ANALYSIS OF UNIVARIATE AND MULTIVARIATE LONGITUDINAL DATA WITH CENSORED AND MISSING RESPONSE WITH COMPLEX COVARIANCE STRUCTURE

by

EMMANUEL TUGLO

(Under the Direction of Daniel Hall)

ABSTRACT

The analysis of univariate and multivariate longitudinal data (U/MLD) with censored and missing response has inspired considerable interest in the statistical community recently. In the case of MLD, estimating the contemporaneous correlation coefficient is of particular interest to applied researchers. In this dissertation, mixed model methodologies are investigated to analyze U/MLD with censored and missing response while accounting for complex features of longitudinal data such as serial correlation and/or heteroscedasticity of the measurement error, and multiple level of nesting in the random effects. We propose a computationally feasible expectation maximization (EM) algorithm for maximum likelihood (ML) and restricted maximum likelihood (REML) estimation for univariate and bivariate linear mixed effects models (U/BLMEMs) with censored and missing response. We implement our proposed methodology in an R function `lmecm`, which computes ML and REML for U/BLMEMs with censored and missing response and fits a wider class of models than existing computational tools. The performance of the proposed methods is evaluated through extensive simulation studies and analysis of real data from environmental and biomedical studies.

Index words: Detection limit; R; `lmecm`; EM algorithm; Maximum likelihood; Heteroscedastic errors; Serial correlation.

ANALYSIS OF UNIVARIATE AND MULTIVARIATE LONGITUDINAL DATA WITH CENSORED AND MISSING RESPONSE WITH COMPLEX COVARIANCE STRUCTURE

by

EMMANUEL TUGLO

B.S., Georgia State University, 2009
M.S., University of Georgia, 2011

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

Analysis of Univariate and Multivariate Longitudinal Data with Censored and Missing Response with Complex Covariance Structure

by

Emmanuel Tuglo

Approved:

Major Professor: Daniel Hall

Committee: Jaxk Reeves
Lynne Seymour
Paul Schliekelman
T.N. Sriram

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
August 2014

Dedication

To my mom.

Acknowledgments

First, I must give my gratitude to God, without whose intervention I would have no doubt be in a very different place, if any at all. Nobody can get a PhD without guidance. I am fortunate to have found a great advisor. I wish to express my appreciation to my major professor, Dr. Daniel Hall, for his valuable support in the direction of this dissertation. He gave me an interesting research topic and advised me at any time I seek. He is such a wonderful advisor: not only knowledgeable, but also truly care for his student. I am truly grateful for his encouragement and for the amount of time he spent helping me. It has been a pleasure to work under him.

Furthermore, I would like to thank to my committee members, Dr. Jaxk Reeves, Dr. Lynne Seymour, Dr. T.N. Sriram, and Dr. Paul Schliekelman for their careful reading and suggestions. I also would like to show my appreciation to Dr. Cheolwoo Park, the faculty, staff and graduate students of Department of Statistics. They gave me a chance to pursue higher degree, and gave me the best place to learn. I also want to thank Dr. Jaxk Reeves and Dr. Kim Love-Myers for recruiting me as a graduate consultant in our Statistical Consulting Center; this is an invaluable experience for me. I would like thank Drs. Femi Adetona and Luke Naeher from the UGA Department of Environmental Health Science for providing us with the motivating data set for this dissertation.

My time in the US was made enjoyable in large part due to my girlfriend Imelda Charles, and my close friend Felix Fenu, who was my host when I first moved to the US. He is been a true friend and look out for me like a brother till this day and I really appreciate everything he has done for me. Last but not least, I would like to thank my family in Ghana, especially my beloved mom, Gbeda Kponyo, for mentally supporting me from far away and for having faith in me.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

With improved computational power over the recent decades, mixed modeling has become one of the most important research areas in statistics in part due to its wide applicability to many disciplines where multiple correlated measurements are made on each unit of interest. Mixed models are prominently used in research involving human and animal subjects in fields ranging from genetics to marketing, and have also been used in industrial statistics.

One problem facing the statistics community today is modeling longitudinal/clustered data with censored response. For example, in biomedical research, biomarkers of disease are often measured repeatedly over time and may be left-censored because of minimum detection limits (DLs) below which the level of biomarkers can not be quantified. Similarly environmental pollutants are often measured repeatedly over time subject to DLs, which induces censoring. Such data poses formidable challenges both in application and theory because they exhibit features such as complex covariance structures in both the random effects and the measurement error which needs to be accounted for to be able to estimate parameters unbiasedly and perform valid statistical inference.

The problem becomes even more complicated when more than one response is measured repeatedly over time subject to DLs. Studying the joint evolution of these multiple response is of interest to researchers. However, analyzing a multivariate longitudinal data (MLD) is more challenging in several respects compared to the univariate case. For example, we have to account for the possibility of different variation in the errors for the different responses, serial correlation across different time points for the different responses and contemporaneous correlation between the multiple responses. Such MLD pose unique challenges to researchers in application because they exhibit complex features. There are many good examples of such MLD data: for example, the data from the Women's Interagency HIV study (WIHS), multicenter longitudinal study of HIV infection in US women. A random sample of 248 HIV seropositive women was selected from the WIHS to study the correlation between HIV viral loads measured in the cervix and in plasma (Xie et al., 2012). Another example is the data that contain the real time measurements of cookstove related carbon monoxide (CO) and particulate matter with aerodynamic diameter less than 2.5 microns ($PM_{2.5}$) from households in Peru who are exposed to these environmental pollutants when preparing their meals (Commodore et al., 2013, see Figure 3.1 in section 3.3).

Estimating the association between these multiple responses is very important to researchers in biomedical and environmental studies but the literature dealing with these data is lacking and most importantly there are no computational tools available to analyze these data while accounting for the complications such as censoring, missingness, serial correlation and heteroscedasticity.

1.2 Literature Review

1.2.1 Linear Mixed Effects Model for Censored Response

Linear mixed effects models (LMEMs) are well established in statistical methodology and practice; for example Pinheiro and Bates (2000), and Jiang (2007). Computationally, there are fast and efficient statistical packages including **nlme** and **lme4** in R/S-PLUS (Pinheiro et al., 2006; Bates and Shakar, 2007), and `PROC MIXED` and `PROC NLMIXED` in SAS that allow model fitting in an efficient manner suitable for multiple model comparison, and statistical simulations and data analysis.

However, in practice censored observations often occur which complicates the fitting of mixed effects models making it difficult to use existing software to correctly analyze the data. The treatment of univariate longitudinal data subject to DLs has been discussed by Hughes (1999), Vaida et al. (2007), and Vaida and Liu (2009). Hughes proposed a Monte Carlo EM algorithm (MCEM) for linear mixed effects models with censored response which involves Monte Carlo simulation at the E-step of the algorithm. Vaida et al. (2007) proposed a hybrid EM with a more efficient implementation of Hughes' algorithm including a numeric computation at the E-step for clusters with one or two censored observations instead of the Monte Carlo approach proposed by Hughes. Even though their approach improves the simulation at the E-step and the numeric implementation at the M-step, just like any MCEM algorithm, MCEM by its nature is an expensive proposition computationally since it combines a Monte Carlo simulation with an iterative procedure like EM (Ruppert, 2005).

Motivated by this drawback, Vaida and Liu (2009) proposed a numeric implementation of the EM algorithm with improved speed and precision compared to previous methods. They show the E-step of the EM algorithm reduces to computing the first two moments of a certain truncated multivariate normal (multinormal) distribution, and therefore can be done without simulation. The general formulas for these moments were derived by Tallis (1961) and Finney (1962). Vaida and Liu's approach may be used to compute either the maximum likelihood (ML) or the restricted maximum likelihood (REML) estimate. They also provide explicit derivations for a wide class of mixed effects models with censored response. These includes the Laird and Ware model (1982) and extension to different variance-covariance structures for random effects, heteroscedastic and autocorrelated errors, and multilevel models. However their R function, `lmec`, which implements their proposed methods, cannot fit the class of mixed effects models with censored response when the error variances are heteroscedastic, autocorrelated or when there are multiple levels of nesting in the random effect.

In the case where multiple response variables are observed repeatedly over time, modeling such multivariate longitudinal data using `PROC MIXED` has been proposed by Hamlett et al. (2004) and Thiebaut et al. (2002). Roy (2006) further generalized the approach in Hamlett et al. (2004) to allow measurement errors to be correlated. However, these approach can not accommodate censored observations. Comulada and Weiss (2010) studied the bivariate longitudinal data but their focus is on sample size and power calculations and their approach can not accommodate censored data as well. Lyles and Chuachoowong (2001) studied the bivariate censoring problem without repeated measurements by optimizing the profile likelihood and drawing inferences with a profile likelihood approach. Hopke et al. (2001) address multivariate repeated measures with left censoring and relatively general covariance structure and proposed multiple imputation (MI) algorithm. However left-censoring does not satisfy the "missing at random" condition (Little and Rubin, 1987) typically assumed in MI.

1.2.2 Detection Limits

A DL is a threshold below which measured values are not considered significantly different from a null signal. Thus the signal is not detected. DLs induce left-censoring when values lying somewhere to the left of the detection threshold are unknown. This is commonly known as a "nondetect" in environmental studies (Helsel, 2005). Occasionally, upper detection limits arise as well, in which case the DL induces right-censoring when values are only known to lie above a specified value. In some cases DLs induce interval-censoring when there is both a lower DL and an upper DL. Interval censoring may also be induced by the presence of both DLs and quantitation limits (QLs), but we defer description of that case until chapter 5. Data that are affected by DLs arise and cause difficulties in several disciplines such as environmental science, epidemiology, and clinical trials. They are encountered in human HIV and other viral studies (Hughes, 1999; Vaida and Liu, 2009), studies of air quality (Rao et al., 1991), and studies of streams and ground water quality (Helsel and Gilliom, 1986), to name just a few. The issue of how best to deal with data subject to DLs is a concern for many scientists. There are three main general approaches that scientists use to deal with a data subject to DL. These include substitution, nonparametric, and maximum likelihood (ML) estimation methods.

Substitution Methods

Substitution methods are naive approaches that involve fabricating values as a function of the recorded DLs and substituting those values in place of the censored values in the data set. For instance substituting half the DLs, DLs/$\sqrt{2}$, or the DLs themselves for the censored values are the most common substitution methods. Statistical analysis and summary statistics, are then calculated using these fabricated numbers, along with uncensored values.

One advantage of substitution methods is that they are very simple and no special software packages are required for statistical analysis. As a result, substitution methods are widely used, even though there is no theoretical basis for the approach and it has been shown to lead to inefficient and/or biased results, especially when there are multiple DLs (Helsel, 2005). In particular, substitution methods fail to incorporate incomplete data uncertainty. That is when a value is censored, there is some uncertainty about the true value. Therefore substituting a single value may produce estimates with under-estimated standard errors (since they do not account for uncertainty of censored data), and hence statistical inference based on these simple substitution methods may be misleading especially when the proportion of censoring is high. We will use the substitution approach as a standard of comparison.

ML Estimation Methods

ML estimation methods are the standard approach for dealing with censored data. A major reason is that ML estimators have attractive asymptotic properties, such as consistency and asymptotic normality, under some regularity conditions. The estimation procedure is based on the likelihood function L, which is the product of an uncensored component and a censored component. For instance, suppose that for subject i, y_i is the measurement or, in the case of a nondetect, the DL, and let c_i be an indicator for whether a detectable measurement was obtained. Then for left-censored data, the likelihood function can be expressed as

$$L = \prod_{i}^{m} p(y_i)^{c_i} F(y_i)^{1-c_i}$$

where p is the probability distribution function (pdf), and F is the cumulative distribution function (cdf) assumed for the measurements had they not been censored. ML estimation and inference such as hypothesis tests, confidence intervals can then be performed based on the likelihood.

Some statistical packages that implement ML for censored data are designed for the right-censored case, which is frequently encountered in survival analysis. However in some disciplines such as environmental science, left-censored data are common. According to Helsel (2005), even though most of these packages are designed only for right-censored data, they can also be used for left-censored data by a simple transformation of the data. This transformation is known as "flipping":

$$\text{Flipped Data} = \text{Constant} - \text{Original Data}$$

where Constant could be any large value, preferably the largest value in the original data set. Since this a linear transformation, the shape of the data does not change other than reversing its ordering.

Apart from the advantage of attractive asymptotic properties of ML estimators, expectation maximization (EM) algorithm can be used to compute ML estimates in the presence of censoring or unobservable. The advantage of using the EM to compute the ML estimates is that convergence is stable and the M-step involves only a complete-data maximization so it is often computationally straightforward. This makes ML methods particularly useful for our problem since the M-step of our EM process involves repeatedly using an already existing function, `lme` for optimization and hence provides us with the ability to estimate all the parameters in our model including the contemporaneous correlation efficiently.

NONPARAMETRIC METHODS

The nonparametric approach uses the relative positions (ranks) of the data. Nonparametric methods are useful for censored data because they efficiently use the available information. For example, in the case of left-censoring, data below the DLs are known to be lower than the uncensored values and hence are ranked lower. Similarly in the case of right-censoring, data above the DLs are known to be higher than uncensored values and hence are ranked higher. An advantage of the nonparametric approach is that censored values do not have to be imputed; they are regarded as tied for the lowest rank (left-censoring), which is an accurate reflection of the available information, rather than an inference about unobserved values from observed values.

The nonparametric approach is appealing in some contexts, but not for our problem because switching to ranks destroys the correlation structure we are interested in. In addition, the nonparametric approach would yield a correlation coefficient that is interpreted in terms of the ranks, rather than the original variables, which is not what is typically desired.

1.2.3 ESTIMATION METHODS IN MIXED MODELS WITH CENSORED AND MISSING RESPONSE

LMEMs have proven to be valuable data analytic tools in a variety of disciplines due to their ability to model multiple sources of heterogeneity and complex patterns of correlation. A general LMEM can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \qquad (1.1) \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}) \end{aligned}$$

where $\mathbf{y}$ is the entire $N-$dimensional response vector, $\mathbf{X}$ and $\mathbf{Z}$ are known design matrices for the fixed and random effects, respectively, $\boldsymbol{\beta}$ is a vector of unknown fixed effects, $\mathbf{b}$ is a vector of random effects, and $\boldsymbol{\epsilon}$ is a vector of random error terms. The most general assumptions for the LMEM are: $\mathrm{E}(\mathbf{b}) = \mathbf{0}, \mathrm{var}(\mathbf{b}) = \mathbf{D}$, $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \mathrm{var}(\mathbf{b}) = \mathbf{R}$, and $\mathrm{cov}(\mathbf{b}, \boldsymbol{\epsilon}) = \mathbf{0}$. The variance-covariance matrices $\mathbf{D}$ and $\mathbf{R}$ are not assumed to be known and are of general form. Estimation and inference are usually likelihood based methods (e.g., maximum likelihood estimation; likelihood ratio, score and Wald tests) based on either the log-likelihood or restricted log-likelihood function (Laird and Ware, 1982; Lindstrom and Bates, 1988). We make the usual assumptions that $\mathbf{b} \sim N(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$. These additional normality assumptions make the LMEM a class of Gaussian linear mixed models. Typically, $\mathbf{D}$ and $\mathbf{R}$ are parameterized by a lower dimension variance-covariance parameter $\boldsymbol{\theta}$. Note that the unknown parameters of the LMEM are $\boldsymbol{\beta}, \boldsymbol{\theta}$. Let $\mathrm{var}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \mathbf{R} = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, then the marginal distribution of the response $\mathbf{y}$ is given by $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Under normality assumptions, the loglikelihood of the LMEM in (1.1) is

$$\log L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log \boldsymbol{\Sigma} - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{1.2}$$

In the special case where $\boldsymbol{\Sigma}$ is known up to a proportionality constant σ^2, the values of $\boldsymbol{\beta}$ and σ^2 that maximize the likelihood are

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} && (1.3) \\ \hat{\sigma}^2 &= \frac{1}{N}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) && (1.4) \end{aligned}$$

which may be used to derived the profile likelihood. Profile likelihood is an alternative way to maximize $\log L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y})$ (Kalbfleisch and Sprott, 1970; McCullagh and Tibshirani, 1990). Instead of solving the estimating equations simultaneously, the method of profile likelihood maximizes $\log L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y})$ with respect to each parameter in turn.

Typically $\boldsymbol{\Sigma}$ is unknown. ML estimates of the unknown parameters in LMEM model (1.1) can be obtained using an iterative algorithm such as the EM algorithm or Newton-Raphson method (Laird and Ware, 1982; Lindstrom and Bates, 1988; Verbeke and Molenberghs, 2001). Pinheiro and Bates (2000) provided computational details. However in some cases the components of the response $\mathbf{y}$ may be incompletely observed as a result of censoring or missingness. In such a situation, the EM algorithm and some of its extensions, such as the Monte Carlo EM algorithm, are especially convenient for obtaining the ML estimate (Hughes, 1999; Vaida and Liu, 2009). In this dissertation, we focus on the application of the EM algorithm for ML and REML estimates of parameters in LMEM while simultaneously accounting for censoring and missingness in the response.

1.2.4 EM ALGORITHM

The EM algorithm is a general purpose, iterative technique for finding ML estimates in a variety of incomplete data problems (Dempster, Laird and Rubin, 1977). It has been successfully applied not only in situations where the incompleteness of the data is obvious such as when there are missing or censored data, but also in a variety of situations where the incompleteness of the data is not natural or evident. For instance, it is useful for fitting mixed effects models where the random effects are viewed as unobservable or "missing data".

Let $\mathbf{Y}$ be the random vector corresponding to the observed data $\mathbf{y}$ and let $\mathbf{z}$ denotes the vector containing the additional data, referred to as the unobservable or missing data. Let $\mathbf{x}$ denote the complete or augmented data, so that $\mathbf{x} = (\mathbf{y}, \mathbf{z})$. If we denote the probability density function (p.d.f) of the complete data as $f_c(\mathbf{x}|\boldsymbol{\psi})$, where $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_d)$ is a vector of unknown parameters on the parameter space $\boldsymbol{\Omega}$, the complete data log likelihood function is given by

$$\log L_c(\boldsymbol{\psi}|\mathbf{x}) = \log f_c(\mathbf{x}|\boldsymbol{\psi}).$$

Given $\boldsymbol{\psi}^{(0)}$ an initial value for $\boldsymbol{\psi}$, then on the first iteration, the E-step requires the calculation of

$$Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(0)}) = E_{\mathbf{z}|\mathbf{y}}\{\log L_c(\boldsymbol{\psi}|\mathbf{x})|\mathbf{y}, \boldsymbol{\psi}^{(0)}\}.$$

The M-step then involves the maximization of $Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(0)})$ with respect to $\boldsymbol{\psi}$ over the parameter space $\boldsymbol{\Omega}$. That is, we choose $\boldsymbol{\psi}^{(1)}$ such that

$$Q(\boldsymbol{\psi}^{(1)}|\boldsymbol{\psi}^{(0)}) \geq Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(0)})$$

for all $\boldsymbol{\psi} \in \boldsymbol{\Omega}$. The next iteration of the E- and M-steps are carried out by replacing $\boldsymbol{\psi}^{(0)}$ with the current estimate $\boldsymbol{\psi}^{(1)}$ and the process is repeated until a convergence criterion is reached.

1.3 SUMMARY

As discussed above, the development of mixed model methodologies has provided us with a powerful set of tools to address challenging problems arising from analyzing a longitudinal/clustered data with complex features. Many application have been seen in biomedical, environmental studies, etc. The contribution in this dissertation is to study the mixed model methodologies in analyzing U/MLD subjects to DLs and missingness.

In Chapter 2, we consider a ULD with censored and missing response. We propose an EM implementation for LMEM with censored data implemented in a computational package, **lmecm**, which fit a wide class of LMEMs with censored response when the error variance are heteroscedastic and/or autocorrelated, and when there are multiple levels of nesting in the random effects. We extend the methodology to accommodate the case where censored and missing values occurred simultaneously. We illustrate the performance of our proposed methodology through extensive simulation studies and data analysis.

In Chapter 3, we develop a BLMEM that allows for the analysis of bivariate longitudinal data subject to DLs and missingness and easily extends to the case of three or more response variables. We develop a computationally feasible algorithm for ML and REML estimation and implement it in **lmecm**. In particular, our implementation accounts for the possibility of the different variation in the errors for the different responses, serial correlation across different time points for the different responses and contemporaneous correlation between the two responses. We also propose and evaluate methods of inference for the contemporaneous correlation coefficient, which is often of primary interest in applications of such models. The performance of our proposed method is illustrated through an extensive simulation studies and data analysis.

In Chapter 4, we provide documentation for the R package **lmecm**. We illustrates the features and capabilities of **lmecm** through empirical results. Concluding remarks and possible extensions of the methodology proposed in this dissertation are provided in chapter 5.

1.4 References

Bates, D. and D. Shakar (2007). *lme4: Linear mixed effects models using S4 classes.* R package version 0.9975-13.

Commodore, A., S. Hartinger, C. Lanata, D. Mausezahl, A. Gil, D. Hall, M. Aguilar-Villalobos, and L. Naeher (2013). A pilot study characterizing real time exposures to particulate matter and carbon monoxide from cookstove related woodsmoke in rural Peru. *Atmospheric Environment 79*, 380–384.

Comulada, W. and R. Weiss (2010). Sample size and power calculations fo correlations between bivariate longitudinal data. *Statistics in Medicine 29*, 2811–2824.

Finney, D. (1962). Cumulants of truncated multinormal distribution. *Journal of the Royal Statistical Society Series B 24*, 535–536.

Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Preceedings of the Twenty-Ninth Annual SAS Users Group International Conference, SAS Institute Inc.,2004*, 198–29:1–7.

Helsel, D. (2005). *Nondetect and data analysis, statistics for censored environmetal data.* New Jersey: John Wiley and Sons.

Helsel, D. and R. Gilliom (1986). Estimation of distributional parameters for censored trace level water quality data, 2. verification and application:. *Water Resources Research 22*, 147–155.

Hopke, P., C. Liu, and D. Rubin (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: time series concentrations of pollutants in arctic. *Biometrics 57*, 23–33.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*, 625–629.

Jiang, J. (2007). *Generalized linear models: theory and applications.* New York: Springer.

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

Lindstrom, M. and D. Bates (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association 83*, 1014–1022.

Little, R. and D. Rubin (1987). *Statitical analysis with missing data.* Wiley: New York.

Lyles, R. and J. W. R. Chuachoowong (2001). Correlating two viral loads assay with know detection limits. *Biometrics 57*, 1238–1244.

Pinheiro, J. and D. Bates (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., D. Bates, S. DeRoy, and D. Shakar (2006). *nlme: Linear and nonLinear mixed effects models*. R package version 3.1-78.

Rao, S., J. Ku, and K. Rao (1991). Analysis of toxic air contaminant Data containing concerntrations below the limit of detection. *Journal of Air and Waste Management Association 41*, 442–448.

Roy, A. (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical Journal 48*, 286–301.

Ruppert, D. (2005). Comments on maximization by parts in likelihood inference. *Journal of Air and Waste Management Association 100*, 1161–1163.

Tallis, G. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society, Series, B 23*, 223–229.

Thiebaut, R., H. Jacqmin-Gadda, G. Chene, C. Leport, and D. Commenges (2002). Bivariate linear mixed models using SAS PROC MIXED. *Computer Methods and Programs in Biomedicine 69*, 249–256.

Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybid EM for linear and nonlinear mixed effects model with censored response. *Computational Statistics and Data Analysis 51*, 5718–5730.

Vaida, F. and L. Liu (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Verbeke, G. and G. Molenberghs (2001). *Linear mixed models for longitudinal data.* New York: Springer.

Xie, X., X. Xue, S. Gange, H. Strickler, M. Kim, and W. H. Study-Group (2012). Estimation and inference on correlations between biomarkers with repeated measures and left-censoring due to minimum detection levels. *Statistics in Medicine 31*(21), 2275–2289.

Chapter 2

Linear Mixed Effects Models with Censored and Missing Response with Non-spherical Errors[1]

---

[1] Tuglo, E., and Hall, D. (2014). Linear Mixed Effects Models with Censored and Missing Response with Non-spherical Errors. To be submitted to *Journal of Computational Statistics and Data Analysis*.

ABSTRACT

In biomedical and environmental studies, measurements are often subject to an upper or, more commonly, lower detection limit (DL). Consequently, the responses are either right or left censored. An additional complication arises when components of the response are completely missing. As a result censored and missing data may arise simultaneously. Longitudinal continuous data are routinely analyzed with linear and nonlinear mixed models (N/LMEMs). In this paper we explore methods for fitting such models that can accommodate both censoring induced by DLs and missingness simultaneously. Extending the work of Hughes (1999) and Vaida et al. (2007), Vaida and Liu (2009a) proposed an EM algorithm for computing the maximum likelihood (ML) and restricted maximum likelihood (REML) estimators in LMEMs and NLMEMs. However, the M-step in their formulation becomes computationally difficult when dealing with non-spherical errors (e.g., any kind of serial and/or heteroscedastic errors) so their approach, while theoretically of broad applicability, has practical limitations. For instance, the R function introduced by Vaida and Liu, which is known as `lmec` (Vaida and Liu, 2009b), can only handle the spherical errors case. Motivated by this limitation, we propose a computationally feasible EM algorithm for ML and REML estimation for LMEM with censored and missing response and implement it in an R function `lmecm`, which offers greater flexibility (a wider class of models) than `lmec`. In particular, our implementation has the ability to handle all kinds of serial correlation structures, heteroscedastic errors, and multiple levels of nesting in random effects while accounting for censoring and missingness.

KEYWORDS: Detection limits; EM algorithm; Maximum likelihood; Heteroscedastic errors; Serial correlation.

2.1 Introduction

In biomedical and environmental studies, longitudinal or repeated measures are routinely analyzed using mixed effect models. Complications arise when the response is subject to lower or upper detection limits resulting in some observations being left or right-censored. In many cases the proportion of censored data may not be negligible (Hughes, 1999), so failure to account for censoring or the use of ad hoc imputation methods such as substituting the DLs, or some simple function of the DL, in the analysis might lead to biased estimators of the fixed effects and variance components (Wu, 2010).

As an example we consider a data set that contains real time measurements of carbon monoxide (CO) and particulate matter with aerodynamic diameter less than 2.5 microns (PM2.5) from wildland fire-fighters who are exposed to these environmental pollutants in their line of work. These response variables (PM2.5 and CO) were collected every 30 seconds for several hours on each of several days in the fire-fighting season on each of several subjects (fire-fighters). These fire-fighters were performing one of four different tasks when these measurements were taken. The response variables are subject to lower DLs. One objective of the investigators is to determine the mean response for the four different tasks.

As an alternative to ad hoc imputation methods, Hughes (1999) proposed a Monte Carlo EM algorithm (MCEM) for linear mixed model with censored response. Vaida et al. (2007) proposed a hybrid EM using a more efficient version of Hughes's algorithm that involved numeric computation at the E-step for clusters with one or two censored observations and extended the methodology to nonlinear mixed effect models with censored response. MCEM is an expensive proposition computationally (Ruppert, 2005), hence Vaida and Liu (2009a) proposed an EM algorithm that uses closed-form expressions at the E-step as opposed to Monte Carlo simulation. However, the M-step in their formulation becomes computationally difficult when dealing with non-spherical errors (any kind of serial and/or heteroscedastic errors) so their approach, while theoretically of broad applicability, has practical limitations.

For instance, **lmec**, the R package introduced by Vaida and Liu to implement their approach can only handle the spherical errors case.

Recently, there have been some developments in dealing with the case where the Gaussian assumption is questionable and the longitudinal responses are subject to DLs. Vock et al. (2011) proposed a mixed model framework for censored longitudinal data in which the random effects are represented by a flexible semiparametric density and illustrated the consequences of misspecifying the random effects distribution when the proportion of non-censored observation within a cluster is small. In a situation where the longitudinal responses are skewed in nature and are also subject to DLs, Bandyopadhyay et al. (2012) proposed a Bayesian analysis of censored linear mixed effects models replacing the Gaussian assumption with skew-normal/independent (SNI) distributions. Matos et al. (2013) extend the work of Vaida and Liu (2009a) and proposed a conditional expectation maximization algorithm (ECM) which enables them to develop diagnostics measures for detecting influential observation in linear and nonlinear mixed effects models with censored response.

Although, several papers have dealt with censored longitudinal data, to the best of our knowledge, except for Vaida and Liu (2009a), no attention has been given to the case of non-sphericity of the within-group error nor are there any packages to implement a suitable methodology. In this paper, we propose a computationally feasible EM algorithm for ML and REML estimation for LMEM with censored response and implement it in an R function `lmecm`, which offers greater flexibility than `lmec`. Our implementation has the ability to handle all kinds of serial correlation structures, heteroscedastic errors, multiple levels of nesting, and complex structure of random effects whiles accounting for censoring. We show that with pseudo data augmentation, the log-likelihood corresponding to LMEM with censored response is identical to the log-likelihood corresponding to LMEM without censored response in case of ML estimation for balanced data hence the M-step objective function can be optimized using an already existing function `lme` in the **nlme** package (Pinheiro et al.,

2006) without any modification. In case of unbalanced data and REML estimation, the M-step objective function corresponding to LMEM with censored response can be optimized with a slight modification of the `lme` function. These provides us the advantage of being able to implement all kinds serial correlation structures, and heteroscedastic errors in the **nlme** pacakge for LMEM with censored response.

The rest of this paper is organized as follows. In section 2, the LMEM with censored response is defined and the EM algorithm for obtaining the ML and REML estimates is described. In section 3, the extension to more general LMEM with censored response is discussed. The methodology is illustrated in section 4 with an application to the real time data and in section 5, three simulation studies are presented: the case of spherical error while accounting for censoring, the case of non-spherical error while accounting for censoring, and the case of non-spherical error while accounting for censoring and missingness simultaneously. Concluding remarks and discussion are given in section 6.

