unfortunately this aspect of LME theory is not as expansive as the corresponding body of work for classical linear models. Pinheiro and Bates (1994) provide some insight into how to compute and interpret model diagnostic plots for LME models.

Their particular observations will be reverted to shortly.

Influence diagnostics are formal techniques that allow the identification observation that heavily influence estimates of the estimates of fixed effects and variance covariance parameters.

Influential points are a set of one or more observations whose removal would cause a different conclusion in the analysis, e.g. substantially affect estimates. Influential points have a large influence on the fit of the model. Influential points are a set of one or more observations whose removal would cause a different conclusion in the analysis, e.g. substantially changes the estimate of the regression coefficients.

The process of carrying out model diagnostic involves several informal and formal techniques, which will mentioned throughout the chapter.

One approach for determining influential points is to compare the fit of the model with and without each observation. The basic rationale behind identifying influential data is that when iteratively single units are omitted from the data, models based on these data should not produce substantially different estimates.

Procedure for Quantifying Influence

Schabenberger (2004) describes a simple procedure for quantifying influence for LME Models. Firstly a model should be fitted to the data, and estimates of the parameters should be obtained.

The second step is that either single or multiple data points, specifically outliers, should be omitted from the analysis, with the original parameter estimates being updated. This is known as *"leave one out"* or *"leave k out"* analysis.

The final step of the procedure is comparing the sets of estimates computed from the entire and reduced data sets to determine whether the absence of observations changed the analysis.

2.2.1 Comparing Influence and Residual Analysis

? compares residual analysis and influence analysis. Cases with high residuals (defined as the difference between the observed and the predicted scores on the dependent variable) or with high standardized residuals (defined as the residual divided by the standard deviation of the residuals) are indicated as outliers.

However, an influential case is not necessarily an outlying residual. On the contrary: a strongly influential case dominates the regression model in such a way, that the estimated regression line lies closely to this case. The analysis of residuals cannot be used for the detection of influential cases (**?**).

Case Deletion

The impact of an observation on a regression fitting can be determined by the difference between the estimated regression coefficient of a model with all observations and the estimated coefficient when the particular observation is deleted.

The subscript (U) is used to denote quantities computed from data with subset of cases U omitted. If the global measure suggests that the points in U are influential, you should next determine the nature of that influence. In particular, the points can affect the estimates of the precision of the fixed effects and covariance parameters, and hence predicted values.

For case-deletion approaches, Preisser (1996) describes two type of diagnostics. When the set consists of only one observation, the type is called '*observation-diagnostics*'. For multiple observations, Preisser describes the diagnostics as '*cluster-deletion*' diagnostics. Consideration of how leave-U-out diagnostics would work in the context of Method Comparison problems is required. Suppose we have two methods of measurement X and Y, each with three measurements for a specific case: $(x_1, x_2, x_3, y_1, y_2, y_3)$

- **Leave One Out** - one observation is omitted (e.g. x_1)

- **Leave Pair Out** - one pair of observation is omitted (e.g. x_1 and y_1)

- **Leave item) Out** - All observations associated with a particular case or subject are omitted. (e.g. $\{x_1, x_2, x_3, y_1, y_2, y_3\}$)

The natural sampling unit is the item or subject, similar to the example provided by Schabenberger (2004). Hence, the third option, henceforth, referred to as "*Leave item out*" will be the option used.

2.2.2 Analyzing Influence in LME Models

Influence can be thought of as consequence of leverage and outlierness. Outliers are the most noteworthy data points in an analysis, and an objective of influence analysis is how influential they are, and the manner in which they are influential. They can point to a model breakdown and lead to development of a better model.

While linear models and GLMS can be studied with a wide range of well-established diagnostic technqiues, the choice of methodology is much more restricted for the case of LMEs. However influence diagnostics for LME Models is an area of active research. Research on diagnostic analyses for LME models are presented in **?**, Christensen et al. (1992), **?**, **?**, **?**, **?**, Demidenko (2004), Zewotir and Galpin (2005) and **??**.

Schabenberger (2004) states that goal of influence analysis is not primarily to mark data points for deletion so that a better model fit can be achieved for the reduced data, although this might be a result of influence analysis. The goal is rather to determine which cases are influential and the manner in which they are important to the analysis.

2.2.3 Measuring of Influence for LME Models

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 (Zewotir and Galpin, 2005).

