INFERENCES IN MIXED-EFFECTS MODELS WITH MISSING COVARIATES AND APPLICATION TO META-ANALYSIS

By

Akshita Chawla

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics - Doctor of Philosophy

2015

ABSTRACT

INFERENCES IN MIXED-EFFECTS MODELS WITH MISSING COVARIATES AND APPLICATION TO META-ANALYSIS

By

Akshita Chawla

This dissertation consists of four chapters. The first chapter motivates problems of interest and gives a brief literature review.

The second chapter investigates popular methods of testing in linear mixed models. The linear mixed model is prominently used in research involving human and animal subjects. Drawing inferences on model parameters is primarily important for explaining the biological outcomes. Unlike the standard linear regression models, the linear mixed model inference is based on asymptotic theory. Thus a study of finite sample performance of the existing procedure could be of practical interest. This chapter reviews the following popular approaches of testing fixed effects in a linear mixed model: (1) likelihood ratio test, (2) restricted maximum likelihood ratio test (3) Bartlett corrected profile likelihood ratio test (4) Bartlett corrected Cox-Reid likelihood ratio test and (5) Kenward-Roger approximate F-test. The performance of these methods are compared based on Type-I error rate via an extensive simulation study. We conclude that the Kenward-Roger test is the best in preserving Type-I error rate.

The third chapter develops a test for fixed effects in small sample linear mixed model with missing covariates. Partially observed variables are common in scientific research. Ignoring the subjects with partial information may lead to biased and or inefficient estimators, and consequently any test based only on the completely observed subjects may inflate the error probabilities. Missing data issue has been extensively considered in the regression model, especially in the independently identically (IID) data setup. Relatively less attention has

been paid for handling missing covariate data in the linear mixed effects model– a dependent data scenario. In case of complete data, Kenward-Roger's F test is a well established method for testing of fixed effects in a linear mixed model. In this chapter, we present a modified Kenward-Roger type test for testing fixed effects in a linear mixed model when the covariates are missing at random. In the proposed method, we attempt to reduce bias from three sources, the small sample bias, the bias due to missing values, and the bias due to estimation of variance components. The operating characteristics of the method is judged and compared with two existing approaches, listwise deletion and mean imputation, via simulation studies.

The fourth chapter applies the random effects model to meta-analysis of rare event data. In clinical trials and many other applications, meta-analysis is mainly conducted to summarize the effect size on primary endpoints or on key secondary endpoints. However, there are times when safety endpoints such as risk of complications in pancreatic surgery or the risk of myocardial infarction (MI) is the main point of interest in a drug study. As these types of safety endpoints are rare in nature, most studies report zero such incidences; the general statistical framework of meta-analysis based on large sample theory falls apart to combine the effect size. As a workaround, either trials with both arms having zero events are deleted or a 0.5 correction is applied. In a randomized control trial (RCT) set-up, Cai, Parast and Ryan, 2010, proposed methods based on Poisson random effects models to draw inferences on relative risks for two arms with rare event data. In this chapter, we give the general framework of how their assumption of having RCT can be further relaxed and utilized for non-RCT studies to draw inferences for two treatments. We also develop two new approaches based on zero inflated Poisson random effects models that are more appropriate with excessive zero counts data.

ACKNOWLEDGMENTS

I wish to express my sincere thanks to my advisor Professor Tapabrata Maiti for guiding and inspiring me through my graduate study. I thank him for suggesting interesting and challenging problems and also for providing me with innumerable opportunities through my time here.

I thank Professor Hira Koul for having given me the opportunity to pursue graduate study at Michigan State University and for always being available for guidance. I also thank him and his family for being so welcoming. I thank Professor Chae Young Lim, Professor Ping-Shou Zhong and Professor Robert Tempelman for serving on my dissertation committee.

I thank my parents for their constant love and support, without which this work would not have been possible.

Finally I would like to thank my friends Abhishek, Shaheen and Atishi for making this journey joyous.

This research has been partly supported by NSF grants DMS-1106450 and SES-0961649, for which I am grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Longitudinal data and repeated measurements arise frequently in applied sciences. In such studies, the observations are collected over time. Linear mixed models (Laird and Ware, 1982) are well suited for such data, which display heterogeneity of responses to treatment. The statistical test procedures for model parameters are usually based on large sample theory. However, very often longitudinal data arising in medical data involve very few number of subjects along with unbalanced data. Therefore, it is essential to investigate the performance of mixed models in the context of small to moderate samples.

In a linear mixed model, the model parameters are typically estimated using ML or REML estimates (Patterson and Thompson, 1977; Harville, 1977). In order to draw inferences on the fixed effects, asymptotic theory is used. Any method that ignores the uncertainty of the estimated variance component in the inference of the fixed effect parameter $\boldsymbol{\beta}$ will result in deflated confidence interval, specially when sample size is small. Therefore, in a mixed model setup, we have two biases to deal with : small sample bias and variance bias. The most commonly used statistical tool is the likelihood ratio test, which does not account for any of the above mentioned biases. The null distribution of the likelihood ratio test statistic is approximated as a χ^2 distribution up to the approximation $O_p(n^{-1})$, where n denotes the sample size. However, when the sample size is small, the inferences using this approximation could be highly inaccurate. Bartlett, 1937 proposed a correction to the likelihood ratio test to obtain a better approximation to the null distribution. Zucker et al.,

2000 proposed an adjustment to the likelihood ratio statistic in case of a linear mixed model, the null distribution of which is approximated to a χ^2 upto the approximation of $O_p(n^{-2})$ (Barndorff-Neilsen and Hall, 1988), thus addressing the first bias due to small sample size. To address the second bias, Cox and Reid, 1987 proposed forms of adjusted likelihood which reduce the effect of nuisance parameters on the estimation of fixed effects. Zucker et al., 2000 also derived the Bartlett correction for the Cox-Reid adjusted likelihood in case of linear mixed effects model, thus addressing both biases.

The first problem of approximating the small sample precision of $\widehat{\boldsymbol{\beta}}$ has also been studied by Kackar & Harville (1984), Harville (1985) and Harville & Jeske (1992). In order to address the second problem of underestimation of variance of $\widehat{\boldsymbol{\beta}}$, the estimator of $\boldsymbol{\beta}$, an approximate t or an F-statistic is used. In general, to test the hypothesis $H_0 : L^T\boldsymbol{\beta} = 0$, a Wald-type test statistic is used which is approximated as an F distribution with the numerator degrees of freedom as rank(L) and the denominator degrees of freedom being estimated from the data. The degrees of freedoms were used to be estimated by the residual method, containment method or the Satterthwaite-type approximation. Later Kenward & Roger (1997), in their seminal paper, proposed a different approach of dealing with those two problems. They combined the results from Kackar and Harville, 1984 to show that the variance of estimated fixed effects can be partitioned into two components and approximated such that the variance bias due to estimation of variance components is accounted for. They constructed a Wald-type test statistic and further reduced the small sample bias by approximating its null distribution to an F distribution up to $O_p(n^{-5/2})$. The test statistic and the denominator degrees of freedom of the F distribution derived by them uses the adjusted estimator of the covariance matrix of the estimated fixed effects. It is a very widely used method of testing in small samples and has been incorporated in the statistical software SAS.

In Chapter 2, we review different approaches of testing fixed effects namely likelihood ratio test and its Bartlett correction, Cox-Reid adjusted likelihood ratio test and its Bartlett-correction and the Kenward-Roger approximate F-test. All these tests are approximate tests that aim at obtaining the most accurate solution by reducing the effect of nuisance parameters and accounting for small sample bias. This chapter addresses the practical problem of choosing the most robust approach to test the fixed effects in a linear mixed model when the sample size is small. Our focus is on the case when the covariance matrix has a linear structure. The parameter of interest varies from a single parameter case to a multi-parameter case. To assess the performance of the above methods, we present an extensive simulation study on a wide range of practical problems.

Partially missing variable is common in clinical studies. Depending on the missingness mechanism missing data can be classified into three main categories, missing completely at random (MCAR) where the missingness mechanism is completely random and does not depend on any variables, missing at random (MAR) where missingness mechanism depends only on the observed values of the variables, and missing not at random (MNAR) where the missingness mechanism may depend on the unobserved values of the variables along with the completely observed variables. The first two mechanisms are ignorable in a likelihood framework while the third is not. Missing values may occur in a response variable, in a covariate or in both. There are several methods of dealing with missing responses depending on the missing mechanism. Among the most common methods are listwise deletion where the subject with any missing value is deleted from the study, mean or multiple imputation and the maximum likelihood method that do not involve modelling the missing mechanism when the missing mechanism is either MCAR or MAR. Padilla and Algina (2004) showed via a simulation study that when the responses are missing (MCAR or MAR) in a small

sample setup, Kenward-Roger test preserves the Type I error rates for a single factor within-subject ANOVA as compared to the Hotelling-Lawley-McKeon procedure (McKeon, 1974). Therefore, Kenward Roger F test is proven to be superior even in case of missing responses. In several aricles, Ibrahim and coworkers considered partially missing response or covariate in regression models (Ibrahim, 1990; Ibrahim & Chen, 2001; Stubbendick & Ibrahim, 2003; Ibrahim et al., 2005; Chen et al., 2008). However, little has been done in case of missing covariates in a linear mixed model.

In Chapter 3, we deal with missing covariate data, and assume that the covariate data are missing at random (MAR). We assume a parametric distribution on the partially missing covariate. Under this set up, we derive a Kenward-Roger type adjusted test for the fixed effects in a linear mixed model in case of small samples. To derive an improved test, we need to overcome three biases: the variance bias (such that the variability of the variance components is taken into account), the small sample bias and the bias due to missingness. Kenward & Roger (1997) considered the first two types of biases in their seminal paper for fully observed data. Here, we need to develop a similar test by accounting for the missingness in the covariates. We consider the bias in the estimation of the covariance matrix of $\widehat{\boldsymbol{\beta}}$ and then propose a new Wald statistic that uses this new adjusted covariance matrix. Further, this Wald statistic is approximated to an F distribution with the degrees of freedom calculated using the new covariance matrix. To demonstrate the accuracy of this method, we compare this method to the listwise deletion and imputation methods through simulation studies.

In Chapter 4, we look at application of random-effects models in meta-analysis of rare events. In meta-analysis, the study specific effect sizes such as mean differences, relative risks or odds ratios are combined using either the fixed effects or random effects models.

In fixed effects models, the study specific effects are assumed to be identical across studies; however, in random effects models this assumption is relaxed. Using the large sample theory, DerSimonian and Laird, 1986 proposed a random effects model for estimating the mean of the random effect distribution. However, the random effects model based on large sample theory does not work well when the study outcomes across studies have low incidences or zero events. In case of zero events, Mantel and Haenszel method, 1959 requires zero-cell 0.5 corrections to obtain the estimates of the effect size. Peto method, 1985 does not require any such corrections; however trials with no events in both arms are not given any weight and are hence excluded from the analysis. For the details we recommend the discussions from Bradburn *et al.*, 2007 and Lane, 2013.