2.2 Linear Mixed Effects with Censored Response

2.2.1 Model Specification and Maximum Likelihood Estimation

Following Laird and Ware (1982), we write the standard LMEM for clustered data as

$$\begin{aligned}\mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots.., m, \\ \mathbf{b}_i &\sim N(0, \sigma^2\mathbf{D}_s), \qquad \boldsymbol{\epsilon}_i \sim N(0, \sigma^2\mathbf{I}),\end{aligned} \tag{2.1}$$

where $\mathbf{y}_i$ is a vector of n_i outcomes on the ith individual, $N = \sum_{i=1}^m n_i$, $\boldsymbol{\beta}$ is a $p-$dimensional vector of fixed effects, $\mathbf{b}_i$ is a $q-$dimensional vector of random effects, $\mathbf{X}_i(n_i \times p)$ and $\mathbf{Z}_i(n_i \times q)$ are known fixed effects and random regressor matrices, and $\boldsymbol{\epsilon}_i$ is an n_i-dimensional within-group error vector with spherical Gaussian distribution. The random effect $\mathbf{b}_i$ and the within-group errors $\boldsymbol{\epsilon}_i$ are assumed to be independent across subjects and of each other, and $\mathbf{D}_s$ is a positive definite symmetric scaled matrix. Let $\boldsymbol{\Delta}$ be the relative precision factor such that $\mathbf{D}_s^{-1} = \boldsymbol{\Delta}^T\boldsymbol{\Delta}$. Let $\boldsymbol{\theta}$ be an unconstrained set of parameters that determines $\boldsymbol{\Delta}$, and denote the combined set of model parameters as $\boldsymbol{\psi} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T$. Put $\mathbf{D} = \sigma^2\mathbf{D}_s$ and $\boldsymbol{\Sigma}_i = \mathrm{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I} = \sigma^2(\mathbf{Z}_i\mathbf{D}_s\mathbf{Z}_i^T + \mathbf{I}) = \sigma^2\boldsymbol{\Sigma}_{s,i}$. Based on the independence of $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ we can express the likelihood function of model (2.1) as

$$L(\boldsymbol{\psi} \mid \mathbf{y}) \;=\; \prod_{i=1}^{m} \int p(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\theta}, \sigma^2) p(\mathbf{b}_i \mid \boldsymbol{\theta}, \sigma^2) d\mathbf{b}_i.$$

Thus the non-observable random effects $\mathbf{b}_i, i = 1, \ldots\ldots, m$, are integrated out of the likelihood. Following Pinheiro and Bates (2000), in the absence of censoring the likelihood may be written as

$$L(\boldsymbol{\psi} \mid \mathbf{y}) \;=\; \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(- \frac{\sum_{i=1}^{m} \| \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \|^2}{2\sigma^2}\right) \prod_{i=1}^{m} \frac{\mathrm{abs}|\boldsymbol{\Delta}|}{\sqrt{| \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i |}}, \quad (2.2)$$

where

$$\widetilde{\mathbf{y}}_i \;=\; \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\mathbf{X}}_i \;=\; \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\mathbf{Z}}_i \;=\; \begin{bmatrix} \mathbf{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix},$$

are augmented data vectors and model matrices (cf. equation (2.10) in Pinheiro and Bates (2000)) and $\hat{\mathbf{b}}_i = (\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})$. The above data augmentation procedure is called a *pseudo data approach* because the contribution of the random effects to the marginal likelihood is changed into extra rows for the response and design matrices. The pseudo data representation of the marginal density $p(\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ is just one way of expressing these density and deriving the likelihood. It follows that the log-likelihood is given by

$$\begin{aligned} \log L(\boldsymbol{\psi} \mid \mathbf{y}) \;&=\; -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\| \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \|^2 \right) \\ &\quad + \sum_{i=1}^{m}\left(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log | \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i | \right). \qquad (2.3) \end{aligned}$$

In the situations of interest in this research, $\mathbf{y}_i$ may be incompletely observed due to left or right censoring imposed by the laboratory methods or some other mechanism. We defer the treatment of completely missing case to section (2.2.2). Instead of y_{ij}, we observe the pair (Q_{ij}, C_{ij}) where Q_{ij} is a vector of uncensored readings or censoring levels and C_{ij} is a vector of censoring indicators. We will assume for simplicity of description that the data are left-censored. Thus we can write

complete data: $(\mathbf{y}_i),\ \ i = 1, \ldots, m$

observed data: $(\mathbf{Q}_i, \mathbf{C}_i),\ \ i = 1, \ldots, m$, where

$$\begin{aligned} \mathrm{y}_{ij} \le Q_{ij} \text{ if } C_{ij} = 1 \\ \mathrm{y}_{ij} = Q_{ij} \text{ if } C_{ij} = 0 \end{aligned}$$

and $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\mathbf{Q}_i = (Q_{i1}, \ldots, Q_{in_i})^T$, $\mathbf{C}_i = (C_{i1}, \ldots, C_{in_i})^T$. Following Magnus and Neudecker (1999) and Wand (2002), let $\mathrm{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_m)$ be the block diagonal matrix with diagonal blocks $\mathbf{B}_1, \ldots, \mathbf{B}_m$. if $\mathbf{B}_1 = \cdots = \mathbf{B}_m = \mathbf{B}$, we write $\mathrm{diag}_m(\mathbf{B})$ for $\mathrm{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_m)$ and let $\mathrm{vec}(\cdot)$ denote the function that stacks the columns of its argument(s) into a single column vector. We write $\mathbf{y} = \mathrm{vec}(\mathbf{y}_1, \ldots, \mathbf{y}_m) = (\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T)^T$, $\mathbf{Q} = \mathrm{vec}(\mathbf{Q}_1, \ldots, \mathbf{Q}_m)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_m)$. Then the complete data log-likelihood in this case is expression (2.3). Let $\boldsymbol{\psi}^{(k)}$ be the current estimate of the parameters in the model. On the (k+1)th iteration, the E-step for the EM formulation then involves computing $E(\log L(\boldsymbol{\psi} \mid \mathbf{y}) | \mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}^{(k)})$ which is given by

$$\begin{aligned} Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= E\left\{ \log L(\boldsymbol{\psi} \mid \mathbf{y}) \middle| \mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}^{(k)} \right\} \\ &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m} E\bigg(\parallel \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \parallel^2 |\ \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)} \bigg) \\ &\quad + \sum_{i=1}^{m}\bigg(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid \bigg). \end{aligned} \tag{2.4}$$

Following Vaida and Liu (2009a), $\parallel \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \parallel^2 = \parallel (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta}) \parallel^2$ where $\mathbf{P}_{\widetilde{\mathbf{Z}}_i} = \widetilde{\mathbf{Z}}_i(\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T$ is the projection matrix onto the space spanned by $\widetilde{\mathbf{Z}}_i$. Then

$$
\begin{aligned}
&E(\|\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) \\
&\quad= E\bigg((\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}\bigg) \\
&\quad= \mathrm{tr}\bigg\{(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})E\bigg[(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}\bigg]\bigg\} \\
&\quad= \mathrm{tr}\bigg\{(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})\bigg\} \\
&\qquad + \mathrm{tr}\bigg\{(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})\mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})\bigg\} \\
&\quad= \bigg\|E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_i\bigg\|^2 + \mathrm{tr}\bigg\{(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})\mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})\bigg\}. \quad (2.5)
\end{aligned}
$$

where $\widetilde{\hat{\mathbf{b}}}_i = (\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})$. Let $\widetilde{\mathbf{S}}_i\widetilde{\mathbf{S}}_i^T = \mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})$, where $\widetilde{\mathbf{S}}_i$ the lower triangular Cholesky factor and such that $\widetilde{\mathbf{S}}_i = (\widetilde{\mathbf{S}}_{i,1}, \ldots, \widetilde{\mathbf{S}}_{i,n_i+q})$ then
$\mathrm{tr}\bigg\{(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})\mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})\bigg\} = \mathrm{tr}\bigg\{\widetilde{\mathbf{S}}_i^T(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i})\widetilde{\mathbf{S}}_i\bigg\}$

$$
= \left\|\begin{pmatrix} \widetilde{\mathbf{S}}_{i,1} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i}\widetilde{\mathbf{S}}_{i,1} \\ \vdots \\ \widetilde{\mathbf{S}}_{i,n_i+q} - \mathbf{P}_{\widetilde{\mathbf{Z}}_i}\widetilde{\mathbf{S}}_{i,n_i+q} \end{pmatrix}\right\|^2 = \left\|\begin{pmatrix} \widetilde{\mathbf{S}}_{i,1} - \mathbf{0}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i(\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{S}}_{i,1} \\ \vdots \\ \widetilde{\mathbf{S}}_{i,n_i+q} - \mathbf{0}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i(\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{S}}_{i,n_i+q} \end{pmatrix}\right\|^2 \quad (2.6)
$$

substituting (2.6) into (2.5) yields

$$
\begin{aligned}
&E(\|\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) \\
&\quad= \left\|\begin{pmatrix} E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_i \\ \widetilde{\mathbf{S}}_{i,1} - \mathbf{0}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,1} \\ \vdots \\ \widetilde{\mathbf{S}}_{i,n_i+q} - \mathbf{0}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,n_i+q} \end{pmatrix}\right\|^2 = \sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \quad (2.7)
\end{aligned}
$$

where

$$
\widetilde{\mathbf{w}}_{i,l} = \begin{cases} E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})^T & l = 1 \\ \widetilde{\mathbf{S}}_{i,l-1} & l = 2, \ldots, n_i + q + 1, \end{cases}
$$

$$\begin{aligned}
\widetilde{\mathbf{X}}_{i,l} &= \begin{cases} \widetilde{\mathbf{X}}_i & l = 1 \\ \\ \mathbf{0} & l = 2, \dots, n_i + q + 1, \end{cases} \qquad (2.8) \\
\widetilde{\mathbf{Z}}_{i,l} &= \widetilde{\mathbf{Z}}_i \ \ \forall \ l, \qquad\qquad\qquad\qquad \text{and} \\
\widetilde{\hat{\mathbf{b}}}_{i,l} &= \begin{cases} \widetilde{\hat{\mathbf{b}}}_i & l = 1 \\ \\ \widetilde{\mathbf{Z}}_i(\widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i)^{-1} \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{S}}_{i,l-1} & l = 2, \dots, n_i + q + 1. \end{cases}
\end{aligned}$$

Thus for each cluster with a censored observation, we introduce an additional $(n_i + q)$ subcluster with the response vector, fixed and random effect design matrices as in (2.8). Substituting (2.7) into (2.4) yields

$$\begin{aligned}
Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} \sum_{l=1}^{n_i+q+1} \| \widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l} \boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l} \widetilde{\hat{\mathbf{b}}}_{i,l} \|^2 \\
&\quad + \sum_{i=1}^{m} \left(\log \operatorname{abs} |\boldsymbol{\Delta}| - \frac{1}{2} \log \mid \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \mid \right). \qquad (2.9)
\end{aligned}$$

Using the profile likelihood approach, for a given $\boldsymbol{\theta}$, we can calculate the conditional estimates $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\hat{\sigma}^2(\boldsymbol{\theta})$ as the values that maximize $Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)})$. Let

$$\begin{aligned}
\boldsymbol{\delta} &= \operatorname{vec}(\mathbf{b}_{1,1}, \dots, \mathbf{b}_{m,1}, \dots, \mathbf{b}_{m,n_1+q+1}, \dots, \mathbf{b}_{m,n_m+q+1}, \boldsymbol{\beta}) \ \ \text{and} \\
\widetilde{\mathbf{M}} &= \begin{pmatrix}
\widetilde{\mathbf{Z}}_1 & & & & & 0 & \widetilde{\mathbf{X}}_1 \\
& \ddots & & & & & \vdots \\
& & & & & & \\
& & \widetilde{\mathbf{Z}}_m & & & & \widetilde{\mathbf{X}}_m \\
& & & \widetilde{\mathbf{Z}}_{m,n_1+q+1} & & & \mathbf{0} \\
& & & & \ddots & & \vdots \\
0 & & & & & \widetilde{\mathbf{Z}}_{m,n_m+q+1} & \mathbf{0}
\end{pmatrix}; \qquad (2.10)
\end{aligned}$$

then M-step updates for $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\hat{\sigma}^2(\boldsymbol{\theta})$ are

$$\hat{\boldsymbol{\delta}} = (\widetilde{\mathbf{M}}^T \widetilde{\mathbf{M}})^{-1} \widetilde{\mathbf{M}}^T \widetilde{\mathbf{w}} \qquad (2.11)$$

$$\hat{\sigma}^2(\boldsymbol{\theta}) \;=\; \frac{\|\widetilde{\mathbf{w}} - \widetilde{\mathbf{M}}\hat{\boldsymbol{\delta}}\|^2}{N} = \frac{1}{N}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2.$$

Substituting these conditional estimators back into $Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)})$ provides the profile log-likelihood

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) = const - \frac{N}{2}\log\left(\hat{\sigma}^2(\boldsymbol{\theta})\right) + \sum_{i=1}^{m}\left(\log \text{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid\right). \tag{2.12}$$

ML estimates are obtained by first optimizing the profile log-likelihood (2.12) over $\boldsymbol{\theta}$ and then replacing $\boldsymbol{\theta}$ with its ML estimate $\hat{\boldsymbol{\theta}}$, to get the ML estimates for $\boldsymbol{\beta}$ and σ^2. The optimization of the profile likelihood can be accomplished by a general purpose optimization function like `optim` and `nlminb` in the **stat** package in R but from a practical perspective, it is important to note that no specialized routines are needed to calculate values $\boldsymbol{\theta}$, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\hat{\sigma}^2(\boldsymbol{\theta})$ as described above. The `lme` function in the **nlme** package in R estimates parameters in the model using the profile likelihood approach described in (2.10)-(2.12) so instead of updating the parameters as described above, we can just use the `lme` function as our M-step. Thus we fit a LMEM using the `lme` function with the response $\widetilde{\mathbf{w}}$ and the design matrices $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Z}}$ for the fixed and random effect respectively. Thus estimating the parameters in the models reduces to repeatedly fitting LMEMs using `lme` function until convergence. Below are the computational details for the E and M-steps of our EM formulation for estimating the parameters in the model.

2.2.2 Computational Details of the E-step

From equation (2.9), it is clear that the E-step reduces to computing $\widetilde{\mathbf{w}}_{i,1}$ and $\widetilde{\mathbf{S}}_i$ which then reduces to computing $E(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \psi^{(k)})$ and $\text{Var}(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \psi^{(k)})$, which are the mean and variance of a truncated multinormal distribution taken at the current parameter $\psi^{(k)}$. Vaida and Liu (2009a) identified closed form expressions for $E(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \psi^{(k)})$ and $\text{Var}(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \psi^{(k)})$ as a functions of multinormal probabilities using a sequence of transformations. The general formulas for these moments were derived by Tallis (1961) and Finney

(1962). The first step is to treat separately the observed and the censored components. Let $\mathbf{y}_i^o$ be n_i^o-vector of observed outcomes with $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$ and $\mathbf{y}_i^c$ be n_i^c-vector of censored observation with $C_{ij} = 1$ for all elements in $\mathbf{y}_i^c$ such that $(n_i = n_i^o + n_i^c)$. After reordering, $\mathbf{y}_i, \mathbf{Q}_i, \mathbf{X}_i, \mathbf{\Sigma}_i$ can be partitioned as follows

$$\mathbf{y}_i = \mathrm{vec}(\mathbf{y}_i^o, \mathbf{y}_i^c), \quad \mathbf{Q}_i = \mathrm{vec}(\mathbf{Q}_i^o, \mathbf{Q}_i^c), \quad \mathbf{X}_i^T = (\mathbf{X}_i^o, \mathbf{X}_i^c), \quad \mathbf{\Sigma}_i = \begin{pmatrix} \mathbf{\Sigma}_i^{oo} & \mathbf{\Sigma}_i^{oc} \\ \mathbf{\Sigma}_i^{co} & \mathbf{\Sigma}_i^{cc} \end{pmatrix}.$$

In the absence of censoring, $\mathbf{y}_i \sim \mathrm{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{\Sigma}_i)$. Then $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \mathbf{\Sigma}_i^{oo})$, $\mathbf{y}_i^c|\mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\boldsymbol{\mu}_i, \mathbf{S}_i)$ where $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{\Sigma}_i^{co}(\mathbf{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \mathrm{Var}(\mathbf{y}_i^c|\mathbf{y}_i^o) = \mathbf{\Sigma}_i^{cc} - \mathbf{\Sigma}_i^{co}(\mathbf{\Sigma}_i^{oo})^{-1}\mathbf{\Sigma}_i^{oc}$. In the presence of censoring, $\mathbf{y}_i^c$ lies within the interval $(-\infty, \mathbf{Q}_i^c)$. Thus, conditionally, $\mathbf{y}_i^c$ has a multinormal distribution right-truncated at $\mathbf{Q}_i^c$. Let $\boldsymbol{\mu}_i^c = E(\mathbf{y}_i^c|\mathbf{y}_i^o, \mathbf{Q}_i^c)$ and $\mathbf{S}_i^c = \mathrm{Var}(\mathbf{y}_i^c|\mathbf{y}_i^o, \mathbf{Q}_i^c)$. It follows that $E(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}) = E(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \mathbf{y}_i^o, \boldsymbol{\psi}) =$ $\mathrm{vec}(\mathbf{y}_i^o, E(\mathbf{y}_i^c|\mathbf{Q}_i, \mathbf{y}_i^o, \boldsymbol{\psi})) = \mathrm{vec}(\mathbf{y}_i^o, \boldsymbol{\mu}_i^c)$ and $\mathrm{Var}(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_i^c \end{pmatrix}$. For more details on the computation of these moments, one may refers to Vaida and Liu (2009a).

Now suppose instead of censoring, portions of the response are ignorably missing in a sense described by Rubin (1976). Let $\mathbf{y}_i^m$ be the missing component with the $\mathbf{Q}_i$, $\mathbf{X}_i$ and $\mathbf{\Sigma}_i$ partitioned appropriately as above. Then $\mathbf{y}_i^m|\mathbf{y}_i^o \sim \mathrm{N}_{n_i^m}(\boldsymbol{\mu}_i, \mathbf{S}_i)$, where $\boldsymbol{\mu}_i$ and $\mathbf{S}_i$ are defined exactly as in the censoring case but with superscript c replaced with superscript m. Similarly, in the presence of missingness, $\mathbf{y}_i^m|\mathbf{y}_i^o \sim \mathrm{N}(\boldsymbol{\mu}_i, \mathbf{S}_i)$ and lies within the interval $\mathbf{y}_i^m|\mathbf{y}_i^o \in (-\infty, \mathbf{Q}_i^c = \infty)$ so $\mathbf{y}_i^m|\mathbf{y}_i^o$ conditional on $-\infty < \mathbf{y}_i^m < \mathbf{Q}_i^c = \infty$ has a multinormal distribution right-truncated at $\mathbf{Q}_i^c = \infty$. Thus in this framework, missingness is a special case of left-censoring with censoring level $\mathbf{Q}_i^c = \infty$. Computationally, we suggest setting $\mathbf{Q}_i^c$ to arbitrary large number say $\max \mathbf{y}_i$. Note that an observation which is distinctly different from the rest of the data or quite inconsistent with majority of the data is considered an outlier so setting $\mathbf{Q}_i^c = \max \mathbf{y}_i$ guarantees that the imputed values are consistent with the rest of the data. Hence dealing with censored and missing response simultaneously reduces to dealing with multiple level of DLs which is easily accommodated in this framework. In a more general setting the relationship between the problems of left-censoring due to DLs and missing data has been discussed in Senn et al. (2012).

Note that equation (2.9) is equivalent to (2.3), the complete data log-likelihood, except that in the second term of equation (2.9), for each cluster with a censored observation, an additional (n_i+q) sub-cluster is introduced with the response vector, fixed and random effect design matrices as in (2.8). Thus the routine for optimizing the log-likelihood (2.3), the `lme` function in the **nlme** package can be used to optimize the M-step objective function (2.9). At the M-step, we fit a LMEM by passing `lme` the response $\mathbf{w}$ and the design matrices $\mathbf{X}$ and $\mathbf{Z}$ for the fixed and random effects respectively as in (2.8). In the case that the data are unbalanced, which is more common in practice, let

$$\ddot{n} = \max_i n_i$$

and

$$\ddot{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \quad \ddot{\mathbf{X}}_i = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}, \quad \ddot{\mathbf{Z}}_i = \begin{bmatrix} \mathbf{Z}_i \\ \mathbf{0} \end{bmatrix},$$

which implies that

$$\widetilde{\ddot{\mathbf{y}}}_i = \begin{bmatrix} \ddot{\mathbf{y}}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\ddot{\mathbf{X}}}_i = \begin{bmatrix} \ddot{\mathbf{X}}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\ddot{\mathbf{Z}}}_i = \begin{bmatrix} \ddot{\mathbf{Z}}_i \\ \boldsymbol{\Delta} \end{bmatrix},$$

where $\ddot{\mathbf{y}}_i, \ddot{\mathbf{X}}_i, \ddot{\mathbf{Z}}_i$ are $\ddot{n}\times 1$, $\ddot{n}\times p$ and $\ddot{n}\times q$, respectively. Thus for each cluster i with cluster size less than $\ddot{n}$, we augment the response vector and the design matrices with $\ddot{n}-n_i$ extra rows of zeros. The goal is to obtain a balanced data set so the methodology described below applies. With the response $\ddot{\mathbf{w}}$, and design matrices $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$ given as arguments, `lme` constructs the objective function

$$\begin{aligned} Q_{lme}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{\widetilde{\ddot{N}}}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{\ddot{n}+q+1} \|\widetilde{\ddot{\mathbf{w}}}_{i,l} - \widetilde{\ddot{\mathbf{X}}}_{i,l}\boldsymbol{\beta} - \widetilde{\ddot{\mathbf{Z}}}_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\ &+ \sum_{i=1}^{m}(\ddot{n}+q+1)\left(\log \operatorname{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\ddot{\mathbf{Z}}}_i^T \widetilde{\ddot{\mathbf{Z}}}_i \mid \right) \end{aligned} \tag{2.13}$$

$$
\begin{aligned}
&= -\frac{\widetilde{\ddot{N}}}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\
&\quad + \sum_{i=1}^{m}(\ddot{n}+q+1)\Big(\log\text{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_{i,l}^T\widetilde{\mathbf{Z}}_{i,l}\ |\Big) \qquad (2.14)
\end{aligned}
$$

where $\widetilde{\ddot{N}} = \sum_{i=1}^{m}\ddot{n}(\ddot{n}+q+1) = m\ddot{n}(\ddot{n}+q+1) = \ddot{N}(\ddot{n}+q+1)$. Note that

$$
\sum_{i=1}^{m}\sum_{l=1}^{\ddot{n}+q+1}\|\widetilde{\ddot{\mathbf{w}}}_{i,l} - \widetilde{\ddot{\mathbf{X}}}_{i,l}\boldsymbol{\beta} - \widetilde{\ddot{\mathbf{Z}}}_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 = \sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2
$$

and

$$
\widetilde{\ddot{\mathbf{Z}}}_i^T\widetilde{\ddot{\mathbf{Z}}}_i = \begin{pmatrix} \mathbf{Z}_i^T & \mathbf{0}^T & \boldsymbol{\Delta}^T \end{pmatrix}\begin{pmatrix} \mathbf{Z}_i \\ \mathbf{0} \\ \boldsymbol{\Delta} \end{pmatrix} = \mathbf{Z}_i^T\mathbf{Z}_i + \mathbf{0}^T\mathbf{0} + \boldsymbol{\Delta}^T\boldsymbol{\Delta} = \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i.
$$

From (2.13)-(2.14), it is clear that, except for the change in dimension, by augmenting the response and the design matrices with extra zeros, we have achieved a balanced data set without changing the contribution of the response and the design matrices to the log-likelihood. Using profile likelihood estimation, for a given $\boldsymbol{\theta}$, `lme` computes the estimates of $\boldsymbol{\delta}$ and σ^2 as in (2.11) except with the divisor $\widetilde{\ddot{N}}$ instead of N. Substituting these conditional estimators back into $Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)})$ yields the profile likelihood

$$
\begin{aligned}
Q_{lme}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) = const + (\ddot{n}+q+1)\Big[-\frac{\ddot{N}}{2}\log(\hat{\sigma}^2(\boldsymbol{\theta})) + \sum_{i=1}^{m}\Big(\log\text{abs}|\boldsymbol{\Delta}| \\
- \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\Big)\Big]. \qquad (2.15)
\end{aligned}
$$

But the profile log-likelihood we want to optimize is (2.12), and (2.15) is equivalent to (2.12) except that N is replaced with $\ddot{N}$. This can be corrected with a minor modification to the `lme` function whereby we subtract N_r from $\ddot{N}$ in the first piece of the profile log-likelihood (2.15) such that $\ddot{N} - N_r = N$. Note that in the case original data are balanced, $N_r = 0$ and $\ddot{N} = N$ so such modification is not needed. For the purpose of optimization the constant in

(2.12) and (2.15) is neglected so $Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)})$ can be written as

$$\begin{aligned} Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) &= (\ddot{n}+q+1)\Big[-\frac{N}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta})\big) + \sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\,\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\,|\,\big)\Big] \\ \Longrightarrow \quad Q(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) &= \frac{1}{\ddot{n}+q+1}Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}). \end{aligned} \tag{2.16}$$

Thus,

$$\arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) = \arg\max_{\boldsymbol{\theta}}\left\{\frac{1}{\ddot{n}+q+1}Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)})\right\}, \tag{2.17}$$

indicating that the $\boldsymbol{\theta}$ that maximizes the profile likelihood, $Q_{lme}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)})$ is the same $\boldsymbol{\theta}$ that maximizes the profile likelihood $Q(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)})$. Since `lme` uses the divisor $\widetilde{\ddot{N}}$ instead of N the parameter σ^2 is underestimated. To correct this, we multiply the estimate of σ^2 from `lme` by $\frac{\ddot{N}(\ddot{n}+q+1)}{N}$. This can be done with the `varFixed` function, a class of variance functions with no parameters and a single covariate, being used when the within-group variance is known up to a proportionality constant. The advantage is that the correction to the estimate of σ^2 is done inside the `lme` function itself at the M-step. Using `lme` as the optimization routine at the M-step has the advantage of offering computational speed and efficiency because `lme` uses computation methods based on the QR decomposition which takes advantage of the sparse nature of the matrix $\widetilde{M}$.

Upon convergence we compute the variance of the fixed effect using Louis' formula (Orchard and Woodbury, 1972; Louis, 1982). An asymptotic approximation of the variance of the fixed effect is given (Hughes, 1999) by

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{m}\{\mathbf{X}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \mathbf{X}_i^T\boldsymbol{\Sigma}_i^{-1}\mathrm{Var}(\mathbf{y}_i|\mathbf{Q}_i,\mathbf{C}_i)\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\}\right)^{-1}. \tag{2.18}$$

Recognizing that the complete data log-likelihood in the presence of censoring is equivalent to the log-likelihood in the absence of censoring with respect to the parameter estimates helps reduce computational burdens when dealing with different variance matrices for the random effects and non-spherical errors in a more general linear mixed effects model. Modifying the `lme` function to implement the M-step provides many advantages including being able to fit wide class of LMEMs with censored response and circumvents the the task of writing special routines to optimize a series of objectives functions as proposed in Vaida and Liu (2009a).

2.2.3 Restricted Maximum Likelihood Estimation

In estimating the parameters in the variance-covariance matrix $\boldsymbol{\Sigma}$ in a linear mixed models, REML estimators are usually preferred to ML estimators (Harville, 1977; Bates and Pinheiro, 1997). REML estimators take into account the estimation of the fixed effects when estimating variance parameters typically resulting in less biased estimators than ML. Suppose the data are completely observed. One computationally convenient way of writing the restricted likelihood is

$$
\begin{aligned}
&L_R(\boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) = \int L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) d\boldsymbol{\beta} \\
&= (2\pi\sigma^2)^{-\frac{(N-p)}{2}} \exp\left(- \frac{\sum_{i=1}^m \|\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2}{2\sigma^2}\right) \Big| \sum_{i=1}^m \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \Big|^{-1/2} \prod_{i=1}^m \left(\frac{\text{abs}|\boldsymbol{\Delta}|}{\sqrt{|\widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i|}} \right)
\end{aligned}
\tag{2.19}
$$

where $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \right)^{-1} \sum_{i=1}^m \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{y}}_i$. Put $\boldsymbol{\xi} = (\boldsymbol{\theta}, \sigma^2)$ so that $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\xi})$. The log-restricted-likelihood is given by

$$
\begin{aligned}
\log L_R(\boldsymbol{\xi}|\mathbf{y}) &= -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m \left(\| \, \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \, \|^2 \right) \\
&\quad + \sum_{i=1}^m \left(\log \text{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log \mid \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \mid \right) - \frac{1}{2} \log \Big| \sum_{i=1}^m \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \Big|.
\end{aligned}
\tag{2.20}
$$

The E-step for the EM formulation, which involves computing $Q_R(\boldsymbol{\psi} \mid \boldsymbol{\xi}_R^{(k)}) =$ $E(\log L_R(\boldsymbol{\xi}|\mathbf{y})|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$, is given by

$$
Q_R(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) = -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m E\left(\| \, \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \, \|^2 | \, \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)} \right)
$$

$$+ \sum_{i=1}^{m} \left(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log \mid \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \mid \right) - \frac{1}{2} \log \Big| \sum_{i=1}^{m} \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \Big|. \quad (2.21)$$

But $E(\log L_R(\boldsymbol{\xi}|\mathbf{y})|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$ involves computing $E(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$ and $\mathrm{Var}(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$. The distribution of $(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi})$ is that of $(\mathbf{y}|\boldsymbol{\xi})$ truncated to the region $\boldsymbol{\Xi} = \{\mathbf{y} : y_{ij} \leq Q_{ij}$ for $C_{ij} = 1, y_{ij} = Q_{ij}$ for $C_{ij} = 0, i = 1, \ldots, m, j = 1 \ldots, n_i\}$:

$$p(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}) \propto p(\mathbf{Q}, \mathbf{C}|\mathbf{y}, \boldsymbol{\xi}) p(\mathbf{y}|\boldsymbol{\xi}) = p(\mathbf{y}|\boldsymbol{\xi}) I[\mathbf{y} \in \boldsymbol{\Xi}].$$

However, as noted in Vaida and Liu (2009a) $(\mathbf{y}|\boldsymbol{\xi})$ does not have a proper distribution, which makes the E-step in the REML case intractable. Instead we follow Vaida and Liu and approximate the necessary moments via $E(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) \approx E(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}^{(k)}), \mathrm{Var}(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) \approx \mathrm{Var}(\mathbf{y}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}^{(k)})$ to simplify the calculations. Thus

$$E\left(\mid\mid \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i \mid\mid^2 \mid \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)} \right) \approx \sum_{l=1}^{n_i+q+1} \| \widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l} \boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l} \widetilde{\hat{\mathbf{b}}}_{i,l} \|^2 \quad (2.22)$$

where $\mathbf{w}_{i,l}, \widetilde{\mathbf{X}}_{i,l}, \widetilde{\mathbf{Z}}_{i,l}$ and $\widetilde{\hat{\mathbf{b}}}_{i,l}$ are defined in (2.8). The analytic derivation of $\sum_{i=1}^{m} E(\|\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i\|^2 | \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)})$ is provided in the appendix. Substituting (2.22) into (2.21) yields

$$\begin{aligned} Q_R(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) &\approx -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} \sum_{l=1}^{n_i+q+1} \| \widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l} \boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l} \widetilde{\hat{\mathbf{b}}}_{i,l} \|^2 \\ &+ \sum_{i=1}^{m} \left(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log \mid \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \mid \right) - \frac{1}{2} \log \Big| \sum_{i=1}^{m} \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \Big|. \end{aligned} \quad (2.23)$$

At the M-step, the conditional estimate of $\hat{\sigma}_R^2$ is given by

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = \frac{\|\widetilde{\mathbf{w}} - \widetilde{\mathbf{M}} \hat{\boldsymbol{\delta}}_R\|^2}{N-p} = \frac{1}{N-p} \sum_{i=1}^{m} \sum_{l=1}^{n_i+q+1} \| \widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l} \boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l} \widetilde{\hat{\mathbf{b}}}_{i,l} \|^2 \quad (2.24)$$

where $\hat{\boldsymbol{\delta}}_R$ is obtained as in the ML case (2.12). Thus we obtain the profile log-restricted-likelihood as

$$Q_R(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) \approx const - \frac{(N-p)}{2} \log \left(\hat{\sigma}_R^2(\boldsymbol{\theta}) \right) + \sum_{i=1}^{m} \left(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log \mid \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \mid \right)$$

$$- \ \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|. \tag{2.25}$$

The evaluation of the restricted maximum likelihood estimates is done by first optimizing the profiled log-restricted likelihood (2.25) with respect to $\boldsymbol{\theta}$ only, and using the resulting REML estimate $\hat{\boldsymbol{\theta}}_R$ to obtain the REML estimates of $\sigma^2, \sigma_R^2(\hat{\boldsymbol{\theta}}_R)$. As indicated earlier in the ML case, with a slight modification to the `lme` function the optimization of the profile likelihood in (2.25) and the parameter estimates in restricted log-likelihood can be implemented using `lme` function. Suppose we pass `lme` the response vector $\ddot{\mathbf{w}}$ and the design matrices $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$ and indicate `method=REML`, then `lme` constructs the objective function

$$\begin{aligned}Q_{R_{lme}}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) = -\frac{\widetilde{\ddot{N}}-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\ + \sum_{i=1}^{m}(\ddot{n}+q+1)\bigg(\log\text{abs}|\mathbf{\Delta}| - \frac{1}{2}\log\mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid\bigg) - \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|.\end{aligned} \tag{2.26}$$

Note that $\sum_{i=1}^{m}\sum_{l=1}^{(n_i+q+1)}\widetilde{\mathbf{X}}_{i,l}^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_{i,l} = \sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i$ so with the objective function in (2.26), for a given $\boldsymbol{\theta}$, `lme` computes $\hat{\boldsymbol{\delta}}_R$ and $\hat{\sigma}_R^2$ given by

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{R_{lme}} &= (\widetilde{\mathbf{M}}^T\widetilde{\mathbf{M}})^{-1}\widetilde{\mathbf{M}}^T\widetilde{\mathbf{w}} \qquad (2.27)\\ \hat{\sigma}^2_{R_{lme}}(\boldsymbol{\theta}) &= \frac{\|\widetilde{\mathbf{w}} - \widetilde{\mathbf{M}}\hat{\boldsymbol{\delta}}_R\|^2}{\widetilde{\ddot{N}}-p} = \frac{1}{\widetilde{\ddot{N}}-p}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2.\end{aligned}$$

It is obvious that `lme` underestimates the parameter σ^2 in (2.27) since it uses the divisor $\widetilde{\ddot{N}}-p$ instead of $N-p$. As noted in the ML case we can correct this by multiplying the estimate of σ^2 from `lme` by $\frac{\widetilde{\ddot{N}}-p}{N-p}$. This can be done with the `varFixed` function. Substituting these conditional estimators back into (2.26) provides the profile log-restricted-likelihood

$$\begin{aligned}Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\psi}^{(k)}) = const - (\ddot{n}+q+1)\bigg[\frac{\ddot{N}-\widetilde{\ddot{p}}}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta})\big) + \\ \sum_{i=1}^{m}\big(\log\text{abs}|\mathbf{\Delta}| - \frac{1}{2}\log\mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid\big) - \frac{1}{2(\ddot{n}+q+1)}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|\bigg]\end{aligned} \tag{2.28}$$

where $\widetilde{\ddot{p}} = \frac{p}{\ddot{n}+q+1}$. But the profiled restricted-log-likelihood that we want to optimize is (2.25). Therefore we implement a modification of the `lme` software by subtracting N_r from

$\ddot{N}$ such that $\ddot{N} - N_r = N$ and multiplying the $\widetilde{\widetilde{p}}$ and the fourth component of the profile restricted-log-likelihood (2.28) by $\ddot{n} + q + 1$.

As indicated in both ML and REML estimation, the modification to `lme` function simply involves multiplying the profile likelihood function with or subtracting a constant from the profile likelihood function constructed by `lme`.

2.2.4 LIKELIHOOD FUNCTION AND CONVERGENCE CRITERIA

Following Jacqmin-Gadda and Thiebaut (2000) and Vaida and Liu (2009a), the observed data log-likelihood is given by

$$\log L_{obs}(\Psi) = \sum_{i=1}^{m} \{\log \boldsymbol{\alpha}_i + \log \boldsymbol{\phi}_{n_i^o}(\mathbf{Q}_i^o - \mathbf{X}_i^o \boldsymbol{\beta}; \boldsymbol{\Sigma}_i^{oo})\} \tag{2.29}$$

where $\boldsymbol{\alpha}_i = P(\mathbf{y}_i^c < \mathbf{Q}_i^c | \mathbf{y}_i^o) = \Phi_{n_i^c}(;)$ is the left-tail probability and $\phi(;)$ is the probability density function of the multinormal distribution. Since $\boldsymbol{\alpha}_i$ is computed at the E-step, the observed data log-likelihood is computed without significant computational burden (Vaida and Liu, 2009a) through the routine `mvtnorm()` available in R; see Genz et al. (2008) and R Development Core Team (2009).

For the purpose of comparing the speed of our implementation with that of the **lmec** package of Vaida and Liu (2009a), we used the same convergence criteria to monitor the convergence of our algorithm. Convergence of the algorithm is monitored using the objective function

$$f_o(\Psi) = -\frac{n}{2}\{1 + \log(2\pi\hat{\sigma}^2(\boldsymbol{\theta})\} + \sum_{i=1}^{m} \left(\log \text{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log |\widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i| \right) \tag{2.30}$$

which is the log-likelihood of the linear mixed effect model without censoring, with $\boldsymbol{\beta}$ profiled out (Pinheiro and Bates (2000), formula 2.13). An analogous objective function is defined for REML estimation. The algorithm runs for a number of steps (say 10 iterations) before the convergence criterion is evaluated. After 10 iterations, the difference between the log-likelihood at the current iteration and the previous iteration is computed. Convergence is declared if the difference is less than a certain preset limit. In our examples and simulation studies we used 0.001.

