Chapter 1

Model Diagnostics

Contents

1 Model Diagnostics **1**

1.1 Model Validation using Residual Diagnostics 3

1.2 Case Deletion Diagnostics 4

1.2.1 Extension of Diagnostic Methods to LME models 4

1.3 Case Deletion Diagnostics for LME models 5

1.4 Cook's Distance 6

1.5 Effects on fitted and predicted values 6

1.6 Exention of Cook's Distance methodology to LME models 7

1.6.1 Cook's Distance 7

1.6.2 Cook's Distance for LMEs 8

1.7 Cook's Distance for LMEs 10

1.7.1 Overall Influence 10

1.7.2 Influence Diagnostics: Basic Idea and Statistics 10

1.7.3 Cook's 1986 paper on Local Influence 11

1.8 Measures of Influence 11

1.8.1 DFFITS 11

1.8.2 PRESS 11

1.9 Influence in LME Models 12

1.9.1 Influence Statistics for LME models 12

1.10 Influence analysis for LME Models 14

1.10.1 Influence Analysis for LME Models 14

1.10.2 Influence Analysis for LME Models 16
1.10.3 Influence Statistics for LME models 16
1.10.4 What is Influence 17
1.10.5 Quantifying Influence 17
1.11 Extension of techniques to LME Models 18
1.11.1 Likelihood Distance 18
1.12 Likelihood Distance 20
1.12.1 Likelihood Distance 20
1.12.2 Restricted Likelihood Distances 21
1.13 Analyzing Influence in LME models 21
1.14 Zewotir Measures of Influence in LME Models 21
1.15 Iterative and non-iterative influence analysis 22
1.15.1 Iterative Influence Analysis 22
1.16 Matrix Notation for Case Deletion 23
1.16.1 Case deletion notation 23

2 Model Diagnostics **24**

2.0.2 Further Assumptions of Linear Models 24
2.0.3 Residuals diagnostics in mixed models 24
2.0.4 Marginal and Conditional Residuals 24

1.1 Model Validation using Residual Diagnostics

In statistical modelling, the process of model validation is a critical step, but also a step that is too often overlooked. A very simple procedure is to examine commonly encountered metrics, such as the R^2 value. However, using a small handful of simple measures and methods is insufficient to properly assess the quality of a fitted model. To do so properly, a full and comprehensive analysis that tests of all of the assumptions, as far as possible, must be carried out.

Residual analysis is a widely used model validation technique. A residual is simply the difference between an observed value and the corresponding fitted value, as predicted by the model. 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. that is to say, if the residuals behave randomly, with no discernible trend, the model has fitted the data well. If some sort of non-random trend is evident in the model, then the model can be considered to be poorly fitted. Statistical software environments, such as the `R` Programming language, provides a suite of tests and graphical procedure sfor appraising a fitted linear model, with several of these procedures analysing the model residuals.

The question of whether or not a point should be considered an outlier must also be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables. The leverage of an observation is a further consideration. Leverage describes an observation with an extreme value on a predictor variable is a point with high leverage. High leverage points can have a great amount of effect on the estimate of regression coefficients.

Influence can be thought of as the product of leverage and outlierness. An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. The `R` programming language has a variety of methods used to study each of the aspects for a linear 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.

For classical linear models, 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.

1.2 Case Deletion Diagnostics

Since the pioneering work of Cook in 1977, deletion measures have been applied to many statistical models for identifying influential observations. Case-deletion diagnostics provide a useful tool for identifying influential observations and outliers.

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.

The computation of case deletion diagnostics in the classical model is made simple by the fact that estimates of β and σ^2, which exclude the $i-$th observation, can be computed without re-fitting the model. Such update formulas are available in the mixed model only if you assume that the covariance parameters are not affected by the removal of the observation in question. However, this is rarely a reasonable assumption.

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.

1.2.1 Extension of Diagnostic Methods to LME models

? noted the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect. **?** develops these techniques in the context of REML.

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. Deletion diagnostics provide a means of assessing the influence of an observation (or groups of observations) on inference on the estimated parameters of LME models. We shall provide a fuller discussion of Cook's Distance in due course.

Demidenko (2004) extends several regression diagnostic techniques commonly used in linear regression, such as leverage, infinitesimal influence, case deletion diagnostics, Cook's distance, and local influence to the linear mixed-effects model. In each case, the proposed new measure has a direct interpretation in terms of the effects on a parameter of interest, and reduces to the familiar linear regression measure when there are no random effects.