A plethora of literature is available that discusses different methods applicable for meta-analysis; however, very few discuss the methods applicable for rare events data (see Bradburn *et al.*, 2007; Daniels and Gatsonis, 1999; Lane, 2013. For the fixed effects model Tian *et al.*, 2009 proposed an exact approach. Using Poisson random effects model, Cai, Parast and Ryan, 2010 proposed an elegant approach to calculate the effect size for rare event data in case of randomized trials. However, the models proposed by them are unable to handle cases beyond a certain threshold of occurrence of zero events. They also require the trials used in their analysis to have two groups per trial. In this chapter, our main contribution is to propose zero-inflated Poisson models. Furthermore, our models will also be capable of analysing trials where both groups are not available for each trial. We show that our models are highly suited for the case of extremely rare events and illustrate their advantages via a comprehensive simulation study.

Chapter 2

Small sample inferences in linear mixed models

We consider the following linear mixed model with the response vector Y. The design matrix for the fixed effects is denoted by X while the design matrix for the random effects is denoted by Z. The model error vector will be denoted by $\boldsymbol{\epsilon}$. The response vector is related to X, Z, $\boldsymbol{\epsilon}$ by the following relation,

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}. \tag{2.0.1}$$

In this relation, $\boldsymbol{\beta} = (\beta_1, ..., .\beta_p)' \in \boldsymbol{R}^p$ is the vector of fixed effects, $\mathbf{b} = (b_1, ..., b_q)' \sim \mathcal{N}_q(0, G)$ is a q-dimensional random effects vector where G involves $q(q+1)/2$ unknown parameters, and $\boldsymbol{\epsilon}$=$(\epsilon_1, ..., \epsilon_n,)'$ is an n-dimensional error vector, the elements of which are assumed to be i.i.d normal with mean 0 and unknown variance σ_e^2, i.e. $\boldsymbol{\epsilon} \sim \mathcal{N}_n(0, \sigma_e^2)$. Furthermore, $\mathbf{b}$ and $\boldsymbol{\epsilon}$ are assumed to be independent. The model (2.0.1) can be rewritten as follows

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \tag{2.0.2}$$

where $\boldsymbol{\epsilon}^* \sim \mathcal{N}_n(0, \Sigma(\boldsymbol{\sigma}))$, $\Sigma(\boldsymbol{\sigma}) = ZGZ' + \sigma_e^2$ and $\boldsymbol{\sigma}$ is an $r = q(q+1)/2 + 1$ dimensional vector of variance-covariance parameters.

In the following, our objective is to test the hypothesis

$$H_0 \quad : L'\boldsymbol{\beta} = 0, \tag{2.0.3}$$

where L' is any $l \times p$ matrix and $\boldsymbol{\beta}$ is as defined in (2.0.1). We consider the case when Σ is linear in $\boldsymbol{\sigma}$. If $\boldsymbol{\sigma}$ is known and fixed, then the problem is trivial since the model reduces to a linear model. However, in the case of linear mixed models, $\boldsymbol{\sigma}$ is unknown and has to be estimated. In the following, we discuss various test procedure for the hypothesis (2.0.3) which are popularly used in the literature.

2.1 Likelihood ratio test (LRT)

The most widely used method for testing is the likelihood ratio test using ML estimation. The maximum likelihood estimates of the parameters of interest are computed by maximizing the profile likelihood which is defined below. It follows from classical likelihood theory, that under standard regularity conditions, the testing of fixed effects is carried out by approximating the likelihood ratio to a chi-square distribution (Cox and Hinkley, 1990) We begin with the likelihood of the model (2.0.2), which can be written as follows

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \frac{1}{2\pi^{n/2}|\Sigma(\boldsymbol{\sigma})|^{1/2}} \exp\left\{\frac{-1}{2}(\mathbf{Y} - X\boldsymbol{\beta})'\Sigma(\boldsymbol{\sigma})^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\right\}.$$

Hence, the log likelihood is given by

$$l(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \frac{-1}{2}\left\{\log|\Sigma(\boldsymbol{\sigma})| + (\mathbf{Y} - X\boldsymbol{\beta})'\Sigma(\boldsymbol{\sigma})^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\right\}. \tag{2.1.4}$$

By maximizing (2.1.4), we obtain the estimate of $\boldsymbol{\beta}$ in terms of $\boldsymbol{\sigma}$ as follows,

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma}) = (X'\Sigma(\boldsymbol{\sigma})^{-1}X)^{-1}X'\Sigma(\boldsymbol{\sigma})^{-1}\mathbf{Y}.$$

Then the profile likelihood is given by

$$\begin{aligned} l_p(\boldsymbol{\sigma}) &= l(\tilde{\boldsymbol{\beta}}, \boldsymbol{\sigma}) \\ &= \frac{-1}{2}\left\{\log|\Sigma(\boldsymbol{\sigma})| + (\mathbf{Y} - X\tilde{\boldsymbol{\beta}})'\Sigma(\boldsymbol{\sigma})^{-1}(\mathbf{Y} - X\tilde{\boldsymbol{\beta}})\right\}. \end{aligned}$$

Maximizing l_p w.r.t $\boldsymbol{\sigma}$ gives the maximum likelihood estimate of $\boldsymbol{\sigma}$ as $\hat{\boldsymbol{\sigma}}_{ML}$. Substituting it back in $\tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma})$ we obtain the maximum likelihood estimate of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma})\big|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}_{ML}}$.

The test statistic for testing (2.0.3) is given by,

$$LR = 2[l(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{ML}) - l(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\sigma}}_0)], \tag{2.1.5}$$

where $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\sigma}}_0) = \underset{\substack{\beta\in H_0 \\ \sigma\in\Omega}}{\text{argmax}}\, l(\beta, \sigma)$.

Under standard regularity conditions and (2.0.3), the LR statistic converges in distribution to a χ^2_l distribution and

$$E(LR) = l + O(n^{-1})$$

where l=rank(L). Therefore, the rejection region is given by

$$LR > \chi^2_{1-\alpha,l}. \tag{2.1.6}$$

2.2 Restricted maximum likelihood ratio test (RM-LRT)

Restricted maximum likelihood (REML) (Patterson and Thompson, 1971) is well established as a method for estimating the parameters of a linear mixed model. In contrast to earlier maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters. The testing procedure using REML is described in Welham and Thompson, 1997. Here the marginal log likelihood is used to estimate $\boldsymbol{\sigma}$ which is given by,

$$l_R(\boldsymbol{\sigma}) = \log \int L(\boldsymbol{\beta}, \boldsymbol{\sigma}) d\boldsymbol{\beta},$$

where

$$\int L(\boldsymbol{\beta}, \boldsymbol{\sigma}) d\boldsymbol{\beta} = \int \frac{1}{(2\pi)^{n/2}} |\Sigma(\boldsymbol{\sigma})|^{-1/2} \exp\big\{\frac{-1}{2}(\mathbf{Y} - X\boldsymbol{\beta})'\Sigma(\boldsymbol{\sigma})^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\big\} d\boldsymbol{\beta}. \tag{2.2.7}$$

Therefore, the restricted maximum likelihood is given by,

$$\begin{aligned} l_R(\boldsymbol{\sigma}) &= -\frac{1}{2}[\log|\Sigma| + (\mathbf{Y} - X\tilde{\boldsymbol{\beta}})'\Sigma^{-1}(\mathbf{Y} - X\tilde{\boldsymbol{\beta}})] - \frac{1}{2}\log|X'\Sigma^{-1}X| + const. \\ &= l_p(\boldsymbol{\sigma}) - \frac{1}{2}\log|A| + const. \end{aligned} \tag{2.2.8}$$

We can maximize (2.2.8) to obtain the REML estimate of $\boldsymbol{\sigma}$ which is denoted by $\hat{\boldsymbol{\sigma}}_{REML}$. The estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (X'\Sigma^{-1}(\hat{\boldsymbol{\sigma}})X)^{-1}X'\Sigma^{-1}(\hat{\boldsymbol{\sigma}})\mathbf{Y}. \tag{2.2.9}$$

The REML likelihood under H_0 is given by

$$l_{R_0}(\boldsymbol{\sigma}) = -\frac{1}{2}[\log|\Sigma| + (\mathbf{Y} - X_0\tilde{\boldsymbol{\beta}}_0)'\Sigma^{-1}(\mathbf{Y} - X_0\tilde{\boldsymbol{\beta}}_0)] - \frac{1}{2}\log|X_0'\Sigma^{-1}X_0| + const. \tag{2.2.10}$$

where X_0 and $\boldsymbol{\beta}_0$ are portions of X and $\boldsymbol{\beta}$ such that H_0 is satisfied. If we consider the difference of the two log-likelihoods (2.2.8) and (2.2.10), there is no common basis of comparison. Therefore Welham and Thompson, 1997 proposed a test by fitting the model under the null as usual, and thereafter looking at the change in the log-likelihood when the full model is fitted i.e.

$$l_{R_1}(\boldsymbol{\sigma}) = -\frac{1}{2}[\log|\Sigma| + (\mathbf{Y} - X\tilde{\boldsymbol{\beta}})'\Sigma^{-1}(\mathbf{Y} - X\tilde{\boldsymbol{\beta}})] - \frac{1}{2}\log|X_0'\Sigma^{-1}X_0| + const. \tag{2.2.11}$$

Therefore, the test statistic based on the REML is given by $LR_{REML} = -2(l_{R_0}(\boldsymbol{\sigma}) - l_{R_1}(\boldsymbol{\sigma}))$ and under standard regularity conditions, LR_{REML} can be approximated to a χ^2_l distribution.

Although, the REML based test statistic is also approximated as a χ^2 distribution up to the order of n^{-1}, but it provides unbiased estimates of the covariance parameters, resulting in better inference as compared to LRT.

2.3 Bartlett corrected likelihood ratio tests

2.3.1 Bartlett-corrected profile likelihood ratio test (BC-LRT)

The LR statistic (2.1.5) is distributed as χ^2 to order $O(n^{-1})$. It is known that the first order approximation does not work well for small sample sizes. So in order to achieve higher accuracy, Bartlett, 1937 proposed multiplying the LR statistic by a constant, thus giving the Bartlett corrected test statistic,

$$LR^* = \frac{LR}{1 + C/l}, \tag{2.3.12}$$

where LR is given in (2.1.5), l is the number of fixed effects to be tested and C is a constant of order n^{-1} such that

$$E(LR^*) = l + O(n^{-3/2}).$$

In order to calculate C, reparametrization is required such that the parameters of interest are orthogonal to the remaining parameters i.e. the expected mixed partial derivative of the likelihood with respect to the parameter of interest and any one of the nuisance parameters is zero. To test for the some fixed effects $\boldsymbol{\psi} = (\beta_1, \beta_2, \cdots, \beta_l)$ i.e. $H_0 : \psi = \mathbf{0}$ against $H_1 : \psi \neq \mathbf{0}$, a reparametrization is used in which the parameters of interest $\boldsymbol{\psi}$ and the nuisance parameters are orthogonal. Let $\tilde{X}_l$ denote the first l columns of X and $\tilde{X}_{n-l}$ be the rest of the columns. After reparametrization, the expression for C is given by

$$C = tr(D^{-1}(-\frac{1}{2}M + \frac{1}{4}P - \frac{1}{2}(\gamma + \nu)\tau')).$$