2.3 More General Linear Mixed Effects Models

The LMEM formulation in (2.1) allows flexibility in the specification of the random effects structure of the model but restricts the within-group error to be distributed as $N(0, \sigma^2\mathbf{I})$. However in many applications involving longitudinal and spatial data, the within-group errors are heteroscedastic and/or are correlated. In this section we provide theoretical and computational details for LMEMs with censored and missing response with non-spherical errors. Consider the extended single level linear mixed model,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, ..., m,$$
$$\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2\mathbf{D}_s), \qquad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda}_i),$$

where the $\boldsymbol{\Lambda}_i$ are positive-definite matrices parameterized by $\boldsymbol{\lambda}$. Due to positive definiteness of $\boldsymbol{\Lambda}_i$, there is an invertible square-root $\boldsymbol{\Lambda}_i^{1/2}$ with an inverse $\boldsymbol{\Lambda}_i^{-1/2}$ such that $\boldsymbol{\Lambda}_i = (\boldsymbol{\Lambda}_i^{1/2})^T\boldsymbol{\Lambda}_i^{1/2}$ and $(\boldsymbol{\Lambda}_i)^{-1} = \boldsymbol{\Lambda}_i^{-1/2}(\boldsymbol{\Lambda}_i^{-1/2})^T$. By defining

$$\begin{aligned} \mathbf{y}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^T\mathbf{y}_i, \qquad \boldsymbol{\epsilon}_i^* = (\boldsymbol{\Lambda}_i^{-1/2})^T\boldsymbol{\epsilon}_i, \\ \mathbf{X}_i^* &= (\boldsymbol{\Lambda}_i^{-1/2})^T\mathbf{X}_i, \qquad \mathbf{Z}_i^* = (\boldsymbol{\Lambda}_i^{-1/2})^T\mathbf{Z}_i, \end{aligned}$$

we can transform the model as follows

$$\mathbf{y}_i^* = \mathbf{X}_i^*\boldsymbol{\beta} + \mathbf{Z}_i^*\mathbf{b}_i + \boldsymbol{\epsilon}_i^*, \qquad i = 1, ..., m \tag{2.31}$$
$$\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2\mathbf{D}_s), \qquad \boldsymbol{\epsilon}_i^* \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Thus, given $\boldsymbol{\lambda}$, $\mathbf{y}^*$ is described by the basic linear mixed model. Since $\mathbf{y}_i^* = (\boldsymbol{\Lambda}_i^{-1/2})^T\mathbf{y}_i$ and $d\mathbf{y}_i^* = |\boldsymbol{\Lambda}_i^{-1/2}|d\mathbf{y}_i$, the log-likelihood function corresponding to the extended linear mixed effects model can be expressed as

$$\log L(\boldsymbol{\psi} \mid \mathbf{y}) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\|\widetilde{\mathbf{y}}_i^* - \widetilde{\mathbf{X}}_i^*\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i^*\hat{\mathbf{b}}_i\|^2\right)$$

$$+\sum_{i=1}^{m}\left(\log\text{abs}|\mathbf{\Delta}|-\frac{1}{2}\log|\,\widetilde{\mathbf{Z}}_i^{*T}\widetilde{\mathbf{Z}}_i^{*}\,|\right)+\sum_{i=1}^{m}\log\text{abs}|\mathbf{\Lambda}_i^{-1/2}|. \tag{2.32}$$

It follows that

$$\begin{aligned} Q^*(\boldsymbol{\psi}\mid\boldsymbol{\psi}^{(k)}) &= -\frac{N}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{m}E\left(\|\widetilde{\mathbf{y}}_i^*-\widetilde{\mathbf{X}}_i^*\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i^*\hat{\mathbf{b}}_i\|^2|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^k\right)\\ &\quad+\sum_{i=1}^{m}\left(\log\text{abs}|\mathbf{\Delta}|-\frac{1}{2}\log|\,\widetilde{\mathbf{Z}}_i^{*T}\widetilde{\mathbf{Z}}_i^{*}\,|\right)+\sum_{i=1}^{m}\log\text{abs}|\mathbf{\Lambda}_i^{-1/2}|. \end{aligned}\tag{2.33}$$

But because $\|\widetilde{\mathbf{y}}_i^*-\widetilde{\mathbf{X}}_i^*\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i^*\hat{\mathbf{b}}_i\|^2=\|\mathbf{\Lambda}_i^{-T/2}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})(\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})\|^2$ we have

$$\begin{aligned} &E\left(\|\widetilde{\mathbf{y}}_i^*-\widetilde{\mathbf{X}}_i^*\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i^*\hat{\mathbf{b}}_i\|^2|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^k\right)\\ &= E\left((\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})^T\mathbf{\Lambda}_i^{-1}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})(\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)}\right)\\ &= \text{tr}\left\{(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})^T\mathbf{\Lambda}_i^{-1}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})E\left[(\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})(\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)}\right]\right\}\\ &= \text{tr}\Bigg\{(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})^T\mathbf{\Lambda}_i^{-1}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})\Big[(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T\\ &\qquad+\ \text{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})\Big]\Bigg\}\\ &= \text{tr}\left\{(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})^T(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})^T\mathbf{\Lambda}_i^{-1}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta})\right\}\\ &\qquad+\ \text{tr}\left\{(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})^T\mathbf{\Lambda}_i^{-1}(\mathbf{I}-\mathbf{P}_{\widetilde{\mathbf{Z}}_i})\text{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})\right\}\\ &= \left\|\mathbf{\Lambda}_i^{-T/2}(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_i)\right\|^2+\left\|\begin{matrix}\mathbf{\Lambda}_i^{-T/2}(\widetilde{\mathbf{S}}_{i,1}-\mathbf{0}\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,1})\\ \vdots\\ \mathbf{\Lambda}_i^{-T/2}(\widetilde{\mathbf{S}}_{i,n_i+q}-\mathbf{0}\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,n_i+q})\end{matrix}\right\|^2\\ &= \left\|\begin{matrix}\mathbf{\Lambda}_i^{-T/2}(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i,\mathbf{C}_i,\boldsymbol{\psi}^{(k)})-\widetilde{\mathbf{X}}_i\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_i)\\ \mathbf{\Lambda}_i^{-T/2}(\widetilde{\mathbf{S}}_{i,1}-\mathbf{0}\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,1})\\ \vdots\\ \mathbf{\Lambda}_i^{-T/2}(\widetilde{\mathbf{S}}_{i,n_i+q}-\mathbf{0}\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,n_i+q})\end{matrix}\right\|^2\\ &= \sum_{l=1}^{n_i+q+1}\|\mathbf{\Lambda}_i^{-T/2}(\widetilde{\mathbf{w}}_{i,l}-\widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l})\|^2=\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}_{i,l}^*-\widetilde{\mathbf{X}}_{i,l}^*\boldsymbol{\beta}-\widetilde{\mathbf{Z}}_{i,l}^*\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2. \end{aligned}\tag{2.34}$$

Substituting (2.34) into (2.33) yields

$$\begin{aligned} Q^*(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}^*_{i,l} - \widetilde{\mathbf{X}}^*_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}^*_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\ &\quad + \frac{1}{2}\sum_{i=1}^{m}\Bigg(\log\text{abs}|\boldsymbol{\Delta}| - \log|\, \widetilde{\mathbf{Z}}_i^{*T}\widetilde{\mathbf{Z}}_i^* \,|\Bigg) + \sum_{i=1}^{m}\log\text{abs}|\boldsymbol{\Lambda}_i^{-1/2}|. \quad (2.35) \end{aligned}$$

Then, for a given $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ the M-step updates for $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ and $\hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\lambda})$ are:

$$\hat{\boldsymbol{\delta}} = (\widetilde{\mathbf{M}}^{*T}\widetilde{\mathbf{M}}^*)^{-1}\widetilde{\mathbf{M}}^{*T}\widetilde{\mathbf{w}}^*. \quad (2.36)$$

$$\hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{\|\widetilde{\mathbf{w}}^* - \widetilde{\mathbf{M}}^*\hat{\boldsymbol{\delta}}\|^2}{N} = \frac{1}{N}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1}\|\widetilde{\mathbf{w}}^*_{i,l} - \widetilde{\mathbf{X}}^*_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}^*_{i}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2.$$

Substituting these conditional estimators back into $Q^*(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)})$ provides the profile log-likelihood

$$\begin{aligned} Q^*((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) &= \mathit{const} - \frac{N}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\lambda})\big) + \sum_{i=1}^{m}\big(\log\text{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\, \widetilde{\mathbf{Z}}_i^{*T}\widetilde{\mathbf{Z}}_i^* \,|\big) \\ &+ \sum_{i=1}^{m}\log\text{abs}|\boldsymbol{\Lambda}_i^{-1/2}|. \quad (2.37) \end{aligned}$$

As indicated earlier in the case of spherical error, with a slight modification to the `lme` function we can use it to estimate the parameters in the objective function (2.37). Suppose we pass `lme` the response vector $\ddot{\mathbf{w}}$, design matrices $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$, and specify the variance and/or correlation structure we want, then `lme` constructs the objective function

$$\begin{aligned} Q^*_{lme}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{\widetilde{\ddot{N}}}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{\ddot{n}+q+1}\|\widetilde{\ddot{\mathbf{w}}}^*_{i,l} - \widetilde{\ddot{\mathbf{X}}}^*_{i,l}\boldsymbol{\beta} - \widetilde{\ddot{\mathbf{Z}}}^*_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\ &+ \frac{1}{2}\sum_{i=1}^{m}\sum_{l=1}^{\ddot{n}+q+1}\Bigg(\log\text{abs}|\boldsymbol{\Delta}| - \log|\, \widetilde{\ddot{\mathbf{Z}}}^{*T}_{i,l}\widetilde{\ddot{\mathbf{Z}}}^*_{i,l} \,|\Bigg) + \sum_{i=1}^{m}\sum_{l=1}^{\ddot{n}+q+1}\log\text{abs}|\ddot{\boldsymbol{\Lambda}}_{i,l}^{-1/2}|. \quad (2.38) \end{aligned}$$

Note that for each i, $\widetilde{\ddot{\mathbf{Z}}}^*_{i,l} = \widetilde{\ddot{\mathbf{Z}}}^*_{i}$, and $\ddot{\boldsymbol{\Lambda}}_{i,l} = \ddot{\boldsymbol{\Lambda}}_i, l = 1, \ldots, \ddot{n} + q + 1$ so the objective function constructed by `lme` can be written as

$$
\begin{aligned}
Q^*_{lme}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{\widetilde{\ddot{N}}}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}^*_{i,l} - \widetilde{\mathbf{X}}^*_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}^*_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2 \\
&+ \frac{1}{2}\sum_{i=1}^{m}(\ddot{n}+q+1)\Big(\log\text{abs}|\boldsymbol{\Delta}| - \log\mid \widetilde{\mathbf{Z}}^{*T}_i\widetilde{\mathbf{Z}}^*_i \mid \Big) + \sum_{i=1}^{m}(\ddot{n}+q+1)\log\text{abs}|\ddot{\boldsymbol{\Lambda}}_i^{-1/2}|. \quad (2.39)
\end{aligned}
$$

With this objective function, for a given $(\boldsymbol{\theta}, \boldsymbol{\lambda})$, `lme` computes the estimates of δ and σ^2 given by:

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}_{lme} &= (\widetilde{\mathbf{M}}^{*T}\widetilde{\mathbf{M}}^*)^{-1}\widetilde{\mathbf{M}}^T\widetilde{\mathbf{w}}^* \qquad (2.40)\\
\hat{\sigma}^2_{lme}(\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \frac{\|\widetilde{\mathbf{w}}^* - \widetilde{\mathbf{M}}^*\hat{\boldsymbol{\delta}}\|^2}{\widetilde{\ddot{N}}} = \frac{1}{\widetilde{\ddot{N}}}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}^*_{i,l} - \widetilde{\mathbf{X}}^*_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}^*_i\widetilde{\hat{\mathbf{b}}}_{i,l}\|^2
\end{aligned}
$$

Substituting these conditional estimators back into $Q^*_{lme}(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)})$ provides the profile log-likelihood

$$
\begin{aligned}
Q^*_{lme}((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) = const + (\ddot{n}+q+1)\Bigg[-\frac{\ddot{N}}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\lambda})\big) + \\
\sum_{i=1}^{m}\big(\log\text{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log\mid \widetilde{\mathbf{Z}}^{*T}_i\widetilde{\mathbf{Z}}^*_i \mid\big) + \sum_{i=1}^{m}\log\text{abs}|\ddot{\boldsymbol{\Lambda}}_i^{-1/2}|\Bigg]. \quad (2.41)
\end{aligned}
$$

Note in the case original data is balanced, $\ddot{N} = N$ and $\ddot{\boldsymbol{\Lambda}}_i = \boldsymbol{\Lambda}_i$ so equation (2.41) is equivalent to (2.37), the profile likelihood that we want to optimize. Since the scaled variance-covariance matrix $\boldsymbol{\Lambda}_i$ depends on i through its dimension, by augmenting the response and the design matrices of the cluster with cluster size less than the $\ddot{n}$ with $\ddot{n} - n_i$ extra rows in the case of unbalanced data, we also change the dimension of $\boldsymbol{\Lambda}_i(n_i \times n_i)$ to dimension of $\ddot{\boldsymbol{\Lambda}}_i(\ddot{n} \times \ddot{n})$. So depending on the structure of $\boldsymbol{\Lambda}_i$, $|\ddot{\boldsymbol{\Lambda}}_i|$ may not be equal to $|\boldsymbol{\Lambda}_i|$. Thus with the pseudo data augmentation, we are inducing a correlation between the original observations and the pseudo observation which is not desired since it will result in under-estimation of the correlation parameter. We can correct this by setting $\ddot{\boldsymbol{\Lambda}}_i = \begin{pmatrix} \boldsymbol{\Lambda}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\ddot{n}-n_i} \end{pmatrix}$ which yields

$|\ddot{\boldsymbol{\Lambda}}_i| = \left|\begin{pmatrix} \boldsymbol{\Lambda}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\ddot{n}-n_i} \end{pmatrix}\right| = |\boldsymbol{\Lambda}_i|$.

Since the correlation between two within-group errors, $\epsilon_{ij}, \epsilon_{ij'}$ is assumed to depend on the corresponding position vectors $\mathbf{p}_{ij}, \mathbf{p}_{ij'}$ only through their absolute difference, to make sure the correlation between the error corresponding to the original and the pseudo observation is zero, we set the position vector corresponding to the pseudo observation to be large such that for $i = 1, \dots, m$, and $j = 1, \dots, n_i, j' = n_i + 1, \dots, \ddot{n}$, $\boldsymbol{\rho}^{|\mathbf{p}_{ij} - \mathbf{p}_{ij'}|} \approx 0$ where $\boldsymbol{\rho}$ is the vector of correlation parameters. For all the correlation structures implemented in **nlme**, except for the compound symmetry structure (`corCompSymm`) and general correlation structure (`corSymm`), we can obtain the structure of $\ddot{\boldsymbol{\Lambda}}_i$ from $\boldsymbol{\Lambda}_i$. Thus

$$\begin{aligned} Q^*_{lme}((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) &= (\ddot{n} + q + 1) Q^*((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) \\ \Longrightarrow \quad Q^*((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) &= \frac{1}{\ddot{n} + q + 1} Q^*_{lme}((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}). \end{aligned} \tag{2.42}$$

Thus

$$\arg\max_{(\boldsymbol{\theta}, \boldsymbol{\lambda})} Q^*((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) = \arg\max_{(\boldsymbol{\theta}, \boldsymbol{\lambda})} \left\{ \frac{1}{\ddot{n} + q + 1} Q^*_{lme}((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)}) \right\} \tag{2.43}$$

indicating that, the optimization of the profile likelihood $Q^*((\boldsymbol{\theta}, \boldsymbol{\lambda}) \mid \boldsymbol{\psi}^{(k)})$ can be accomplished by the `lme` function. Analogous to the spherical error case, the underestimation of the parameter σ^2 in (2.40) can be corrected by multiplying the estimate of σ^2 from `lme` by $\frac{\ddot{N}(\ddot{n}+q+1)}{N}$ which is implemented by the `varFixed` function. In both spherical and non-spherical error cases, when the data are balanced, no modification to the `lme` function is needed to estimate the parameters in the model except in the case of REML estimation.

2.3.1 Multilevel Models

The methodology described in section (2.2) can easily be extended to multilevel models. Bates and Pinheiro (1997) provide a representation that reduces a multilevel model to a single level model in the absence of censoring. Vaida and Liu (2009a) extends this representation to the case where some elements of the response are censored. In the notation of Pinheiro and Bates

(2000) and Vaida and Liu (2009a), consider the nested model

$$\begin{aligned} \mathbf{y}_{ij} &= \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad i = 1, ..., m, \quad j = 1, ..., m_i, \qquad (2.44) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \sigma^2\mathbf{D}_{s1}), \quad \mathbf{b}_{ij} \sim N(\mathbf{0}, \sigma^2\mathbf{D}_{s2}), \quad \boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \end{aligned}$$

where $\mathbf{b}_i, \mathbf{b}_{ij}, \boldsymbol{\epsilon}_{ij}$ are all independent for different i, j. Let $\mathbf{Q}$ be the vector of observed values $\mathbf{y}$ or censoring limits, and $\mathbf{C}$ the censoring indicator. We can define $\mathbf{b}_i^* = \text{vec}(\mathbf{b}_i, \mathbf{b}_{i1} \ldots, \mathbf{b}_{im_i})$ for each i; $\mathbf{X}_i = \text{vec}(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})$; $\mathbf{Z}_i = (\text{vec}(\mathbf{Z}_{i,1}, \ldots, \mathbf{Z}_{i,m_i}), \text{diag}(\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{im_i}))$; $\mathbf{y}_i = \text{vec}(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{im_i})$; $\boldsymbol{\epsilon}_i = \text{vec}(\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{im_i})$. Model (2.44) then can be written as

$$\begin{gathered} \mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i^* + \boldsymbol{\epsilon}_i, \qquad i = 1, ..., m, \qquad (2.45) \\ \mathbf{b}_i^* \sim N(\mathbf{0}, \sigma^2\mathbf{D}_s^*), \qquad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \end{gathered}$$

which is identical to the the Laird-Ware model (2.1) except the matrix $\mathbf{D}_s^*$ is structured as a block diagonal matrix with blocks $\mathbf{D}_{s1}$ and $\text{diag}_{m_i}(\mathbf{D}_{s2})$. That is, $\mathbf{D}_s^* = \text{diag}(\mathbf{D}_{s1}, \text{diag}_{mi}(\mathbf{D}_{s2}))$. In the presence of censoring, the derivation and formulas for $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ in section (2.2) all apply.

2.4 Example

The example with which we will illustrate our model fitting methodology is the real time data set mentioned in the introduction. The original data set contains N = 43,700 observations with the response variables CO and $PM_{2.5}$ measured repeatedly on 15 fire-fighters over 10 days at different times of the day. In this data set task is a covariate that indicates which of several activities the fire fighters were performing when the measurements were taken. The possible job tasks include holding (H), lighting (L), mop-up (M), and other (O) activities that do not belong to these major groupings. Briefly, holding involves the maintenance of fire within boundary lines, mop-up entails the extinguishing of smoldering fire after the major burning phase, and lighting is the fire igniting process (for controlled burns).

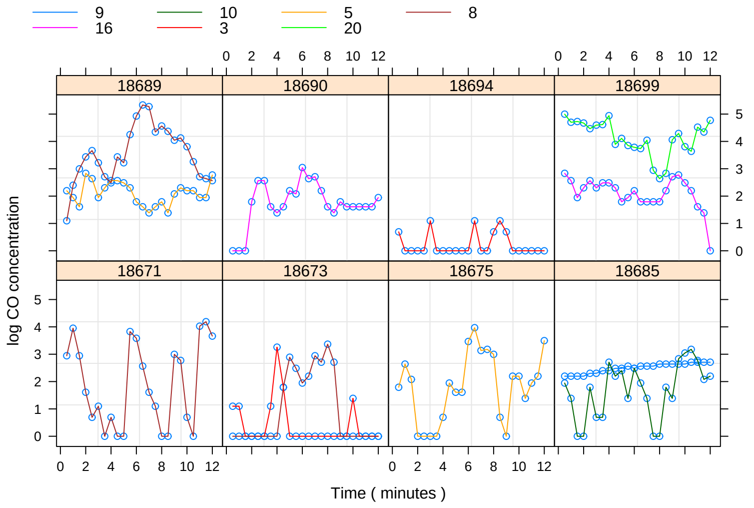

Figure 2.1: Trajectory of the log CO concentration of the 7 fire-fighters performing task H over 8 days. The lower limit of detection of the log CO concentration is 0.

Since not all the subjects work each of the 10 days, the data are unbalanced. Overall 40.57% of the $PM_{2.5}$ measurements are missing and 0.10% are censored. The CO responses also have missing and censored values of 0.30% and 52.81% respectively. Some components of the covariate, task, are also missing. One objective of the investigators is to determine the mean response for the four different tasks.

The extreme sparsity of the original data set, coupled with its size, makes it unsuitable for model fitting. Therefore, to ensure sufficient available information, we use only a sub-sample of the original data. As depicted in Figure 2.1, the sub-sample is selected as a specific time window during the day for all days that each fire-fighter worked. Here, we analyze the data only for the univariate response CO for those fire-fighters who were performing task H during the selected time-frame. Thus the sub-sample includes data from 7 fire-fighters who were working on different combinations of 8 different days. Out of the 288 observation in our

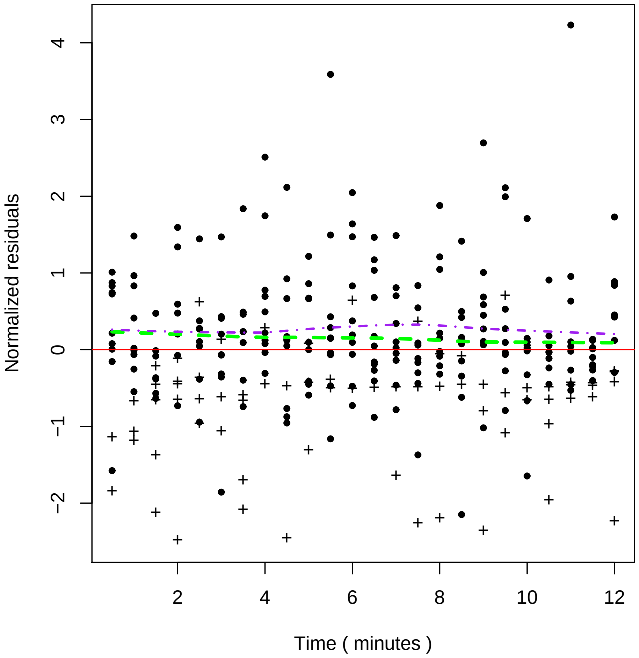

Figure 2.2: Lowess smoother fit to the normalized residuals from (i) homogeneous model (-.-); (iii) ARMA(1,1) model (- -). The residuals from model (iii) appear as points; the left-censored residuals appear as ”+”.

sub-sample, 70 (24.3%) observations are below the detection limit of 1. The response CO is strongly skewed to the right so it was log-transformed, which also leads to a log-transformed detection limit of $\log(1) = 0$.

We consider three models for these data. We started by fitting a nested random effect model (day nested within subject) with homogeneous within-group errors. Let $y_{ijk\ell}$ be the log CO concentration response from the ith fire fighter at the kth time point at the jth day for the ℓth task. The first model can be expressed as

$$y_{ijk\ell} \quad = \quad \beta_\ell + b_i + d_{j(i)} + \epsilon_{ijk\ell}, \tag{2.46}$$

$$b_i \sim N(0,\sigma_b^2),\ d_{j(i)} \sim N(0,\sigma_d^2),\ \epsilon_{ijk\ell} \sim N(0,\sigma^2).$$

The plot of the residuals against time shows a relatively good fit (Figure 2.2) but the lowess smoother fit to the residuals shows there is slight oscillating pattern over time that is not captured by the model. This is confirmed by the sample variogram (computed from the observed squared differences between pairs of residuals, see Diggle et al. 2002, page 49) plotted in Figure 2.3, which suggests the within group correlation is not properly accounted for by the spherical error model. This could either be attributed to the bias in the modelling of the mean or a true serial autocorrelation beyond the random component. The variogram rises steadily with increasing time up to 4 minutes, and begins to decrease to 8 minutes before stabilizing around 0.25.

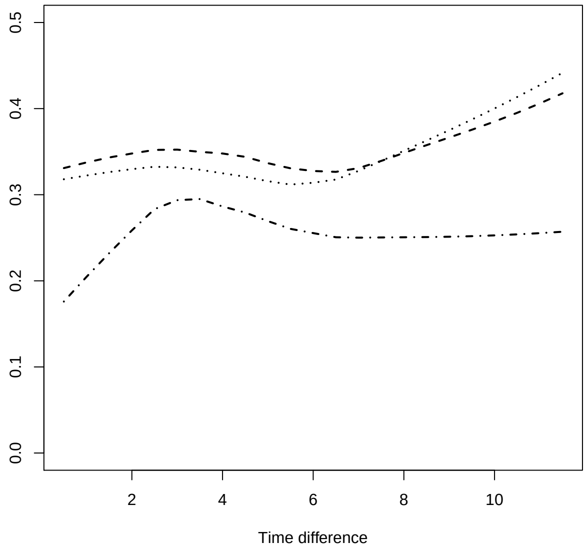

Figure 2.3: Sample variogram estimates corresponding to the normalized residuals from (i) homogeneous model (-.-);(ii) AR(1) model (..); (iii) ARMA(1,1) model(- -).

To address the concern of unaccounted serial correlation depicted in the variogram plot, we consider a heteroscedastic version of model (2.46) with AR(1) within-group serial correlation:

$$\begin{aligned} y_{ijk\ell} &= \beta_\ell + b_i + d_{j(i)} + \epsilon_{ijk\ell}, \qquad (2.47) \\ b_i &\sim N(0,\sigma_b^2),\ \ d_{j(i)} \sim N(0,\sigma_d^2),\ \ \epsilon_{ijk\ell} \sim N(0,\sigma^2\Lambda_{ijk\ell}) \end{aligned}$$

The variogram plot (Figure 2.3) indicates that the model with AR(1) error fits better than the spherical error model; however after 8 minutes, the variogram of the normalized residuals seems to be increasing slightly with time. This suggests we consider a third model. We refit model (2.47) but with ARMA(1,1) within-group error correlation. The model fits slightly better than the AR(1) model according to a likelihood ratio test (LRT). The LRT statistic, 2(-337.033 + 345.42) = 8.39 with p-value 0.003 supporting the ARMA(1,1) model.

Table 2.1: Parameter estimates for the three models fit to the log CO concentrations.

	Homogeneous model		AR(1) model		ARMA(1,1) model	
	Estimates	SE	Estimates	SE	Estimates	SE
β_H	1.75	0.38	1.73	0.39	1.73	0.39
σ_b	0.60		0.48		0.54	
σ_d	1.16		1.09		1.13	
σ	1.18		1.31		1.26	
ϕ			0.74		0.53	
θ					0.44	

Parameter estimates are in Table 2.1. Based on the fitted model we estimate the average log CO concentration to be 1.73 for task H between the time of the day 2:23:03 PM to 2:35:03 PM, with 95% CI (0.97, 2.49). As expected, we find substantially different variance components at the subject and day level across the responses CO, and that variance from subject to subject and day to day are 0.54 and 1.13 respectively for task H. The within-subject standard deviation is 1.26 and the estimates of the ARMA(1,1) correlation parameters are $\hat{\phi} = 0.53$, $\hat{\theta} = 0.44$.

2.5 Simulation Study

We conducted a simulation study to evaluate our proposed methodology. The simulation study is divided into three parts. Part I focus on the case of spherical errors (which is the case implemented in Vaida and Liu's **lmec** package). For this part, we demonstrate that our proposed method can fit the same models as in Vaida and Liu's **lmec** package but with improved computational time. Part II of the simulation study is to demonstrate that our proposed approach can compute ML and REML for a wider class of linear mixed effects models with censored response (than the models implemented in the **lmec** package of Vaida and Liu), including models with non-spherical error (heteroscedasticity and/or serial correlation). As mentioned earlier in the introduction, censored and missing data may arise simultaneously. Part III of the simulation study is to demonstrate that our proposed approach can fit all class of LMEMs whiles accounting for censoring and missingness simultaneously.

To separate the effects of censoring from small sample inaccuracies of ML and REML in linear mixed models, we also include results from the complete data LME in all three cases. Comparison of the three methods (complete data LME, our proposed closed-form EM with pseudo data augmentation and closed-form EM by Vaida and Liu) is based on average estimates of the parameters, simulation-based mean square error (MSE), relative bias $\hat{E}(\hat{\theta} - \theta)/|\theta|$, coverage of 95% confidence intervals $\hat{Pr}(\theta \in \hat{I})$ and average computational time. We will refer to the three algorithms compared in this section (2.5) by the name of the R function that was used to implement them: `lme, lmecm, lmec` respectively.

2.5.1 SPHERICAL ERROR CASE

We consider the LMEM (2.48) with random intercept b_{0i} and random slope b_{1i} which is studied in Vaida and Liu's (2009) paper:

$$\begin{aligned} y_{ij} &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, I, \quad j = 1, \ldots n_i \qquad (2.48) \\ & (b_{0i}, b_{1i}) \sim^{iid} \mathrm{N}(0, D), \quad \epsilon_{ij} \sim^{iid} \mathrm{N}(0, \sigma^2). \end{aligned}$$

Here, y_{ij} is the response from the ith subject at time t_{ij}. We simulated 1000 data sets, each with 60 subjects (I = 60) and 5 observations ($n_i = 5$) per subject measured at time points 1, 2, 3, 4, 5. This gives a total of $N = 60 \times 5 = 300$ observations per data set. The data were generated from model (2.48) with the following parameters, $\beta_0 = 5$, $\beta_1 = 2$, $D = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} = \begin{pmatrix} 1.7 & 0.5 \\ 0.5 & 2.3 \end{pmatrix}$, and $\sigma = 2.3$. In addition, 20% of the observations were censored. Both ML and REML estimation were used but we only present the results for ML estimation. REML results were similar and are included in the appendix. Comparisons with the substitution methods are also included in the appendix.

Table 2.2: Spherical error case. Estimated parameters from simulation based on 1000 data sets (mean, MSE) and elapsed CPU time in seconds (mean). Three algorithm are compared: `lme`; `lmecm`; and `lmec`. Maximum likelihood estimation is used throughout.

		`lme`		`lmecm`		`lmec`	
	True Value	Mean	MSE	Mean	MSE	Mean	MSE
Time		0.10		8.51		11.14	
β_0	5.0	5.00	0.12	4.98	0.15	4.99	0.15
β_1	2.0	2.01	0.04	2.01	0.05	2.00	0.05
D_{11}	1.7	1.75	1.68	2.19	2.67	2.22	2.05
D_{12}	0.5	0.46	0.36	0.48	0.45	0.34	0.46
D_{22}	2.3	2.29	0.27	2.26	0.30	2.34	0.32
σ	2.3	2.29	0.01	2.30	0.02	2.27	0.02

Table 2.3: Spherical error case: Coverage probability(%) and relative bias (%) based on a simulation study of 1000 data sets.

		`lme`	`lmecm`	`lmec`
Coverage probability	β_0	95.2	95.6	95.4
	β_1	96.3	95.9	96.1
Relative bias	β_0	0.0	-0.4	-0.1
	β_1	0.3	0.4	0.2
	D_{11}	2.9	28.5	30.5
	D_{12}	-8.3	-3.6	-31.8
	D_{22}	-0.3	-1.9	1.8
	σ	-0.6	-0.2	-1.4

Table 2.2 presents the average estimates for parameters in model (2.48) and their simulation-based mean squared error (MSE). Table 2.3 contain the coverage probability of 95% Wald confidence intervals for the fixed effects and the relative bias for all the estimated parameters. The three algorithms: `lme`, `lmecm`, and `lmec` performed similarly. The average estimates of the parameters were very close to each other and close to the true value except the variance components of the random effect (D_{11},D_{12}). Compared to `lme`, both `lmecm` and `lmec` overestimated D_{11} with bias of 28.5 for `lmecm`, 30.5 for `lmec`. `lmecm` performed better in estimating D_{12} compared to `lme` and `lmec` which underestimated D_{12} with bias of -8.3 and -31.8 respectively. The MSE is similar for the censored data algorithms (`lmecm` and `lmec`); however, as expected, both had larger MSE compared to `lme`. All three methods perform very well in terms of coverage probability. The coverage probabilities of β_0 were 95.2% for `lme`, 95.6% for `lmecm`, and 95.4% for `lmec`; the coverage probabilities of β_1 were 96.3% for `lme`, 95.9% for `lmecm`, and 96.1% for `lmec`. The estimated relative information loss due to censoring for all six parameters, $1 - \text{var}_{\texttt{lme}}(\hat{\boldsymbol{\psi}})/\text{var}_{\texttt{lmecm}}(\hat{\boldsymbol{\psi}})$ ranged from 5.5% for β_1 to 30.9% for D_{11}.

`lmecm` slightly improved computational time compared to `lmec` even though this was not expected. This improvement can be attributed to the difference in optimization routines used at the M-step of the EM algorithm. On average, it took 11.14 sec for `lmec` to converge, however, it took 8.51 sec for `lmecm` to converge, about 1.3 times faster than `lmec`.

2.5.2 Non-spherical Error Case

The second part of the simulation study is to demonstrate that our proposed methodology can fit a class of linear mixed model with non-spherical error. We consider a more general LMEM

$$\begin{aligned} y_{ijk} &= \mu_1 + \delta_1 \mathrm{I}(j=2) + (\beta_1 + \delta_2 \mathrm{I}(j=2)) t_{ijk} + b_{ij} + \epsilon_{ijk}, \qquad (2.49) \\ & i = 1, \ldots, I, \;\; j = 1, 2, \;\; k = 1, \ldots, n_{ij}, \\ & b_{ij} \sim^{iid} \mathrm{N}(0, \sigma_b^2), \quad \epsilon_{ijk} \sim^{iid} \mathrm{N}(0, \sigma_j^2 \Lambda_{ijk}) \end{aligned}$$

where y_{ijk} is the response from the ith subject in the jth treatment ($j = 2$ correspond to treatment 2) at time t_{ijk} and the $\mathrm{var}(\epsilon_{ijk}) = \sigma_j^2 \Lambda_{ijk}$ is heteroscedastic AR(1). Thus the model involves different variance parameters for the two groups $\sigma_j^2, j = 1, 2$, and we assume the $(h, l)^{\mathrm{th}}$ element of Λ_{ijk} is given by $\rho_{ijhl} = \rho^{d(t_{ijh}, t_{ijl})}$, with AR(1) correlation parameter ρ. According to model (2.49), the fixed effect parameters to estimate are $(\mu_1, \delta_1, \beta_1, \delta_2)$ where $\mu_2 = \mu_1 + \delta_1, \beta_2 = \beta_1 + \delta_2$. The data are generated based on model (2.49) but under the null hypothesis $H_0 : \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = \begin{pmatrix} \mu_2 - \mu_1 \\ \beta_2 - \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. The simulation is based on 1000 data sets and for each data set, 20% of the observation in each treatment group were censored. The data were generated with the following values: $(\mu_1, \delta_1, \beta_1, \delta_2, \sigma_b, \sigma_1, \sigma_2, \rho) =$ (6.4, 0, 2.0, 0, 2.0, 2.3, 5.2, 0.6).