The new measures that are proposed by Demidenko (2004) are explicitly defined functions and do not require re-estimation of the model, especially for cluster deletion diagnostics. The basis for both the cluster deletion diagnostics and Cook's distance is a generalization of Miller's simple update formula for case deletion for linear models. Furthermore Demidenko (2004) shows how Pregibon's infinitesimal case deletion diagnostics is adapted to the linear mixed-effects model.

1.3 Case Deletion Diagnostics for LME models

Data from single individuals, or a small group of subjects may influence non-linear mixed effects model selection. Diagnostics routinely applied in model building may identify such individuals, but these methods are not specifically designed for that purpose and are, therefore, not optimal. We describe two likelihood-based diagnostics for identifying individuals that can influence the choice between two competing models.

Cook's distance

In the study of Linear Model Diagnostics, Cook proposed a measure that combines the information of leverage and residual of the observation, now known simply as the Cook's Distance. **?** have adapted this measure for the analysis of LME models.

1.4 Cook's Distance

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's Distance , denoted as$D_{(i)}$, is a well known diagnostic technique used in classical linear models, used as an overall measure of the combined impact of the $i-$th case of all estimated regression coefficients. Cook's key observation was the effects of deleting each observation in turn could be calculated with little additional computation. That is to say, $D_{(i)}$ 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 focus of this analysis is related to the estimation of point estimates (i.e. regression coefficients). It must be pointed out that 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.

As well as individual observations, Cook's Distance can be used to analyse the influence of observations in subset U on a vector of parameter estimates (Cook, 1977).

1.5 Effects on fitted and predicted values

$$\hat{e}_{i(U)} = y_i - x\hat{\beta}_{(U)} \tag{1.1}$$

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)} \tag{1.2}$$

1.6 Exention of Cook's Distance methodology to LME models

Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance was extended from classical linear models to LME models. For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by definining.

$$C_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

1.6.1 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

1.6.2 Cook's Distance for LMEs

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

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.

1.7 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. Cook (1986) gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

1.7.1 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].

1.7.2 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.

1.7.3 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 quitedifferent from the case deletion approach, comparisons are of interest.

1.8 Measures of Influence

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 measure DFBETA is the studentized value of this difference.

1.8.1 DFFITS

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. It is closely related to the studentized residual.

$$DFFITS = \frac{\widehat{y_i} - \widehat{y_{i(k)}}}{s_{(k)}\sqrt{h_{ii}}}$$

1.8.2 PRESS

The prediction residual sum of squares (PRESS) is an value associated with this calculation. When fitting linear models, PRESS can be used as a criterion for model selection, with smaller values indicating better model fits.

$$PRESS = \sum (y - y^{(k)})^2 \tag{1.3}$$

- $e_{-Q} = y_Q - x_Q\hat{\beta}^{-Q}$
- $PRESS_{(U)} = y_i - x\hat{\beta}_{(U)}$

DFBETA

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \quad (1.4)$$

$$= B(Y - Y_{\bar{a}} \quad (1.5)$$

1.9 Influence in 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.

1.9.1 Influence Statistics for LME models

Influence statistics can be coarsely 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 CovTrace
- 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

Beckman, Nachtsheim and Cook (1987) Beckman et al. (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.

1.10 Influence analysis for LME Models

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 β and θ. A common technique is to refit the model with an observation or group of observations omitted.

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'.

1.10.1 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. Christensen et al. (1992) 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 (Zewotir and Galpin, 2005). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

Influence arises at two stages of the LME model. Firstly when V is estimated by $\hat{V}$, and subsequent estimations of the fixed and random regression coefficients β and u, given $\hat{V}$.

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. Christensen et al 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 (Zewotir). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

Influence Statistics for LME models

Influence statistics can be coarsely 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 CovTrace
- 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

1.10.2 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. Christensen et al. (1992) 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 (Zewotir and Galpin, 2005). Studentized residuals, error contrast matrices and the inverse of the response variance covariance matrix are regular components of these tools.

1.10.3 Influence Statistics for LME models

Influence statistics can be coarsely 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 CovTrace
- 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

1.10.4 What is Influence

Broadly defined, influence is understood as the ability of a single or multiple data points, through their presence or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model. The 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 (Schabenberger, 2004).