Where the elements of D, M and P are given by,

$$\begin{aligned} D &= \frac{1}{2}tr(\dot{\Sigma}^j\dot{\Sigma}_k), \\ M &= tr\big((\tilde{\tilde{X}}_l{}'\dot{\Sigma}^j\tilde{\tilde{X}}_l)^{-1}(\tilde{\tilde{X}}_l{}'\ddot{\Sigma}^{jk}\tilde{\tilde{X}}_l + 2\dot{\tilde{X}}_l{}'\dot{\Sigma}^j\tilde{\tilde{X}}_l)\big), \\ P &= tr\big((\tilde{\tilde{X}}_l{}'\dot{\Sigma}^j\tilde{\tilde{X}}_l)(\tilde{\tilde{X}}_l{}'\Sigma^{-1}\tilde{\tilde{X}}_l)^{-1}(\tilde{\tilde{X}}_l{}'\dot{\Sigma}^k\tilde{\tilde{X}}_l)(\tilde{\tilde{X}}_l{}'\Sigma^{-1}\tilde{\tilde{X}}_l)^{-1}\big), \end{aligned}$$

and τ, γ and ν are vectors whose j^{th} elements are given by $tr((\tilde{\tilde{X}}_l{}'\Sigma^{-1}\tilde{\tilde{X}}_l)^{-1}(\tilde{\tilde{X}}_l{}'\dot{\Sigma}^j\tilde{\tilde{X}}_l))$, $tr(D^{-1}A^{(j)})$ and $tr((\tilde{X}'_{n-l}\Sigma^{-1}\tilde{X}_{n-l})^{-1}(\tilde{X}'_{n-l}\dot{\Sigma}^j\tilde{X}_{n-l}))$ respectively. In the above expressions,

$$\begin{aligned} \dot{\Sigma}_j &= \frac{\partial\Sigma}{\partial\boldsymbol{\sigma}_j}, \qquad \dot{\Sigma}^j = \frac{\partial\Sigma^{-1}}{\partial\boldsymbol{\sigma}_j} = -\Sigma^{-1}\dot{\Sigma}_j\Sigma^{-1}, \\ \ddot{\Sigma}_{jk} &= \frac{\partial^2\Sigma}{\partial\sigma_j\partial\sigma_k}, \qquad \ddot{\Sigma}^{jk} = \frac{\partial^2\Sigma^{-1}}{\partial\sigma_j\partial\sigma_k} = -2\dot{\Sigma}^k\dot{\Sigma}_j\Sigma^{-1} - \Sigma^{-1}\ddot{\Sigma}_{jk}\Sigma^{-1}, \\ \tilde{\tilde{X}}_l &= [I - \tilde{X}_{n-l}(\tilde{X}'_{n-l}\Sigma^{-1}\tilde{X}_{n-l})^{-1}\tilde{X}'_{n-l}\Sigma^{-1}]\tilde{X}_l \text{ and } \dot{\tilde{X}}_l = \frac{\partial\tilde{\tilde{X}}_l}{\partial\sigma_j}. \end{aligned}$$

It has been shown that LR^* is χ^2_l distributed, up to an error of order n^{-2} in Barndorff-Neilsen and Hall[6]. Therefore, it is expected to be more accurate than profile likelihood ratio test especially for small samples.

2.3.2 Bartlett-corrected Cox-Reid profile likelihood ratio test (BC-CRT)

Bartlett corrected LRT takes care of the bias due to small sample size to some extent. However, in the linear mixed model, specially in small samples, inference for the fixed effects is highly affected by the estimation of nuisance parameters. To reduce the influence of

nuisance parameters, Cox and Reid[7] proposed an adjusted form of the profile likelihood. A transformation of the parameter vector is required such that the parameters of interest are orthogonal to the nuisance parameters. Let ψ denote the $(l \times 1)$ vector of parameter of interest and ϕ be the transformed nuisance parameter vector. The form of Cox-Reid adjusted log-likelihood is given by

$$l_{CR}(\psi) = l(\psi, \hat{\phi}(\psi)) - \frac{1}{2} log\{| - l_{\phi\phi}(\hat{\phi}(\psi))|\}$$

where $\hat{\phi}(\psi)$ is the maximum likelihood estimator of ϕ for a fixed value of ψ and $l_{\phi\phi}$ is the matrix of second derivatives of l with respect to ϕ. When testing for the entire fixed effects vector in a linear mixed models, the Cox-Reid adjusted log-likelihood is the same as the REML log-likelihood given in (2.2.8). The Cox-Reid adjusted likelihood ratio statistic for testing $H_0 : \psi = \psi_0 = \mathbf{0}$ is given by

$$LR_{CR} = 2\{l_{CR}(\tilde{\psi}) - l_{CR}(\psi_0)\}. \tag{2.3.13}$$

Under standard regularity conditions, the LR_{CR} converges in distribution to χ^2_l up to an error of $O(n^{-1})$. DiCiccio and Stern, 1994 gave the general Bartlett corrected likelihood for Cox-Reid likelihood and Zucker et al., 2000 worked out the Bartlett corrected test statistic for testing $H_0 : \psi = \mathbf{0}$ in case of linear mixed models, which is given by

$$LR^{\star}_{CR} = \frac{LR_{CR}}{1 + C^{\star}/p}, \tag{2.3.14}$$

where

$$C^{\star} = tr(D^{-1}(-M + \frac{1}{4}P + \gamma^{\star}\tau')).$$

Here D, M, P and τ are same as given as section 2.3.1 and the j^{th} element of $\gamma^{\star}$ is given by $-tr(D^{-1}tr(\dot{\Sigma}^{k}\dot{\Sigma}_{j}\Sigma^{-1}\dot{\Sigma}_{l}))$.

The details of the proof can be found in Melo et al., 2009 and Zucker et al., 2000. The $LR^{\star}_{CR}$ is approximated as χ^2_l distributed up to an error of $O(n^{-2})$. Therefore Bartlett corrected Cox-Reid likelihood ratio test addresses both issues; it reduces the effect of nuisance parameters while taking care of small sample bias.

2.4 Kenward-Roger approximate F-test (KRT)

The above four methods were based on the likelihood ratio test statistic. Kenward and Roger, 1997 propose a Wald-type test statistic. They use the REML estimation to estimate the covariance parameters, described in section (2.2). The estimate of $\boldsymbol{\beta}$ that uses the REML estimate of Σ is given by (2.2.9). To make inference for the fixed effects $\boldsymbol{\beta}$, the covariance matrix of its asymptotic distribution is used in a Wald-type test statistic. However the variability in the estimate of Σ is not taken into account. This has serious repercussions in small sample studies in the sense that it can impact the precision of the fixed effects significantly. Kenward and Roger, 1997 used an estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$ which is adjusted for small sample bias. They suggested that the variability of $\hat{\boldsymbol{\beta}}$ can be partitioned into two components,

$$V[\hat{\boldsymbol{\beta}}] = \Phi_A = \Phi + \Lambda,$$

where $\Phi(\sigma)$ is the covariance matrix of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, the estimate of which is given by $\hat{\Phi} = (X'\Sigma^{-1}(\hat{\boldsymbol{\sigma}})X)^{-1}$ and Λ accounts for the variability in $\hat{\boldsymbol{\sigma}}$. Let $\boldsymbol{\sigma}$=$(\sigma_1, \ldots, \sigma_r)'$. Kackar and Harville, 1984 show that the value of Λ can be approximated by

$$\Lambda = \Phi \left(\sum_{i=1}^{r} \sum_{j=1}^{r} w_{ij}(Q_{ij} - P_i \Phi P_j) \right) \Phi + O(n^{-5/2}) \tag{2.4.15}$$

where

$$P_i = X' \frac{\partial \Sigma^{-1}}{\partial \sigma_i} X = X' \dot{\Sigma}^i X \text{ and } Q_{ij} = X' \frac{\Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\Sigma^{-1}}{\partial \sigma_j} X = X' \dot{\Sigma}^i \Sigma \dot{\Sigma}^j X$$

and w_{ij} is the $(i,j)^{th}$ element of $V(\hat{\boldsymbol{\sigma}})$. The bias in $\hat{\Phi}$ is approximated using a Taylor series expansion about $\hat{\boldsymbol{\sigma}}$ which gives

$$\begin{aligned} E[\hat{\Phi}] &= \Phi + \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{\partial^2 \Phi}{\partial \sigma_i \partial \sigma_j} + O(n^{-5/2}) \\ &\approx \Phi + \Phi \left(\sum_{i=1}^{r} \sum_{j=1}^{r} w_{ij}(P_i \Phi P_j - Q_{ij}) \right) \Phi + \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{r} w_{ij} \Phi R_{ij} \Phi \end{aligned} \tag{2.4.16}$$

Hence, using 2.4.15 and 2.4.16, the new estimator of the covariance matrix of $\hat{\beta}$ is given by

$$\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Phi} \left[\sum_{i=1}^{r} \sum_{j=1}^{r} \hat{w}_{ij}(\hat{Q}_{ij} - \hat{P}_i \hat{\Phi} \hat{P}_j - \frac{1}{4} \hat{R}_{ij}) \right] \hat{\Phi}$$

where

$$E[\hat{\Phi}_A] = Var(\hat{\boldsymbol{\beta}}) + O(n^{-5/2})$$

Further, to draw inferences about fixed effects; l linear combinations of $(\boldsymbol{\beta})$: $\boldsymbol{L\beta}$, $\mathbf{L}$ being a $(l \times p)$ fixed matrix, Kenward and Roger approximated the distribution of the Wald-type test statistic which uses the new adjusted covariance matrix. The Wald type pivot is given

by $F = \frac{1}{l}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' L (L' \hat{\Phi}_A L)^{-1} L' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. With the help of a Taylor series expansion, the expected value and the variance of this test statistic is given by

$$\begin{aligned} E[F] &= 1 + \frac{A_2}{l} + O(n^{-3/2}) \\ Var[F] &= \frac{2}{l}(1 + B) + O(n^{-3/2}), \end{aligned}$$

where $B = \dfrac{1}{2l}(A_1 + 6A_2)$,

$A_1 = \displaystyle\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} tr(\hat{\Theta}\hat{\Phi}\hat{P}_i\hat{\Phi}) tr(\hat{\Theta}\hat{\Phi}\hat{P}_j\hat{\Phi}),\ A_2 = \displaystyle\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} tr(\hat{\Theta}\hat{\Phi}\hat{P}_i\hat{\Phi}\hat{\Theta}\hat{\Phi}\hat{P}_j\hat{\Phi})$

for $\Theta = L(L'\hat{\Phi}L)^{-1}L'$. (2.4.17)

In order to estimate the denominator degrees of freedom (m) and the scale factor (λ), such that $F^\star = \lambda F \sim F(l, m)$, the first two moments of $F^\star$ are matched to those of the approximating F distribution. Therefore the following results are obtained,

$$m = 4 + \frac{2 + l}{l\rho - 1} \text{ where } \rho = \frac{V[F]}{2E[F]^2},$$
$$\text{and } \lambda = \frac{m}{E[F](m-2)}.$$

The detailed proofs can be found in Kenward and Roger, 1997 and Alnosaier, 2007. The overall procedure can be applied to construct tests and confidence intervals for fixed effects.

On comparison of the Bartlett corrected LRT and the Kenward Roger test, we observe that the Bartlett corrected test uses the traditional likelihood ratio test and approximates

it to a χ^2 approximation which is rescaled using its expected value. On the other hand, Kenward-Roger test uses a Wald-type statistic and approximates it to an F distribution while making use of the new adjusted covariance matrix. Closely observing the two formulas we get,

$$E(LR) = 1 + C/l + O(n^{-3/2}),$$
$$E(F) = 1 + A_2/l + O(n^{-3/2})$$

where both C and A_2 are trace of matrices constructed using first and second order derivatives of Σ.