Table 2.4 presents the average estimates of the parameters in model (2.49), and their simulation-based MSE. Table 2.5 contains the estimated coverage probability of 95% Wald

Table 2.4: Non-spherical case. Estimated parameters from simulation based on 1000 datasets (mean, MSE), rejection rate corresponding to the null hypothesis and elapsed CPU time in seconds (mean). Three algorithm are compared: `lme`; `lmecm`; and `rank`. Maximum likelihood estimation is used throughout.

		`lme`		`lmecm`		`rank`
	True Value	Mean	MSE	Mean	MSE	Mean
Time		0.14		15.61		
$\mu 1$	6.4	6.39	0.44	6.39	0.53	23.30
δ_1	0.0	-0.04	2.13	0.01	2.41	6.49
β_1	2.6	2.60	0.02	2.60	0.02	37.07
δ_2	0.0	0.01	0.10	0.00	0.11	-3.00
σ_b	2.0	1.86	0.38	1.85	0.44	18.47
σ_1	2.3	2.27	0.06	2.28	0.07	34.57
σ_2	5.2	5.12	0.24	5.10	0.29	69.00
ρ	0.6	0.59	0.01	0.58	0.01	0.54

confidence intervals for the fixed effects and the relative bias for the estimated parameters. The three algorithms compared in this section are `lme`, `lmecm`, and the nonparametric approach of using the rank of the response (`rank`). Nonparametric methods are commonly used to analysis data subject to detection limit and are especially useful for censored data because the method does not require values to be imputed since censored data retain some rank information (left-censored data are all tied for smallest). The comparison of parametric methods (`lme` and `lmecm`) with nonparametric method is strictly based on the inference on δ_1 and δ_2 since the parameters estimates are not comparable.

The two parametric algorithms perform similarly. The average estimates of the parameters were very close to each other and both are very close to the true value. Both `lme` and `lmecm` underestimated σ_b with bias -6.79% and -7.50% respectively. The estimated relative loss of information due to censoring ranged from 10% to 20%. The coverage probability for

Table 2.5: Non-spherical case. Coverage probability(%) and relative bias (%) based on a simulation study of 1000 data sets.

	parameters	`lme`	`lmecm`	`rank`
Coverage probability	μ_1	95.2	95.2	36.1
	δ_1	95.2	95.3	80.1
	β_1	93.7	94.3	00.0
	δ_2	93.3	94.2	82.6
Relative bias	μ_1	0.21	-0.13	
	$\delta_1{}^\star$	-0.04	0.01	
	β_1	0.03	0.01	
	$\delta_2{}^\star$	0.01	0.00	
	σ_b	-6.79	-7.50	
	σ_1	-1.26	-1.07	
	σ_2	-1.49	-1.93	
	ρ	-2.50	-3.48	

*Estimated actual bias is reported.

the fixed effects $(\mu_1, \delta_1, \beta_1, \delta_1)$ for `lmecm` ranged from 94.2% - 95.2% and is similar to the coverage probability for `lme` which range from 93.3%-95.2%

In order to answer the question of which of the three methods get the inference correct, we compute the rejection rate for testing the null hypothesis H_0. In Table 2.5, with the rejection rate (1 - 0.952 = 0.048, 1 - 0.933 = 0.067) for `lme` and (1 - 0.953 = 0.047, 1 - 0.942 = 0.058) for `lmecm`, the methods performed similarly and are very close to the nominal value of 0.05. For the nonparametric methods, the rejection rate for δ_1 and δ_2 are 1 - 0.801 = 0.199 and 1 - 0.826 = 0.174 respectively. Thus the rejection rate for δ_1 and δ_2 is about 4 times lager than the nominal value. Thus using nonparametric approach based on the rank to account censoring whiles testing the hypothesis of whether there is a difference in the two groups have resulted in erroneous conclusion. We are 4 times more likely to erroneously reject H_0 when it is true. As expected the simulation based MSE for the censored data algorithm, `lmecm` is larger than `lme`.

2.5.3 MISSINGNESS AND CENSORING SIMULTANEOUSLY

The third part of the simulation study is to demonstrate that our proposed method can fit the full class of linear mixed effects model whiles simultaneously accounting for censoring and missingness in the response. We consider here a more general LMEM

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 x_{ij} + b_{0i} + \epsilon_{ij}, \quad i = 1, \ldots, 50, \quad j = 1, \ldots 6 \qquad (2.50) \\ & b_{0i} \sim^{iid} \mathrm{N}(0, \sigma_b^2), \quad \epsilon_{ij} \sim^{iid} \mathrm{N}(0, \sigma^2 |x_{ij}|^{2\delta}) \end{aligned}$$

where y_{ij} and x_{ij} is the response and the covariate respectively from the ith subject at time j. Here, the variance increases with the absolute value of the covariate. The simulation is based on 1000 data sets and for each data set, 22% of the observation are unobserved (missing and censored).

Table 2.6: Missingness and censoring case. Estimated parameters from simulation based on 1000 datasets (mean, MSE), elapsed CPU time in seconds (mean). Two algorithm are compared: `lme`; and `lmecm`. Maximum likelihood estimation is used throughout.

		`lme`		`lmecm`	
	True Value	Mean	MSE	Mean	MSE
Time		0.08		11.28	
β_0	3.2	3.19	0.09	3.20	0.10
β_1	1.5	1.51	0.04	1.51	0.05
σ_b	1.2	1.16	0.05	1.16	0.07
σ	2.3	2.28	0.02	2.28	0.03
δ	0.7	0.70	0.01	0.70	0.01

Out of the 22% unobserved values, 12% is missing completely at random and 10% is left-censored. The data were generated with the following values; $(\beta_0, \beta_1, \sigma_b, \sigma, \delta)$ = (3.2, 1.5, 1.2, 2.3, 0.7). Table 2.6 presents the average estimates of the parameters in model (2.50), and their simulation-based MSE. Table 2.7 contains the estimated coverage probability of

Table 2.7: Missingness and censoring case. Coverage probability(%) and relative bias (%) based on a simulation study of 1000 data sets.

	parameters	`lme`	`lmecm`
Coverage probability	β_0	95.0	95.3
	β_1	94.8	95.1
Relative bias	β_0	-0.34	-0.13
	β_1	0.99	0.46
	σ_b	-3.56	-3.61
	σ	-0.66	-0.77
	δ	0.89	0.70

95% Wald confidence intervals for the fixed effects and the relative bias for all the estimated parameters.

Our proposed method, `lmecm` still performed very similar to the complete data `lme`; the average parameter estimates are very close and both are very close to the true value. Compared to `lme`, the biases of `lmecm` in estimating the parameters were very similar and both estimated the parameters in the model with maximum bias of approximately 4%. As expected the MSEs of `lmecm` for $(\beta_0, \beta_1, \sigma_b, \sigma, \delta)$ were larger than MSEs using the complete data and `lme`. The relative loss of information due to censoring and missingness ranges from 15%-18% for the fixed effect and 23% for σ_b, 28% for σ and 31% for δ. The coverage probability for the fixed effects β_i (i = 0, 1) using `lme` and `lmecm` were similar and both were very close to the target 95%. The range of the coverage probability was 94.8%-95.0% for `lme` and 95.1%-95.3% for `lmecm`.

2.6 Conclusion

This article serves as an extension of Vaida and Liu (2009a) by providing a feasible implementation of censored LMEMs with non-spherical errors. We have developed an EM algorithm for computing ML and REML for LMEM with censored and missing response. We have shown that with pseudo data augmentation, the E-step objective function can be optimized at the M-step with an already existing function, `lme`, with a slight modification. The use of `lme` at the M-step in our implementation makes it feasible to fit several extensions of the Laid-Ware model including models allowing heteroscedasticity and/or serial correlation, multilevel models, and random effects with complex structure. Moreover, these extensions can be handled with improve computational speed without having to write special routines to optimize the series of objective functions proposed by Vaida and Liu for the case of non-spherical errors. The use of `lme` at the M-step make it feasible to implement all the correlation, variances and random effects structures in the **nlme** package.

The proposed methods can be extended to multivariate mixed-effects model with censored responses. Multivariate longitudinal data (MLD) arise often in practice and studying the joint evolution of multiple response variables over time is of interest. However, the analysis of MLD can be challenging compared to the univariate case in several respects. For instance, the variance of the errors can be different for the different responses, the errors are likely to be correlated for a particular response across different time points, and the errors are likely to be correlated across the multiple response variables at the same time point (contemporaneous correlation) (Gao et al., 2009). The analysis of MLD becomes even more complicated when the multiple response variables are subject to DLs. This issue is currently under investigation and we hope to report these findings in a future paper.

2.7 References

Bandyopadhyay, D., V. Lachos, M. Castro, and D. Dey (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal 54*, 405–425.

Bates, D. M. and J. Pinheiro (1997). Computational methods for multilevel modelling. Technical report, Bell Labs, Murray Hill, NJ.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. Zeger (2002). *Analysis of longitudinal data.* Oxford, England: Oxford University Press.

Finney, D. (1962). Cumulants of truncated multinormal distribution. *Journal of the Royal Statistical Society Series B 24*, 535–536.

Gao, F., P. Thompson, C. Xiong, and J. Miller (2009). Analyzing multivariate longitudinal data using SAS. *Proceedings of the Thirty-first Annual SAS Users Group International Conference, SAS Institute Inc.,2009*, 187–31.

Genz, A., F. Bretz, T. Horthon, T. Miwa, X. Mi, F. Leisch, and F.Scheip (2008). *mvtnorm: Multivariate normal and t distribution.* R package version 0.9-2.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Assoiation 72*, 320–340.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*, 625–629.

Jacqmin-Gadda, H. and R. Thiebaut (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics 1*, 355–368.

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

Louis, T. (1982). Finding the oberserved information matrix when using EM algorithm. *Journal of Royal Statistical Society, Ser. B 44*, 226–233.

Magnus, J. and H. Neudecker (1999). *Matrix differential calculus with application in statistics and econometrics*. Chichester: Wiley.

Matos, L., V. Lachos, N.B.Balakrishnan, and F. Labra (2013). Influence diagnostics linear and nonlinear mixed-effects models with censored data . *Computational Statistics and Data Analysis 57*, 450–464.

Orchard, T. and M. Woodbury (1972). A missing information princple, theory and application. *in Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability 1*, 679–715.

Pinheiro, J. and D. Bates (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., D. Bates, S. DeRoy, and D. Shakar (2006). *nlme: Linear and nonLinear mixed effects models*. R package version 3.1-78.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R foundation for statistical computing. ISBN 3-900051-07-0.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Ruppert, D. (2005). Comments on maximization by parts in likelihood inference. *Journal of Air and Waste Management Association 100*, 1161–1163.

Senn, S., N. Holford, and H. Hockey (2012). The ghost of departed quantities: approaches to dealing with observations below the limit of quantifications. *Statiscs in Medicine 31*, 4285–4295.

Tallis, G. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society, Series, B 23*, 223–229.

Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybid EM for linear and nonlinear mixed effects model with censored response. *Computational Statistics and Data Analysis 51*, 5718–5730.

Vaida, F. and L. Liu (2009a). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Vaida, F. and L. Liu (2009b). *lmec: Linear mixed effects models with censored responses.* R package version 1.0.

Vock, D. M., M. Davidian, and A. A. Tsiatis (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics 0*, 1–13.

Wand, M. (2002). Vector differential calculus in statistics. *The American Statistician 56*, 55–62.

Wu, L. (2010). *Mixed effects models for complex data.* Chapman and Hall CRC Boca Raton FL.

2.8 Appendix

2.8.1 Additional Simulation Results: Substitution Methods

In the presence of censoring, substitution methods are widely used because they are very simple and no special software packages are required for statistical analysis. We present here simulation studies to illustrate the impact of these substitution methods on the parameter estimates and inferences. The substitution methods considered here are; substituting half the DL (`halfsub`), substituting $\frac{\text{DL}}{\sqrt{2}}$ (`sqrtsub`), and substituting the DL (`sub`). After the substitution, the model for the complete data is fitted using `lme`.

SPHERICAL ERROR

Table 2.8: ML estimation. Spherical error case: The comparison is based on the averaged estimates of the parameters (Mean), the simulation based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		β_0	β_1	D_{11}	D_{12}	D_{22}	σ
Truth		5	2	1.7	0.5	2.3	2.3
`lme`	Mean	5.00	2.01	1.75	0.46	2.29	2.29
	MSE	0.12	0.04	1.69	0.36	0.27	0.01
	RB	-0.0	-0.3	2.9	-8.3	-0.3	-0.6
	CP	95.2	96.3				
`lmecm`	Mean	4.98	2.01	2.19	0.48	2.26	2.30
	MSE	0.15	0.05	2.67	0.45	0.30	0.02
	RB	-0.4	0.4	28.5	-3.6	-1.9	0.2
	CP	95.6	95.9				
`lmec`	Mean	4.99	2.00	2.22	0.34	2.34	2.27
	MSE	0.15	0.05	2.05	0.46	0.32	0.02
	RB	-0.1	-0.2	30.5	-31.8	1.8	-1.4
	CP	95.4	96.1				
`sub`	Mean	5.68	1.96	0.54	-0.18	2.09	1.98
	MSE	0.60	0.03	1.77	0.70	0.21	0.11
	RB	-13.5	-1.80	-68.4	-136.3	-9.3	-13.9
	CP	39.3	96.6				
`halfsub`	Mean	4.67	2.11	2.22	0.51	2.03	2.35
	MSE	0.23	0.06	2.25	0.37	0.26	0.02
	RB	-6.6	5.4	30.6	2.6	-11.3	-5.7
	CP	87.0	92.0				
`sqrtsub`	Mean	5.09	2.05	1.24	0.26	2.04	2.17
	MSE	0.12	0.04	1.28	0.34	0.24	0.03
	RB	1.8	2.4	-27.1	-47.1	-11.3	-5.7
	CP	93.2	95.6				

The data are generated from model (2.48) with 20% of the data censored. We present here the results for both ML and REML estimations. According to Table 2.8, as expected, the model based methods (`lmecm` and `lmec`) performed better according to all four measures (Mean, MSE, RB, CP) except the MSE when compared to substitution methods. Specifically, the relative biases of the three substitution methods (`sub`, `halfsub`, `sqrtsub`) for all the parameters are orders of magnitude larger than the model based methods (`lme`, `lmecm`, `lmec`) except for the parameter D_{11}. The relative bias of `sub` for estimating the random

Table 2.9: REML estimation. Spherical error case: The comparison is based on the averaged estimates of the parameters (Mean), the simulation based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		β_0	β_1	D_{11}	D_{12}	D_{22}	σ
Truth		5	2	1.7	0.5	2.3	2.3
`lme`	Mean	5.00	2.00	1.84	0.49	2.30	2.28
	MSE	0.13	0.05	1.68	0.37	0.26	0.01
	RB	0.1	0.0	8.4	-2.5	0.1	-0.7
	CP	93.9	94.6				
`lmecm`	Mean	4.97	2.01	2.34	0.52	2.26	2.30
	MSE	0.15	0.05	2.82	0.42	0.32	0.02
	RB	-0.6	0.3	37.4	3.2	-1.7	0.2
	CP	94.5	94.1				
`lmec`	Mean	4.99	2.00	2.21	0.41	2.34	2.27
	MSE	0.15	0.05	2.03	0.43	0.35	0.02
	RB	-0.2	0.0	30.2	-18.6	1.7	-1.2
	CP	94.2	94.6				
`sub`	Mean	5.69	1.96	0.55	-0.16	2.10	1.98
	MSE	0.62	0.04	1.75	0.68	0.21	0.11
	RB	14.7	-2.1	-67.5	-133.0	-8.7	-13.9
	CP	37.0	95.1				
`halfsub`	Mean	4.67	2.10	2.31	0.55	2.03	2.35
	MSE	0.23	0.06	2.27	0.35	0.27	0.02
	RB	-6.6	5.2	35.7	9.6	-11.8	2.2
	CP	87.6	89.2				
`sqrtsub`	Mean	5.09	2.04	1.28	0.29	2.05	2.17
	MSE	0.12	0.05	1.24	0.32	0.25	0.03
	RB	1.8	2.2	-24.7	-41.1	-11.0	-5.6
	CP	92.8	93.7				

component D_{11} and D_{12} is approximately 24 and 16 times larger respectively when compared to `lme`, 2 and 38 times larger respectively when compared `lmecm` and 2 and 4 times larger respectively when compared to `lmec`. The MSE of the three substitution methods in estimating the parameters were smaller compared to the MSE of `lmecm` and `lmec`. This is because by imputing a single value in place of the censored values (in substitution methods), the variability of the observations are artificially reduced resulting in smaller MSE.

The coverage probability (CP) of 95% Wald confidence interval for the fixed effect parameter β_1 for `sub` and `halfsub` differ largely from target. Specifically the CP for β_1 is 39.3% and 87% respectively. Among the three substitution methods, overall, `sqrtsub` performed better in estimating the average parameter and the CP of the fixed effects which are similar to the model based methods. The results of the REML estimation presented in Table 2.9 are similar to the ML estimation.

NON-SPHERICAL ERROR

Here we provide simulation results to illustrate the impact of substitution methods on the parameter estimates and inferences in the case the within-group error is non-spherical. The data are generated from model (2.49) and 20% of the observation are censored.

Table 2.10: ML estimation. Non-spherical error case: The comparison is based on the averaged estimates of the parameters (Mean), the simulation based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		μ_1	δ_1	β_1	δ_2	σ_b	σ_1	σ_2	ρ
Truth		6.4	0	2.6	0	2	2.3	5.2	0.6
`lme`	Mean	6.39	-0.04	2.60	0.01	1.86	2.27	5.12	0.59
	MSE	0.44	2.13	0.02	0.10	0.38	0.06	0.24	0.01
	RB	-0.2	-3.7*	0.03	1.1 *	-6.8	-1.3	-1.5	-2.5
	CP	95.2	95.2	93.7	93.3				
`lmecm`	Mean	6.39	0.01	2.60	0.00	1.85	2.28	5.10	0.58
	MSE	0.53	2.41	0.02	0.11	0.44	0.07	0.29	0.01
	RB	-0.1	1.3*	-0.01	0.1*	-7.5	-1.1	-1.9	-3.5
	CP	95.2	95.3	94.3	94.2				
`sub`	Mean	8.55	0.38	2.18	-0.05	0.97	2.25	4.55	0.60
	MSE	5.11	2.11	0.19	0.09	1.42	0.05	0.59	0.00
	RB	33.7	38.3*	-16.0	-4.8*	-51.5	-2.4	-12.4	0.1
	CP	7.1	90.4	8.6	93.8				
`halfsub`	Mean	4.68	1.50	2.92	-0.29	1.89	2.83	5.07	0.51
	MSE	3.31	3.78	0.12	0.18	0.57	0.36	0.22	0.02
	RB	-26.9	49.6*	12.2	-29.5*	-5.6	0.23.0	-2.6	-14.7
	CP	37.1	88.2	44.0	86.0				
`sqrtsub`	Mean	6.26	1.05	2.62	-0.19	1.61	2.42	4.77	0.54
	MSE	0.38	2.75	0.02	0.12	0.59	0.09	0.38	0.01
	RB	-2.1	104.6*	0.6	-19.4*	-19.3	5.4	-8.3	-9.3
	CP	96.0	89.8	95.5	94.2				

*Estimated actual bias is reported.

According to Table 2.10, as expected, the model based method `lmecm` still performed better than the substitution methods. Similarly the relative biases of the three substitution methods for all the parameters are orders of magnitude larger than the model based method `lmecm`. The three substitution methods (`sub`, `halfsub`, and `sqrtsub`) overestimate the parameter δ_1 with bias 38.3%, 149.6% and 104.6% respectively. Compared to `lmecm`, `sub` overestimate μ_1 with bias 33.7% and `halfsub` underestimate with bias 26.7%.

The coverage probability for the fixed effects parameters $\mu_1, \delta_1, \beta_1, \delta_2$ ranges from 7.1% – 93.8% for sub, 37.1%-88.2% for `halfsub` and 89.8%-96% for `sqrtsub`. Compared to `lmecm`, `sqrtsub` performed similar in terms of CP. Compared to `lmecm`, MSE is bigger for the substitution methods when the bias is bigger and the MSE of substitution method is smaller compared to the `lmecm` when bias is smaller. Amongs the three substitution methods, `sqrtsub` still performed better in terms average parameter estimates and CP of the fixed effect. The results of the REML estimation presented in Table 2.11 is similar to the ML estimation.

Table 2.11: REML estimation. Non-spherical error case: The comparison is based on the averaged estimates of the parameters (Mean), the simulation based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		μ_1	δ_1	β_1	δ_2	σ_b	σ_1	σ_2	ρ
Truth		6.4	0	2.6	0	2	2.3	5.2	0.6
`lme`	Mean	6.41	-0.01	2.60	0.00	1.88	2.30	5.19	0.59
	MSE	0.46	2.06	0.02	0.09	0.39	0.06	0.24	0.01
	RB	0.2	-1.3*	-0.1	0.3*	-6.0	-0.1	-0.3	-1.5
	CP	93.5	94.8	94.0	95.6				
`lmecm`	Mean	6.40	-0.03	2.60	0.01	1.87	2.31	5.19	0.59
	MSE	0.57	2.49	0.02	0.10	0.45	0.07	0.30	0.01
	RB	0.0	-3.4*	-0.0	0.6*	-6.4	0.4	-0.1	-2.0
	CP	94.0	95.0	94.6	94.8				
`sub`	Mean	8.58	0.39	2.18	-0.05	0.98	2.27	4.62	0.61
	MSE	5.23	2.21	0.19	0.08	1.40	0.05	0.52	0.00
	RB	34.0	39.1*	-16.1	-5.4*	-51.1	-1.4	-1.1	1.2
	CP	6.2	90.3	7.9	94.2				
`halfsub`	Mean	4.67	1.50	2.92	-0.30	1.94	2.86	5.11	0.52
	MSE	3.33	3.83	0.12	0.18	0.55	0.39	0.21	0.01
	RB	-27.0	149.5*	12.3	-29.7*	-3.0	24.2	-1.8	-14.1
	CP	36.7	87.4	44.4	87.7				
`sqrtsub`	Mean	6.27	1.05	2.62	-0.20	1.65	2.45	4.82	0.55
	MSE	0.40	2.82	0.02	0.12	0.54	0.08	0.35	0.01
	RB	-2.0	104.9*	0.6	-19.7*	-17.4	6.4	-7.3	-8.5
	CP	95.7	89.9	95.5	90.1				

*Estimated actual bias is reported.

As part of assessing the impact of substitutions methods on parameter estimates and inference in case of non-spherical error, we also generate data from model (2.50). We simulated 1000 data sets, each simulated data have $N = 50 \times 6 = 300$ observation with 20% observation left-censored. Table 2.12 presents the average estimates of the parameters, simulation-based MSE, RB, and CP. According to table 2.12, `lmecm` and `lme` performed similarly and the parameter estimates are close to the true value. Compared to `lme` and `lmecm`, all the three substitution methods overestimate the variance parameter δ with bias over 100%. The REML estimation presented in Table 2.13 is similar to the ML estimation.

Table 2.12: ML estimation. Additional simulation with non-spherical error: The comparison is based on the average estimates of the parameters (Mean), the simulation-based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		β_0	β_1	σ_b	σ	δ
Truth		3.2	1.5	1.2	2.3	0.7
`lme`	Mean	3.19	1.51	1.19	2.29	0.70
	MSE	0.05	0.02	0.02	0.01	0.00
	RB	-0.4	0.4	-1.1	-0.5	0.4
	CP	95.0	96.0			
`lmecm`	Mean	3.18	1.51	1.19	2.29	0.70
	MSE	0.05	0.02	0.03	0.01	0.01
	RB	-0.5	0.4	-0.7	-0.3	0.5
	CP	95.0	96.0			
`sub`	Mean	3.43	1.58	0.93	1.89	1.41
	MSE	0.09	0.02	0.09	0.18	0.51
	RB	7.3	5.2	-22.2	-17.9	101.1
	CP	71.0	92.0			
`halfsub`	Mean	3.07	1.63	1.13	2.24	1.42
	MSE	0.06	0.03	0.03	0.01	0.52
	RB	-4.0	8.7	-5.8	-2.5	102.4
	CP	92.0	86.0			
`sqrtsub`	Mean	3.22	1.61	1.05	2.08	1.42
	MSE	0.04	0.03	0.04	0.06	0.52
	RB	0.7	7.3	-12.4	-9.5	102.5
	CP	94.0	88.0			

The simulations studies in this section (2.8.3) indicates that substitution methods leads to bias in estimates of the fixed effects and most importantly bias in the variances-covariances component and the estimate of the variance function parameters. As illustrated in Table 2.8 and 2.9, the crude imputation methods performed worse in estimating variance-covariance matrix especially when the dimension is greater than 1. This is also true in case of estimating parameters in the variance function (δ in Table 2.12).

Table 2.13: REML estimation. Additional simulation with non-spherical error: The comparison is based on the averaged estimates of the parameters (Mean), the simulation based mean square error (MSE), relative bias (RB)(%) and coverage probability (CP)(%) for the 95% confidence interval.

		β_0	β_1	σ_b	σ	δ
Truth		3.2	1.5	1.2	2.3	0.7
`lme`	Mean	3.20	1.50	1.19	2.30	0.70
	MSE	0.05	0.02	0.02	0.01	0.00
	RB	-0.1	-0.1	-0.7	-0.2	-0.3
	CP	94.4	95.3			
`lmecm`	Mean	3.19	1.50	1.19	2.30	0.70
	MSE	0.05	0.02	0.03	0.01	0.00
	RB	-0.2	-0.1	-0.5	-0.1	-0.3
	CP	94.8	94.8			
`sub`	Mean	3.45	1.57	0.93	1.89	1.40
	MSE	0.10	0.02	0.09	0.18	0.50
	RB	7.7	4.5	-22.4	-17.8	100.3
	CP	67.3	91.5			
`halfsub`	Mean	3.08	1.62	1.13	2.25	1.41
	MSE	0.06	0.03	0.03	0.01	0.51
	RB	-3.6	8.1	-5.6	-2.3	101.5
	CP	91.8	85.2			
`sqrtsub`	Mean	3.23	1.60	1.05	2.09	1.41
	MSE	0.04	0.03	0.04	0.05	0.51
	RB	1.1	6.6	-12.4	-9.3	101.7
	CP	93.7	87.5			

2.8.2 ADDITIONAL MATERIAL ON REML ESTIMATION

We present here the technical details of the derivation of the objective function for the REML estimation without an approximation of the E-step and the modification that can be done to `lme` to implement the M-step. As noted earlier, the complete data log-restricted-likelihood is given by

$$\begin{aligned}\log L_R(\boldsymbol{\xi} \mid \mathbf{y}) \quad = \quad & -\frac{N-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\Bigg(\| \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \|^2 \Bigg) \\ & + \sum_{i=1}^{m}\Bigg(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log | \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i | \Bigg) - \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|. (2.51)\end{aligned}$$

The E-step for the EM formulation, which involves computing $Q_R(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) = E(\log L_R(\boldsymbol{\xi} \mid \mathbf{y})|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$ is given by

$$Q_R(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) \quad = \quad -\frac{N-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m} E\Bigg(\| \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \|^2 | \, \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)} \Bigg)$$

$$+ \sum_{i=1}^{m} \left(\log \operatorname{abs}|\boldsymbol{\Delta}| - \frac{1}{2} \log |\, \widetilde{\mathbf{Z}}_i^T \widetilde{\mathbf{Z}}_i \,| \right) - \frac{1}{2} \log \Big| \sum_{i=1}^{m} \widetilde{\mathbf{X}}_i^T \boldsymbol{\Sigma}_{s,i}^{-1} \widetilde{\mathbf{X}}_i \Big|. \tag{2.52}$$

But

$$\sum_{i=1}^{m} \left(||\, \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i \,||^2 \right) = \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}} \hat{\boldsymbol{\delta}}\|^2 = \|(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}})\widetilde{\mathbf{y}}\|^2$$

where

$$\widetilde{\mathbf{H}} \;=\; \begin{pmatrix} \widetilde{\mathbf{Z}}_1 & & & \widetilde{\mathbf{X}}_1 \\ & \ddots & & \vdots \\ & & \widetilde{\mathbf{Z}}_m & \widetilde{\mathbf{X}}_m \end{pmatrix} \quad \text{and} \quad \mathbf{P}_{\widetilde{\mathbf{H}}} = \widetilde{\mathbf{H}}(\widetilde{\mathbf{H}}^T \widetilde{\mathbf{H}})^{-1} \widetilde{\mathbf{H}}^T.$$

Thus

$$\begin{aligned} &\textstyle\sum_{i=1}^{m} E\left(||\, \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i \,||^2 |\, \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)} \right) \\ =\;& E\left(\|(\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}})\widetilde{\mathbf{y}}\|^2 | \mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)} \right) = E\left(\|\widetilde{\mathbf{y}}^T (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}})\widetilde{\mathbf{y}}\|^2 | \mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)} \right) \\ =\;& \operatorname{tr}\left\{ E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}^{(k)})^T (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}})(E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}^{(k)})) \right\} + \operatorname{tr}\left\{ (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \operatorname{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}^{(k)}) \right\} \\ =\;& \|E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) - \widetilde{\mathbf{H}} \widetilde{\hat{\boldsymbol{\delta}}}_R\|^2 + \operatorname{tr}\left\{ (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \operatorname{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) \right\} \end{aligned} \tag{2.53}$$

where $\widetilde{\hat{\boldsymbol{\delta}}}_R = (\widetilde{\mathbf{H}}^T \widetilde{\mathbf{H}})^{-1} \widetilde{\mathbf{H}}^T E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$. Let $\widetilde{\mathbf{S}}_R \widetilde{\mathbf{S}}_R^T = \operatorname{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$, where $\widetilde{\mathbf{S}}_R$ is the lower triangular Cholesky factor such that $\operatorname{vec}(\widetilde{\mathbf{S}}_R) = (\widetilde{\mathbf{S}}_{.1_R}^T, \ldots, \widetilde{\mathbf{S}}_{.N'}^T)^T$. Note that $N' = \sum_{i=1}^{m} (n_i + q) = N + mq$. Analogous to the ML case,

$$\begin{aligned} &\operatorname{tr}\left\{ (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \operatorname{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) \right\} \\ =\;& \operatorname{tr}\left\{ \widetilde{\mathbf{S}}_R^T (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \widetilde{\mathbf{S}}_R \right\} = \left\| \begin{pmatrix} (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \widetilde{\mathbf{S}}_{.1_R} \\ \vdots \\ (\mathbf{I} - \mathbf{P}_{\widetilde{\mathbf{H}}}) \widetilde{\mathbf{S}}_{.N'_R} \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \widetilde{\mathbf{S}}_{.1_R} - \widetilde{\mathbf{H}}(\widetilde{\mathbf{H}}^T \widetilde{\mathbf{H}})^{-1} \widetilde{\mathbf{H}}^T \widetilde{\mathbf{S}}_{.1_R} \\ \vdots \\ \widetilde{\mathbf{S}}_{.N'_R} - \widetilde{\mathbf{H}}(\widetilde{\mathbf{H}}^T \widetilde{\mathbf{H}})^{-1} \widetilde{\mathbf{H}}^T \widetilde{\mathbf{S}}_{.N'_R} \end{pmatrix} \right\|^2 \\ =\;& \left\| \begin{pmatrix} \widetilde{\mathbf{S}}_{.1_R} - \widetilde{\mathbf{H}} \widetilde{\hat{\boldsymbol{\delta}}}_{1_R} \\ \vdots \\ \widetilde{\mathbf{S}}_{.N'_R} - \widetilde{\mathbf{H}} \widetilde{\hat{\boldsymbol{\delta}}}_{N'_R} \end{pmatrix} \right\|^2, \end{aligned} \tag{2.54}$$

where $\widetilde{\hat{\boldsymbol{\delta}}}_{l_R} = (\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}})^{-1}\widetilde{\mathbf{H}}^T\widetilde{\mathbf{S}}_{.l_R}, l = 1, \dots, N'$. Substituting (2.54) into (2.53) gives

$$\sum_{i=1}^{m} E\Bigg(\mid\mid \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \mid\mid^2 \mid \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)}\Bigg) = \left\| \begin{pmatrix} E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) - \widetilde{\mathbf{H}}\widetilde{\hat{\boldsymbol{\delta}}}_R \\ \widetilde{\mathbf{S}}_{.1_R} - \widetilde{\mathbf{H}}\widetilde{\hat{\boldsymbol{\delta}}}_{1_R} \\ \vdots \\ \widetilde{\mathbf{S}}_{.N'_R} - \widetilde{\mathbf{H}}\widetilde{\hat{\boldsymbol{\delta}}}_{N'_R} \end{pmatrix} \right\|^2 . \qquad (2.55)$$