Zewotir and Galpin (2005) 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, and the Andrews-Prebigon statistic.

Zewotir and Galpin (2005) remarks the development of efficient computational formulas is crucial making deletion diagnostics useable, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model. A number of approaches to model diagnostics are described, including variance components, dixed effects parameters, prediction of the response variable and of random effects, and the likelihood function. Influence statistics can be grouped by the aspect of estimation that is their primary target:

- **Overall measures compare changes in objective functions**: (restricted) likelihood distance (Cook and Weisberg 1982, Ch. 5.2)
- **Influence on parameter estimates**: Cook's (Cook 1977, 1979), MDFFITS (Belsley, Kuh, and Welsch 1980, p. 32)
- **Influence on precision of estimates**: CovRatio and , item **Influence on fitted and predicted values**: PRESS residual, PRESS statistic (Allen 1974), DFFITS (Belsley, Kuh, and Welsch 1980, p. 15),
- **Outlier properties**: internally and externally studentized residuals, leverage

For example, if observations primarily affect the precision of the covariance parameters without exerting much influence on the fixed effects, then their presence in the data may not distort hypothesis tests or confidence intervals about β. Schabenberger (2004) notes that removing observations or sets of observations affects fixed effects and covariance parameter estimates.

2.3 Deletion Diagnostics

Deletion diagnostics provide a means of assessing the influence of an observation (or groups of observations) on parameters inferences for a fitted model. For classical linear models, **?** greatly expands the study of residuals and influence 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.

Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. Cook proposed a measure that combines the information of leverage and residual of the observation, now known simply as the Cook's Distance, $D_{(i)}$, which can be calculated without fitting a new regression coefficient each time an observation is deleted. Consequently deletion diagnostics have become an integral part of assessing linear models.

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

Christensen et al. (1992) notes the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect. Christensen et al. (1992) developed their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted resgression problem (conditional on the estimated covariance matrix) for fixed effects.

Calculation of case deletion diagnostics in the OLS model is made simple by the fact that estimates of β and σ^2, which exclude the ith observation, can be computed without re-fitting the model. Such update formulas are available in the LME model only if you assume that the covariance parameters are not affected by the removal of the observation in question. This is rarely a reasonable assumption, and fundamentally undermines the use of many proposed procedures for method comparison.

2.3.1 Cook's Distance

Cooks Distance (D_i) is a diagnostic technique used in classical linear models, that functions as an overall measure of influence of a subset of observations U on the regression coefficients, and consequently the fitted values. Cook's Distance as a measure of the influence of observations in subset U on a vector of parameter estimates is given below (?)

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}.$$

Observations, or sets of observations, that have high Cook's distance usually have high residuals, although this is not necessarily the case.

If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

Cook's distance can be used in several ways: to indicate data points that are particularly worth checking for validity; to indicate regions of the design space where it would be good to be able to obtain more data points.

Large values for Cook's distance indicate observations for special attention. Although informal heuristics exist, use of threshold values for Cook's Distance is discouraged (?). ? advises the use of diagnostic plotting and to examine in closer details the points with *"values of D that are substantially larger than the rest"*, and that thresholds should feature only to enhance graphical displays.

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

Christensen et al. (1992) develops case deletion diagnostics, in particular the equivalent of Cook's distance for diagnosing influential observations when estimating the fixed effect parameters and variance components, adapting the Cook's Distance measure for the analysis of LME models.

For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ. Zewotir and Galpin (2005) gives a detailed discussion of the various formulation for Cook's distances for LME Models.

Such update formulas are available in the LME model only if one can assume that the covariance parameters are not affected by the removal of the observation in question. This is rarely a reasonable assumption, and fundamentally undermines the use of many proposed procedures for method comparison.