2.5 Simulation Study

The performance of the profile likelihood ratio test (LRT), restricted maximum likelihood ratio test (RM-LRT), Bartlett-corrected profile likelihood ratio test (BC-LRT), Bartlett-corrected Cox Reid adjusted likelihood ratio test (BC-CRT) and Kenward-Roger approximate F test (KRT) in testing the fixed effects of a linear mixed model is assessed through simulation studies on the basis of Type-I error rates. We have conducted simulation studies for two different settings. In both the settings, the covariance matrix of random effects has a linear structure and so does Σ. In each study, we will run the simulations for two cases: (i) testing for the whole vector of fixed effects (ii) testing for a single fixed effect. For each case, 5000 data sets are generated.

2.5.1 Study I: longitudinal study

The first simulation study is a longitudinal study where the data consists of a growth data for 11 girls and 16 boys. Measurements are taken over four time intervals (age of the person): 8, 10, 12 and 14. The data set up is taken from Verbeke and Molenberghs, 2000. Let x_i denote the indicator variable for sex and t_j be the time points at which the observation is recorded. The model can be expressed as

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_{10}t_j + \beta_{11}t_j x_i + b_0 + b_1 t_j + \epsilon_{ij}.$$

In the matrix notation $Y_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b} + \epsilon_i$, the fixed effect design matrix X_i and the random-effects design matrix can be written as

$$X_i = \begin{pmatrix} 1 & x_i & 8 & 8x_i \\ 1 & x_i & 10 & 10x_i \\ 1 & x_i & 12 & 12x_i \\ 1 & x_i & 14 & 14x_i \end{pmatrix} \quad Z_i = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}$$

The measurement error structure $\Sigma_i = \sigma^2 I_4$. We assume an unstructured covariance matrix G for the random effects b_i given by

$$G = \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_2 & \omega_3 \end{pmatrix}$$

Therefore, the resulting covariance matrix of Y_i is given by $Z_i G Z_i' + \sigma^2 I_4$, hence the covariance matrix has a linear structure. We focus on $\boldsymbol{\beta}$, the full vector of fixed effects, and on β_{11}, the

effect of gender on slope. To simulate data, we assume $x_i = 1$ for girls and $x_i = 0$ for boys. We keep σ^2 fixed at 0.05 and ω_1 fixed at 1. We consider multiple cases by choosing different values of ω_2 and ω_3 ranging from 0 to 1. For each setting, 5000 data sets are generated. We compare the five test procedures on the basis of simulated size based on a 0.05 nominal level. Table 2.1 gives the Type-I error rates of the different test statistics when testing for $H_0 : \boldsymbol{\beta} = 0$ while Table 2.2 gives the same for $H_0 : \beta_{11} = 0$.

Table 2.1 Type-I error rates for $\boldsymbol{\beta}$=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.075	0.071	0.055	0.053	0.053
0	1	0.089	0.074	0.072	0.062	0.056
0.25	0.5	0.060	0.057	0.048	0.046	0.049
0.25	1	0.068	0.065	0.052	0.049	0.047

Table 2.2 Type-I error rates for β_{11}=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.065	0.060	0.047	0.045	0.053
0	1	0.056	0.052	0.045	0.044	0.046
0.25	0.5	0.057	0.054	0.042	0.040	0.047
0.25	1	0.060	0.061	0.058	0.038	0.049

By looking at Tables 2.1 and 2.2, we observe that in general all testing procedures perform

better when testing for single parameter as compared to a multi-parameter case. Also, while testing $H_0 : \boldsymbol{\beta} = 0$, for all values of ω_2 and ω_3, Bartlett Corrected Cox-Reid adjusted likelihood ratio test and Kenward-Roger test preserve the size of the test at 0.05 nominal level best as compared to the other three methods. For testing of a single parameter i.e. $\beta_{11} = 0$, KRT preserves the nominal level while BC-CRT fails to do so when $\omega_2 \neq 0$.

To investigate further, we also consider a set up of smaller size i.e. n=14. A random sample of size fourteen is drawn from the original data set to fix the gender and observation times. Again, simulated size is calculated based on 0.05 nominal level. Tables 2.3 and 2.4 give the comparison of the five test procedures.

Table 2.3 Type-I error rates for $\boldsymbol{\beta}$=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.121	0.090	0.057	0.053	0.050
0	1	0.118	0.098	0.077	0.058	0.056
0.25	0.5	0.102	0.076	0.063	0.055	0.056
0.25	1	0.096	0.084	0.063	0.056	0.049

Table 2.4 Type-I error rates for β_{11}=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.080	0.067	0.054	0.053	0.051
0	1	0.075	0.069	0.059	0.055	0.052
0.25	0.5	0.082	0.078	0.060	0.056	0.056
0.25	1	0.070	0.065	0.074	0.055	0.048

Due to the decrease in sample size, the Bartlett corrections indicate a much better improvement as compared to the standard likelihood ratio tests. However, the performance of all testing procedures except KRT deteriorate in terms of preservation of Type-I error rates. The simulated size of KRT is still very close to the nominal level.

Next, we consider an unbalanced longitudinal study. Little and Rubin, 1987 deleted 9 observations from the data set resulting in 9 (out of 27) incomplete subjects. Deletion is confined to measurements taken at age 10. So, we repeat the simulation study for incomplete data. Table 2.5 and Table 2.6 give the Type-I error rates of the testing procedures.

Table 2.5 Simulated size and simulated power for $\boldsymbol{\beta}$=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.061	0.055	0.047	0.045	0.046
0	1	0.090	0.080	0.073	0.061	0.054
0.25	0.5	0.071	0.062	0.050	0.048	0.048
0.25	1	0.076	0.068	0.057	0.053	0.051

Table 2.6 Type-I error rates for β_{11}=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ω_2	ω_3	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0	0.5	0.067	0.061	0.039	0.038	0.046
0	1	0.063	0.057	0.049	0.046	0.052
0.25	0.5	0.065	0.060	0.058	0.054	0.048
0.25	1	0.066	0.061	0.050	0.046	0.054

The observation remains the same that for a multiparameter case both KRT and BC-CRT seem to be work equally well for all values of ω_2 and ω_3, but for a single parameter case KRT seems to be the best choice. Again, from the incomplete data set, fourteen subjects are chosen at random. Therefore we have a few incomplete subjects in the data. A similar simulation study was conducted and the results were similar to those obtained in the previous cases (not shown here for the sake of brevity). Therefore, for a longitudinal study (balanced or unbalanced) with small sample sizes, we recommend the approximate F-test proposed by

Kenward and Roger.

2.5.2 Study II: balanced incomplete block design

Next, following a similar simulation setup as in Alnosaier, 2007, we consider a balanced incomplete block design where we have n observations in all: q blocks, p treatments and s treatments per block. The model of the block design is given by

$$y_{ij} = \mu + \alpha_i + b_j + e_{ij} \tag{2.5.18}$$

for $i = 2, \ldots, p$, $j = 1, \ldots, q$, where μ is the general mean, α_i are the treatment effects (difference of treatments from treatment 1), b_j are the block effects, $b_j \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, and b_j's and e_{ij}'s are independent. The covariance matrix of Y_i is given by $\Sigma_i = \sigma_e^2 I_s + \sigma_b^2 \mathbf{1}_s$ where $\mathbf{1}_s$ is an s dimensional vector of 1's. Our main objective is to test the following hypothesis (i) $H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$ and (ii) $H_0 : \alpha_2 = 0$. To compute the simulated size of tests, we generate the data under the null hypothesis i.e. assuming there is no significant treatment effect 5000 times and apply all five methods of testing on it. We also calculate the simulated power for all the five tests. We choose different values of $\rho = \frac{\sigma_b^2}{\sigma_e^2}$ ranging from 0.25 to 4.

Case 1: p=6, q=10, n=30

The data set up is given as follows

Table 2.7 Case 1

Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.	Block	Treat.
1	1,2,5	3	1,3,4	5	1,4,5	7	2,3,5	9	3,5,6
2	1,2,6	4	1,3,6	6	2,3,4	8	2,4,6	10	4,5,6

The simulation results are as follows:

Table 2.8 Type-I error rates for $\boldsymbol{\alpha}$=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.105	0.091	0.067	0.062	0.049
0.5	0.104	0.085	0.069	0.067	0.048
1	0.105	0.088	0.075	0.069	0.051
2	0.098	0.080	0.067	0.065	0.046
4	0.108	0.087	0.075	0.071	0.051

Table 2.9 Type-I error rates for α_2=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.087	0.079	0.055	0.053	0.044
0.5	0.088	0.078	0.056	0.052	0.049
1	0.089	0.076	0.048	0.045	0.048
2	0.092	0.079	0.048	0.047	0.048
4	0.096	0.085	0.053	0.049	0.052

Table 2.10 Type-I error rates for α_2=0 and α_3=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.120	0.105	0.074	0.069	0.060
0.5	0.120	0.103	0.069	0.064	0.057
1	0.104	0.086	0.048	0.040	0.050
2	0.115	0.095	0.056	0.046	0.053
4	0.120	0.102	0.060	0.049	0.058

From the results in Tables 2.8, 2.9 and 2.10 we observe that LRT and RM-LRT do a very poor job on preserving the Type-I error rates, since the effective sample size is only ten. KRT continues to preserve the size very accurately for all values of ρ.

A similar simulation study is conducted for six blocks and six treatments, three treatments per block.

Case 2: n=18, p=6, q=6

The data set up is as follows

Table 2.11 Case 2

Block	Treat.	Block	Treat.
1	1,2,5	4	2,3,4
2	1,2,6	5	3,5,6
3	1,3,4	6	4,5,6

The simulated size of the tests is given in Tables 2.12, 2.13 and 2.14.

Table 2.12 Type-I error rates for $\boldsymbol{\alpha}$=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.183	0.143	0.108	0.085	0.053
0.5	0.184	0.142	0.113	0.087	0.053
1	0.176	0.131	0.107	0.092	0.054
2	0.175	0.126	0.105	0.095	0.047
4	0.174	0.121	0.105	0.094	0.050

Table 2.13 Type-I error rates for α_2=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.138	0.123	0.082	0.071	0.052
0.5	0.133	0.113	0.069	0.055	0.048
1	0.145	0.125	0.067	0.060	0.051
2	0.139	0.114	0.060	0.053	0.046
4	0.136	0.115	0.065	0.059	0.049

Table 2.14 Type-I error rates for α_2=0 and α_3=0. Here LRT, RM-LRT, BC-LRT, BC-CRT and KRT represent likelihood ratio test, restricted maximum likelihood ratio test, Bartlett-corrected profile likelihood ratio test, Bartlett-corrected Cox Reid adjusted likelihood ratio test and Kenward-Roger approximate F test, respectively.

ρ	LRT	RM-LRT	BC-LRT	BC-CRT	KRT
0.25	0.138	0.123	0.082	0.071	0.052
0.5	0.133	0.113	0.069	0.055	0.048
1	0.145	0.125	0.067	0.060	0.051
2	0.139	0.114	0.060	0.053	0.046
4	0.136	0.115	0.065	0.059	0.049

In the above case, the effective sample size is only six, thus it has a significant impact on the performance of all tests. KRT still manages to preserve the nominal level.