Note that the m subjects are assumed independent so $\mathrm{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)})$ has a block diagonal structure. Thus $\mathrm{Var}(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\xi}_R^{(k)}) = \mathrm{diag}\Big(\mathrm{Var}(\widetilde{\mathbf{y}}_1|\mathbf{Q}_1, \mathbf{C}_1, \boldsymbol{\xi}_R^{(k)}), \dots, \mathrm{Var}(\widetilde{\mathbf{y}}_m|\mathbf{Q}_m, \mathbf{C}_m, \boldsymbol{\xi}_R^{(k)})\Big)$ hence $\widetilde{\mathbf{S}}_R\widetilde{\mathbf{S}}_R^T = \mathrm{diag}(\widetilde{\mathbf{S}}_{1_R}\widetilde{\mathbf{S}}_{1_R}^T, \dots, \widetilde{\mathbf{S}}_{m_R}\widetilde{\mathbf{S}}_{m_R}^T)$ where $\widetilde{\mathbf{S}}_{i_R}\widetilde{\mathbf{S}}_{i_R}^T = \mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)})$ such that $\mathrm{vec}(\widetilde{\mathbf{S}}_{i_R}) = (\widetilde{\mathbf{S}}_{i,1_R}^T, \dots, \widetilde{\mathbf{S}}_{i,n_i+q_R}^T)^T$. It is desirable to write equation (2.55) at the cluster level. Note that

$$\|E(\widetilde{\mathbf{y}}|\mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}_R^{(k)}) - \widetilde{\mathbf{H}}\widetilde{\hat{\boldsymbol{\delta}}}_R\|^2 = \sum_{i=1}^{m} \|E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)}) - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i_R}\|^2$$

where $\hat{\boldsymbol{\beta}}$ is defined as in section (2.2.3) except $\mathbf{y}_i$ is replaced with $E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)})$ and $\widetilde{\hat{\mathbf{b}}}_{i_R} = (\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T(E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)}) - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}})$. Also $\widetilde{\mathbf{S}}_{.1_R} = (\widetilde{\mathbf{S}}_{1,1_R}^T, \mathbf{0}^T, \dots, \mathbf{0}^T)^T$ where $\widetilde{\mathbf{S}}_{1,1_R}^T$ is the first column of the $\widetilde{\mathbf{S}}_{1R}$, the Cholesky decomposition of $\mathrm{Var}(\widetilde{\mathbf{y}}_1|\mathbf{Q}_1, \mathbf{C}_1, \boldsymbol{\xi}_R^{(k)})$ so similarly

$$\begin{aligned}
&\|\widetilde{\mathbf{S}}_{.1_R} - \widetilde{\mathbf{H}}(\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}})^{-1}\widetilde{\mathbf{H}}^T\widetilde{\mathbf{S}}_{.1_R}\|^2 \\
&\qquad = \|\widetilde{\mathbf{S}}_{1,1_R} - \widetilde{\mathbf{X}}_1\hat{\boldsymbol{\beta}}_1 - \widetilde{\mathbf{Z}}_1\widetilde{\hat{\mathbf{b}}}_{1,1_R}\|^2 \\
&\qquad\quad + \|\mathbf{0} - \widetilde{\mathbf{X}}_2\hat{\boldsymbol{\beta}}_1 - \widetilde{\mathbf{Z}}_2\widetilde{\hat{\mathbf{b}}}_{2,1_R}\|^2 + \cdots + \|\mathbf{0} - \widetilde{\mathbf{X}}_m\hat{\boldsymbol{\beta}}_1 - \widetilde{\mathbf{Z}}_m\widetilde{\hat{\mathbf{b}}}_{m,1_R}\|^2 \\
&\qquad = \sum_{i=1}^{m} \|\widetilde{\mathbf{S}}_{i,1_R} - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_1 - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,1_R}\|^2
\end{aligned}$$

where $\widetilde{\mathbf{S}}_{i,1_R} = \mathbf{0}$ if $i \neq 1$. Thus equation (2.55) can be written as

$$\sum_{i=1}^{m} E\Bigg(\mid\mid \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \mid\mid^2 \mid \mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)}\Bigg) = \sum_{l=1}^{N'+1}\sum_{i=1}^{m} \|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}}_l - \widetilde{\mathbf{Z}}_{i,l}\widetilde{\hat{\mathbf{b}}}_{i,l_R}\|^2 \quad (2.56)$$

where

$$\widetilde{\mathbf{w}}_{i,l_R} = \begin{cases} E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\xi}_R^{(k)})^T & l = 1 \\ \widetilde{\mathbf{S}}_{i,l-1_R} & l = 2, \dots, N'+1, \end{cases}$$

$$\begin{aligned}
\widetilde{\mathbf{X}}_{i,l} &= \widetilde{\mathbf{X}}_i \ \ \forall \ l, \\
\widetilde{\mathbf{Z}}_{i,l} &= \widetilde{\mathbf{Z}}_i \ \ \forall \ l, \qquad\qquad\qquad\qquad \text{and} \\
\tilde{\hat{\mathbf{b}}}_{i,l_R} &= \begin{cases} \tilde{\hat{\mathbf{b}}}_{i_R} & l=1 \\ \tilde{\hat{\mathbf{b}}}_{i,l-1_R} & l=2,\ldots,N'+1. \end{cases}
\end{aligned}$$

Substituting (2.56) into (2.52) yields

$$\begin{aligned}
Q_R(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) &= -\frac{N-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{l=1}^{N'+1}\sum_{i=1}^{m} \|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}}_l - \widetilde{\mathbf{Z}}_{i,l}\tilde{\hat{\mathbf{b}}}_{i,l_R}\|^2 \\
&+ \sum_{i=1}^{m}\left(\log \operatorname{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid \right) - \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|. \quad (2.57)
\end{aligned}$$

The M-step update for σ^2, the conditional estimate is given by

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = \frac{1}{N-p}\sum_{l=1}^{N'+1}\sum_{i=1}^{m}\|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}}_l - \widetilde{\mathbf{Z}}_i\tilde{\hat{\mathbf{b}}}_{i,l_R}\|^2 \quad (2.58)$$

where $\hat{\boldsymbol{\delta}}_R$ is obtained the same way as in ML case. We obtain the profile log-restricted-likelihood as

$$\begin{aligned}
Q_R(\boldsymbol{\theta} \mid \boldsymbol{\xi}_R^{(k)}) &= const - \frac{(N-p)}{2}\log\left(\hat{\sigma}_R^2(\boldsymbol{\theta})\right) + \sum_{i=1}^{m}\left(\log \operatorname{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid \right) \\
&- \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|. \quad (2.59)
\end{aligned}$$

Suppose we pass `lme` the response vector $\mathbf{w}_R$ and the design matrices $\mathbf{X}$ and $\mathbf{Z}$, `lme` construct the objective function

$$\begin{aligned}
Q_{R_{lme}}(\boldsymbol{\xi} \mid \boldsymbol{\xi}_R^{(k)}) &= -\frac{\widetilde{N}-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{l=1}^{N'+1}\sum_{i=1}^{m}\|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_{i,l}\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_{i,l}\tilde{\hat{\mathbf{b}}}_{i,l_R}\|^2 \\
&+ \sum_{l=1}^{(N'+1)}\sum_{i=1}^{m}\left(\log \operatorname{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_{i,l}^T\widetilde{\mathbf{Z}}_{i,l} \mid \right) \\
&- \frac{1}{2}\log\Big|\sum_{l=1}^{(N'+1)}\sum_{i=1}^{m}\widetilde{\mathbf{X}}_{i,l}^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_{i,l}\Big| \\
&= -\frac{\widetilde{N}-p}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{l=1}^{N'+1}\sum_{i=1}^{m}\|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\tilde{\hat{\mathbf{b}}}_{i,l_R}\|^2
\end{aligned}$$

$$
+ (N'+1)\sum_{i=1}^{m}\Bigg(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\Bigg)
$$

$$
- \frac{1}{2}\log\left|(N'+1)\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\right|. \tag{2.60}
$$

For a given $\boldsymbol{\theta}$, `lme` compute the estimate of $\boldsymbol{\delta}_R$ and σ_R^2 given by:

$$
\hat{\sigma}^2_{R_{lme}}(\boldsymbol{\theta}) = \frac{1}{\widetilde{N}-p}\sum_{l=1}^{N'+1}\sum_{i=1}^{m}\|\widetilde{\mathbf{w}}_{i,l_R} - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\widetilde{\hat{\mathbf{b}}}_{i,l_R}\|^2. \tag{2.61}
$$

Substituting these conditional estimators back into (2.60) provides the profile log-restricted-likelihood

$$
\begin{aligned}
Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_R^{(k)}) \quad &= \quad const - \frac{\widetilde{N}-p}{2}\log\big(\hat{\sigma}^2_R(\boldsymbol{\theta})\big) + (N'+1)\sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\big) \\
&\quad - \frac{1}{2}\log\left|(N'+1)\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\right|.
\end{aligned}
$$

In case of balance data, $\widetilde{N} = \sum_{i=1}^{m} n_i(\sum_{i=1}^{m}(n_i+q)+1) = mn(mn+mq+1) = N(N'+1)$ so $Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_R^{(k)})$ can be written

$$
\begin{aligned}
&Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_R^{(k)}) \\
&= const - \frac{N(N'+1)-p}{2}\log\big(\hat{\sigma}^2_R(\boldsymbol{\theta})\big) + (N'+1)\sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\big) \\
&\quad - \frac{1}{2}\log\left|(N'+1)\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\right| \\
&= const - \frac{N(N'+1)-p}{2}\log\big(\hat{\sigma}^2_R(\boldsymbol{\theta})\big) + (N'+1)\sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\big) \\
&\quad - \frac{1}{2}\log\left\{(N'+1)^p\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big|\right\} \\
&= const - \frac{N(N'+1)-p}{2}\log\big(\hat{\sigma}^2_R(\boldsymbol{\theta})\big) + (N'+1)\sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\big) \\
&\quad - \underbrace{\frac{1}{2}p\log(N'+1)}_{const} - \frac{1}{2}\log\Big|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\boldsymbol{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i\Big| \\
&= const - \frac{N(N'+1)-p}{2}\log\big(\hat{\sigma}^2_R(\boldsymbol{\theta})\big) + (N'+1)\sum_{i=1}^{m}\big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log|\ \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\ |\big)
\end{aligned}
$$

$$
\begin{aligned}
& - \frac{1}{2}\log|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i| \\
= \; & const - (N'+1)\Bigg[\frac{N-\widetilde{p}}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta})\big) + \sum_{i=1}^{m}\big(\log\mathrm{abs}|\mathbf{\Delta}| - \frac{1}{2}\log|\,\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\,|\,\big) \\
& - \frac{1}{2(N'+1)}\log|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s,i}^{-1}\widetilde{\mathbf{X}}_i|\Bigg] \qquad (2.62)
\end{aligned}
$$

where $\widetilde{p} = \frac{p}{N'+1}$. But the profiled restricted-log-likelihood that we want to optimize is (2.59). We then modify (2.62), $Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_R^{(k)})$ to be the same as (2.59), $Q_R(\boldsymbol{\psi} \mid \boldsymbol{\xi}_R^{(k)})$ by multiplying the the $\widetilde{p}$ in the first component of the (2.62) and the fourth component of (2.62) by $N'+1$. We then modified the `lme` function appropriately.

In case of unbalanced data, we use the pseudo data approach described in ML case for unbalance data. Analogous to the ML case, `lme` estimates the parameters in the model as in (2.11) but with $\widetilde{N}$ replaced with $\widetilde{\ddot{N}}$ and the log-restricted profiled likelihood given by:

$$
\begin{aligned}
Q_{R_{lme}}(\boldsymbol{\theta} \mid \boldsymbol{\xi}^{(k)}) \;=\; & const - (\ddot{N}'+1)\Bigg[\frac{\ddot{N}-\widetilde{\ddot{p}}}{2}\log\big(\hat{\sigma}^2(\boldsymbol{\theta})\big) + \sum_{i=1}^{m}\big(\log\mathrm{abs}|\mathbf{\Delta}| \\
& - \frac{1}{2}\log|\,\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i\,|\,\big) - \frac{1}{2(\ddot{N}'+1)}\log|\sum_{i=1}^{m}\widetilde{\mathbf{X}}_i^T\mathbf{\Sigma}_{s_i}^{-1}\widetilde{\mathbf{X}}_i|\Bigg] \qquad (2.63)
\end{aligned}
$$

where $\widetilde{\ddot{p}} = \frac{p}{\ddot{N}'+1}$. The profile restricted-log-likelihood in (2.63) can be modified to be the same as (2.59) by subtracting N_r from $\ddot{N}$ such that $\ddot{N} - N_r = N$ and multiplying the $\widetilde{\ddot{p}}$ and the fourth component of the profile restricted-log-likelihood piece by $\ddot{N}'+1$.

Chapter 3

Analysis of Bivariate Longitudinal Data with Censored Response[⋆]

[⋆] Tuglo, E., and Hall, D. (2014+). Analysis of Bivariate Longitudinal Data with Censored Response. To be submitted.

Abstract

Estimating correlation between bivariate longitudinal continuous outcomes is of particular interest to researchers in environmental and biomedical studies. Analysis of bivariate longitudinal data can be challenging when the responses are subject to left or right-censoring as result of lower or upper detection limits (DL) respectively. The use of a Monte Carlo expectation maximization (MCEM) algorithm for computing the maximum likelihood estimates for bivariate linear mixed-effects models (BLMEM) with censored responses has been proposed by Xie et al. (2012) but their approach can only handle bivariate random intercept models. In this paper we propose a methodology for likelihood-based estimation in a wider class of BLMEMs than that in Xie et al. (2012). In particular, we propose a computationally feasible EM algorithm that uses closed form expressions at the E-step, as opposed to Monte Carlo simulation for maximum likelihood (ML) and restricted maximum likelihood (REML) estimation for BLMEM with censored response. The R function `lmecm`, which implements our methedology, has the ability to handle all kinds of serial correlation structures within each response, heteroscedastic errors within each response, and multiple levels of nesting in random effects within each response while accounting for censoring and contemporaneous correlation across the responses.

Keywords: Detection limit; Bivariate; EM algorithm; Maximum likelihood; Heteroscedastic errors; Serial correlation.

3.1 Introduction

Pollutants and other variables are often measured repeatedly over time in environmental studies. Such multivariate longitudinal data (MLD) provide a unique opportunity for researchers to study the joint evolution of multiple response variables over time. However, the analysis of MLD can be challenging compared to the univariate case in several respects.

For instance, the variance of the errors can be different for the different responses, the errors are likely to be correlated for a particular response across different time points, and the errors are likely to be correlated across the multiple response variables at the same time point (Gao et al., 2009). In addition, the analysis of MLD becomes more complicated when there are nested random effects and/or when components of the multiple response variables are missing and/or subject to detection limits (DLs). These complications are common in practice in environmental and epidemiological studies where biomarkers of disease or environmental pollutants are often measured repeatedly over time and are subject to minimum DLs.

For example, the methodology in this paper is motivated by a data set that contains real time measurements of cookstove-related carbon monoxide (CO) and particulate matter with aerodynamic diameter less than 2.5 microns ($PM_{2.5}$) from 19 subjects (households) in Peru who are exposed to these environmental pollutants when preparing their meals. These response variables ($PM_{2.5}$ and CO) were collected every 30 seconds for several hours in each household. The response variables are subject to lower DLs. The investigators want to determine the contemporaneous (time-specific) correlation between these two variables.

This sort of a problem - quantifying the association between real time measurements of environmental variables - is an important one in environmental science. Since these response variables ($PM_{2.5}$ and CO) are collected in real time, they are subject to complications such as complex serial correlation as well as an evolving mean structure. Also since the response variables are subject to lower DLs which induce left-censoring, these complications must be accounted for to be able to quantify the association properly. Even in the absence aforementioned complications, sample Pearson correlation coefficients can not be used for such data because repeated measures from the same subjects tend to be correlated (Bland and Altman, 1994). Lorenz et al. (2011) proposed an approach to estimate correlation between bivariate continuous clustered variables but their focus is on computing the marginal correlation. The use of mixed effects models implemented by SAS PROC MIXED has been proposed by the Thiebaut et al. (2002) and Hamlett et al. (2004) to estimates correlation between multiple responses involving repeated measures, but these approaches cannot accommodate censored observations.

To address the problem of left-censoring, naive and crude imputation approaches of substituting DLs or some function of the DLs in place of the censored observations are widely used even though the validity of these approaches is questionable; these methods have been shown to lead to biased estimators (Wu, 2010). As noted in Lynn (2001) and Lau and Gange (2004), multiple imputation (MI) is another potential approach for handling left-censored data. Rather than filling in the censored observations with single value, several imputed values based on the imputation model are imputed for the censored value and standard methods for the completed data can be used to analyse the imputed data sets. But, left-censoring induced by DL does not satisfied the 'missing at random' condition typically assumed by MI. This is because an observation is left-censored only when the value is lower than the DL hence the pattern of missingness depends on the missing values so the performance of MI in case of left-censoring needs to be examined (Xie et al., 2012).

As alternative to ad hoc imputation methods, maximum likelihood (ML) and restricted maximum likelihood (REML) approach that incorporate censoring information into the likelihood of the observed data have been proposed by Hughes (1999), Vaida et al. (2007), Vaida and Liu (2009) and Tuglo and Hall (2014). But these approaches only estimate the association of the univariate response with the predictor (e.g, treatment) variables and the association of the observation within the univariate response. As noted in Xie et al. (2012), little research has been done to address the association/correlation between multiple response variables with repeated measurements while accounting for censoring. Thiebaut et al. (2000) studied correlation between HIV RNA viral load and CD4 counts subject to left-censoring and informative drop out. However only one biomarker, HIV RNA viral load, was subject to a DL and it was assumed that correlation between the two biomakers was completely explained by the random effects rather than allowing correlation between the measurement errors.

Xie et al. (2012) proposed ML estimation but their methodology can handle only the bivariate random-intercept model. Essentially this is a generalization of the compound symmetry covariance structure to bivariate longitudinal observations. Thus their methodology cannot accommodate the possibility of the within-group errors being correlated across different time points for each response or any form of hetorescedascity for the within-group errors for each response. In this paper we develop a bivariate linear mixed model that allows the analysis of bivariate longitudinal data (BLD) subject to DLs. We develop a computationally feasible algorithm for ML and REML estimation and implement it in an R function called `lmecm` that can handle a wider class of bivariate linear models than proposed by Xie et al. (2012). In particular, our implementation has the ability to handle all kinds of serial correlation structures, multiple levels of nesting in random effects and can easily be extended to analyze MLD with three or more response variables. We apply this methodology to the aforementioned real time cookstove data to quantify the contemporaneous correlation between CO and $\mathrm{PM}_{2.5}$.

The rest of this paper is organized as follows. In section 3.2, the BLMEM is defined and the estimation and inference procedures are described. The methodology is illustrated in section 3.3 with an application to the cookstove smoke data. Two simulation studies corresponding to the case of spherical and non-spherical error are presented in section 3.4. A discussion and concluding remarks are made in section 3.5.

3.2 Bivariate Linear Mixed Effects Models for Censored Response

3.2.1 Model Specification

Let $\mathbf{y}_h$, $h = 1, 2$, denote the response vector for the hth response type (e.g, CO or PM2.5), and write the jth response for the ith subject, $i = 1, \ldots, m$ for the two response types as $\mathrm{y}_{1ij}, j = 1, \ldots, n_{1i}$ and $\mathrm{y}_{2ij}, j = 1, \ldots, n_{2i}$, respectively, taken at associated covariate values $\mathbf{X}_{1ij}$ and $\mathbf{X}_{2ij}$. Let $\mathbf{y}_{1i} = (\mathrm{y}_{1i1}, \mathrm{y}_{1i2}, \ldots, \mathrm{y}_{1in_{1i}})$ and $\mathbf{y}_{2i} = (\mathrm{y}_{2i1}, \mathrm{y}_{2i2}, \ldots, \mathrm{y}_{2in_{2i}})$ denote the $(n_{1i} \times 1)$ and $(n_{2i} \times 1)$ vectors of observations for the ith cluster for the two response types. In case there are no missing values, $n_{1i} = n_{2i} = n_i$. By indexing the corresponding error vector, covariate vector and random vector similarly, we can write separate (unlinked) univariate linear mixed models for the two response as

$$\begin{aligned}
\mathbf{y}_{1i} &= \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}_{1i}\mathbf{b}_{1i} + \boldsymbol{\epsilon}_{1i}, \;\; \mathbf{b}_{1i} \sim N(0, \mathbf{D}_1), \quad \boldsymbol{\epsilon}_{1i} \sim N(0, \sigma^2_{\epsilon_1}\boldsymbol{\Lambda}_{1i}) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.1) \\
\mathbf{y}_{2i} &= \mathbf{X}_{2i}\boldsymbol{\beta}_2 + \mathbf{Z}_{2i}\mathbf{b}_{2i} + \boldsymbol{\epsilon}_{2i}, \;\; \mathbf{b}_{2i} \sim N(0, \mathbf{D}_2), \quad \boldsymbol{\epsilon}_{2i} \sim N(0, \sigma^2_{\epsilon_2}\boldsymbol{\Lambda}_{2i}),
\end{aligned}$$

where , for each h and i, $\mathbf{X}_{hi}$ is an $n_{hi} \times p_h$ design matrix, $\boldsymbol{\beta}_h$ is a p_h-vector of fixed effects, $\mathbf{Z}_{hi}$ is a $n_{hi} \times q_h$ random effect design matrix. Notice that model (3.1) allows heteroscedastic and/or correlated within-group errors for each component of the model. However when the two response are not independent, fitting these two models separately would fail to account for the possible correlation between the two responses.

Suppose there are no missing values. Then the total number of measurements on individual i is $2n_i$, and let the $(2n_i \times 1)$ vector $\mathbf{y}_i$ denote the vector consisting of $\mathbf{y}_{1i}$ and $\mathbf{y}_{2i}$ "stacked" with jth element $\mathbf{y}_{ij}$, so that j now indexes position in this vector and let $\boldsymbol{\epsilon}_i$ and its elements denote the associated within-individual errors. Index the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ similarly.

To take account of correlation between the two responses, one could use the following bivariate linear mixed model:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \\ \mathbf{b}_i &= \begin{pmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\epsilon}_i = \begin{pmatrix} \boldsymbol{\epsilon}_{1i} \\ \boldsymbol{\epsilon}_{2i} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R}_i) \end{aligned} \tag{3.2}$$

where

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix}, \ \mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{1i} & 0 \\ 0 & \mathbf{X}_{2i} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \ \mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{1i} & 0 \\ 0 & \mathbf{Z}_{2i} \end{pmatrix},$$

the random effect $\mathbf{b}_i$ and the the within-group errors $\boldsymbol{\epsilon}_i$ are assumed to be independent. The covariance matrices of the random effects and measurement errors are defined by

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_{12} \\ \mathbf{D}_{12} & \mathbf{D}_2 \end{pmatrix}, \qquad \mathbf{R}_i = \begin{pmatrix} \sigma_{\epsilon_1}^2 \boldsymbol{\Lambda}_{1i} & \sigma_{\epsilon_{12}}\mathbf{I} \\ \sigma_{\epsilon_{12}}\mathbf{I} & \sigma_{\epsilon_2}^2 \boldsymbol{\Lambda}_{2i} \end{pmatrix}.$$

Within $\mathbf{D}$ we allow for arbitrary and possibly different random effects variance covariance structures $\mathbf{D}_1$ and $\mathbf{D}_2$ for the two responses, and correlation across the response-specific random effects $\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$ via possibly non-zero matrices $\mathbf{D}_{12} = \mathbf{D}_{21}^T$. Similarly, the structure of the covariance matrix $\mathbf{R}_i$ accounts for the possibility of different error variances for the two responses, the errors being correlated for a particular response across different time points, and the errors being correlated across the two response variables at the same time points (contemporaneous correlation). Note that the two response are independent if $\mathbf{D}_{12} = \mathbf{D}_{21}^T = \mathbf{0}$ and $\sigma_{\epsilon_{12}} = 0$, in which case model (3.2) reduces to (3.1).

From model (3.2), it is obvious there are subject level and time level correlation structure between the two responses: The quantity $\mathbf{P}_b = [\text{diag}(\mathbf{D}_1)]^{-1/2}\mathbf{D}_{12}[\text{diag}(\mathbf{D}_2)]^{-1/2}$ measures the correlation between the response because of the subject-specific random variation, $\rho_\epsilon = \frac{\sigma_{\epsilon_{12}}}{\sqrt{\sigma_{\epsilon_1}^2 \sigma_{\epsilon_2}^2}}$ measures the correlation between the two responses within the same subject at the same time point because of the time-specific random variation. In case the $\mathbf{b}_i$ are subject-specific random intercept for both responses then $\mathbf{D} = \begin{pmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{12}} & \sigma^2_{b_1} \end{pmatrix}$ and $\mathbf{P}_b = \rho_b = \frac{\sigma_{b_{12}}}{\sqrt{\sigma^2_{b_1}\sigma^2_{b_2}}}$ then ρ, the overall correlation when both subject-specific and time-specific random variation is taken into account is given by

$$\rho = \mathrm{Corr}(y_{1ij}, y_{2ij}) = \frac{\mathrm{Cov}(y_{1ij}, y_{2ij})}{\sqrt{\mathrm{Cov}(y_{1ij}, y_{1ij})\mathrm{Cov}(y_{2ij}, y_{2ij})}} = \frac{\sigma_{b_{12}} + \sigma_{\epsilon_{12}}}{\sqrt{(\sigma^2_{b_1} + \sigma^2_{\epsilon_1})(\sigma^2_{b_2} + \sigma^2_{\epsilon_2})}}. \tag{3.3}$$

In this research, our main goal is to estimate and do inference on the contemporaneous correlation, the correlation above and beyond the random effect component in the model. It is convenient to reparametrize in terms of the contemporaneous correlation parameter. Note that with the parametrization use in model (3.2), the contemporaneous correlation is given by

$$\mathrm{Corr}(\mathrm{y}_{1ij}, \mathrm{y}_{2ij} | \mathbf{b}_i) = \rho_\epsilon \tag{3.4}$$

which implies $\sigma_{\epsilon_{12}} = \rho_\epsilon \sqrt{\sigma^2_{\epsilon_1}\sigma^2_{\epsilon_2}}$ so the covariance matrix $\mathbf{R}_i$ can be reparametrized as $\mathbf{R}_i = \begin{pmatrix} \sigma^2_1 \mathbf{\Lambda}_{1i} & \rho_\epsilon \sqrt{\sigma^2_{\epsilon_1}\sigma^2_{\epsilon_2}}\mathbf{I} \\ \rho_\epsilon \sqrt{\sigma^2_{\epsilon_1}\sigma^2_{\epsilon_2}}\mathbf{I} & \sigma^2_{\epsilon_2}\mathbf{\Lambda}_{2i} \end{pmatrix}$. Because ρ_ϵ is a correlation parameter, $\hat{\rho}_\epsilon$ may not converge quickly to normality. Therefore for the purpose of inference, we reparametrize again using Fisher's Z-transformation. Let $\tau = \frac{1}{2} \log \frac{1+\rho_\epsilon}{1-\rho_\epsilon}$ then $\rho_\epsilon = \tanh \tau$. The matrix $\mathbf{R}_i$ can be decomposed as $\mathbf{R}_i = \mathbf{V}_i^{1/2}\mathbf{K}_i\mathbf{V}_i^{1/2}$ where, $\mathbf{V}_i = \begin{pmatrix} \sigma^2_{\epsilon_1}\mathbf{I} & 0 \\ 0 & \sigma^2_{\epsilon_2}\mathbf{I} \end{pmatrix}$ is a diagonal matrix that describes the variance and $\mathbf{K}_i = \begin{pmatrix} \mathbf{\Lambda}_{1i} & \rho_\epsilon \mathbf{I} \\ \rho_\epsilon \mathbf{I} & \mathbf{\Lambda}_{2i} \end{pmatrix}$ is a correlation matrix that described correlation of the within-group error $\boldsymbol{\epsilon}_i$. Note that in the special case where the within-group error for each response is spherical, $\mathbf{\Lambda}_{1i} = \mathbf{\Lambda}_{2i} = \mathbf{I}$ hence $\mathbf{K}_i = \begin{pmatrix} \mathbf{I} & \rho_\epsilon \mathbf{I} \\ \rho_\epsilon \mathbf{I} & \mathbf{I} \end{pmatrix} = G \otimes \mathbf{I}$ where $\mathbf{G} = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix}$ and $\otimes$ is the Kronecker product.

Here, $\rho_\epsilon = \tanh \tau$ and we first estimate the unconstrained parameter τ and then estimate the contemporaneous correlation via $\hat{\rho}_\epsilon = \tanh \hat{\tau}$. An approximate $100(1-\alpha)\%$ Wald confidence interval for ρ_ϵ can be obtained as $(\tanh L, \tanh U)$ where (L, U) are the endpoints of the interval $\hat{\tau} \pm z_{1-\alpha/2}\mathrm{se}(\hat{\tau})$.

The decomposition of $\mathbf{R}_i$ into a variance structure component $\mathbf{V}_i$ and the correlation structure component $\mathbf{K}_i$ is convenient computationally. Note that the structure of $\mathbf{V}_i$ implies we estimate different variances for the two response types. This can be accomplished by `varIdent` variance function in the **nlme** library. The **nlme** library also allows new correlation structure (`corStruct`) classes, representing user-defined correlation structures to be added to the set of standard `corStruct` classes in the **nlme** library that can be used for modeling. Taking advantage of this facility, we can write new `corStruct` classes to represent the correlation structure component $\mathbf{K}_i$. For instance, if we assume the correlation within each response is autoregressive of order 1 (AR(1)), then in the case of 3 repeated measures, the correlation structure component $\mathbf{K}_i$ in the bivariate model has the following structure:

$$\begin{pmatrix} 1 & \phi_1 & \phi_1^2 & \rho_\epsilon & 0 & 0 \\ \phi_1 & 1 & \phi_1 & 0 & \rho_\epsilon & 0 \\ \phi_1^2 & \phi_1 & 1 & 0 & 0 & \rho_\epsilon \\ \rho_\epsilon & 0 & 0 & 1 & \phi_2 & \phi_2^2 \\ 0 & \rho_\epsilon & 0 & \phi_2 & 1 & \phi_2 \\ 0 & 0 & \rho_\epsilon & \phi_2^2 & \phi_2 & 1 \end{pmatrix}, \tag{3.5}$$

where ϕ_1 and ϕ_2 represents the AR(1) correlation parameter for the first and second response respectively, and ρ_ϵ represent the contemporaneous correlation between the two responses. Similarly suppose the correlation structure within each response is compound symmetry, the correlation structure component $\mathbf{K}_i$ in the bivariate model would the have the following structure:

$$\begin{pmatrix} 1 & \phi_1 & \phi_1 & \rho_\epsilon & 0 & 0 \\ \phi_1 & 1 & \phi_1 & 0 & \rho_\epsilon & 0 \\ \phi_1 & \phi_1 & 1 & 0 & 0 & \rho_\epsilon \\ \rho_\epsilon & 0 & 0 & 1 & \phi_2 & \phi_2 \\ 0 & \rho_\epsilon & 0 & \phi_2 & 1 & \phi_2 \\ 0 & 0 & \rho_\epsilon & \phi_2 & \phi_2 & 1 \end{pmatrix}, \tag{3.6}$$

where ϕ_1 and ϕ_2 specifies the intraclass correlation parameter for compound symmetry structure for the first and second response, respectively. In a simple case where for each response the within-group error is spherical, the correlation structure component $\mathbf{K}_i$ in the bivariate model would have the following structure:

$$\begin{pmatrix} 1 & 0 & 0 & \rho_\epsilon & 0 & 0 \\ 0 & 1 & 0 & 0 & \rho_\epsilon & 0 \\ 0 & 0 & 1 & 0 & 0 & \rho_\epsilon \\ \rho_\epsilon & 0 & 0 & 1 & 0 & 0 \\ 0 & \rho_\epsilon & 0 & 0 & 1 & 0 \\ 0 & 0 & \rho_\epsilon & 0 & 0 & 1 \end{pmatrix}. \tag{3.7}$$

The standard errors of the fixed effects parameters were obtained based on the observed information using Louis' (1982) formula. To obtain the standard error of the contemporaneous correlation, ρ_ϵ, in the final model, we use the property that in large samples from regular models for which the log-likelihood is quadratic in the parameters, likelihood ratio and the Wald test of significance of individual parameters are asymptotically equivalent (Dietz and Böhning, 1995; Aitkin, 1999), so that the deviance change in omitting ρ_ϵ is approximately equal to the square of the t-statistic. Thus

$$\left(\frac{\hat{\rho}_\epsilon}{\mathrm{se}(\hat{\rho}_\epsilon)}\right)^2 \approx 2(\ell_{full} - \ell_{reduce}) \implies \mathrm{se}(\hat{\rho}_\epsilon) \approx \frac{\mathrm{abs}(\hat{\rho}_\epsilon)}{\sqrt{2(\ell_{full} - \ell_{reduce})}} \tag{3.8}$$

where ℓ_{full} and ℓ_{reduce} is the log-likelihood of the model with and without ρ_ϵ respectively. We can repeat this process (fitting a full and reduce model) for any variable for which assessing the significance of the corresponding parameter is of interest.

The approach to fitting a BLMEM emphasized here is to stack the response vectors and fit the model as a univariate LMEM. By doing so all of the univariate LMEM methodology for censored and/or missing data presented in Tuglo and Hall (2014) applies immediately to the bivariate case. Thus we can perform inference on the contemporaneous correlation between the bivariate response while accounting for a variety of complex data features such as censoring, missingness, heteroscedasticity, and autocorrelation. Conversely, we may instead be primarily interested in other features of the model such as the mean response, variance components, or autocorrelation parameters which, in this formulation, may be inferred while accounting for the contemporaneous correlation typical of multivariate data.

3.3 EXAMPLE

The example with which we will illustrate our model fitting methodology is the cookstove smoke data set mentioned in the introduction. In this study 19 households were conveniently selected and the subjects were measured for cookstove related $PM_{2.5}$ and CO exposures in real time. The original data set contains N = 11,242 pairs of observation. Overall 11.3% and 65.9% of the $PM_{2.5}$ and CO are left censored respectively. 0.6% of $PM_{2.5}$ are right-censored and no CO observations are right-censored. Approximately 90% - 100% of the CO response for 13 subjects were left-censored.