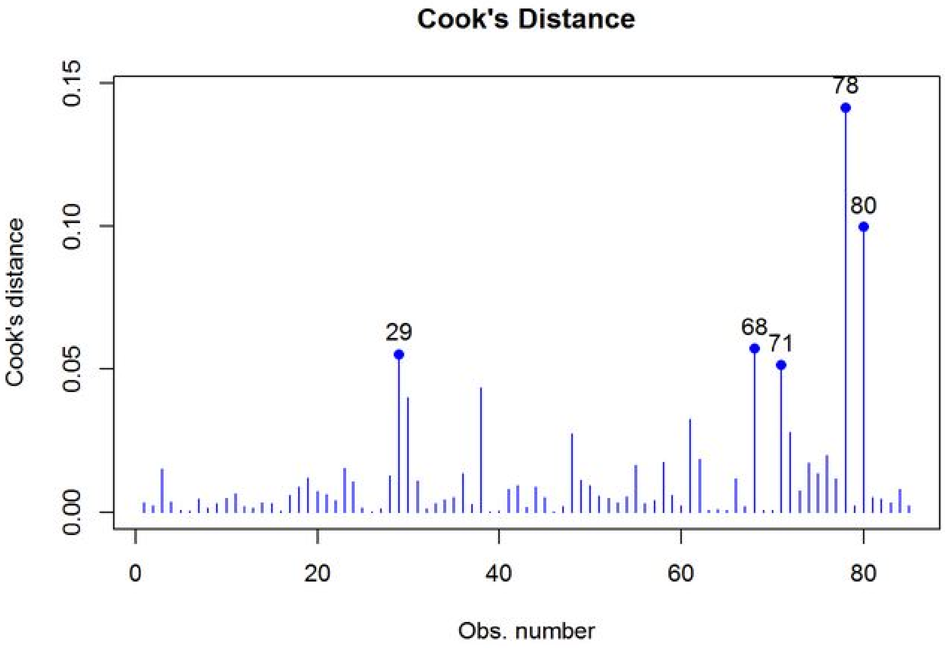


Figure 2.3:

Christensen et al. (1992) notes the case deletion diagnostics techniques have not been applied to LME models and seeks to develop methodologies in that respect. Christensen et al. (1992) developed their global influences for the deletion of single observations in two steps: a one-step estimate for the REML (or ML) estimate of the variance components, and an ordinary case-deletion diagnostic for a weighted regression problem (conditional on the estimated covariance matrix) for fixed effects.

2.3.2 Local Influence

Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models, introducing methods for local influence assessment for classical linear models. 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.

? applied the local influence method of Cook (1986) to the analysis of the LME model. Other authors such as **?** have also extended these idea to LME models.

While the concept of influence analysis is straightforward, implementation in LME models is more complex. Update formulae for fixed effects models are available only when the covariance parameters are assumed to be known. As such the local influence approach are not particularly useful in the context of method comparison, and so will not be considered further.

2.3.3 Iterative and Non-Iterative Influence Analysis

For linear models, the implementation of influence analysis is straightforward, but for LME models the process is more complex. Schabenberger (2004) examines the use and implementation of influence measures in LME models. Schabenberger (2004) highlights some of the issue regarding implementing LME model diagnostics, describing the choice between iterative influence analysis and non-iterative influence analysis. Schabenberger (2004) considers several important aspects of the use and implementation of influence measures in LME models, noting that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates. Closed-form expressions for computing the change in important model quantities might not be available.

Schabenberger (2004) describes the scenario wherein a data point is removed and the new estimate of the G matrix is not positive definite. This may occur if a variance component estimate now falls on the boundary of the parameter space (Schabenberger, 2004).

For classical linear models, it is not necessary to refit the model after removing a data point in order to measure the impact of an observation on the model. The change in fixed effect estimates, residuals, residual sums of squares, and the variance-covariance matrix of the fixed effects can be computed based on the fit to the full data alone, using update formulas (**??**).

However, in LME models several important complications arise. Data points can affect not only the fixed effects but also the covariance parameter estimates on which the fixed-effects estimates depend. However update formulas are available only if one assumes that the covariance parameters are not affected by the removal of the observation in question. However, this is rarely a reasonable assumption. For LME models, non-iterative methods are computationally efficient, but require the rather strong assumption that all covariance parameters are known, and thus are not updated, with the exception of the profiled residual variance.

Update formulas for "*leave-U-out*" estimates typically fail to account for changes in covariance parameters. As the influence that each item would have on the variance estimate of a method comparison model is crucial, this substanitally negates their usefulness for Roy's Model.

Iterative influence diagnostics requiring fitting the model without the observations in question. Computation time is substantially longer, although this is balanced by algorithmic simplicity, with no assumptions beyond those used for the original model. A measure of total influence requires updates of all model parameters. This can only be achieved in general is by omitting observations or cases, then refitting the model.

An iterative analysis may seem computationally expensive. Computing iterative influence diagnostics for n observations requires $n + 1$ mixed models to be fitted iteratively. The execution times for iterative procedures are longer relative to non-iterative procedures, but are not so long that they would dissuade an analyst from using them. Despite the addition execution time of iteratives approaches, they are preferable for method comparison problems, as they can facilitate several complementary analyses concurrently.