From the above simulation study we observe that LRT and RM-LRT are not well suited for a small sample study. Theoretically, we would expect the Kenward-Roger method to perform better than the Bartlett corrected tests, in terms of preserving the size of the test. This is because, Bartlett corrected tests rescale the χ^2 distribution using the expected value. This would imply that the distribution will be more accurate in the median region as compared to the tail region. On the other hand, KR method uses an F distribution with suitable degrees of freedom such that the second moment of the statistics match as well; hence providing a good fit in the tail regions. This fact has been reflected in the simulation study results as well.

2.6 Real data example

We use the methods described on a study of 20 preadolescent girls reported by Goldstein, 1979, whose height are measured on a yearly basis from age 6 to 10. The girls are classified according to the height of their mother which is a discrete variable grouped from 1 (short in height) to 3 (tall in height). The measurements for the fifth girl were reported as 114.5, 112, 126.4, 131.2 and 135. The second measurement seems inaccurate and has been replaced by 122 as done in Verbeke and Molensberghs, 2000. Overall there are 100 measurements. We use the following model

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_{20} G_{2i} + \beta_{30} G_{3i} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}$$

with $i = 1, 2, \cdots, 20$ and $j = 1, \cdots, 4$, where Y_{ij} is the height of girls, t_{ij} is the j^{th} time point at which i^{th} subject's height was recorded, G_{2i} is a dummy variable indicating that mother's height is in group 2 and G_{3i} is the indicator that mother's height is in group 3. Assuming $(b_{0i}, b_{1i})' \overset{iid}{\sim} N_2(0, D)$, where D is the covariance matrix and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$ independent of b_i.

The goal of the study is to investigate the effect of mother's height on the height of the girl, i.e. we want to test the following hypothesis:

$$H_0 : (\beta_{20}, \beta_{30})' = 0 \text{ against } H_0 : (\beta_{20}, \beta_{30})' \neq 0 \tag{2.6.19}$$

The test statistic and p-value for testing the above hypothesis 2.6.19 for the different test procedures are (i) LRT: 8.658 (p-value= 0.0132), (ii) RM-LRT: 8.496 (p-value= 0.0143) (iii) BC-LRT: 7.273 (p-value= 0.0263) (iv) BC-CRT: 8.611 (p-value= 0.0137) and (v) KRT:

8.225 (p-value:0.0032). The p-values for (i)-(iv) were based on a χ^2_2 distribution while (v) was based on the F-distribution. All tests yield the same result that the null hypothesis is rejected at 5% significance level i.e. the height of mother has a significant impact on the height of girls.

However, in the hope of drawing the same conclusion from a subset of this data set, we randomly sampled 15 subjects from the data set (75 overall measurements), and used all the five testing procedures again to test the hypothesis given in (2.6.19). We obtain the following test statistics and p-values for the tests: (i) LRT: 5.576 (p-value= 0.0616), (ii) RM-LRT: 5.347 (p-value= 0.0690) (iii) BC-LRT: 4.544 (p-value= 0.103) (iv) BC-CRT: 5.416 (p-value= 0.0666) and (v) KRT: 4.576 (p-value:0.0333). From the above results, only KRT rejects the null hypothesis at 5% significance level.

Therefore, we conclude that the height of the mother has a significant impact on the height of girls. This conclusion is consistent with the results obtained by Verbeke and Molenberghs, 1997.

Chapter 3

Kenward-Roger approximation for linear mixed models with missing covariates

Consider a linear mixed model with m groups and n_i measurements in the i^{th} group. Denote $n = \sum_{i=1}^{m} n_i$ the total number of observations. For each group, we observe an $n_i \times 1$ vector of responses $\mathbf{Y}_i$, let X_i be an $n_i \times p$ fixed-effects design matrix whose first column is a vector of ones to account for the intercept, Z_i be an $n_i \times q$ random-effects design matrix, $\boldsymbol{\beta}$ be a $p \times 1$ vector of fixed-effect coefficients and $\mathbf{b}_i$ be a group-specific vector of random regression coefficients. The model can be written as

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,\ i = 1, \cdots, m,$$

where $\boldsymbol{\epsilon}_i \sim N_{n_i}(0, \sigma_e^2 I_{n_i})$, $\mathbf{b_i} \sim N_q(0, V)$ and $\boldsymbol{\epsilon}_i$ and $\mathbf{b_i}$ are independently distributed. We can deduce that $\mathbf{Y}_1, \cdots, \mathbf{Y}_m$ are independent and $\mathbf{Y}_i \sim N_{n_i}(X_i\boldsymbol{\beta}, \Sigma_i)$ with $\Sigma_i = Z_i V Z_i^T + \sigma_e^2 I_{n_i}$. Denote the stacked vectors $\mathbf{Y} = (\mathbf{Y}_1^T, \cdots, \mathbf{Y}_m^T)^T$, $\mathbf{b} = (\mathbf{b}_1^T, \cdots, \mathbf{b}_m^T)^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \cdots, \boldsymbol{\epsilon}_m^T)^T$ and the stacked matrices $X_{n\times p} = (X_1^T, \cdots, X_n^T)^T$, $Z_{n\times mq} = diag(Z_1, \cdots, Z_m)$

and $\Sigma_{n\times n} = diag(\Sigma_1, \cdots, \Sigma_n)$. Then the model can be written as

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^{\star},$$

where $\boldsymbol{\epsilon}^{\star} = Z\mathbf{b} + \boldsymbol{\epsilon}$ such that $\boldsymbol{\epsilon}^{\star} \sim N_n(0, \Sigma(\boldsymbol{\sigma}))$. The covariance matrix Σ is a function of $1+q(q+1)/2$ parameters, where the first parameter is the variance of the model errors i.e., σ_e^2 and the rest $q(q+1)/2$ parameters characterize the random effect covariance matrix V. These parameters are represented by the vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_r)^T$ where $r = 1 + q(q+1)/2$. Let $\mathbf{X_{(1)}} = (X_{1(1)}, X_{2(1)}, \cdots, X_{n(1)})^T$ be the first column of the covariate matrix X. Then $X_{n\times p}$ can be written as $X = (\mathbf{X_{(1)}} : X_{(-1)})$ where $\mathbf{X_{(1)}}$ is of dimension $n \times 1$ and $X_{(-1)}$ is a $n \times (p-1)$ matrix of the covariates other than the first. Without loss of generality, assume that $X_{(1)}$ contains partially missing covariates, and define the missing indicator as

$$B_j = \begin{cases} 0 & \text{if } X_{j(1)} \text{ is missing} \\ 1 & \text{otherwise,} \end{cases}$$

for $j = 1, \cdots, n$. Assume that the partially missing covariate $\mathbf{X_{(1)}}$ follows a parametric model $f(\mathbf{X_{(1)}}|\mathbf{X_{(-1)}}, \boldsymbol{\gamma})$ that is known upto a finite dimensional parameter $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k)^T$. In the presence of missing data, we write

$$\mathbf{Y} = X^{\star}\boldsymbol{\beta} + \boldsymbol{\epsilon}^{\star},$$

where $X^{\star}_{n\times p} = (\mathbf{X^{\star}}_{(1)} : X_{(-1)})$, $\mathbf{X^{\star}_{(1)}} = (X^{\star}_{1(1)}, \cdots, X^{\star}_{n(1)})^T$ such that,

$$X^{\star}_{j(1)} = \begin{cases} E(\mathbf{X_{j(1)}}|\mathbf{X_{j(-1)}}, \boldsymbol{\gamma}) = \mu(\boldsymbol{\gamma}) & \text{if } B_j = 0 \\ X_{j(1)} & \text{if } B_j = 1, \end{cases}$$

for $j = 1, \cdots, n$. This approach is very commonly used, for eg., Little & Rubin (2002).

Now we can estimate $\boldsymbol{\gamma}$ by maximizing the likelihood

$$L = \prod_{j=1}^{n} f(X_{j(1)}|X_{j(-1)}, \boldsymbol{\gamma}) \ \forall \ \{j : B_j = 1\}. \tag{3.0.1}$$

The standard error of the parameter estimates can be determined from the Hessian matrix. We will be working with the following model setup in the rest of the chapter. Now the n-dimensional $\mathbf{Y}$ follows a multivariate normal distribution,

$$\mathbf{Y} \sim \text{Normal}(X^{\star}\boldsymbol{\beta}, \Sigma),$$

where $X^{\star}$ is an $n \times p$ matrix of known/estimated covariates, assumed to be of full rank. $\boldsymbol{\beta}$ is a $p\times 1$ vector of unknown parameters and Σ is an unknown variance covariance matrix whose elements are functions of $\boldsymbol{\sigma}$=$(\sigma_1, \cdots, \sigma_r)^T$ which in turn is a function of $\boldsymbol{\gamma}$=$(\gamma_1, \cdots, \gamma_k)^T$. This is because the covariate X is a function of $\boldsymbol{\gamma}$ and from the REML equations we know that the estimates of $\boldsymbol{\sigma}$ depend on X.

The REML based estimated least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}\{\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})\}X^{\star}(\widehat{\boldsymbol{\gamma}})^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}\{\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})\}\mathbf{Y}, \tag{3.0.2}$$

where $\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})$ is the REML estimator of $\boldsymbol{\sigma}$. In our study, $\boldsymbol{\beta}$ is the parameter of interest and $\Sigma\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}$ is the nuisance parameter. We will test the hypothesis $H_0 : L^T\boldsymbol{\beta} = 0$, where L is a $p \times l$ matrix.

Remark: The method can be generalized for the case of two or more partially missing independent covariates. Without loss of generality, assume that $X_{(1)}$ and $X_{(2)}$ contain partially missing covariates and that they follow parametric distributions $f(\mathbf{X_{(1)}}|\mathbf{X_{(-1)}}, \boldsymbol{\tau}_1)$ and $f(\mathbf{X_{(2)}}|\mathbf{X_{(-2)}}, \boldsymbol{\tau}_2)$ respectively. The missing values for each of the covariates can be imputed using similar maximum likelihood procedure as described above. Then the REML based estimator of $\boldsymbol{\beta}$ is given by $\widehat{\boldsymbol{\beta}}$ given in (3.0.2) where $\boldsymbol{\gamma} = (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)^T$.

3.1 Inference for the fixed effects

To make inferences about fixed effects such as confidence interval estimation and testing of hypothesis, we need to define an approximate pivot and compute its distribution in small sample settings.