1.10.5 Quantifying Influence

The basic procedure for quantifying influence is simple as follows:

- Fit the model to the data and obtain estimates of all parameters.
- Remove one or more data points from the analysis and compute updated estimates of model parameters.
- Based on full- and reduced-data estimates, contrast quantities of interest to determine how the absence of the observations changes the analysis.

Cook (1986) introduces powerful tools for local-influence assessment and examining perturbations in the assumptions of a model. In particular the effect of local perturbations of parameters or observations are examined.

1.11 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) Beckman et al. (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.

1.11.1 Likelihood Distance

The likelihood distance gives the amount by which the log-likelihood of the full data changes if one were to evaluate it at the reduced-data estimates. The important point is that $l(\psi_{(U)})$ is not the log-likelihood obtained by fitting the model to the reduced data set.

It is obtained by evaluating the likelihood function based on the full data set (containing all n observations) at the reduced-data estimates.

The likelihood distance is a global, summary measure, expressing the joint influence of the observations in the set U on all parameters in ψ that were subject to updating.

$$LD(\boldsymbol{\omega} = 2[L\boldsymbol{th\hat{e}ta} - \boldsymbol{th\hat{e}ta}_{\boldsymbol{\omega}}$$

Large values indicate that $\boldsymbol{th\hat{e}ta}$ and $\boldsymbol{th\hat{e}ta}_{\boldsymbol{\omega}}$ differ considerably.

1.12 Likelihood Distance

The likelihood distance gives the amount by which the log-likelihood of the full data changes if one were to evaluate it at the reduced-data estimates. The important point is that $l(\psi_{(U)})$ is not the log-likelihood obtained by fitting the model to the reduced data set.

It is obtained by evaluating the likelihood function based on the full data set (containing all n observations) at the reduced-data estimates.

The likelihood distance is a global, summary measure, expressing the joint influence of the observations in the set U on all parameters in ψ that were subject to updating.

1.12.1 Likelihood Distance

The likelihood distance is a global, summary measure, expressing the joint influence of the observations in the set U on all parameters in ϕ that were subject to updating.

1.12.2 Restricted Likelihood Distances

1.13 Analyzing Influence in LME models

1.14 Zewotir Measures of Influence in LME Models

Zewotir and Galpin (2005) describes a number of approaches to model diagnostics, investigating each of the following;

- Variance components
- Fixed effects parameters
- Prediction of the response variable and of random effects
- likelihood function

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,
- the Andrews-Prebigon statistic.

1.15 Iterative and non-iterative influence analysis

Schabenberger (2004) highlights some of the issue regarding implementing mixed model diagnostics.

1.15.1 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.

Schabenberger (2004) describes the choice between iterative influence analysis and non-iterative influence analysis.

1.16 Matrix Notation for Case Deletion

1.16.1 Case deletion notation

For notational simplicity, $\boldsymbol{A}(i)$ denotes an $n \times m$ matrix $\boldsymbol{A}$ with the i-th row removed, a_i denotes the i-th row of $\boldsymbol{A}$, and a_{ij} denotes the $(i, j)-$th element of $\boldsymbol{A}$.

Chapter 2

Model Diagnostics

2.0.2 Further Assumptions of Linear Models

As with fitted models, the assumption of normality of residuals and homogeneity of variance is applicable to LMEs also.

Homoscedascity is the technical term to describe the variance of the residuals being constant across the range of predicted values. Heteroscedascity is the converse scenario : the variance differs along the range of values.

2.0.3 Residuals diagnostics in mixed models

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

2.0.4 Marginal and Conditional Residuals

A marginal residual is the difference between the observed data and the estimated (marginal) mean, $r_{mi} = y_i - x_0'\hat{b}$ A conditional residual is the difference between the

observed data and the predicted value of the observation, $r_{ci} = y_i - x_i'\hat{b} - z_i'\hat{\gamma}$

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 $\hat{y}_i$.

$$\hat{epsilon}_i = y_i - \hat{y}_i = y_i - (X_i\hat{beta} + Z_i\hat{b}_i)$$

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.

Bibliography

Beckman, R., C. Nachtsheim, and R. Cook (1987). Diagnostics for mixed-model analysis of variance. *Technometrics 29*(4), 413–426.

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

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

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.

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

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

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