Iterative methods retain the potential for useful analyses, if applied at different stage of the modelling process. Diagnostic measures, specifically the DFBETA, have characteristics that would make them very useful at the exploratory stage of the method comparison process. Implicitly various assumptions about variance are used, but simultaneously an approach based on DFBETA can be used to assess if these assumptions are valid.

2.3.4 Likelihood Distance

In LME models fit by maximum likelihood (ML) or restricted maximum likelihood (REML), an overall influence measure is the likelihood distance (**?**).

West et al. (2007) 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*. The likelihood distance is a global summary measure that expresses the joint influence of the subsets of observations, U, on all parameters that were subject to updating. Schabenberger (2004) points out that the likelihood distance $LD(\psi_{(U)})$ is not the log-likelihood obtained by fitting the model to the reduced data set. Instead it is obtained by evaluating the likelihood function based on the full data set (containing all n observations) at the reduced-data estimates.

The procedure requires the calculation of the full data estimates $\hat{\psi}$ and estimates based on the reduced data set $\hat{\psi}_{(U)}$. The likelihood distance is given by determining

$$LD_{(U)} = 2\{l(\hat{\psi}) - l(\hat{\psi}_{(U)})\}$$

$$RLD_{(U)} = 2\{l_R(\hat{\psi}) - l_R(\hat{\psi}_{(U)})\}$$

Large values indicate that $\hat{\theta}$ and $\hat{\theta}_\omega$ differ considerably.

Bibliography

Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research 8*(2), 135–160.

Carstensen, B. (2004). Comparing and predicting between several methods of measurement. *Biostatistics 5*(3), 399–413.

Carstensen, B., J. Simpson, and L. C. Gurrin (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics 4*(1).

Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics 34*(1), 38–45.

Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological) 48*(2), 133–169.

Demidenko, E. (2004). *Mixed Models: Theory And Application.* Dartmouth College: Wiley Interscience.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the em algorithme. *ournal of the Royal Statistical Society. Series B 39*(1), 1–38.

Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms.* Oxford University Press.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics 3*(1), 1–21.

Fisher, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh 2*, 399–433.

Gałecki, A. and T. Burzykowski (2013). *Linear mixed-effects models using R: A step-by-step approach.* Springer Science & Business Media.

Green, P. J. and B. W. Silverman (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach.* Chapman & Hall Ltd.

Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Proceedings of the Statistics and Data Analysis Section, SAS Users Group International 198-229*, 1–7.

Hartley, H. and J. Rao (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika 54*(1/2), 93–108.

Henderson, C. (1953). Estimation of variance and covariance components. *Biometrics 9*(2), 226–252.

Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1963). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1973). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Henderson, C., O. Kempthorne, S. Searle, and C. von Krosigk (1984). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics 21*, 309–310.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and genetics Symposium in Honor of Dr Jay L. Lush*, pp. 10–41. Champaign, Illinois: Americian Society of Animal Science and American Dairy Science Association.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*(4), 963–974.

Lam, M., K. Webb, and D. O'Donnell (1999). Correlation between two variables in repeated measurements. *American Statistical Association, Proceedings of the Biometric Session*, 213–218.

Lee, Y., J. Nelder, and Y. Pawitan (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood.* Chapman and Hall CRC.

Lee, Y. and J. A. Nelder (1996). Hierarchical generalized linear models (Disc: P656-678). *Journal of the Royal Statistical Society, Series B: Methodological 58*, 619–656.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford University Press.

Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.

Preisser, J. S. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika 83*(3), 551–5562.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (Disc: P32-51). *Statistical Science 6*, 15–32.

Roy, A. (2009). An application of the linear mixed effects model to ass the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics 19*, 150–173.

Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.

Singer, J. D. (1998). Using sas proc mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics 24*(4), 323–355.

Tippett, L. (1931). *The Methods of Statistics* (1st ed.). London: Williams and Norgate.

West, B., K. Welch, and A. Galecki (2007). *Linear Mixed Models: a Practical Guide Using Statistical Software*. Chapman and Hall CRC.

Zewotir, T. and J. S. Galpin (2005). Influence diagnostics for linear mixed models. *Journal of Data Science 3*(2), 153–177.