3.1.1 Estimation of the variance of the estimator $\widehat{\boldsymbol{\beta}}$

Under the linear mixed model framework, the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ is given by $\widehat{\Phi} = \Phi\{\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})\} = X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}\{\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})\}X(\widehat{\boldsymbol{\gamma}})^{-1}$, where $\widehat{\boldsymbol{\gamma}}$ is the estimator obtained after maximizing the likelihood given in (3.0.1) and $\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})$ is the REML estimate of $\boldsymbol{\sigma}$ obtained after using the imputed data set. There are three sources of bias when $\widehat{\Phi}$ is used as an estimator for the variance of $\widehat{\boldsymbol{\beta}}$ for small samples: $\widehat{\Phi}$ is a biased estimator of $\Phi(\boldsymbol{\sigma}(\boldsymbol{\gamma}))$, Φ does not take the variability of $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$ into account. We consider an approximation to the small sample covariance matrix of $\widehat{\boldsymbol{\beta}}$ under missing covariates, which accounts for the variability in

both $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$, thus reducing the small sample bias and any bias due to the missingness. For complete data, Kackar and Harville (1984) addressed the second source of bias by partitioning the variance of the estimated $\boldsymbol{\beta}$ into two components: $\Phi + \Lambda$. Further Kenward and Roger (1997) addressed the first source of bias and combined both adjustments thus proposing a new estimator of Φ,

$$\widehat{\Phi}_{KR} = \widehat{\Phi} + \widehat{\Lambda},$$

where the estimators of Φ and Λ are such that

$$E(\widehat{\Phi}) = \Phi - \widetilde{\Lambda} + R^* + O(n^{-5/2}), \tag{3.1.3}$$

$$E(\widehat{\Lambda}) = \widetilde{\Lambda} + O(n^{-5/2}), \tag{3.1.4}$$

with

$$\Phi = [X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}X^{\star}(\boldsymbol{\gamma})]^{-1}, \tag{3.1.5}$$

$$\widetilde{\Lambda} = \sum_{l=1}^{r}\sum_{m=1}^{r} \mathrm{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi(Q_{lm} - P_l\Phi P_m)\Phi, \tag{3.1.6}$$

and $O(n^r)$ denotes $O(n^r)/n^r$ is a bounded number as $n \to \infty$, for some r. In the above expressions

$$\begin{aligned}
P_l &= -X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}\frac{\partial \Sigma}{\partial \sigma_l}\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}X^{\star}(\boldsymbol{\gamma}), \\
Q_{lm} &= X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}\frac{\partial \Sigma}{\partial \sigma_l}\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}\frac{\partial \Sigma}{\partial \sigma_m}\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}X^{\star}(\boldsymbol{\gamma}), \\
R_{lm} &= X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}\frac{\partial^2}{\partial \sigma_l \partial \sigma_m}\Sigma^{-1}\{\boldsymbol{\sigma}(\boldsymbol{\gamma})\}X^{\star}(\boldsymbol{\gamma}), \\
R^* &= \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}\mathrm{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi R_{lm}\Phi.
\end{aligned}$$

We now consider the bias in $\widehat{\Phi}$ as an estimator of $\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))$ thus adjusting for missingness.

Theorem 1 *Under the assumptions given in Appendix A1, the adjusted variance of* $\widehat{\boldsymbol{\beta}}$ *can be approximated as*

$$\widehat{\Phi}_A = \widehat{\Phi} + \widehat{\Lambda} + \widehat{\Psi}.$$

The estimators of Φ, Λ *and* Ψ *are given by* $\widehat{\Phi}$, $\widehat{\Lambda}$ *and* $\widehat{\Psi}$ *respectively such that (3.1.3)-(3.1.6) hold and* $E(\widehat{\Psi}) = \Psi + O(n^{-2})$, *where*

$$\Psi = \sum_{i=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i var(\widehat{\gamma}_i)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i cov(\widehat{\gamma}_i, \widehat{\gamma}_j)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})).$$

Proof: The proof is detailed in Appendix A2.

3.1.2 Approximating the distribution of the test statistic

Suppose we are interested in making inferences about l linear combinations of the elements $\boldsymbol{\beta}$. In other words, we are interested in testing $H_0 : L^T\boldsymbol{\beta} = 0$ where L^T is a fixed matrix of dimension $(l \times p)$. A common statistic to test H_0 is the Wald statistic given by

$$F = \frac{1}{l}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1}L^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{3.1.7}$$

where $\widehat{\Phi}_A$ is an adjusted covariance matrix for $\widehat{\boldsymbol{\beta}}$. We will follow a similar procedure as used in Kenward & Roger (1997) and Alnosaier (2007) by approximating the first two moments of the F statistic given in (3.1.7) and match the moments to approximate the scaling factor and the denominator degrees of freedom of the test statistic i.e., λ and d such that $\lambda F \sim F(l, d)$ in distribution. Here $F(l, d)$ denotes F-distribution with degrees of freedom (l, d).

Theorem 2 *Under the assumptions given in Appendix A1, the first two moments of the test statistic (3.1.7) under* $H_0 : L^T\boldsymbol{\beta} = 0$ *are*

$$E(F) = 1 + \frac{A_2}{l} - \frac{A_4}{l} + O(n^{-3/2}), \tag{3.1.8}$$

$$var(F) = \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2}, \tag{3.1.9}$$

where $\widehat{\Theta} = L(L^T\widehat{\Phi}L)^{-1}L^T$,

$A_1 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Phi}P_l\widehat{\Phi})trace(\widehat{\Theta}\widehat{\Phi}P_m\widehat{\Phi})$,

$A_2 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Phi}P_l\widehat{\Phi}\widehat{\Theta}\widehat{\Phi}P_m\widehat{\Phi})$,

$A_3 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace\{\widehat{\Theta}\Phi(Q_{lm} - P_l\widehat{\Phi}P_m - R_{lm}/4)\}$,

$A_4 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Psi})$,

and using (3.1.8) and (3.1.9), we get $\widetilde{d} = 4 + (2 + l)/(l\widetilde{\rho} - 1)$, *where* $\widetilde{\rho} = \widetilde{V}/2\widetilde{E}^2$, *and*

$\widetilde{\lambda} = \widetilde{d}/\{\widetilde{E}(\widetilde{d}-2)\}$, *where* $\widetilde{E}$ *and* $\widetilde{V}$ *are defined in Appendix A3.*

Proof: The expectation and variance are approximated using the following conditional arguments that $E(F) = E\{E(F|\widehat{\boldsymbol{\sigma}})\}$ and $\text{var}(F) = E\{\text{var}(F|\widehat{\boldsymbol{\sigma}})\} + \text{var}\{E(F|\widehat{\boldsymbol{\sigma}})\}$. After using the Taylor series expansion we obtain

$$E(F) = \widetilde{E} + O(n^{-3/2}), \text{ and } \text{var}(F) = \widetilde{V} + O(n^{-3/2}),$$

where the explicit expressions of $\widetilde{E}$ and $\widetilde{V}$ are derived in Appendix A3.In order to determine λ and d, we match the first two moments of λF with those of $F(l,d)$ i.e.,

$$E\{F(l,d)\} = \frac{d}{d-2}, \text{ and } \text{var}\{F(l,d)\} = 2\left(\frac{d}{d-2}\right)^2 \frac{l+d-2}{l(d-4)},$$

to obtain the expressions for $\widetilde{d}$ and $\widetilde{\lambda}$.

3.2 Simulation

To explore the behaviour of the statistics derived in the previous sections, we conducted simulation studies for two different settings.

Design 1: Random Coefficient Regression Model

We consider a longitudinal study where the data consists of measurements on n subjects. Measurements are taken over four time intervals (say): 8, 10, 12 and 14. Let t_j be the time points at which the observation is recorded. The model can be expressed as

$$Y_i = X_i\beta + Z_i\mathbf{b} + \epsilon_i,$$

where X_i is the $(4 \times p)$ fixed effects matrix X_i and the random-effects design matrix Z_i can be written as

$$Z_i = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}$$

The measurement error structure $\Sigma_i = \sigma^2 I_4$. We assume an unstructured covariance matrix G for the random effects b_i given by

$$G = \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_2 & \omega_3 \end{pmatrix}$$

Therefore, the resulting covariance matrix of Y_i is given by $Z_i G Z_i' + \sigma^2 I_4$. To simulate the data, we consider $p = 5$ and generated the covariate matrix X from a multivariate normal distribution with mean zero and covariance matrix whose $(i, j)^{th}$ element is given by $|0.2|^{i-j}$. We introduced 20% and 40% missingness in one of the columns of X (i.e., missingness in one covariate) using a bernoulli distribution i.e. assuming the missing mechanism to be MCAR. Then we maximized the observed likelihood of the partially missing covariate to get the estimate of the mean and the variance, and replaced the missing values in the covariate matrix by the estimated mean, thus obtaining the imputed data set. Finally, we deployed our method on the imputed data set to test the hypothesis $H_0 : \boldsymbol{\beta} = 0$ vs $H_a : \boldsymbol{\beta} \neq 0$. We keep σ^2 fixed at 0.05 and ω_1 fixed at 1 and consider multiple cases by choosing different values of ω_2 and ω_3 ranging from 0 to 1. For each setting, 5000 data sets are generated and

Type-I error rates are noted.

Design 2: Block Design

In the next simulation study, we considered a block design with one random effect ($q = 1$), whose variance is σ_b^2. Here we considered the number of fixed effects i.e., $p = 5$. To simulate the data, as usual, we generated the covariate matrix X from a gaussian distribution and introduced missingness under MCAR mechanism. Further, to relax the MCAR assumption, we considered another case to mimic the MAR mechanism where we introduced missingness in one of the covariates such that the probability of an observation being missing depends on the value of another covariate. Our objective is to test the hypothesis $H_0 : \boldsymbol{\beta} = 0$ vs $H_a : \boldsymbol{\beta} \neq 0$. To test this, we generated $\mathbf{Y}$ from $N(0, \Sigma)$ for different values of $\rho = \sigma_b^2/\sigma_e^2$ ranging from 0.25 to 4. For each setting we generated 5000 data sets and computed the simulated size (Type-I error probabilities).

Implementation: To be able to run the simulation, we need to calculate the following terms: $\partial X^\star(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$ and $\partial\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma}$. The first term is $\partial X^\star(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma} = I(X^\star = \mu(\boldsymbol{\gamma}))$, and using the chain rule of of derivative the second term is $\partial\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma} = \sum_{i=1}^{r}(\partial\Sigma/\partial\widehat{\sigma}_1)\times(\partial\widehat{\sigma}_1/\partial\boldsymbol{\gamma})$.The detailed expression of these terms are given in Appendix A4.

In our simulation study, we will compare our testing procedure i.e., Kenward Roger adjustment for missing covariates (KRM) to the following methods which are used to deal with missing covariates in a linear mixed model setup on the basis of Type-I error rates: 1) listwise deletion followed by likelihood ratio test (LD-LRT), 2) mean imputation followed by likelihood ratio test (MI-LRT), 3) mean imputation followed by Wald test.

Results: For Design 1, Tables 3.1 and Table 3.2 present the Type-I error rates when the missing mechanism is MCAR for $n = 27$ and $n = 13$ respectively

Table 3.1 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $n = 27$ at 40% missingness. Here KRM, LD-LRT, MI-LRT and MI-WT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test and mean imputation followed by Wald Test.

ω_2	ω_3	KRM	LD-LRT	MI-LRT	MI-WT
0	0.5	0.059	0.085	0.068	0.084
0	1	0.070	0.118	0.122	0.096
0.25	0.5	0.065	0.091	0.077	0.095
0.25	1	0.068	0.119	0.145	0.097

Table 3.2 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $n = 13$ at 40% missingness. Here KRM, LD-LRT, MI-LRT and MI-WT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test and mean imputation followed by Wald Test.

ω_2	ω_3	KRM	LD-LRT	MI-LRT	MI-WT
0	0.5	0.075	0.157	0.103	0.135
0	1	0.062	0.162	0.124	0.109
0.25	0.5	0.070	0.130	0.116	0.113
0.25	1	0.055	0.144	0.154	0.103

For Design 2, Tables 2.8 and 2.9 provide the Type-I error rates when the missing mechanism is MCAR for sample sizes 15 and 30 respectively, Table 2.10 provides the results under the MAR mechanism.

Table 3.3 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 15$. Here KRM, LD-LRT and MI-LRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test and mean imputation followed by likelihood ratio test.