In order to ensure sufficient available information, we use only a small sub-sample of the original data for our model fitting. Since we are interested in contemporaneous correlation, the sub-sample data set is selected in such way that subjects do not have an entire response censored and since 13 subjects have approximately 90%-100% of the entire CO response censored, these 13 subjects are not included in the sub-sample data. For the remaining 6

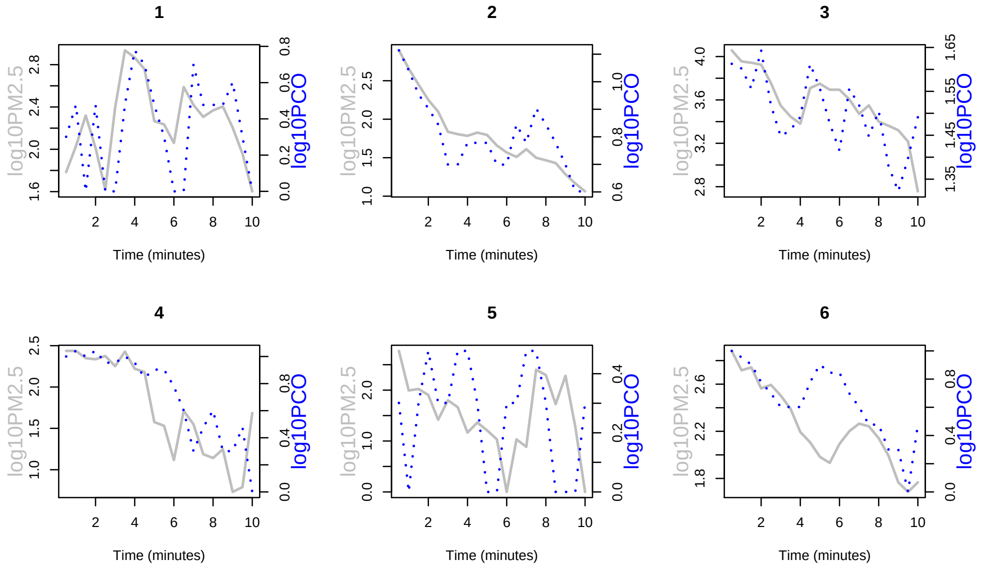

Figure 3.1: Trajectory of the $\log_{10}$ $PM_{2.5}$ and $\log_{10}$ CO concentration of the subjects.

subjects, the sub-sample is selected such that within a specific time window of the day, the 6 subjects are preparing their meals and that subjects do not have all observation censored within that time window. There are 20 pairs of observation per subject resulting in N = 120 pairs observations in our sub-sample data, and 1.6% and 13.3% of the response $PM_{2.5}$ and CO are left-censored respectively.

In estimating the contemporaneous correlation between the CO and $PM_{2.5}$ in our cookstove smoke data, we assumed the correlation within each response type is AR(1) so we wrote a `corStruct` class called `BivarCorAR1` which represent the correlation structure of the form (3.5) for the bivariate model. Since is it of interest to test $H_0 : \rho_\epsilon = 0$, we wrote another `corStruct` class called `BivarCorAR1Diag`. This `corStruct` is identical to `BivarCorAR1` but with $\rho_\epsilon = 0$. The responses, CO and $PM_{2.5}$ are strongly skewed to the right so both are log-transformed, which also leads to a log-transformed detection limit of $\log_{10}(1) = 0$.

Let y_{hij} be the response from the ith subject, $i = 1, \ldots m = 6$ taken at distinct times t_{ij} for $j = 1, \ldots, n_i = 20$ for each subject for outcome $h = 1, 2$. The model we considered can be written as

$$y_{hij} \;=\; \beta_{0h} + \beta_{1h} t_{ij} + b_{hi} + \epsilon_{hij}, \tag{3.9}$$

$$b_{hi} = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N(0, \mathbf{D}),\ \boldsymbol{\epsilon}_{ij} = \begin{pmatrix} \boldsymbol{\epsilon}_{1ij} \\ \boldsymbol{\epsilon}_{2ij} \end{pmatrix} \sim N(0, \mathbf{R}_{ij}),$$

$$\mathbf{D} = \begin{pmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{12}} & \sigma^2_{b_2} \end{pmatrix},\ \mathbf{R}_{ij} = \begin{pmatrix} \sigma^2_{\epsilon_1} \Lambda_{1ij}(\phi_1) & \sigma_{\epsilon_{12}} \mathbf{I} \\ \sigma_{\epsilon_{12}} \mathbf{I} & \sigma^2_{\epsilon_2} \Lambda_{2ij}(\phi_2) \end{pmatrix}.$$

In model (3.9), we can see that β_{0h} and β_{1h} are the fixed effects for intercept and slope for each of the two response type, and random effect b_{hi} for each subject (household). The contemporaneous correlation, ρ_ϵ is captured through the error variance-covariance $\mathbf{R}_{ij}$. Model (3.9) is considered appropriate because of the the defining features of the sub-sampled data set. In particular, the degree of within-subject correlation/between subject heterogeneity is expected to differ across response variables, hence the different subject variance components across h. In addition, due to oscilating patterns in the plots in Figure 1, we considered three different correlation structures for the data set.

We started by fitting model (3.9) with the correlation structure `BivarCorIdent`, representing the correlation structure of the form (3.7). According to results in Table 3.1, the estimate of the contemporaneous correlation is $\hat{\rho}_\epsilon = 0.20$, indicating that the two response variables, CO and $\text{PM}_{2.5}$, go up and down together with a moderate degree of positive association. As expected, we find substantially different variance components at the subject level across the two responses, and that variance from subject to subject appears substantial for the two responses.

The plot of the residuals against time and the lowess smoother fit to the residuals shows a relatively good fit (Figure 3.2), but it suggests the model does not fully capture the time

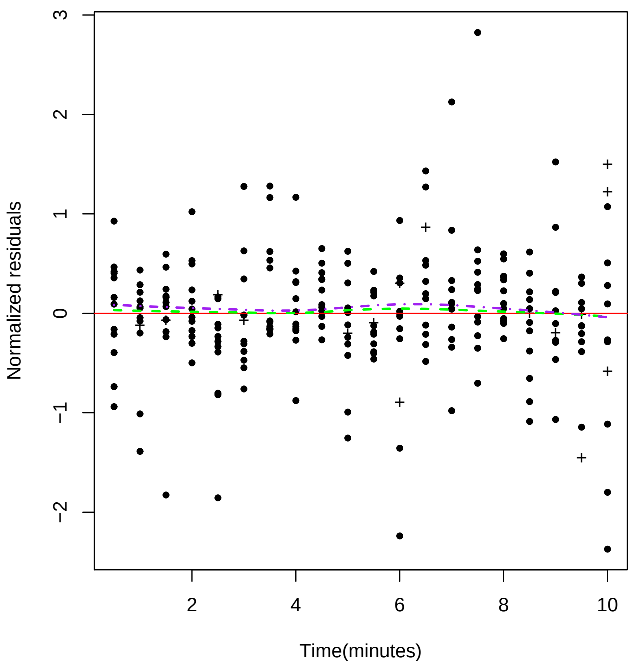

Figure 3.2: Lowess smoother fit to the normalized residuals from (i) Model 1 (-.-); (iii) Model 3 (- -). The residuals from model (iii) appear as points; the left-censored residuals appear as "+"

trend in the data. The strong oscillating pattern in the plot in Figure 3.1 also suggests we consider a correlation structure that accounts for serial correlation above the random effect term in the model. Therefore, we re-fit model (3.9) with the `BivarCorAR1` correlation structure and conduct a likelihood ratio test (LRT). The LRT statistic is 2(-45.28 + 86.13) = 81.7 with corresponding p-values less than 0.0001 supporting the conclusion that the model with the `BivarCorAR1` correlation structure fits better than the model with the `BivarCorIdent` correlation structure. The improved fit can be seen in Figure 3.2 where the lowess smoother for the second model is more nearly flat and closer to zero. However the estimates of the contemporaneous correlation, $\hat{\rho}_\epsilon = -0.07$, suggests no or negligible correlation between CO and $PM_{2.5}$ so we test $H_0 : \rho_\epsilon = 0$ which involves fitting a reduced model with the `BivarCorAR1Diag` correlation structure. The resulting LRT statistic is 2(-45.28 + 45.78) = 1.00 with corresponding p-value 0.3173, which provides little evidence against H_0 so we conclude ρ_ϵ is not statistically significant.

Among the models considered, the LRT and the diagnostics plots support Model 3 (the model with correlation structure, `BivarCorAR1Diag`) to be the best model for the data. This result supports the conclusion that after controlling for serial autocorrelation within each response, and household level heterogeneity, there is negligible contemporaneous correlation between $\log_{10}$ $PM_{2.5}$ and $\log_{10}$ CO.

Table 3.1: Parameter estimates of the three models fit to the cookstove smoke data. The three models differ by the assumed within-household correlation structure.

	Model 1		Model 2		Model 3	
	Estimates	SE	Estimates	SE	Estimates	SE
β_{01}	2.72	0.28	2.81	0.30	2.81	0.30
β_{11}	-0.05	0.01	-0.06	0.01	-0.06	0.01
β_{01}	0.90	0.17	0.91	0.18	0.91	0.18
β_{11}	-0.02	3.6×10^{-3}	-0.02	5.7×10^{-3}	-0.02	5.7×10^{-3}
σ_{b_1}	0.67		0.62		0.62	
σ_{b_2}	0.41		0.41		0.41	
ρ_b	0.82		0.94		0.93	
σ_{ϵ_1}	0.40		0.47		0.48	
σ_{ϵ_2}	0.21		0.22		0.22	
ρ_ϵ	0.20	0.09	-0.07	0.07		
ϕ_1			0.72		0.72	
ϕ_2			0.56		0.56	
ρ	0.67		0.63		0.65	
logLik	-86.13		-45.28		-45.78	

In addition to accounting for serial correlation and household level heterogeneity, Model 1-3 account for simple linear time trends in each response. For example, according to Model 3, the average $\log_{10}$ CO and $\log_{10}$ $\text{PM}_{2.5}$ concentration at time zero is $\hat{\beta}_{01} = 2.81$(95% CI: 2.22–3.40) and $\hat{\beta}_{02} = 0.91$(95% CI: 0.56–1.26) respectively. The average slope was negative, $\hat{\beta}_{11} = -0.05$ and $\hat{\beta}_{12} = -0.02$ between the time of the day 4:58 PM to 5:08 PM with 95% CI (-0.08, -0.04) and (-0.03, -0.01) respectively. As expected, we find substantially different variance components at the subject level across the responses $\text{PM}_{2.5}$ and CO, and that standard deviation from subject to subject are $\hat{\sigma}_{b_1} = 0.07$, and $\hat{\sigma}_{b_2} = 0.15$ respectively. The within-subject standard deviations are 0.48 and 0.22 and the estimates of the correlation parameters are $\hat{\phi_1} = 0.72$, $\hat{\phi_2} = 0.56$ for $\log_{10}$ $\text{PM}_{2.5}$ and $\log_{10}$ CO respectively. According to the final model, overall, there is a strong positive correlation between $\log_{10}$ $\text{PM}_{2.5}$ and $\log_{10}$ CO, $\hat{\rho} = 0.65$, but this correlation is due to the subject-specific random variation.

3.4 SIMULATION STUDY

To assess the performance of our proposed algorithm, two simulations are conducted. The goal of the simulation is to demonstrate ML and REML estimation of BLMEMs with censored response in spherical and non-spherical error settings and examine the finite sample properties of the resulting estimators and inference methods for the contemporaneous correlation coefficient. The first part of the simulation focuses on the case where the within-group error for each response is spherical. Here we consider the bivariate random-intercept linear model considered by Xie et al. (2012) and replicate their results. Part II of the simulation study focuses on a wider class of BLMEMs with non-spherical within-group errors (heteroscedasticity and/or serial correlation) for the two responses.

To separate the effects of censoring from small sample inaccuracies of ML and REML in linear mixed models, we also include results from the complete data LME in both cases. Comparison of the three methods (complete data LME, our proposed closed-form EM with pseudo data augmentation, and the MCEM approach by Xie et al) is based on average estimates of the parameters, relative mean square error (RelMSE), mean relative bias (RelBias) and coverage of 95% Wald confidence intervals $\hat{Pr}(\psi \in \hat{I})$ (CovProb) for a given parameter $\boldsymbol{\psi}$. The RelBias and RelMSE are defined as

$$\text{RelBias}(\boldsymbol{\psi}) = \frac{1}{R}\sum_{i=1}^{R}(\hat{\boldsymbol{\psi}}^{(i)} - \boldsymbol{\psi})/\boldsymbol{\psi} \quad \text{and} \quad \text{RelMSE}(\boldsymbol{\psi}) = \frac{1}{R}\sum_{i=1}^{R}(\hat{\boldsymbol{\psi}}^{(i)} - \boldsymbol{\psi})^2/\boldsymbol{\psi}^2$$

where $\hat{\psi}^{(i)}$ is the estimate of $\boldsymbol{\psi}$ from the ith data set. For simplicity we will refer to the three algorithms, complete data LME, our proposed closed-form EM with pseudo data augmentation and MCEM approach compared in this paper by `lme`, `lmecm` and `censre4c` respectively which are the R functions that are used to fit the models.

3.4.1 Bivariate Random Intercept Model

Let y_{hij} be the response from the ith subject, $i = 1, \dots m$ taken at distinct times t_{ij} for $j = 1, \dots, n_i = 4$ for each subject for outcome $h = 1, 2$, the model we considered can be written as

$$y_{hij} = \beta_{0h} + b_{hi} + \epsilon_{hij}, \tag{3.10}$$

$$b_{hi} = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N(0, \mathbf{D}),\ \boldsymbol{\epsilon}_{ij} = \begin{pmatrix} \boldsymbol{\epsilon}_{1ij} \\ \boldsymbol{\epsilon}_{2ij} \end{pmatrix} \sim N(0, \mathbf{R}_{ij}),$$

$$\mathbf{D} = \begin{pmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{12}} & \sigma^2_{b_2} \end{pmatrix},\ \mathbf{R}_{ij} = \begin{pmatrix} \sigma^2_{\epsilon_1} & \sigma_{\epsilon_{12}} \\ \sigma_{\epsilon_{12}} & \sigma^2_{\epsilon_2} \end{pmatrix} \otimes \mathbf{I}_{n_i}.$$

where β_{0h} is the regression coefficients for the response $h = 1, 2$. Six sets of simulations were conducted based on model (3.10) with different censoring proportions and sample sizes. For comparison purposes, the model parameters and censoring proportion were set to the same values as in Xie et al. (2012), paper. The model parameters were set as follows: $(\beta_1, \beta_2) = (1.2, 2.0)$, $\mathbf{D} = \begin{pmatrix} 2.0 & 1.3 \\ 1.3 & 1.5 \end{pmatrix}$, $\mathbf{R}_{ij} = \begin{pmatrix} 2.3 & 0.5 \\ 0.5 & 0.9 \end{pmatrix} \otimes \mathbf{I}$.

It follows that $\rho_b = 0.75$, $\rho_e = 0.36$ and $\rho = 0.56$. Three sets of censoring proportions for the first and second response were assumed: (0.20, 0.10), (0.35, 0.30) and (0.60, 0.50).

Table 3.2: ML Estimation. Comparison of three methods: Complete data LME (`lme`), our proposed closed-form EM with pseudo data augmentation (`lmecm`) and MCEM approach (`censre4c`) in simulation studies (number of subjects, $m = 100$)

			β_{01}	β_{02}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ρ
Cens.Prop[†]	Criterion[⋆]	Method	1.2[‡]	2.0	2.0	1.5	0.75	2.3	0.9	0.35	0.56
(0%,0%)	Est	`lme`	1.20	2.00	1.97	1.47	0.75	2.29	0.90	0.34	0.56
	RelBias		0.1	-0.1	-1.4	1.7	-0.2	-0.6	0.0	-1.2	-0.8
	RelMSE		1.7	0.4	3.1	2.6	0.7	0.7	0.7	2.1	0.7
	CovProb		97.6	95.2						95.8	
(20%,10%)	Est	`censre4c`	1.20	2.00	1.96	1.48	0.76	2.31	0.89	0.35	0.56
		`lmecm`	1.19	2.00	1.98	1.46	0.75	2.31	0.89	0.35	0.56
	RelBias	`censre4c`	0.2	-0.1	-1.9	-1.0	0.8	0.5	-0.6	-0.6	-0.1
		`lmecm`	-0.8	-0.1	-1.1	-2.4	-0.2	0.6	-0.8	-0.2	-0.7
	RelMSE	`censre4c`	1.8	0.4	4.2	3.1	0.8	1.0	0.8	2.9	0.6
		`lmecm`	2.0	0.4	3.7	2.8	0.8	0.9	0.8	2.4	0.7
	CovProb	`censre4c`	93.4	96.2						94.8	
		`lmecm`	93.3	95.1						96.8	
(35%,30%)	Est	`censre4c`	1.20	2.00	1.96	1.48	0.76	2.31	0.90	0.35	0.56
		`lmecm`	1.20	2.00	1.98	1.46	0.75	2.29	0.89	0.33	0.55
	RelBias	`censre4c`	0.4	0.0	-2.0	-1.3	1.3	0.3	-0.4	-1.1	-0.1
		`lmecm`	0.1	0.0	-1.1	-2.4	-0.2	-0.3	-0.7	-3.7	-1.6
	RelMSE	`censre4c`	2.0	0.4	4.7	3.5	0.9	1.3	1.1	3.6	0.7
		`lmecm`	1.7	0.5	3.9	3.2	0.8	1.0	1.0	2.8	0.7
	CovProb	`censre4c`	92.6	96.2						93.6	
		`lmecm`	95.8	95.2						95.7	
(60%,50%)	Est	`censre4c`	1.23	1.99	1.85	1.50	0.78	2.29	0.90	0.34	0.56
		`lmecm`	1.19	2.00	1.95	1.45	0.75	2.34	0.89	0.30	0.54
	RelBias	`censre4c`	2.5	-0.3	-7.6	0.2	3.7	-0.3	0.1	-2.5	0.0
		`lmecm`	-0.7	0.0	-2.6	-3.1	0.1	1.9	-0.6	-12.4	-4.5
	RelMSE	`censre4c`	2.8	0.5	7.5	4.4	1.5	2.3	1.7	5.7	0.9
		`lmecm`	2.5	0.5	5.6	3.9	1.0	1.9	1.4	4.9	1.0
	CovProb	`censre4c`	93.2	96.2						92.8	
		`lmecm`	91.2	94.8						93.2	

[†] Cens.Prop.,censoring proportions (left-censrong) for response 1 and 2

[⋆] RelBias(%), RelMSE(%) and CovProb(%)

[‡] The numbers in the header row are true parameter values

We generate the data based on model (3.10) with the true parameter values and censoring proportions above for the case where $m = 100$ and $m = 300$. Both ML and REML estimation were used but because the MCEM approach of Xie et al. (2012) only implements ML estimation, we present here the results for ML estimation for the purpose of comparison. REML results were similar and are included in the appendix. Comparison with the substitution methods and the non-parametric approach of using the rank are also included in the appendix.

Table 3.2 presents the average estimates for the parameters, RelBias, RelMSE for all the estimated parameters and the coverage probability of 95% Wald confidence intervals (CovProb) for the fixed effects and the contemporaneous correlation, ρ_ϵ. The three algorithms: `lme`, `lmecm`, and `censre4c` performed similarly. The average estimates of the parameters were very close to each other and close to the true value except the variance components of the random effect $\sigma^2_{b_1}$ and the contemporaneous correlation ρ_ϵ in case of high censoring proportion, (0.60, 0.50). Compared to `lme`, with high censoring proportion, (0.60, 0.50), `censre4c` underestimated $\sigma^2_{b_1}$ with bias of -7.6 and `lmecm` underestimated ρ_ϵ with bias -12.4. The relative MSE is similar for the censored data algorithms (`lmecm` and `censre4c`); however, as expected, both had larger relative MSE compared to `lme`. All three methods perform very well in terms of coverage probability. The coverage probabilities of `lmecm` for $\beta_1, \beta_2, \rho_\epsilon$ vary around the target 95% with minimum coverage probability of 91.2% for β_1 when the censored proportion is (0.60, 0.50) and maximum coverage probability of 97.8% for β_1 when the censored proportion is ((0.20, 0.10).

With a larger sample size ($m = 300$), the average parameter estimates from the three methods remain similar to those observed with the smaller sample size. As expected the relative bias and the relative MSE is smaller compared to the results from the smaller sample size for the three methods. The coverage probability is also similar.

Overall the the two censoring based methods, `lmecm` and `censre4c` perform similar in terms average parameter estimates, relative bias, relative MSE and coverage probability across all value of censoring proportions and sample size.

Table 3.3: ML Estimation. Comparison of three methods: Complete data LME (`lme`), our proposed closed-form EM with pseudo data augmentation (`lmecm`) and MCEM approach (`censre4c`) in simulation studies (number of subjects, $m = 300$)

			β_{01}	β_{02}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ρ
Cens.Prop	Criterion	Method	1.2	2.0	2.0	1.5	0.75	2.3	0.9	0.35	0.56
(0%,0%)	Est	`lme`	1.20	2.00	2.01	1.51	0.75	2.30	0.90	0.35	0.56
	RelBias		0.0	0.1	0.3	0.6	0.4	0.1	0.2	-0.1	0.2
	RelMSE		0.7	0.2	1.3	1.0	0.1	0.3	0.2	0.8	0.2
	CovProb		96.1	94.2						95.8	
(20%,10%)	Est	`censre4c`	1.20	2.00	1.97	1.50	0.75	2.30	0.90	0.35	0.56
		`lmecm`	1.20	2.00	2.01	1.51	0.75	2.31	0.90	0.35	0.56
	RelBias	`censre4c`	0.4	-0.1	-1.3	-0.3	0.2	0.0	0.0	0.3	0.0
		`lmecm`	-0.1	0.1	0.4	0.6	0.3	0.2	0.2	-0.7	0.0
	RelMSE	`censre4c`	0.7	0.2	1.4	1.1	0.3	0.3	0.3	0.9	0.2
		`lmecm`	0.7	0.2	1.3	1.0	0.2	0.3	0.2	0.8	0.2
	CovProb	`censre4c`	91.6	93.4						96.2	
		`lmecm`	95.2	94.8						96.8	
(35%,30%)	Est	`censre4c`	1.21	2.00	1.98	1.50	0.75	2.30	0.90	0.35	0.56
		`lmecm`	1.20	2.00	2.00	1.50	0.75	2.32	0.89	0.34	0.55
	RelBias	`censre4c`	0.4	-0.1	-1.1	0.0	0.2	-0.1	0.1	0.4	0.0
		`lmecm`	-0.1	0.2	-0.2	0.0	-0.3	0.8	-0.6	-2.4	-1.1
	RelMSE	`censre4c`	0.8	0.2	1.6	1.2	0.3	0.4	0.4	1.0	0.3
		`lmecm`	0.7	0.2	1.4	1.0	0.2	0.3	0.4	0.9	0.2
	CovProb	`censre4c`	93.0	92.8						94.8	
		`lmecm`	95.0	94.9						97.2	
(60%,50%)	Est	`censre4c`	1.21	2.00	1.93	1.50	0.76	2.29	0.90	0.35	0.56
		`lmecm`	1.20	2.00	1.99	1.47	0.75	2.33	0.90	0.30	0.54
	RelBias	`censre4c`	1.2	-0.2	-3.5	0.3	1.2	-0.4	0.2	-0.6	0.1
		`lmecm`	-0.3	-0.1	-0.4	-2.1	0.1	1.3	-0.3	-12.6	-3.9
	RelMSE	`censre4c`	1.0	0.2	2.3	1.5	0.4	0.7	0.5	1.4	0.3
		`lmecm`	0.7	0.2	2.0	1.4	0.3	0.6	0.4	2.5	0.4
	CovProb	`censre4c`	93.2	91.8						96.8	
		`lmecm`	95.2	93.6						99.4	

3.4.2 BIVARIATE RANDOM INTERCEPT AND SLOPE WITH AR(1) ERROR

For the second simulation, we consider model (3.9) with the within-group correlation `BivarCorAR1` but with $m = 300$ and $n_i = 4$. The data are generated with the following true parameters:

$$\begin{aligned}(\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}) &= (2.4, 1.3, 4.3, 1.5) \\ \begin{pmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{12}} & \sigma^2_{b_2} \end{pmatrix} &= \begin{pmatrix} 3.2 & 1.9 \\ 1.9 & 1.7 \end{pmatrix}\end{aligned}$$

$$\begin{pmatrix} \sigma^2_{\epsilon_1} & \sigma_{\epsilon 12} \\ \sigma_{\epsilon_{12}} & \sigma^2_{\epsilon_2} \end{pmatrix} = \begin{pmatrix} 2.5 & 0.7 \\ 0.7 & 1.2 \end{pmatrix}$$

$$(\phi_1, \phi_2) = (0.6, 0.2).$$

It follows that $\rho_b = 0.81$, $\rho_e = 0.40$ and $\rho = 0.64$. Table 3.4 presents the average estimates for the parameters, RelBias, RelMSE for all the estimated parameters and the coverage probability for the fixed effects and the contemporaneous correlation, ρ_ϵ. Two algorithms, `lmecm` and `lme` for the complete data were compared in this section.

Table 3.4: ML Estimation. Comparison of two methods: Complete data LME (`lme`), and our proposed closed-form EM with pseudo data augmentation (`lmecm`) in simulation studies (number of subjects, $m = 300$)

	Method													
		β_{01}	β_{11}	β_{02}	β_{12}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ϕ_1	ϕ_2	ρ
Cens.Prop	Criterion	2.40	1.30	4.30	1.50	3.20	1.70	0.81	2.50	1.20	0.40	0.60	0.20	0.64
	`lme`													
(0%,0%)	Est	2.39	1.30	4.30	1.50	3.17	1.69	0.82	2.51	1.20	0.41	0.60	0.20	0.64
	RelBias	-0.4	0.0	-0.1	0.0	-0.8	-0.4	0.6	0.3	-0.2	0.5	-0.4	-0.3	0.1
	RelMSE	0.4	0.1	0.1	0.0	2.0	1.0	0.4	1.2	0.3	0.6	0.5	3.8	0.2
	CovProb	95.6	95.6	94.0	94.2						95.0			
	`lmecm`													
(20%,10%)	Est	2.38	1.30	4.30	1.50	3.16	1.69	0.82	2.59	1.18	0.40	0.59	0.20	0.64
	RelBias	-0.8	0.1	0.0	0.0	-1.4	-0.5	1.3	3.4	-1.8	-2.2	-1.0	1.6	-0.6
	RelMSE	0.4	0.1	0.1	0.0	2.3	1.1	0.5	2.0	0.4	0.8	0.7	4.7	0.2
	CovProb	95.6	94.2	96.6	95.8						94.8			
(35%,30%)	Est	2.34	1.31	4.33	1.49	3.28	1.66	0.81	2.58	1.13	0.38	0.55	0.21	0.62
	RelBias	-2.5	1.0	0.8	-0.5	2.6	-2.5	-1.1	3.1	-5.8	-5.0	-8.6	5.7	-2.4
	RelMSE	0.6	0.1	0.1	0.1	2.3	1.4	0.5	1.7	0.8	1.0	1.5	6.9	0.2
	CovProb	94.0	94.0	93.4	94.4						93.4			
(60%,50%)	Est	2.18	1.34	4.36	1.48	3.44	1.65	0.81	2.89	1.11	0.32	0.48	0.23	0.59
	RelBias	-9.0	3.2	1.4	-1.0	7.6	-3.1	-1.1	15.5	-7.6	-20.6	-19.4	13.3	-7.0
	RelMSE	1.8	0.3	0.2	0.1	3.0	1.6	0.4	4.4	1.4	4.9	4.7	14.1	0.7
	CovProb	86.6	93.8	90.8	92.8						53.2			

For low (20%, 10%) to medium (35%, 30%) censoring proportions, `lmecm` performed very similar to the complete data `lme`. The average parameter estimates were very close to each other and both were very close to the true parameter. Compared to `lme`, `lmecm` underestimated ϕ_1 with relative bias of -8.6 in the case of medium censoring proportion. The coverage probability with `lmecm` for the fixed effect parameters and ρ_ϵ in the case of low to medium censoring proportion generally vary around 95% with minimum 93.4% for ρ_ϵ and a maximum of 96.6% for β_{02}. As expected, the relative MSE is larger for the censored algorithm, `lmecm`, compared to the complete data `lme` for the three censoring proportions.

As the censoring proportion increases from medium to high (60%,50%), the performance of `lmecm` is detrimentally affected, but it still remains within acceptable range except for the parameter ρ_ϵ. For example, the relative mean bias for ρ_ϵ increase from -5% to -20.6% with the corresponding coverage probability decreasing from 93.4% to 53.2%. Compared to `lme`, at the highest censoring proportion, `lmecm` overestimates σ_{ϵ_1} with relative bias 15.5% and underestimates ϕ_1 with relative bias -19.9%.

Overall, `lmecm`, performed very well in terms of average parameter estimates, relative bias, relative MSE and coverage probability when compared to complete data `lme`, in case of low to medium censoring proportion. The performance of `lmecm` declined in case of high censoring proportion but still remains within acceptable range except for the parameter ρ_ϵ.

3.5 Conclusion

In this paper we proposed a computationally feasible EM algorithms for BLMEMs with censored response that uses closed form expressions at the E-step and serves as an alternative to the MCEM proposed by Xie et al. (2012). Our implementation is flexible and can fit BLMEMs with censored response with more complex random effect structure (e.g, random effect for intercept and slope, nested random effects), complex error covariance structures that accounts for all kinds of heteroscedasticity and/or serial correlation (e.g. AR(1), compound symmetry structure), and easily extends to the case of more than two response variables.

We demonstrated our approach can fit wider class of BLMEM with censored response and have shown in our data analysis the effect of misspecification of the measurement error covariance structure. For example, in Table 3.1, we would have concluded there is a moderate contemporaneous correlation between CO and $\text{PM}_{2.5}$ had we not accounted for the possibility of the serial correlation within each response. Consequently, this would have resulted in estimates of the population parameters being wrong as well.

The issues of how the degree of censoring affects estimation and inference has been addressed in our simulation study but it is also of interest to investigate the effects of different patterns of censoring on the parameter estimates. We leave this as a future work.

The issue of misspecification with regards to distribution assumptions have been addressed by Vock et al. (2011) and Bandyopadhyay et al. (2012) but the focus is on univariate response. It is of interest to extend this to the case of BLMEM with censored response. Another potential area of research is assessing the goodness of fit of a BLMEM with censored response which can be challenging considering the complex covariance structures for bivariate repeated measures.

3.6 References

Aitkin, M. (1999). A General maximum likelihood analysis of variance components in the generalized linear models. *Biomatrics 55*, 117–128.

Bandyopadhyay, D., V. Lachos, M. Castro, and D. Dey (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal 54*, 405–425.

Bland, J. and D. Altman (1994). Correlation, regression, and repeated data. *British Medical Journal 1994*, 308:896.

Dietz, E. and D. Böhning (1995). Statistical inference based on a general model of unobserved heterogeneity. *In Statistical Modelling*. B. J. Francis, R. Hatzinger, G. U. H. Seeber, and G. Steckel-Berger (eds). New York: Springer-Verlag.

Gao, F., P. Thompson, C. Xiong, and J. Miller (2009). Analyzing multivariate longitudinal data using SAS. *Proceedings of the Thirty-first Annual SAS Users Group International Conference, SAS Institute Inc.,2009*, 187–31.

Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Preceedings of the Twenty-Ninth Annual SAS Users Group International Conference, SAS Institute Inc.,2004*, 198–29:1–7.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*, 625–629.

Lau, B. and S. Gange (2004). Methods for the analysis of continuous biomaker assay data with increased sensitivity. *Epidemiology 15*, 724–732.

Liu, H. and W. Tu (2012). A Semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. *Annals of Applied Statistics 6*, 1861–1882.

Lorenz, D., S. Datta, and S. Harkema (2011). Marginal association measure for clustered data. *Statistics in Medicine 30*, 3181–3191.

Louis, T. (1982). Finding the oberserved information matrix when using EM algorithm. *Journal of Royal Statistical Society, Ser. B 44*, 226–233.

Lynn, H. (2001). Maximum likelihood inference for left-censored HIV RNA data. *Statistics in medicine 20*, 33–45.

Thiebaut, R., H.Jacqmin-Gadda, C. Leport, C. Katlama, D. Costagliola, V. Le Moing, P. Morlat, G. Chene, and the APROCO Study Group (2000). Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left-censoring of HIV RNA measures. *Biostatistics 13*, 271–282.

Thiebaut, R., H. Jacqmin-Gadda, G. Chene, C. Leport, and D. Commenges (2002). Bivariate linear mixed models using SAS PROC MIXED. *Computer Methods and Programs in Biomedicine 69*, 249–256.

Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybid EM for linear and nonlinear mixed effects model with censored response. *Computational Statistics and Data Analysis 51*, 5718–5730.

Vaida, F. and L. Liu (2009). *lmec: Linear mixed effects models with censored responses.* R package version 1.0.

Vock, D. M., M. Davidian, and A. A. Tsiatis (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics 0*, 1–13.

Wu, L. (2010). *Mixed effects models for complex data.* Chapman and Hall CRC Boca Raton FL.

Xie, X., X. Xue, S. Gange, H. Strickler, M. Kim, and W. H. Study-Group (2012). Estimation and inference on correlations between biomarkers with repeated measures and left-censoring due to minimum detection levels. *Statistics in Medicine 31*(21), 2275–2289.

3.7 Appendix

3.7.1 Additional Simulation Results

Simulation Study: REML Estimation

We present here the simulation results for REML estimation of BLMEMs with censored response. Table 3.5 and 3.6 presents REML results for the simulation studies base on model (3.10) for the case where $m = 100$ and $m = 300$ respectively. As indicated earlier, this results are similar to the ML results in Table 3.2 and 3.3 respectively. Note that Table 3.5 and 3.6

does not include the REML results for the MCEM approach because the MCEM approach of Xie et al. (2012) only implements ML estimation. The REML results for the case where the within-group errors are non-spherical for each response is presented in Table 3.7 and are similar to the ML results in Table 3.4.

