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 only can only indicate agreement of lack thereof, and the presence of inter-method bias. To consider more complex questions, more complex LME models are 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).

In part this is due to the increased profile of LME models, and furthermore the availability of capable software. Additionally a great understanding of residual analysis and influence analysis for LME models has been adchieved 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 D 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

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

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

$$\left(\begin{matrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & X'\Sigma^{-1}X + D^{-1} \end{matrix} \right) \left(\begin{matrix} \beta \\ b \end{matrix} \right) = \left(\begin{matrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{matrix} \right). \tag{1.6}$$

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 D and Σ are known.

4

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 $DZ'V^{-1}(y - X\beta)$. 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 D 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

6

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

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

7

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.

$$\Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

σ_A^2 is the variance of variable A, σ_B^2 is the variance of variable B and σ_{AB} is the covariance of the two variable. It is assumed that Σ 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

$$y_i = X_i\beta + Z_i b_i + \epsilon_i,$$

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

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

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

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

$$\text{var}\begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

where D and Σ 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 $V = ZDZ' + \Sigma$. This implies $y \sim (X\beta, V) = (X\beta, \boldsymbol{ZDZ'} + \Sigma)$.

$\boldsymbol{R}_i$ 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 $\boldsymbol{D}$ 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 ,

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

It is assumed that $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$, $\boldsymbol{\epsilon}_i$ 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 $\boldsymbol{D}$ and $\boldsymbol{R}_i$ will be discussed in due course.

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

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

The distribution of the random effects is described as $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$. 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 $V = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' + \Sigma$. This implies $y \sim (X\beta, V) = (X\beta, ZDZ' + \Sigma)$.

By letting $\mathrm{var}(b) = D$ (i.e $\mathbf{b}\ N(0, \mathbf{D})$), this becomes $\boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$. This specifies the covariance due to random effects. The residual covariance matrix $\mathrm{var}(e)$ 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:

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

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.

1.3.1 Correlation Terms

Hamlett et al. (2004) demonstrated how the between-subject and within subject vari-

abilities can be expressed in terms of correlation terms with D and Σ specified as follows:

$$D = \begin{pmatrix} \sigma_A^2 \rho_A & \sigma_A \sigma_b \rho_{AB} \delta \\ \sigma_A \sigma_b \rho_{AB} \delta & \sigma_B^2 \rho_B \end{pmatrix} \qquad l\Sigma = \begin{pmatrix} \sigma_A^2(1 - \rho_A) & \sigma_{AB}(1 - \delta) \\ \sigma_{AB}(1 - \delta) & \sigma_B^2(1 - \rho_B) \end{pmatrix}.$$

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.

Bibliography

Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet i*, 307–310.

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.

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.

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.

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.