Percentage of missing data	ρ	KRM	LD-LRT	MI-LRT	MI-WT
20	0.25	0.049	0.148	0.113	0.094
	0.5	0.055	0.246	0.122	0.087
	1	0.059	0.152	0.120	0.099
	2	0.057	0.188	0.127	0.095
	4	0.063	0.173	0.129	0.094
40	0.25	0.042	0.298	0.112	0.094
	0.5	0.057	0.205	0.126	0.089
	1	0.052	0.375	0.123	0.093
	2	0.058	0.375	0.130	0.098
	4	0.063	0.415	0.124	0.088

Table 3.4 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 30$ and $n_i = 2$ for all $i = 1, \ldots, 30$. Here KRM, LD-LRT and MI-LRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test and mean imputation followed by likelihood ratio test.

Percentage of missing data	ρ	KRM	LD-LRT	MI-LRT	MI-WT
20	0.25	0.044	0.085	0.078	0.067
	0.5	0.055	0.083	0.087	0.068
	1	0.050	0.094	0.082	0.066
	2	0.055	0.106	0.083	0.073
	4	0.047	0.088	0.079	0.063
40	0.25	0.047	0.100	0.073	0.065
	0.5	0.048	0.143	0.081	0.063
	1	0.048	0.104	0.072	0.066
	2	0.051	0.217	0.079	0.066
	4	0.048	0.169	0.076	0.069

Table 3.5 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 15$ when the missing mechanism is MAR. Here KRM, LD-LRT and MI-LRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test and mean imputation followed by likelihood ratio test.

ρ	KRM	LD-LRT	MI-LRT	MI-WT
0.25	0.058	0.103	0.102	0.094
0.5	0.053	0.114	0.091	0.087
1	0.050	0.164	0.103	0.093
2	0.049	0.117	0.089	0.087
4	0.063	0.141	0.099	0.095

From the results in Tables 3.1- 3.5, we observe that for the case of listwise deletion, the Type-I error rates are close to 0.1 when there is 20% missingness in the data and it gets worse when the percentage of missingness increases to 40%. As the sample increases, there is some decrease in the error rates, however, they are still significantly large. Mean imputation seems a viable option since it provides a complete data to work with. It works better than listwise deletion as expected, but the standard likelihood ratio (MI-LRT) used on the imputed data set still does not work well, since the χ^2 test is based on large sample approximations. Further, as seen from results of design 1, Wald-test performs better than LD-LRT and MI-LRT but the size is nowhere close to the nominal level. We conducted another ad-hoc test procedure where we used the original Kenward Roger method on the imputed data set without reducing the bias due to missingness (results not shown here). It performs reasonably well, all Type- I error rates were between 0.045 to 0.075; however it is not theoretically justified. Further, we also compared Wald type tests using two different covariance matrices of $\widehat{\boldsymbol{\beta}}$ namely (i) $\widehat{\Phi}$ which does not take the variability of $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$ into

account and (ii) $\widehat{\Phi}_{KR}$ which takes the variability in only $\widehat{\boldsymbol{\sigma}}$ into account. These are also ad-hoc tests and in these two cases, no approximation is made to the denominator degrees of freedom of F distribution (we use the residual degrees of freedom). To avoid redundancy, the results from methods (i) and (ii) are not shown here. Finally, our proposed method (KRM) outperforms the other procedures for all values of ρ, n and the type and percentage of missingness. It preserves the Type-I error rates close to 0.05 for each case. We also tested for single fixed effect parameter i.e. $H_0 : \beta_1 = 0$ (results not shown here) in which case the results are he same.

Robustness: In practice, the distribution of the partially missing covariate is not known. Therefore, to check the robustness of our procedure, we generated the partially missing covariate from heavy tailed distributions like the Laplace distribution with mean parameter as zero and variance as eight, and the Student's t-distribution with one degree of freedom. We ensured 20% missingness in the covariate. This is done for Design 2. The corresponding results are given in Table 3.6.

Table 3.6 Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 15$ when the partially missing (20% missing) covariate is generated from heavy tailed distributions. Here KRM, LD-LRT, MI-LRT and MI-KRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test and mean imputation followed by likelihood ratio test.

Distribution	ρ	KRM	LD-LRT	MI-LRT	MI-WT
Laplace(0,2)	0.25	0.052	0.106	0.095	0.093
	0.5	0.056	0.115	0.102	0.094
	1	0.049	0.185	0.094	0.089
	2	0.049	0.170	0.099	0.090
	4	0.051	0.199	0.100	0.095
Student's t(1)	0.25	0.050	0.153	0.099	0.085
	0.5	0.051	0.105	0.096	0.093
	1	0.047	0.137	0.104	0.098
	2	0.055	0.142	0.105	0.093
	4	0.059	0.139	0.107	0.094

Table 3.6 shows that our testing procedure is robust. The Type-I error rates for the KRM procedure are contained between 0.049 and 0.059 even when the distribution of the covariate is misspecified. While, as expected the error probabilities for LD-LRT, MI-LRT and MI-WT are inflated.

3.3 Real data examples

3.3.1 Example 1

We deploy the method detailed in section 3.1 on the data mentioned in the introduction. The data is described as the vagal tone data in Weiss (2005). Vagal tone is a measure of the activity in the parasympathetic nervous system (PNS). It is supposed to be high but it gets lower in response to stress. This data sets consists of five vagal tone measurements taken before and after the catheterization, two were taken prior to the surgery (time points: pre1 and pre2) and three were taken post surgery (time points: post1, post2 and post3). The response variable at first two time points is approximately the same, then it decreases at post1 and increases back to normal for the last two measurements. In our analysis, we consider three measurements (at time points pre2, post1 and post2). The covariates in the data are gender, age in months, duration of the catheterization in minutes and a measure of illness severity. The data is highly unbalanced. In addition, we introduce 30% missingness in the covariate age. The model can be expressed as follows.

$$Y_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 t_{ij} + b_{0i} + \epsilon_{ij}$$

with $i = 1, \cdots, 21$ and $j = 1, 2, 3$, where Y_{ij} is the vagal tone measurement for i^{th} subject at the j^{th} time point, X_{1i}, X_{2i}, X_{3i} and X_{4i} indicate the gender, age, duration of surgery and severity of illness respectively for the i^{th} subject. Assuming $(b_{0i}) \overset{iid}{\sim} N(0, \sigma_b^2)$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$ is independent of b_{0i}.

Our objective in this example is to quantify the effect of age of the infant on the vagal tone measurement. Therefore we test $H_0 : \beta_2 = 0$. The test statistic, demoninator degrees

of freedom (ddf) and the p-value for different test procedures are: (i) KRM: 1.626 (ddf = 25.11, p-value = 0.214) (ii) LD-LRT: 2.807 (p-value=0.093) and (iii) MI-LRT: 1.994 (p-value = 0.157). Listwise deletion concludes marginal significance of age on the vagal tone measurement while the MI-LRT suggests a change in the direction of the conclusion. KRM confirms the change in direction by producing a p-value of 0.214 since it is able to account for the bias due to small sample size, the bias due to estimation of variance parameters and the bias due to missingness resulting in a more accurate testing procedure.

3.3.2 Example 2

We consider the data from a randomized, double-blind, study of AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 $cells/mm^3$). The data description is as in the datasets section in Fitzmaurice (2004): Patients in AIDS Clinical Trial Group (ACTG) Study 193A were randomized to dual or triple combinations of HIV-1 reverse transcriptase inhibitors. Specifically, patients were randomized to one of four daily regimens containing 600mg of zidovudine: zidovudine alternating monthly with 400mg didanosine; zidovudine plus 2.25mg of zalcitabine; zidovudine plus 400mg of didanosine or zidovudine plus 400mg of didanosine plus 400mg of nevirapine (triple therapy). The patients characteristics like age and sex were also recorded. The response variable i.e. measurements of CD4 counts were scheduled to be collected at baseline and at 8-week intervals during follow-up. However, due to skipped visits and dropouts, the measurements could not be taken at regular intervals therefore we have the weeks as another variable. For our use, we use the log transformed CD4 counts log(CD4counts+1) as the response variable. The data is available for 1313 patients, however, since we are testing for small sample studies, we will truncate the data by randomly choosing 40 patients from the study. We introduce 30% missingness

in the covariate age and deploy the discussed methods on the obtained data. The model under consideration is

$$Y_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 t_{ij} + \beta_{20} G_{2i} + \beta_{30} G_{3i} + \beta_{40} G_{4i} + b_{0i} + \epsilon_{ij} \qquad (3.3.10)$$

with $i = 1, 2, \cdots, 40$ and $j = 1, \cdots, n_i$ n_i varying from 1 to 9, where Y_{ij} is the log transformed CD4 counts, G_{2i} is a dummy variable indicating that the patient was in treatment group 2 and G_{3i} is the indicator that patient was in treatment group 3 and G_{4i} is defined likewise, X_{1i} represents the gender of the patient (1=male, 0=female), X_{2i} represents the age of the patient in years and t_{ij} is the j^{th} time point (in weeks) at which i^{th} patient's measurement was recorded. Assuming $(b_{0i}) \overset{iid}{\sim} N(0, \sigma_b^2)$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$ is independent of b_{0i}. In the above model (3.3.10), we introduce missingness in the variable X_{2i}.

The goal of our study is to investigate if there is any significant difference between the impact of the four treatment groups on the CD4 counts. Therefore, we will test $H_0 : (\beta_{20}, \beta_{30}, \beta_{40})^T = 0$ against $H_0 : (\beta_{20}, \beta_{30}, \beta_{40})^T \neq 0$. When we test the hypothesis using all the data available i.e. 1313 patients, we obtain a p-value of approximately 0.008 which suggests that treatment group type is significant in the study. However, to test our methodology, we randomly select a sample of size 40. The test statistic and p-value for testing the above hypothesis for the different test procedures are (i) KRM: 3.201 (p-value= 0.029), (ii) LD-LRT: 8.068 (p-value= 0.044) (iii) MI-LRT: 8.775 (p-value= 0.033). The p-value for (i) is based on the approximate F-distribution with numerator degrees of freedom as three and the denominator degrees of freedom as 62.50 which are calculated from the data and the p-values for (ii) and (iii) are based on a χ_3^2 distribution while the rest are based on the F-distribution with different degrees of freedom. All tests yield the result that

the null hypothesis is rejected at 5% significance level i.e. the treatment group type has a significant impact on the CD4 counts. Although the conclusion from all methods is the same, the p-value of the KRM test is the least and we can have more confidence in this test since it is theoretically justified.

3.4 Proofs of Chapter 3

3.4.1 Assumptions

We impose the following assumptions about the model as done in Alnosaier (2007)

C.1 Σ is a block diagonal and non singular matrix. Also $\Sigma^{-1}, \partial\Sigma/\partial\sigma_l, \partial\Sigma/\partial\gamma_i, \partial^2\Sigma/\partial\sigma_l\partial\sigma_m$ and $\partial^2\Sigma/\partial\gamma_i\partial\gamma_j$ are bounded.

C.2 $E(\widehat{\boldsymbol{\sigma}}) = \boldsymbol{\sigma} + O(n^{-3/2})$.

C.3 The possible dependence between $\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})$ and $\widehat{\boldsymbol{\beta}}$ is ignored.

C.4 $\Phi = (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1} = O(n^{-1}),\ (L^T\Phi L) = O(n)$.