Table 3.5: REML Estimation. Comparison of two methods: Complete data LME (`lme`), and our proposed closed-form EM with pseudo data augmentation (`lmecm`) in simulation studies for the case where the within-group error is spherical for each response (number of subjects, $m = 100$)

			β_{01}	β_{02}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ρ
Cens.Prop†	Criterion*	Method	1.2‡	2.0	2.0	1.5	0.75	2.3	0.9	0.35	0.56
(0%, 0%)	Est	`lme`	1.20	2.00	1.98	1.49	0.75	2.30	0.90	0.35	0.56
	RelBias		0.3	-0.2	-1.0	-0.5	0.4	0.0	-0.2	-1.3	-0.2
	RelMSE		1.7	0.4	3.4	2.6	0.7	0.6	0.6	2.2	0.6
	CovProb		95.1	95.8						95.3	
(20%, 10%)	Est	`lmecm`	1.20	2.00	1.99	1.49	0.75	2.30	0.90	0.34	0.56
	RelBias		0.2	-0.2	-0.7	-0.6	0.3	0.0	-0.3	-1.8	-0.4
	RelMSE		1.7	0.4	3.6	2.7	0.7	0.8	0.7	2.5	0.6
	CovProb		95.8	95.7						96.1	
(35%, 30%)	Est	`lmecm`	1.19	2.00	2.01	1.50	0.75	2.30	0.90	0.34	0.55
	RelBias		-0.9	-0.2	0.7	-0.1	-0.6	0.2	0.4	-3.8	-1.6
	RelMSE		2.0	0.4	4.1	3.4	0.8	1.0	0.9	3.1	0.7
	CovProb		95.3	97.7						95.1	
(60%, 50%)	Est	`lmecm`	1.19	2.00	2.01	1.49	0.75	2.35	0.90	0.31	0.54
	RelBias		-0.6	0.2	0.3	-0.8	-0.5	2.0	-0.20	-12.7	-4.4
	RelMSE		2.8	0.5	6.0	3.5	1.0	1.8	1.3	4.9	1.0
	CovProb		90.3	94.6						94.8	

Table 3.6: REML Estimation. Comparison of two methods: Complete data LME (`lme`), and our proposed closed-form EM with pseudo data augmentation (`lmecm`) in simulation studies for the case where the within-group error is spherical for each response (number of subjects, $m = 300$)

			β_{01}	β_{02}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ρ
Cens.Prop[†]	Criterion[⋆]	Method	1.2[‡]	2.0	2.0	1.5	0.75	2.3	0.9	0.35	0.56
(0%, 0%)	Est	`lme`	1.19	2.00	2.01	1.51	0.75	2.30	0.90	0.35	0.56
	RelBias		-0.7	-0.1	0.6	0.9	0.2	-0.1	-0.1	-0.4	0.3
	RelMSE		0.6	0.2	1.0	0.9	0.2	0.2	0.2	0.6	0.2
	CovProb		94.8	93.2						96.9	
(20%, 10%)	Est	`lmecm`	1.19	2.00	2.02	1.51	0.75	2.30	0.90	0.35	0.56
	RelBias		-0.7	-0.1	0.8	1.0	0.1	0.0	0.0	-1.1	0.1
	RelMSE		0.6	0.2	1.2	0.9	0.2	0.3	0.3	0.7	0.2
	CovProb		95.1	93.4						97.6	
(35%, 30%)	Est	`lmecm`	1.20	2.00	2.00	1.50	0.75	2.32	0.90	0.34	0.55
	RelBias		-0.3	0.1	0.2	0.1	-0.2	0.9	-0.2	-2.9	-0.9
	RelMSE		0.6	0.2	1.4	1.1	0.3	0.4	0.3	1.1	0.3
	CovProb		94.8	94.4						95.2	
(60%, 50%)	Est	`lmecm`	1.19	2.00	2.02	1.48	0.75	2.34	0.89	0.31	0.54
	RelBias		-0.5	0.0	1.0	-1.1	0.0	1.9	-0.7	-11.9	-3.5
	RelMSE		0.8	0.2	1.9	1.4	0.3	0.7	0.4	2.3	0.4
	CovProb		92.6	92.4						99.2	

Table 3.7: REML Estimation. Comparison of two methods: Complete data LME (`lme`), and our proposed closed-form EM with pseudo data augmentation (`lmecm`) in simulation studies (number of subjects, $m = 300$)

	Method	β_{01}	β_{11}	β_{02}	β_{12}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ϕ_1	ϕ_2	ρ
Cens.Prop	Criterion	2.40	1.30	4.30	1.50	3.20	1.70	0.81	2.50	1.20	0.40	0.60	0.20	0.64
	`lme`													
(0%,0%)	Est	2.40	1.30	4.30	1.50	3.17	1.71	0.82	2.54	1.20	0.40	0.60	0.20	0.64
	RelBias	0.0	0.1	0.0	-0.1	-1.0	0.6	0.7	1.5	0.0	-0.1	0.2	-0.4	-0.1
	RelMSE	0.4	0.1	0.1	0.0	2.0	1.2	0.4	1.3	0.3	0.5	0.5	3.7	0.2
	CovProb	95.4	96.0	93.2	95.0						96.0			
	`lmecm`													
(20%,10%)	Est	2.39	1.30	4.31	1.50	3.16	1.71	0.82	2.62	1.18	0.39	0.60	0.20	0.63
	RelBias	-0.4	0.3	0.2	-0.2	-1.3	0.4	1.2	4.8	-1.5	-2.8	-0.4	1.8	-0.8
	RelMSE	0.4	0.1	0.1	0.0	2.3	1.3	0.4	2.0	0.4	0.7	0.6	4.5	0.2
	CovProb	94.2	94.8	96.0	97.4						96.4			
(35%,30%)	Est	2.34	1.31	4.33	1.49	3.34	1.70	0.81	2.60	1.13	0.38	0.55	0.21	0.63
	RelBias	-2.4	1.0	0.7	-0.5	4.5	0.1	-0.7	3.9	-6.2	-5.1	-8.2	4.0	-2.0
	RelMSE	0.7	0.1	0.1	0.1	3.0	1.3	0.4	2.1	0.9	1.1	1.6	6.9	0.2
	CovProb	93.4	93.6	92.6	94.0						94.0			
(60%,50%)	Est	2.18	1.34	4.35	1.49	3.41	1.65	0.81	2.91	1.11	0.32	0.49	0.23	0.59
	RelBias	-9.1	3.0	1.1	-0.7	6.4	-3.1	0.8	16.4	-7.5	-21.0	-18.6	13.1	-7.2
	RelMSE	1.8	0.3	0.2	0.1	3.1	1.5	0.4	4.4	1.6	5.1	4.4	14.4	0.8
	CovProb	85.6	92.0	88.6	91.8						54.2			

Simulation Study: Substitution Methods

Here we provide simulation results to illustrate the impact of substitution methods on the parameter estimates and inferences. The data is generated from model (3.10). The substitution methods considered here are; substituting half the DL (`halfsub`), substituting $\frac{\text{DL}}{\sqrt{2}}$ (`sqrtub`) and substituting the DL (`sub`). After the substitution, the model for the complete data is fitted using `lme`.

Table 3.8 presents the average estimates for the parameters, RelBias, RelMSE for all the estimated parameters and the coverage probability of 95% Wald confidence intervals (CovProb) for the fixed effects and the contemporaneous correlation, ρ_ϵ. With a low censoring proportion (20%, 10%), the relative biases of the three substitution methods (`sub`, `halfsub`, `sqrtsub`) are orders of magnitude larger than the model based methods (`censre4c`, `lmecm`). The relative biases of the substitution methods in estimating the contemporaneous correlation, ρ_ϵ, is about eight to ten times the relative biases of the model based methods. The relative MSE for the substitution methods are relatively larger than the model based method. In particular, the relative MSE of $\sigma^2_{b_1}$ for substitution methods is approximately four times the relative MSE for model based methods. Compared to the model based methods, in the case of low censoring proportion, the three substitution methods perform well in terms coverage probability for ρ_ϵ but perform poorly in the case of coverage probability for the fixed effects parameter, β_{01}. In particular, the coverage probability for the three substitution methods for β_{01} ranges from 40.8%-57.1%.

As the censoring proportion increase from low to medium (35%, 30%), the performance of the substitution methods is detrimentally affected. For example, coverage probability for β_{01} reduces from 57.1% to 5.1%, 40.8% to 7.3% and 46.4% to 5.1% for the three substitution methods, `sub`, `halfsub`, and `sqrtsub` respectively. Compared to the model based methods, the relative biases of substitution methods for β_{01} is approximately 35 to 40 times larger. Similarly, as the censoring proportion increase from medium to high (60%, 50%), the sub-

Table 3.8: Substitution methods. Comparison of three methods: sub, halfsub, sqrtsub in simulation studies (m = 100)

			β_{01}	β_{02}	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	ρ_b	$\sigma^2_{\epsilon_1}$	$\sigma^2_{\epsilon_2}$	ρ_ϵ	ρ
Cens.Prop	Criterion	Method	1.2	2.0	2.0	1.5	0.75	2.3	0.9	0.35	0.56
(20%,10%)	Est	`sub`	1.42	2.07	1.31	1.21	0.74	1.67	0.77	0.33	0.54
		`halfsub`	1.48	3.07	1.20	1.21	0.74	1.57	0.77	0.33	0.54
		`halfsqrt`	1.45	2.07	1.24	1.21	0.74	1.61	0.77	0.33	0.54
		`rank`	347.5	453.5	26004.73	23979.1	0.74	34705.1	16112.2	0.32	0.53
	RelBias	`sub`	18.4	3.5	-34.7	-19.4	-1.0	-27.4	-14.7	-5.0	-3.9
		`halfsub`	23.0	3.4	-40.1	-19.2	-1.3	-31.6	-14.3	-6.1	-4.7
		`halfsqrt`	21.1	3.4	-37.9	-19	-1.2	-30.0	-14.5	-5.6	-4.9
	RelMSE	`sub`	5.4	0.6	13.8	5.8	0.9	8.0	2.7	2.7	0.9
		`halfsub`	6.9	0.5	17.6	5.5	0.9	10.4	2.6	2.8	1.0
		`halfsqrt`	6.2	0.5	16.0	5.6	0.9	9.4	2.7	2.7	0.9
	CovProb	`sub`	57.1	87.3						93.8	
		`halfsub`	40.8	90.3						92.1	
		`halfsqrt`	46.4	89.2						92.3	
		`rank`	0.0	0.0						90.1	
(35%,30%)	Est	`sub`	1.69	2.29	0.90	0.78	0.73	1.25	0.55	0.31	0.51
		`halfsub`	1.62	2.11	0.99	1.07	0.74	1.35	0.74	0.31	0.51
		`halfsqrt`	1.65	2.19	0.95	0.94	0.74	1.30	0.65	0.31	0.52
		`rank`	331.7	469.3	26281.1	18181.0	0.74	38256.5	13417.8	0.29	0.50
	RelBias	`sub`	40.6	14.6	-15.1	-48.0	-2.2	-45.9	-38.4	-11.3	-11.3
		`halfsub`	34.7	5.6	-50.5	-28.9	-1.7	-41.5	-18.2	-11.0	-7.3
		`halfsqrt`	37.2	9.3	-52.4	-37.3	-1.8	-43.4	-27.8	-10.6	-7.4
	RelMSE	`sub`	18.2	2.6	31.4	24.2	1.1	21.3	15.1	3.8	1.5
		`halfsub`	13.2	0.7	26.6	9.7	1.0	17.5	3.9	3.6	1.3
		`halfsqrt`	15.1	1.3	28.6	15.2	1.0	19.1	8.2	3.6	1.4
	CovProb	`sub`	5.1	22.2						85.8	
		`halfsub`	7.3	82.7						86.3	
		`halfsqrt`	5.1	54.9						88.2	
		`rank`	0.0	0.0						82.3	
(60%,50%)	Est	`sub`	2.30	2.61	0.38	0.45	0.72	0.67	0.36	0.28	0.47
		`halfsub`	1.79	2.11	0.70	0.93	0.73	1.14	0.74	0.28	0.47
		`halfsqrt`	2.00	2.32	0.55	0.71	0.73	0.91	0.55	0.28	0.48
		`rank`	311.0	490.0	21393.1	12625.0	0.73	39649.1	11080.4	0.25	0.46
	RelBias	`sub`	91.8	30.6	81.0	-70.1	-3.6	-70.9	-59.6	-20.3	-16.2
		`halfsub`	49.0	5.6	-65.2	-37.9	-2.4	-50.6	-17.3	-20.6	-14.2
		`halfsqrt`	66.7	16.0	-72.4	-52.9	-2.9	-60.3	-38.7	-19.1	-14.0
	RelMSE	`sub`	86.5	9.8	65.9	49.7	2.1	50.4	35.7	7.8	3.8
		`halfsub`	25.1	0.6	43.3	15.6	1.4	25.9	3.9	7.3	3.0
		`halfsqrt`	46.0	2.9	52.9	28.9	1.6	36.6	15.4	6.8	3.0
	CovProb	`sub`	0.0	0.0						72.1	
		`halfsub`	0.0	83.1						70.4	
		`halfsqrt`	0.0	11.2						71.8	
		`rank`	0.0	0.0						54.8	

stitution methods perform significantly worse. The coverage probability for β_{01} is 0% for the three methods and the relative biases also increase significantly compared to the model based methods.

Overall, the substitution methods perform poorly in all the four assessments: average parameter estimates, the relatives bias, relative MSE, and coverage probability when compared to the model based methods, `censre4c` and `lmecm`.

The nonparametric method of using the rank of the data is one the common approach for dealing with censored data. Table 3.8 includes the average parameter estimates and the coverage probability for the `rank` approach. Here the focus is on estimation and inference on the contemporaneous correlation, ρ_ϵ. According to Table 3.8, for low censoring proportion, the `rank` approach perform well in terms of the average estimate of the contemporaneous rank correlation ρ_ϵ and the coverage probability. Specifically, the coverage probability for ρ_ϵ is 90.1%. However, the performance of the `rank` approach is detrimentally affected with an increase in the proportion of censoring. Specifically the coverage probability for ρ_ϵ reduces from 82.3% to 54.8% when the proportion of censoring increase from (35%, 30%) to (60%, 50%). Overall the `rank` approach perform well in estimating the rank correlation parameters ρ_b, ρ_ϵ, and ρ when the proportion of censoring is low (20%, 10%).

CHAPTER 4

LMECM: R PACKAGE FOR FITTING LINEAR MIXED MODEL WITH CENSORED AND MISSING RESPONSE

Tuglo, Emmanuel (2014+). **lmecm**: R package for fitting linear mixed model with censored and missing response. To be submitted.

Abstract

The class of mixed effects models is commonly used by practitioners to model longitudinal and repeated measures data. In practice, censored and missing data occur often when measurements are observed repeatedly over time. In addition, more than one response may be observed repeatedly over time and, if so, it is of interest to estimate the association between these multiple response whiles accounting for complications such as censoring and missingness. This paper describe **lmecm**, an R package for fitting univariate and bivariate linear mixed effects models with censored and missing response while accommodating serial correlation and/or heteroscedastic errors.

Keywords: Detection limit; `lme`; Maximum likelihood; Variance structures; Correlation structures; R.

4.1 Introduction

In practice, censored and missing data occur often when measurements are observed repeatedly over time. Analyzing these repeated measures data while accounting for censoring and missingness is challenging considering the limited computational tools available for these problems. The common practice is to impute the censored value with some function of the detection limit (DL) or a single value and then use standard software to fit the model for the complete data. This substitution approach nonetheless has several limitations (Wu, 2010).

The `lme` function in the **nlme** package (Pinheiro et al., 2006) and the `lme4` function in the **lme4** package (Bates et al., 2014) are the two most popular R functions for fitting linear mixed models to repeated measures data. For repeated measures data with censored response, to the best of our knowledge, the `lmec` function in the **lmec** package (Vaida and Liu, 2009b) is the only R function that can fit linear mixed effects models. However, `lmec` cannot accommodate complex covariance structures for the model errors which are highly useful for longitudinal data. For example, `lmec` can not accommodate any kind of serial correlation and/or heteroscedasticity in the errors.

Motivated by this limitation, we develop **lmecm**, an R package for fitting linear mixed effects models (LMEMs) with censored response that can accommodate serial correlation and/or heteroscedastic errors. It can also fit mixed models when the response is completely "missing at random" in a sense described by Rubin (1976) and can accommodate both censoring and missingness simultaneously. In some applications, more than one response are observed repeated over time and it is of interest to study the joint evolution of these multiple responses over time. `lmecm` can easily estimate the association between these multiple response whiles accounting for the aforementioned complications.

The main purposes of this article are: (1) to introduce **lmecm** to R users, (2) to presents background necessary for the appropriate use of the package, and (3) to demonstrate the utility of the software by presenting results on a variety of test problems. The rest of the paper unfolds as follows: Section 2 outlines the theory underlying the B/LMEM with censored response before section 3 describes its implementation in R. Section 3 provides various empirical applications, illustrating various capabilities and features of `lmecm`. Finally, section 4 contains concluding remarks and directions for future research and implementation.

4.2 METHODOLOGY BACKGROUND

The approach for estimating parameters in LMEMs with censored and missing response and the extension to bivariate response were described in Tuglo and Hall (2014), and Tuglo and Hall (2014+). We summarise the main points here.

Consider the standard LMEM for clustered data

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \qquad \mathbf{b}_i \sim N(0, \sigma^2\mathbf{D}_s), \qquad (4.1) \\ \boldsymbol{\epsilon}_i &\sim N(0, \sigma^2\mathbf{I}), \quad \text{with independence over} \quad i = 1, ..., m. \end{aligned}$$

Following Pinheiro and Bates (2000), the log-likelihood corresponding to model (4.1) can be expressed as

$$\begin{aligned}\log L(\boldsymbol{\psi} \mid \mathbf{y}) &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\Big(\mid\mid \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i \mid\mid^2 \Big) \\ &\quad + \sum_{i=1}^{m}\Big(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid \Big) \qquad (4.2)\end{aligned}$$

where

$$\widetilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}, \quad \widetilde{\mathbf{Z}}_i = \begin{bmatrix} \mathbf{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix},$$

$\mathbf{y}_i$ is a vector of n_i outcomes on the ith individual, $N = \sum_{i=1}^{m} n_i$, $\boldsymbol{\beta}$ is a $p-$dimensional vector of fixed effects, $\mathbf{b}_i$ is a $q-$dimensional vector of random effects, $\mathbf{X}_i(n_i \times p)$ and $\mathbf{Z}_i(n_i \times q)$ are known fixed effects and random regressor matrices, $\hat{\mathbf{b}}_i = (\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T(\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\boldsymbol{\beta})$, $\boldsymbol{\Delta}$ is the relative precision factor such that $\mathbf{D}_s^{-1} = \boldsymbol{\Delta}^T\boldsymbol{\Delta}$, $\boldsymbol{\theta}$ is unconstrained set of parameters that determines $\boldsymbol{\Delta}$, and $\boldsymbol{\psi} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T$ denote the combined set of model parameters.

Suppose $\mathbf{y}_i$ is incompletely observed due to censoring, Instead of y_{ij}, we observe the pair (Q_{ij}, C_{ij}) where Q_{ij} is a vector of uncensored readings or censoring levels and C_{ij} is a vector of censoring indicators. We will assume for simplicity of description that the data are left-censored. Thus we can write

complete data: $(\mathbf{y}_i),\ \ i = 1, \ldots, m$

observed data: $(\mathbf{Q}_i, \mathbf{C}_i),\ \ i = 1, \ldots, m$, where

$$\begin{aligned} \mathrm{y}_{ij} &\le Q_{ij} \text{ if } C_{ij} = 1 \\ \mathrm{y}_{ij} &= Q_{ij} \text{ if } C_{ij} = 0 \end{aligned}$$

and $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\mathbf{Q}_i = (Q_{i1}, \ldots, Q_{in_i})^T$, $\mathbf{C}_i = (C_{i1}, \ldots, C_{in_i})^T$. If $\boldsymbol{\psi}^{(k)}$ is the current estimate of the parameter in the model then on the (k+1)th iteration, the E-step for the EM formulation then involves computing $Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) = E(\log L(\boldsymbol{\psi} \mid \mathbf{y})|\mathbf{Q}, \mathbf{C}, \boldsymbol{\psi}^{(k)})$ which is given by

$$\begin{aligned} Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{l=1}^{n_i+q+1} \|\widetilde{\mathbf{w}}_{i,l} - \widetilde{\mathbf{X}}_{i,l}\boldsymbol{\beta} - \widetilde{\mathbf{Z}}_{i,l}\widetilde{\widehat{\mathbf{b}}}_{i,l}\|^2 \\ &\quad + \sum_{i=1}^{m}\left(\log \mathrm{abs}|\boldsymbol{\Delta}| - \frac{1}{2}\log \mid \widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i \mid \right), \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} \widetilde{\mathbf{w}}_{i,l} &= \begin{cases} E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})^T & l = 1, \\ \widetilde{\mathbf{S}}_{i,l-1} & l = 2, \ldots, n_i + q + 1, \end{cases} \qquad \widetilde{\mathbf{X}}_{i,l} = \begin{cases} \widetilde{\mathbf{X}}_i & l = 1, \\ \mathbf{0} & l = 2, \ldots, n_i + q + 1, \end{cases} \\ \widetilde{\mathbf{Z}}_{i,l} &= \widetilde{\mathbf{Z}}_i \ \ \forall \ l, \qquad \widetilde{\widehat{\mathbf{b}}}_{i,l} = \begin{cases} \widetilde{\widehat{\mathbf{b}}}_i & l = 1, \\ \widetilde{\mathbf{Z}}_i(\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{Z}}_i)^{-1}\widetilde{\mathbf{Z}}_i^T\widetilde{\mathbf{S}}_{i,l-1} & l = 2, \ldots, n_i + q + 1, \end{cases} \\ \widetilde{\mathbf{S}}_i\widetilde{\mathbf{S}}_i^T &= \mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)}) \qquad \widetilde{\mathbf{S}}_i = (\widetilde{\mathbf{S}}_{i,1}, \ldots, \widetilde{\mathbf{S}}_{i,n_i+q}). \end{aligned} \tag{4.4}$$

Note that equation (4.3) is equivalent to (4.2), the complete data log-likelihood, except that in the second term of equation (4.3), for each cluster with a censored observation, an additional (n_i+q) sub-cluster is introduced with the response vector, fixed and random effect design matrices as in (4.4). Thus in the routine for optimizing the log-likelihood (4.2), the `lme` function in the **nlme** package can be used to optimize the M-step objective function (4.3) with a slight modification. The detailed description of the modification to the `lme`, the extension to more generalized linear models for both maximum likelihood (ML) and restricted maximum likelihood (REML) estimation is given in Tuglo and Hall (2014). A summary description of the algorithm implemented in `lmecm` is given below:

1. At the E-step compute $E(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})^T$ and $\mathrm{Var}(\widetilde{\mathbf{y}}_i|\mathbf{Q}_i, \mathbf{C}_i, \boldsymbol{\psi}^{(k)})^T$, the mean and variance of the truncated multinormal distribution. Detailed description on computing the truncated mean and variance is given by Vaida and Liu (2009a).

2. Augment the response vector $\mathbf{w}$ and the fixed and random effects design matrices $\mathbf{X}$ and $\mathbf{Z}$ as in (4.4).

3. At the M-step, optimize the log-likelihood using the modified `lme` function.

4. Repeat 1-3 until convergence.

4.3 Implementation in R

The main model-fitting function in the package **lmecm** is `lmecm()` which takes a fairly standard approach for implementing ML and REML for LMEMs with censored and missing response in R: `fixed` and `random` is arguments are used for fixed and random effects model specifications, `data` for the data specification, `weights` and `correlation` are used for variance and correlation structure specifications, and `cens` specifies the censoring indicator.

The model-fitting function `lmecm()` and its associated class are designed to be similar to `lme()` in the **nlme** package (Pinheiro et al., 2006) and `lmec()` in the **lmec** package (Vaida and Liu, 2009b) for fitting LMEMs with censored response. Compared to `lme`, an important difference is that: (1) `lmecm` has one extra argument, `cens`, which specifies the censoring indicator; (2) the names of the variables used in the fixed and the random component should be different and unique because of the different data augmentations for the fixed and random effects design matrix (see equation (4.4)). The arguments and the output from the `lmecm` are given below.

Usage

```
lmecm(fixed, data, random, correlation = NULL, weights = NULL,
                              na.action = na.fail, method = "ML", cens)
```

Arguments

`fixed`	a two-sided linear formula object describing the fixed-effects part of the model, with the response on the left of a ~ operator and the covariates, separated by + operators, on the right. A fixed effect formula of type

	`yL ~ x1+....+xp` specify `yL`, the observed left-censored response vector and `x1,...,xp` specify the fixed effects covariates. It needs includes a column of 1's if intercept is present.
`data`	a `groupedData` object containing the variables named in `fixed`, `random`, `correlation`, and `weights`.
`random`	a named list of formulas or `pdMat` objects with a one-sided formula of the form `~x1+...+xn`, with the grouping factors as names. For example, a random effect formula of type `random=list(id=pdSymm(∼intercept + x1))` specifies `id` as the grouping variable for a general positive-definite matrix `pdSymm` of the `pdMat` class. The `intercept` is a vector of 1's and `x1` is a covariate. for details on `pdMat` classes, one may refer to the R documentation of **nlme** package. Note that, different data augmentation are used for the fixed effect component and the random effect component so it important that the variables names used in the fixed effects component are different from the variable names used in the random effects.
`correlation`	an optional `corStruct` object describing the within-group correlation structure. In case of univariate response, `lmecm` can fit LMEM with censored response with all the `corStruct` classes in **nlme** package. In case of bivariate response, `lmecm` can fit BLMEM with censored response with correlation structure `BivarCorAR1`, `BivarCorIdent`, `BivarCorAR1Diag`. Detailed descriptions of these three `corStruct` classes are given in section (4.3). The default is `correlation = NULL`, corresponding to no within-group correlation.
`weights`	an optional `varFunc` object or one-sided formula describing the within-group heteroscedasticity structure. `lmecm` fits B/LMEM with censored

response with all the variance structure in the **nlme** package. Defaults to NULL, corresponding to homoscedastic within-group errors.

`Method`	a character string. If `"REML"` the model is fit by maximizing the restricted loglikelihood. If `"ML"` the log-likelihood is maximized. Defaults to `"ML"`.
`na.action`	a function that indicates what should happen when the data contain `NAs`. The default action `(na.fail)` causes `lmecm` to print an error message and terminate if there are any incomplete observations.
`cens`	censoring indicator, if `yL` $\leq$ DL, then `cens` $= 1$.

Value

`beta`	a numeric vector of estimated fixed effects.
`sigma`	residual standard deviation.
`Psi`	estimated variance matrix of the random effects.
`Delta`	matrix such that `t(Delta)%*%Delta` = `sigma2*solve(Psi)`.
`corrParam`	estimates of correlation parameters.
`varParam`	estimates of variance parameters.
`logLik`	observed data (restricted) log-likelihood value.
`varFix`	variance matrix for the fixed effects.
`step`	a numeric scalar of number of EM iterations before convergence.
`likseq`	a vector of length step giving the log-likelihood of the complete data, with $\boldsymbol{\beta}$ profiled out.

Functions	Descriptions
`print()`	simple printed display with coefficient estimates.
`summary()`	returns an object of class `"summary.lmecm"` containing the relevant summary statistics (which has a print() method). These relevant standard regression output includes a data frame with columns `estimates`, `stdErr`, `t.values`, `p.values` representing respectively the fixed effects estimates, their approximate standard errors, the ratios between the estimates and their standard errors, and the associated p-values from the t distribution. Rows correspond to the different fixed effects.
`coef()`	extract `lmecm` coefficients (the fixed effects, random effects, variance and correlation parameter estimates). The estimated coefficients are returned as a list.
`logLik()`	returns the (restricted) log-likelihood of the fitted model.
`AIC()`	compute Akaike information criteria.
`vcov()`	returns the variance-covariance matrix of the fixed effects parameters of the fitted model object.
`confint()`	computes confidence intervals for fixed effects parameters in a fitted model.
`ranef()`	extract the estimated random effects, represented as a data frame with rows given by the different groups at that columns given by the random effects.
`fitted()`	the fitted values are obtained by adding together the population fitted values (based only on the fixed effects estimates) and the estimated contributions of the random effects to the fitted values at grouping levels.
`residuals()`	returns the residuals of the fitted model. An optional character string specifying the `type` of residuals to be used. If `"response"` the `"raw"` residual (observed - fitted) are used; else, if `"pearson"`, the standardized residuals (raw residuals divided by the corresponding standard errors) are used; else, if `"normalized"`, the normalized residuals (standardized residuals pre-multiplied by the inverse square-root factor of the estimated error correlation matrix) are used. Defaults to `raw` residuals.

Table 4.1: Functions and methods for objects of class `"lmecm"`.

The returned fitted-model object of class `"lmecm"` is a list and some of its elements are listed above. A set of standard extractor functions for fitted-model objects is also available for objects of class `"lmecm"`, including the usual `summary()` method that includes tests of significance for all the fixed effects coefficients. See Table 4.1 for a list of all available methods. Most of these are standard generic methods in base R.

4.4 EXAMPLE

4.4.1 EXAMPLE 1: LMEM WITH CENSORED RESPONSE WITH HOMOGENEOUS ERROR

UTI data

To illustrate the usage of `lmecm` in practice, we replicate the analyses of the UTI data previously analysed in Vaida and Liu (2009a). The data set is available in the R package **lmec**. The UTI data comes from a study of 72 perinatally HIV-infected children and adolescents (Saitoh et al., 2008; Vaida and Liu, 2009a). Unstructured treatment interruption (UTI) is common in this population, due mainly to treatment fatigue. Suboptimal adherence to the treatment protocol can lead to the antiretroviral (ARV) resistance and diminished treatment options in the future. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI were studied. The number of observations at baseline (month 0) and at the 6 post-baseline measurement occasions are 71, 62, 58, 57, 43, 24, 24, and 13, respectively. Out of the 362 observations, 26 (7%) observations are below the DLs, 50 or 400 copies/mL (two DLs), and were left-censored at these values. Following Vaida and Liu (2009a), we consider a LMEM with random intercepts b_i of the form

$$y_{ij} = \beta_j + b_i + \epsilon_{ij} \tag{4.5}$$

where y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_j. The data profiles are plotted in Figure 4.1

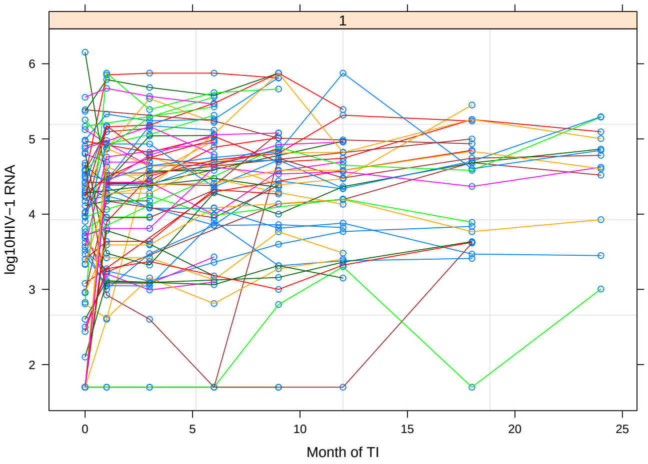

Figure 4.1: UTI data: Trajectory of the individual profiles.

First, we load the UTI data and construct the `groupedData` object which is then given as argument to the `lmecm` function.

```
> library(lmec);  library(lattice); data(UTIdata)
> UTIdata <- subset(UTIdata, !is.na(RNA))
> o <- order(UTIdata$Patid, UTIdata$Fup)
> UTIdata <- UTIdata[o,]
> UTIdata$cens = (UTIdata$RNAcens==1)+0
> UTIdata$yL = log10(UTIdata$RNA)
> X <- cbind((UTIdata$Fup==0)+0, (UTIdata$Fup==1)+0, (UTIdata$Fup==3)+0,
+       (UTIdata$Fup==6)+0, (UTIdata$Fup==9)+0, (UTIdata$Fup==12)+0,
+        (UTIdata$Fup==18)+0, (UTIdata$Fup==24)+0)
> UTIdata$X0 <- X[,1]; UTIdata$X1 <- X[,2]; UTIdata$X2 <- X[,3]
> UTIdata$X3 <- X[,4]; UTIdata$X4 <- X[,5]; UTIdata$X5 <- X[,6]
> UTIdata$X6 <- X[,7]; UTIdata$X7 <- X[,8]
> UTIdata$cluster <- as.numeric(UTIdata$Patid)
> UTIdata$intercept <- rep(1, dim(UTIdata)[1])
> UTIdata$time <- UTIdata$Fup
> #defining a group data object and plot
> UTIdata.grp <- groupedData(yL ~ Fup|cluster, data = UTIdata,
+                            order.groups=F)
```

```
>  plot(UTIdata.grp, outer=~1, aspect="fill", xlab="Month of TI",
+         ylab="log10HIV-1 RNA")     # Figure  1
```

In **lmecm**, model (4.5) can be fitted via

```
> # fitting the model
> m1 <- lmecm(fixed = yL ~X0+X1+X2+X3+X4+X5+X6+X7-1,
+        random =list(cluster = pdIdent(~intercept -1)),
+        cens = UTIdata.grp$cens, data=UTIdata.grp,
+          na.action = na.fail)
> summary(m1)

Linear mixed-effects model with censored response fit by maximum likelihood

Call:
lmecm.formula(fixed = yL ~ X0 + X1 + X2 + X3 + X4 + X5 + X6 +
    X7 - 1, random = list(cluster = pdIdent(~intercept - 1)),
    weights = NULL, correlation = NULL, method = "ML",
    cens = UTIdata.grp$cens, na.action = na.fail, data = UTIdata.grp)

Log-likelihood: -412.0415

Fixed effects:
   Estimate    StdErr  t.value p.value
X0 3.618782 0.1252514 28.89214       0
X1 4.181549 0.1284388 32.55675       0
X2 4.256518 0.1303395 32.65718       0
X3 4.375555 0.1306986 33.47821       0
X4 4.581557 0.1397291 32.78886       0
X5 4.584641 0.1484415 30.88518       0
X6 4.692767 0.1645971 28.51063       0
X7 4.809051 0.2017704 23.83427       0

Random effects:
Random effects variance covariance matrix
          intercept
intercept   0.76502
  Standard Deviations: 0.87465

sigma 0.5844072

Number of Observations: 362
```

```
Number of Groups:  72
Censoring proportion:  7.2%
```

The results confirm those presented in Vaida and Liu (2009a). We can obtain the 95% confidence interval for the fixed effects parameters via

```
> confint(m1)
Approximate 95% confidence interval
Fixed effects:
      lower estimates    upper
X0 3.373228  3.618782 3.864336
X1 3.929746  4.181549 4.433351
X2 4.000990  4.256518 4.512047
X3 4.119322  4.375555 4.631788
X4 4.307620  4.581557 4.855495
X5 4.293624  4.584641 4.875659
X6 4.370077  4.692767 5.015458
X7 4.413482  4.809051 5.204619
```

4.4.2 Example 2: LMEM with Censored Response with Non-spherical Error

To illustrate the capabilities of the `lmecm` function in fitting LMEMs with censored response with serial correlation and/or heteroscedastic errors, we use a classic data set, Potthoff and Roy (1964) orthodontic growth curve data, available in R as the `groupedData` object `Orthodont` in the **nlme** package. The data contain measurements of the distance from the pituitary gland to the pterygomaxillary fissure taken every two years from 8 years of age until 14 years of age on a sample of 27 children - 16 males and 11 females. For the purpose of illustration we created artificial left-censoring. The 23rd percentile of the observed response is computed and the observations equal to or below that percentile are left-censored at that value. The LMEM we consider can be written as

$$\text{distance}_{ij} = \beta_0 + \beta_1\text{Sex} + \beta_2(\text{age}_j - 11) + \beta_3(\text{age}_j - 11)\text{Sex} + b_{0i} + b_{1i}(\text{age} - 11) + \epsilon_{ij}. (4.6)$$

The data with the left-censored values are shown in Figure 4.2

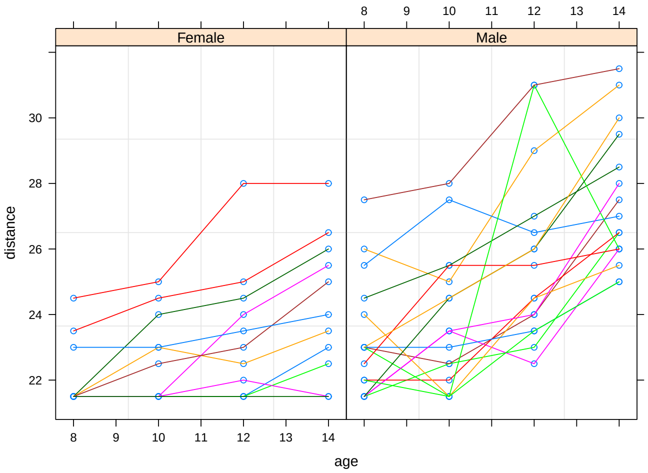

Figure 4.2: Distance from the pituitary to the pterygimaxillary fissure versus age for a sample of 16 boys (subject M01 to M16) and 11 girls (subjects F01 to F11). The paneled plot was chosen to facilitate comparison.

```
> library(nlme)
> data(Orthodont)
> LOD <- quantile(Orthodont$distance,prob=0.23)#compute the 23th percentile
> Orthodont$cens <- ifelse(Orthodont$distance<=LOD,1,0)#censoring indicator
> Orthodont$distance[Orthodont$cens == 1] <- LOD
> Orthodont$sex <- factor(rep(1:2,  c(64, 44)))
> Orthodont$subject <- rep(1:27,  table(Orthodont$Subject))
```

To remove correlation between the intercept and the slope, the age variable is centered.

```
> Orthodont$centeredAge <- Orthodont$age-11  #age centered
> Orthodont$centeredAge1 <-Orthodont$centeredAge
> Orthodont$intercept <- rep(1, dim(Orthodont)[1])
> Orthodont$intercept1 <- Orthodont$intercept
> Orthodont$sexFemale <- (Orthodont$Sex=="Female")+0
> Orthodont$centeredAgeFemale <- c(rep(0,
+            length(Orthodont$centeredAge[Orthodont$Sex=="Male"])),
+             Orthodont$centeredAge[Orthodont$Sex=="Female"])
> Orthodont.grp1 <- groupedData(distance ~ age|Subject,
+                   data=Orthodont, order.group=F)
> plot(Orthodont.grp1,outer = ~ Sex, aspect="fill") #Figure 2
```

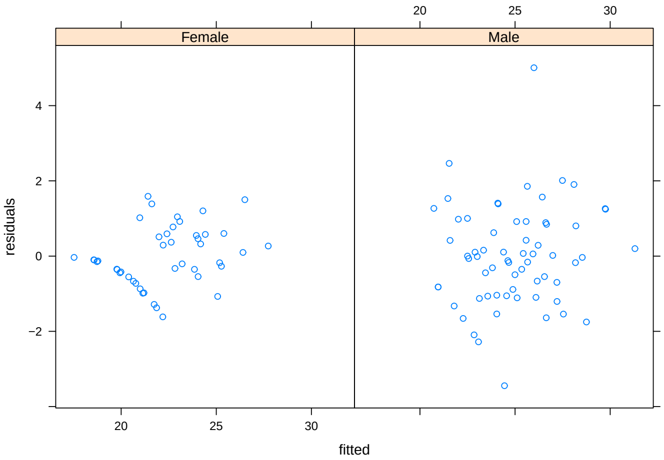

Figure 4.3: Scatter plot of the residuals versus fitted values for the model with homogeneous error.

As other authors have found, the variability in the response for boys is greater than girls. This difference is apparent in Figure 4.3, the plot of residuals versus the fitted values for the model with the homogeneous error. For the orthodontic growth data it is clear that a more general LMEM that allows different variance by gender for the within-group error is appropriate. Just like the `lme`, the `lmecm` allows modelling of of heteroscedasticity of the within groups error through the `weights` argument. The `varIdent` variance function structure in the **nlme** package can be used.

```
> Orthodont.grp <- groupedData(distance ~ 1|Subject,
+                   data=Orthodont, order.group=F)

> m2 <- lmecm(fixed = distance~intercept+sexFemale+centeredAge
+             + centeredAgeFemale -1,
+       random=list(subject=pdSymm(~intercept1+centeredAge1-1)),
+       weights=varComb(varIdent(form=~1|sex)),
+       data = Orthodont.grp, cens=Orthodont$cens)
```

```
> summary(m2)
Linear mixed-effects model with censored response fit by maximum likelihood

Call:
lmecm.formula(fixed = distance ~ intercept + sexFemale + centeredAge +
    centeredAgeFemale - 1, random = list(subject = pdSymm(~intercept1 +
    centeredAge1 - 1)), weights = varComb(varIdent(form = ~1 |
    sex)), correlation = NULL, method = "ML", cens = Orthodont$cens,
    data = Orthodont.grp)

Log-likelihood: -180.3368

Fixed effects:
                   Estimate     StdErr   t.value p.value
intercept         24.9593831 0.49653732 50.266883  0.0000
sexFemale         -2.4605622 0.76204368 -3.228899  0.0016
centeredAge        0.7814700 0.09611629  8.130464  0.0000
centeredAgeFemale -0.2266814 0.12867716 -1.761629  0.0789

Random effects:
Random effects variance covariance matrix
             intercept1 centeredAge1
intercept1     3.347100     0.036336
centeredAge1   0.036336     0.026478
  Standard Deviations: 1.8295 0.16272

sigma 1.516389

Variance structure formula: varComb(varIdent(form = ~1 | sex))
Parameter estimate(s):
      A.2
0.4975093

Number of Observations: 108
Number of Groups:  27
Censoring proportion:  23.1%
```

For the orthodontic data, according to fitted model `m2`, the error standard deviation for the girls is about 50% (`0.4975093`, is the ratio between the error standard deviation for boys and girls) of that for the boys. Suppose in addition to the different variance for the boys and girls, there is also a serial correlation that needs to be accounted for (say `corAR1`), `lmecm` can fit

LMEM with censored response whiles accounting for various combination of correlation and variance classes. The different correlation structures available for the `correlation` arguments are described in Pinheiro and Bates (2000). For example in case the correlation is AR(1) (`corAR1`),

```
> m3 <- lmecm(fixed = distance~intercept+sexFemale+centeredAge
+             +centeredAgeFemale -1,
+       random=list(subject=pdSymm(~intercept1+centeredAge1-1)),
+       weights =varComb(varIdent(form=~1|sex)),
+        correlation=corAR1(form=~1|subject),
+       data = Orthodont.grp, cens=Orthodont$cens)

> coef(m3)$corrParam
       Phi
-0.4774965
```

We can compare model `m2` and `m3` via a likelihood ration test

```
>  1-pchisq(2*(as.numeric(logLik(m3))-as.numeric(logLik(m2))),1)
[1] 0.04144933
```

which provides enough evidence against the null hypothesis hence we conclude model `m3` fits better. Assuming instead of different variance for the two group, the variability is increasing or decreasing with age for the two groups. We can account for that using the `varPower` variance function class.

```
> m4 <- lmecm(fixed = distance~intercept+sexFemale+centeredAge
+             +centeredAgeFemale -1,
+       random=list(subject=pdSymm(~intercept1+centeredAge1-1)),
+       weights = varComb(varPower(form=~age|sex)),
+       data = Orthodont.grp, cens=Orthodont$cens)

> coef(m4)$varParam
       A.1        A.2
-0.2940385 -0.5882195
```

We can compare model `m3` and `m4` using AIC via

```
> AIC(m3)
[1] 376.516
> AIC(m4)
[1] 377.9784
```

which supports the conclusion that model `m3` is better than `m4`. Detailed presentation of the use of variance functions for linear mixed effects model is given in Pinheiro and Bates (2000).

4.4.3 BLMEM with Censored Response

The **lmecm** package provides a set of classes of correlation structures that are used to specify the within-group correlation for BLMEMs. For now there are three bivariate `corStruct` classes in the **lmecm** package. The two main arguments to the `corStruct` classes are `value` and `form` which are the same as the `corStruct` classes in the **nlme** package. We now describe and illustrate the three bivariate `corStruct` classes in the **lmecm** package.

`BivarCorAR1`

This class implements bivariate autoregressive structure of order 1 (BAR(1)). This is a generalization of AR(1) to bivariate response. The argument `value` initializes the correlation parameters ϕ_1, ϕ_2 and, ρ which all take values between -1 and 1, and by default, are set to 0. The parameter ϕ_1 specifies the AR(1) correlation for the first response, ϕ_2 specifies the AR(1) correlation for the second response and ρ specifies the contemporaneous correlation between the two responses. The argument `form` is a one-sided formula specifying the position variable (i.e., time index) and grouping variable. By default, `form = ~ 1`, implying that the order of the observation within the group be used as the position variable. For example a BAR(1) correlation structure with $\phi_1 = 0.5$, $\phi_2 = 0.2$ and, $\rho = 0.6$, with position variable given by observation order within group, and grouping variable `id`, is specified as

```
cs1 <- BivarCorAR1(c(.5,.2, 0.6),form=~1|id)
cs1 <- Initialize(cs1, data=painTran.grp)
```

```
corMatrix(cs1)[[1]]

     [,1] [,2] [,3]  [,4]  [,5] [,6] [,7]  [,8]
[1,] 1.000 0.50 0.25 0.125 0.600 0.00 0.00 0.000
[2,] 0.500 1.00 0.50 0.250 0.000 0.60 0.00 0.000
[3,] 0.250 0.50 1.00 0.500 0.000 0.00 0.60 0.000
[4,] 0.125 0.25 0.50 1.000 0.000 0.00 0.00 0.600
[5,] 0.600 0.00 0.00 0.000 1.000 0.20 0.04 0.008
[6,] 0.000 0.60 0.00 0.000 0.200 1.00 0.20 0.040
[7,] 0.000 0.00 0.60 0.000 0.040 0.20 1.00 0.200
[8,] 0.000 0.00 0.00 0.600 0.008 0.04 0.20 1.000
```

`BivarCorAR1Diag`

This implements a correlation structure similar to `BivarCorAR1` except the argument `value` takes two parameters ϕ_1 and ϕ_2. This correlation structure assume the contemporaneous correlation, ρ, is 0. For example, a BAR(1) correlation structure with $\phi_1 = 0.6$ and $\phi_2 = 0.3$ with position variable given by observation order within group, and grouping variable `id`, is specified below:

```
> cs2 <- BivarCorAR1Diag(c(.6,.3),form=~1|id)
> cs2 <- Initialize(cs2, data=painTran.grp)
> corMatrix(cs2)[[1]]
      [,1] [,2] [,3]  [,4]  [,5] [,6] [,7]  [,8]
[1,] 1.000 0.60 0.36 0.216 0.000 0.00 0.00 0.000
[2,] 0.600 1.00 0.60 0.360 0.000 0.00 0.00 0.000
[3,] 0.360 0.60 1.00 0.600 0.000 0.00 0.00 0.000
[4,] 0.216 0.36 0.60 1.000 0.000 0.00 0.00 0.000
[5,] 0.000 0.00 0.00 0.000 1.000 0.30 0.09 0.027
[6,] 0.000 0.00 0.00 0.000 0.300 1.00 0.30 0.090
[7,] 0.000 0.00 0.00 0.000 0.090 0.30 1.00 0.300
[8,] 0.000 0.00 0.00 0.000 0.027 0.09 0.30 1.000
```

This correlation structure is mostly used in fitting a reduced model when testing the significance of the contemporaneous correlation parameter.

`BivarCorIdent`

This class implements a bivariate correlation structure for which within each response the errors are uncorrelated but the errors at the same time points across the two responses are correlated. This is an extension of homogeneous errors to the bivariate response. The argument `value` initializes a single correlation parameter ρ. Its `form` argument is defined as in `BivarCorAR1`. For example to specify a `BivarCorIdent` correlation structure with $\rho = 0.7$, with position variable given by observation order within group, and grouping variable `id`, we use

```
> cs3 <- BivarCorIdent(c(.7),form=~1|id)
> cs3 <- Initialize(cs3, data=painTran.grp)
> corMatrix(cs3)[[1]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]  1.0  0.0  0.0  0.0  0.7  0.0  0.0  0.0
[2,]  0.0  1.0  0.0  0.0  0.0  0.7  0.0  0.0
[3,]  0.0  0.0  1.0  0.0  0.0  0.0  0.7  0.0
[4,]  0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.7
[5,]  0.7  0.0  0.0  0.0  1.0  0.0  0.0  0.0
[6,]  0.0  0.7  0.0  0.0  0.0  1.0  0.0  0.0
[7,]  0.0  0.0  0.7  0.0  0.0  0.0  1.0  0.0
[8,]  0.0  0.0  0.0  0.7  0.0  0.0  0.0  1.0
```

EXAMPLE 3: USING BIVARIATE CORRELATION STRUCTURE WITH `lmecm`

The example used here come from Weiss (2010) and is available in the **mlmmm,** R package of Yucel (2010). The `pain` data frame consists of up to four observations on 64 children 8 to 10 years of age. The response is the length of time in seconds that the child can tolerate keeping his or her arm in very cold water, a proxy measure of pain tolerance. Missing data exist due to absenteeism, broken arms, or other reasons. 6% of each response are missing. Two measures were taken during a first visit followed by two more measures during a second visit after a two-week gap. During the first visit, the dominant arms of the children were used during the trial and the non-dominant arms were used during the second visit. Also, the children were classified into one of the two groups, attenders and distracters according to their style of coping (CS) with pain during the first visit. A treatment (TMT) was administered prior to the fourth trial. The treatment consisted of ten minutes counselling intervention to either attend (A), distract(D), or no advice (N). Detailed description of the data is given in Weiss (2010). For the purpose of illustration we created artificial left-censoring. The 10th percentile of the observed responses are computed and the observations equal to or below that percentile are left-censored at that value. Thus, censoring and missingness occur simultaneously and need to be accounted for in order to estimate the parameters unbiasedly.

Let y_{hij} be the response from the ith subject, $i = 1, \ldots, 64$ taken at distinct times $j = 1, \ldots, 4$ for each subject for outcome $h = 1, 2$. The model we considered can be written as

$$y_{hij} = \beta_{0h} + \beta_{1h}\text{arm}_{ij} + \beta_{2h}\text{cs}_{ij} + \beta_{3h}\text{tmt1}_{ij} + \beta_{4h}\text{tmt2}_{ij} + b_{hi} + \epsilon_{hij}. \tag{4.7}$$

First, we load the `pain` data, and transform it to a `groupedData` object for the `lmecm` function.

```
> library(mlmmm)
> data(pain)
> y <- as.numeric(c(pain$y.tol,pain$y.rat))
> X1 <- rbind(as.matrix(pain[3:7]), matrix(0,  dim(pain)[1], 5))
> X2 <- rbind(matrix(0,  dim(pain)[1], 5),as.matrix(pain[3:7]))
> id <- c(pain$subj,pain$subj)
> painTran <- as.data.frame(cbind(id ,y, X1, X2), row.names=F)
> names(painTran) <- c("id", "y", "int1","x11","x12", "x13" ,"x14",
+                       "int2" ,"x21", "x22", "x23","x24")
> cens1 <- ifelse(is.na(painTran$y),1, 0)
> painTran$respid <- rep(1:2, each=dim(pain)[1])
> painTran$y[is.na(painTran$y)] <- max(na.omit(painTran$y))
> LOD <- quantile(painTran$y, prob=0.10)
> painTran$y[painTran$y <= LOD] <- LOD
> cens2  <- ifelse(painTran$y <= LOD,1, 0)
> painTran$cens <- cens1+cens2
> painTran$int11 <- painTran$int1
> painTran$int22 <- painTran$int2
> painTran.grp <- groupedData(y ~ 1|id, data=painTran)
```

The bivariate linear mixed model (4.7) is then fitted using the `lmecm` function. The code for fitting this model appears below along with the summary of the fitted model.

```
> m5 <- lmecm(fixed = y~int1+x11+x12+x13+x14+int2+x21+x22+x23+x24-1,
+        random= list(id = pdSymm(~int11 + int22 -1)),
+        correlation = BivarCorIdent(form = ~1|id),
+        weights <- varComb(varIdent(form=~1|respid)),
+        data=painTran.grp, cens=painTran.grp$cens)

> summary(m5)
Linear mixed-effects model with censored response fit by maximum likelihood

Call:
lmecm.formula(fixed = y ~ int1 + x11 + x12 + x13 + x14 + int2 +
    x21 + x22 + x23 + x24 - 1, random = list(id = pdSymm(~int11 +
    int22 - 1)), weights = weights <- varComb(varIdent(form = ~1 |
    respid)), correlation = BivarCorIdent(form = ~1 | id),
     cens = painTran.grp$cens, data = painTran.grp)

Log-likelihood: -693.5933

Fixed effects:
        Estimate    StdErr    t.value p.value
int1  2.65385768 0.2204440 12.0386938  0.0000
x11   0.31395132 0.1784759  1.7590682  0.0786
x12   0.43907802 0.1715067  2.5601216  0.0105
x13   0.21967208 0.2058218  1.0672926  0.2859
x14   0.39291149 0.2016690  1.9482985  0.0514
int2  5.90610919 0.5524904 10.6899762  0.0000
x21  -0.34211422 0.4511503 -0.7583154  0.4483
x22   0.09833007 0.4331682  0.2270021  0.8204
x23  -0.48244271 0.5171081 -0.9329630  0.3509
x24  -0.14786458 0.5076939 -0.2912475  0.7709

Random effects:
Random effects variance covariance matrix
         int11   int22
int11  0.37522 -0.1594
int22 -0.15940  2.0261
  Standard Deviations: 0.61255 1.4234

sigma 0.4369169
```

```
Correlation structure formula: BivarCorIdent(form = ~1 | id)
Parameter estimate(s):
       rho
-0.1850981

Variance structure formula: weights <- varComb(varIdent(form = ~1 | respid))
Parameter estimate(s):
     A.2
3.795283

Number of Observations: 512
Number of Groups:  64
Censoring proportion:  16.2%
```

The p-value corresponding to the fixed effects parameters suggest we remove those variables from the model but for the purpose of illustration and since the focus in on the contemporaneous correlation, we leave those variable in the model. According to model `m5`, the estimated contemporaneous correlation is -0.19 and the estimated subject-specific correlation is -0.18. The overall correlation between the two response, pain tolerance and pain rating is -0.18, suggesting that as pain tolerance increases, we expect pain rating to decrease.

To assess the fit of model `m5`, we consider a slightly more complex model and compare the two model using likelihood ratio test. We fit model `m6` which is similar to `m5` except `correlation = BivarCorAR1(form =  1|id)`, which allows AR(1) correlation over time within each response. The code for fitting this model appears below along with the fitted correlation parameters.

```
> m6 <- lmecm(fixed = y~int1+x11+x12+x13+x14+int2+x21+x22+x23+x24-1,
+         random= list(id = pdSymm(~int11 + int22 -1)),
+         correlation = BivarCorAR1(form = ~1|id),
+         weights <- varComb(varIdent(form=~1|respid)),
+         data=painTran.grp, cens=painTran.grp$cens)
> summary(m6)
...
Correlation structure formula: BivarCorAR1(form = ~1 | id)
Parameter estimate(s):
```

```
      phi1       phi2       rho
 0.1622474  0.3280500 -0.4084515
...
```

The estimate of other parameters in the model are similar to those in `m5`. The likelihood ratio test confirms that model `m6` fits the data better.

```
> pvalue <- 1-pchisq(2*(as.numeric(logLik(m6))-as.numeric(logLik(m5))),2)
> pvalue
[1] 0.02224754
```

According to model `m6`, the contemporaneous correlation is -0.41 indicating at a specific time, there is a very strong negative relationship between the pain tolerance and pain rating. The overall correlation between the two response is -0.32. We test the signicance of the contemporanous correlation by fitting a reduce model `m7` with `correlation = BivarCorAR1Diag(form = ~1|id)` and conduct the LRT test. With p-value 0.0069, we reject the null and conclude that contemporaneous correlation is significant and that model `m6` is the best model for the data among the model we considered.

4.5 Conclusion

This paper addresses R implementation of univariate and bivariate linear mixed effects models with censored and missing response. The most relevant features and the capabilities of the `lmecm` function have been introduced through empirical examples. Future research and implementation will focus on diagnostics methods and additional bivariate correlations structures such bivariate ARMA correlation structures.

4.6 REFERENCES

Bates, D., M. Maerchler, B. Bolker, S. Walker, R. H. B. Christensen, and H. Singmann (2014). *Linear mixed effects models using Eigen and S4.* R package version 1.1-6.

Box, G., G. Jenkins, and G. Reinsel (2004). *Time series analysis: forecasting and control.* 3rd ed., Holden-Day, San Francisco.

Carroll, R. and D. Ruppert (1988). *Transformation and weighting in regression.* Chapman & Hall, New York.

Cook, R., L. Zeng, and G. Yi (2004). Marginal analysis of incomplete longitudinal binary data: a cautionalry note on LOCF imputation . *Biometrics 60*, 820–828.

Davidian, M. and D. Giltinan (1995). *Nonlinear models for repeated measurement data.* Boca Raton, FL:Chapman and Hall.

Fitzmaurice, G. M., N. Laird, and J. H. Ware (2004). *Applied longitudinal analysis.* Wiley: New Jersey.

Little, R., G. Milliken, W. Stroup, and R. Wolfinger (1996). *SAS system for mixed effects models.* SAS Intitutes Inc., Cary NC.

Pinheiro, J. and D. Bates (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer.

Pinheiro, J., D. Bates, S. DeRoy, and D. Shakar (2006). *nlme: Linear and nonLinear mixed effects models.* R package version 3.1-78.

Potthoff, R. and S. Roy (1964). A generalized multivariate analsysis of variance model useful especially for growth curve problems. *Biometrika 51*, 313–326.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Saitoh, A., M. Foca, R. M. Viani, S. Hefferman-Vaca, S. Vaida, F. Lujan-Zilbermann, J. Emmanuel, P. J. Deville, J. G., and S. A. Spector (2008). Clinical Outcome in perinatally acquired HIV-infected childrens and adolescents after unstructured treatment interruption . *Pediatrics 121*, e513–e521.

Thiebaut, R., H. Jacqmin-Gadda, G. Chene, C. Leport, and D. Commenges (2002). Bivariate linear mixed models using SAS PROC MIXED. *Computer Methods and Programs in Biomedicine 69*, 249–256.

Vaida, F. and L. Liu (2009a). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Vaida, F. and L. Liu (2009b). *Linear mixed effects models with censored response.* R package version 1.0.

Venables, W. and B. Ripley (1999). *Moden applied statistics with S-PLUS.* 3rd ed., Springer-Verlag, New York.

Weiss, R. E. (2010). *Modeling longitudinal data.* New York, Springer.

Wu, L. (2010). *Mixed effects models for complex data.* Chapman and Hall CRC Boca Raton FL.

Xie, X., X. Xue, S. Gange, H. Strickler, M. Kim, and W. H. Study-Group (2012). Estimation and inference on correlations between biomarkers with repeated measures and left-censoring due to minimum detection levels. *Statistics in Medicine 31*(21), 2275–2289.

Yucel, R. (2010). *ML estimation under multivariate linear mixed models with missing values.* R package version 0.3-1.2.

Chapter 5

Conclusion

In this dissertation, mixed model methods for analyzing univariate and multivariate longitudinal data with censored and missing response have been developed. The proposed methodology is computationally feasible and it is implemented in an R function `lmecm` which can fit a wider class of B/LMEMs than existing computational tools. In Particular, `lmecm` can fit B/LMEMs with censored and missing response and can accommodate serial correlation and heteroscedasticity of the error which is one of the common features of longitudinal data.

There are still open questions with regards to accounting for censoring induced by detection limits. In terms of censored data, there are cases where there is the presence of both DLs and quantitation limits (QLs) such that DLs $\leq$ QLs. The QL is the lowest concentration at which a particular contaminant can be quantitatively measured with acceptable uncertainty (precision and accuracy). Thus the QL is the threshold above which quantitative results may be obtained with a specified degree of confidence. Labs report single numerical values between the DL and QL, qualified by remarks that mean something like "user beware" which is an indication that the relative error for these measurements are high.

In most cases, users either ignore these remarks and incorporate these values (values between DL and QL) as if they were just as valid as values above the QL or they ignore the QL by re-setting QL as the DL. Both approaches have limitations. For instance, using the values between DL and QL as if they are values above QL may result in biased parameter estimates and underestimated standard errors because of failure to account for the extra uncertainty of those measurements. In addition, treating the QL as the DL means ignoring values between these limits, which could result in decreased efficiency because of not using all the available information in the data.

For further research beyond the scope of this dissertation, I would like to extend our methodology to accommodate both DL and QL via

- a model that uses measurements between the DL and QL but accounts for its extra uncertainty as compared with values above the QL via some kind of weighted analysis.
- a model that ignores the reported values between DL and QL but accounts for interval censoring.

BIBLIOGRAPHY

Aitkin, M. (1999). A General maximum likelihood analysis of variance components in the generalized linear models. *Biomatrics 55*, 117–128.

Bandyopadhyay, D., V. Lachos, M. Castro, and D. Dey (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal 54*, 405–425.

Bates, D., M. Maerchler, B. Bolker, S. Walker, R. H. B. Christensen, and H. Singmann (2014). *Linear mixed effects models using Eigen and S4*. R package version 1.1-6.

Bates, D. and D. Shakar (2007). *lme4: Linear mixed effects models using S4 classes.* R package version 0.9975-13.

Bates, D. M. and J. Pinheiro (1997). Computational methods for multilevel modelling. Technical report, Bell Labs, Murray Hill, NJ.

Bland, J. and D. Altman (1994). Correlation, regression, and repeated data. *British Medical Journal 1994*, 308:896.

Box, G., G. Jenkins, and G. Reinsel (2004). *Time series analysis: forecasting and control.* 3rd ed., Holden-Day, San Francisco.

Carroll, R. and D. Ruppert (1988). *Transformation and weighting in regression.* Chapman & Hall, New York.

Commodore, A., S. Hartinger, C. Lanata, D. Mausezahl, A. Gil, D. Hall, M. Aguilar-Villalobos, and L. Naeher (2013). A pilot study characterizing real time exposures to particulate matter and carbon monoxide from cookstove related woodsmoke in rural Peru. *Atmospheric Environment 79*, 380–384.

Comulada, W. and R. Weiss (2010). Sample size and power calculations fo correlations between bivariate longitudinal data. *Statistics in Medicine 29*, 2811–2824.

Cook, R., L. Zeng, and G. Yi (2004). Marginal analysis of incomplete longitudinal binary data: a cautionalry note on LOCF imputation . *Biometrics 60*, 820–828.

Davidian, M. and D. Giltinan (1995). *Nonlinear models for repeated measurement data.* Boca Raton, FL:Chapman and Hall.

Dietz, E. and D. Böhning (1995). Statistical inference based on a general model of unobserved heterogeneity. *In Statistical Modelling.* B. J. Francis, R. Hatzinger, G. U. H. Seeber, and G. Steckel-Berger (eds). New York: Springer-Verlag.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. Zeger (2002). *Analysis of longitudinal data.* Oxford, England: Oxford University Press.

Finney, D. (1962). Cumulants of truncated multinormal distribution. *Journal of the Royal Statistical Society Series B 24*, 535–536.

Fitzmaurice, G. M., N. Laird, and J. H. Ware (2004). *Applied longitudinal analysis.* Wiley: New Jersey.

Gao, F., P. Thompson, C. Xiong, and J. Miller (2009). Analyzing multivariate longitudinal data using SAS. *Proceedings of the Thirty-first Annual SAS Users Group International Conference, SAS Institute Inc.,2009*, 187–31.

Genz, A., F. Bretz, T. Horthon, T. Miwa, X. Mi, F. Leisch, and F.Scheip (2008). *mvtnorm: Multivariate normal and t distribution.* R package version 0.9-2.

Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Preceedings of the Twenty-Ninth Annual SAS Users Group International Conference, SAS Institute Inc.,2004*, 198–29:1–7.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Assoiation 72*, 320–340.

Helsel, D. (2005). *Nondetect and data analysis, statistics for censored environmetal data.* New Jersey: John Wiley and Sons.

Helsel, D. and R. Gilliom (1986). Estimation of distributional parameters for censored trace level water quality data, 2. verification and application:. *Water Resources Research 22*, 147–155.

Hopke, P., C. Liu, and D. Rubin (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: time series concentrations of pollutants in arctic. *Biometrics 57*, 23–33.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*, 625–629.

Jacqmin-Gadda, H. and R. Thiebaut (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics 1*, 355–368.

Jiang, J. (2007). *Generalized linear models: theory and applications.* New York: Springer.

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

Lau, B. and S. Gange (2004). Methods for the analysis of continuous biomaker assay data with increased sensitivity. *Epidemiology 15*, 724–732.

Lindstrom, M. and D. Bates (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association 83*, 1014–1022.

Little, R., G. Milliken, W. Stroup, and R. Wolfinger (1996). *SAS system for mixed effects models.* SAS Intitutes Inc., Cary NC.

Little, R. and D. Rubin (1987). *Statitical analysis with missing data.* Wiley: New York.

Liu, H. and W. Tu (2012). A Semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. *Annals of Applied Statistics 6*, 1861–1882.

Lorenz, D., S. Datta, and S. Harkema (2011). Marginal association measure for clustered data. *Statistics in Medicine 30*, 3181–3191.

Louis, T. (1982). Finding the oberserved information matrix when using EM algorithm. *Journal of Royal Statistical Society, Ser. B 44*, 226–233.

Lyles, R. and J. W. R. Chuachoowong (2001). Correlating two viral loads assay with know detection limits. *Biometrics 57*, 1238–1244.

Lynn, H. (2001). Maximum likelihood inference for left-censored HIV RNA data. *Statistics in medicine 20*, 33–45.

Magnus, J. and H. Neudecker (1999). *Matrix differential calculus with application in statistics and econometrics.* Chichester: Wiley.

Matos, L., V. Lachos, N.B.Balakrishnan, and F. Labra (2013). Influence diagnostics linear and nonlinear mixed-effects models with censored data . *Computational Statistics and Data Analysis 57*, 450–464.

Orchard, T. and M. Woodbury (1972). A missing information princple, theory and application. *in Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability 1*, 679–715.

Pinheiro, J. and D. Bates (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., D. Bates, S. DeRoy, and D. Shakar (2006). *nlme: Linear and nonLinear mixed effects models*. R package version 3.1-78.

Potthoff, R. and S. Roy (1964). A generalized multivariate analsysis of variance model useful especially for growth curve problems. *Biometrika 51*, 313–326.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R foundation for statistical computing. ISBN 3-900051-07-0.

Rao, S., J. Ku, and K. Rao (1991). Analysis of toxic air contaminant Data containing concerntrations below the limit of detection. *Journal of Air and Waste Management Association 41*, 442–448.

Roy, A. (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical Journal 48*, 286–301.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Ruppert, D. (2005). Comments on maximization by parts in likelihood inference. *Journal of Air and Waste Management Association 100*, 1161–1163.

Saitoh, A., M. Foca, R. M. Viani, S. Hefferman-Vaca, S. Vaida, F. Lujan-Zilbermann, J. Emmanuel, P. J. Deville, J. G., and S. A. Spector (2008). Clinical Outcome in perinatally acquired HIV-infected childrens and adolescents after unstructured treatment interruption . *Pediatrics 121*, e513–e521.

Senn, S., N. Holford, and H. Hockey (2012). The ghost of departed quantities: approaches to dealing with observations below the limit of quantifications. *Statiscs in Medicine 31*, 4285–4295.

Tallis, G. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society, Series, B 23*, 223–229.

Thiebaut, R., H.Jacqmin-Gadda, C. Leport, C. Katlama, D. Costagliola, V. Le Moing, P. Morlat, G. Chene, and the APROCO Study Group (2000). Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left-censoring of HIV RNA measures. *Biostatistics 13*, 271–282.

Thiebaut, R., H. Jacqmin-Gadda, G. Chene, C. Leport, and D. Commenges (2002). Bivariate linear mixed models using SAS PROC MIXED. *Computer Methods and Programs in Biomedicine 69*, 249–256.

Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybid EM for linear and nonlinear mixed effects model with censored response. *Computational Statistics and Data Analysis 51*, 5718–5730.

Vaida, F. and L. Liu (2009a). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Vaida, F. and L. Liu (2009b). *Linear mixed effects models with censored response.* R package version 1.0.

Vaida, F. and L. Liu (2009c). *lmec: Linear mixed effects models with censored responses.* R package version 1.0.

Venables, W. and B. Ripley (1999). *Moden applied statistics with S-PLUS.* 3rd ed., Springer-Verlag, New York.

Verbeke, G. and G. Molenberghs (2001). *Linear mixed models for longitudinal data.* New York: Springer.

Vock, D. M., M. Davidian, and A. A. Tsiatis (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics 0*, 1–13.

Wand, M. (2002). Vector differential calculus in statistics. *The American Statistician 56*, 55–62.

Weiss, R. E. (2010). *Modeling longitudinal data.* New York, Springer.

Wu, L. (2010). *Mixed effects models for complex data.* Chapman and Hall CRC Boca Raton FL.

Xie, X., X. Xue, S. Gange, H. Strickler, M. Kim, and W. H. Study-Group (2012). Estimation and inference on correlations between biomarkers with repeated measures and left-censoring due to minimum detection levels. *Statistics in Medicine 31*(21), 2275–2289.

Yucel, R. (2010). *ML estimation under multivariate linear mixed models with missing values.* R package version 0.3-1.2.