C.5 $\partial\widetilde{\boldsymbol{\beta}}/\partial\sigma_l = O(n^{-1/2}), \partial^2\widetilde{\boldsymbol{\beta}}/\partial\sigma_l\partial\sigma_m = O(n^{-1/2})$ where
$\widetilde{\boldsymbol{\beta}} = \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\mathbf{Y}$.

C.6 The expectation of $\widehat{\boldsymbol{\beta}}$ exists.

3.4.2 Estimates of Ψ, Φ and Λ in Theorem 1

Let $\boldsymbol{\sigma}=(\sigma_1,\cdots,\sigma_r)^T$ and $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_k)^T$. We begin by expanding the REML based estimate of $\boldsymbol{\beta}$ for the imputed data set which is given by

$$\widehat{\boldsymbol{\beta}} = (X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\widehat{\boldsymbol{\gamma}}))^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))\mathbf{Y} = C_1+C_2+C_3+C_4+C_5+C_6$$

where

$$\begin{aligned}
C_1 =&(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\widehat{\boldsymbol{\gamma}}))^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})(\Sigma(\hat{\sigma}(\hat{\gamma}))^{-1}-\Sigma(\hat{\sigma}(\gamma))^{-1})\mathbf{Y},\\
C_2 =&(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\widehat{\boldsymbol{\gamma}}))^{-1}(X^{\star T}(\hat{\gamma})-X^{\star T}(\boldsymbol{\gamma}))\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y},\\
C_3 =&(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\widehat{\boldsymbol{\gamma}}))^{-1}-(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star T}(\boldsymbol{\gamma}))^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y},\\
C_4 =&(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\boldsymbol{\gamma}))^{-1}-(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star T}(\boldsymbol{\gamma}))^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y},\\
C_5 =&(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma}))^{-1}-(X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star T}(\boldsymbol{\gamma}))^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y},\\
C_6 =&(X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma}))^{-1}X^{\star T}(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y}.
\end{aligned}$$

For notational convenience, we will denote $X^{\star}$ as X. By the Taylor series expansion around $\boldsymbol{\gamma}$,

$$\Sigma(\hat{\sigma}(\hat{\gamma}))^{-1}-\Sigma(\hat{\sigma}(\gamma))^{-1} \approx \sum_{i=1}^{k}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial\gamma_i}(\widehat{\gamma}_i-\gamma_i)+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial\gamma_i\partial\gamma_j}(\widehat{\gamma}_i-\gamma_i)(\widehat{\gamma}_j-\gamma_j).$$

Since $E(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma} + O_p(n^{-1/2})$ and $E(\mathbf{Y}) = X\boldsymbol{\beta}$, we get

$$\begin{aligned}
&\Sigma(\hat{\sigma}(\hat{\gamma}))^{-1} - \Sigma(\hat{\sigma}(\gamma))^{-1} \\
&\qquad = \sum_{i=1}^{k} \frac{\partial \Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i) + O_p(n^{-1}) \\
&\qquad = -\sum_{i=1}^{k} \Sigma(\widehat{\boldsymbol{\sigma}})^{-1} \frac{\partial \Sigma(\widehat{\boldsymbol{\sigma}})}{\partial \gamma_i} \Sigma(\widehat{\boldsymbol{\sigma}})^{-1}(\widehat{\gamma}_i - \gamma_i) + O_p(n^{-1}) \qquad (3.4.11)
\end{aligned}$$

Substituting (3.4.11) in the expression for C_1 we get $C_1 = \widetilde{C_1} + O(n^{-1/2})$, where

$\widetilde{C_1} = \{X^T(\boldsymbol{\gamma})\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\sum_{i=1}^{k}\{-\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}\{\partial\Sigma(\widehat{\boldsymbol{\sigma}})/\partial\gamma_i\}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}})(\widehat{\gamma}_i - \gamma_i)\}X\boldsymbol{\beta}$.

Let $\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) = \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}$, then

$$C_1 = \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\sum_{i=1}^{k}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}(\widehat{\gamma}_i - \gamma_i)X\boldsymbol{\beta} + O_p(n^{-1/2}). \qquad (3.4.12)$$

Consider $C_2 = (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\widehat{\boldsymbol{\gamma}}))^{-1}(X^T(\hat{\gamma}) - X^T(\boldsymbol{\gamma}))\Sigma(\hat{\sigma}(\gamma))^{-1}Y$.

By the Taylor series expansion,

$$X^T(\widehat{\boldsymbol{\gamma}}) - X^T(\boldsymbol{\gamma}) = \sum_{i=1}^{k}\frac{\partial X^T}{\partial\gamma_i}(\widehat{\gamma}_i - \gamma_i) + O_p(n^{-1}).$$

Therefore,

$$\begin{aligned}
C_2 &= \widetilde{C_2} + O_p(n^{-1/2}), \text{ where} \\
\widetilde{C_2} &= \Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})\sum_{i=1}^{k}\frac{\partial X^T}{\partial\gamma_i}(\widehat{\gamma}_i - \gamma_i)\Sigma(\hat{\sigma}(\gamma))^{-1}X\boldsymbol{\beta}. \qquad (3.4.13)
\end{aligned}$$

Consider

$C_3 = \{(X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\widehat{\boldsymbol{\gamma}}))^{-1} - (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\boldsymbol{\gamma}))^{-1}\}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}Y$.

Again, using the Taylor series expansion,

$$(X^T(\widehat{\boldsymbol{\gamma}})\Sigma(\hat{\sigma}(\gamma))^{-1}X(\widehat{\boldsymbol{\gamma}}))^{-1} - (X^T(\widehat{\boldsymbol{\gamma}})\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma}))^{-1}$$
$$= \sum_{i=1}^{k} \frac{\partial}{\partial \gamma_i}(X^T(\widehat{\boldsymbol{\gamma}})\Sigma(\hat{\sigma}(\hat{\gamma}))^{-1}X(\boldsymbol{\gamma}))^{-1}(\widehat{\gamma}_i - \gamma_i) + O_p(n^{-1}).$$

Hence,

$$\begin{aligned}
C_3 &= \widetilde{C}_3 + O_p(n^{-1/2}), \text{ where} \\
\widetilde{C}_3 &= \sum_{i=1}^{k} -(X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta} \\
&= -\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\sum_{i=1}^{k} X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}. \qquad (3.4.14)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\widetilde{C}_4 &= \sum_{i=1}^{k} -\Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})X^T(\boldsymbol{\gamma})\frac{\partial \Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \gamma_i}X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta} \\
&= \sum_{i=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}, \qquad (3.4.15) \\
\widetilde{C}_5 &= \sum_{i=1}^{k} -\Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})\frac{\partial X^T}{\partial \gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}, \qquad (3.4.16) \\
C_6 &= (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y} = \breve{\beta}.
\end{aligned}$$

Let $H(\widehat{\boldsymbol{\sigma}}) = \widetilde{C}_1 + \widetilde{C}_2 + \widetilde{C}_3 + \widetilde{C}_4 + \widetilde{C}_5$. Further, we need to compute the variance of $\widehat{\boldsymbol{\beta}}$.

$$V(\widehat{\boldsymbol{\beta}}) = V(H(\widehat{\boldsymbol{\sigma}})) + V(\breve{\beta}).$$

To take into account the variability in $\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})$ when $\boldsymbol{\gamma}$ is fixed, we follow the same proof as in

Kenward & Roger (1997) and Alnosaier (2007),

$$
\begin{aligned}
V(\breve{\beta}) &= V(\widetilde{\boldsymbol{\beta}}) + V(\breve{\beta} - \widetilde{\boldsymbol{\beta}}), \text{where } \widetilde{\boldsymbol{\beta}} = (X^T(\boldsymbol{\gamma})\Sigma(\boldsymbol{\sigma}(\boldsymbol{\gamma}))^{-1}X(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\boldsymbol{\sigma}(\boldsymbol{\gamma}))^{-1}\mathbf{Y}. \\
V(\breve{\beta}) &= \Phi + \Lambda.
\end{aligned}
$$

The estimates of Φ and Λ are given by $\widehat{\Phi}$ and $\widehat{\Lambda}$ such that

$$
\begin{aligned}
E(\widehat{\Phi}) &= \Phi - \widetilde{\Lambda} + R^* + O(n^{-5/2}), \\
E(\widehat{\Lambda}) &= \widetilde{\Lambda} + O(n^{-5/2}), \text{ where} \\
\Phi &= (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1}, \\
\widetilde{\Lambda} &= \sum_{l=1}^{r}\sum_{m=1}^{r} \operatorname{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi(Q_{lm} - P_l\Phi P_m)\Phi.
\end{aligned}
$$

In the above expressions

$$
\begin{aligned}
P_l &= -X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\
Q_{lm} &= X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\
R_{lm} &= X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial^2}{\partial\sigma_l\partial\sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\
R^* &= \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r} \operatorname{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi R_{lm}\Phi.
\end{aligned}
$$

Now, from (3.4.12)-(3.4.16)

$$\begin{aligned} H(\widehat{\boldsymbol{\sigma}}) &= \sum_{i=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\big[\{-X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} \\ &\quad +\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} - \{X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial\gamma_i}\} \\ &\quad +\{X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} \\ &\quad -\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\}\big]\boldsymbol{\beta}(\widehat{\gamma}_i-\gamma_i) \\ &= \sum_{i=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i(\widehat{\gamma}_i-\gamma_i), \end{aligned}$$

where

$$\begin{aligned} K_i &= \big[\{-X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} + \{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} \\ &\quad -\{X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial\gamma_i}\} + \{X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\} \\ &\quad -\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\}\big]\boldsymbol{\beta}, \end{aligned}$$

and

$$\begin{aligned} \Psi = V(H(\widehat{\boldsymbol{\sigma}})) &= \sum_{i=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_iV(\gamma_i)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\ &\quad + \sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i\mathrm{cov}(\gamma_i,\gamma_j)K_j^T\Phi^T(\widehat{\boldsymbol{\sigma}},\boldsymbol{\gamma}). \end{aligned}$$

Now using the Taylor series expansion about $\boldsymbol{\sigma}$ we get,

$$\Psi(\widehat{\boldsymbol{\sigma}}) = \Psi(\boldsymbol{\sigma}) + \sum_{l=1}^{r}\frac{\partial\Psi}{\partial\sigma_l}(\widehat{\sigma}_l-\sigma_l) + \sum_{l=1}^{r}\sum_{m=1}^{r}\frac{\partial^2\Psi}{\partial\sigma_l\partial\sigma_m}(\widehat{\sigma}_l-\sigma_l)(\widehat{\sigma}_m-\sigma_m) + \cdots . \quad (3.4.17)$$

Consider

$$\begin{aligned}\frac{\partial \Psi(\boldsymbol{\sigma})}{\partial \sigma_l} &= \frac{\partial}{\partial \sigma_l}\Big(\sum_{i=1}^{k}\Phi K_i \text{var}(\widehat{\gamma}_i)K_i^T\Phi^T + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi K_i \text{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)K_j^T\Phi^T\Big)\\ &= \sum_{i=1}^{k}\frac{\partial \Phi K_i}{\partial \sigma_l}\text{var}(\widehat{\gamma}_i)K_i^T\Phi^T + \sum_{i=1}^{k}\Phi K_i \text{var}(\widehat{\gamma}_i)\frac{\partial K_i^T\Phi^T}{\partial \sigma_l}\\ &\quad + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\frac{\partial \Phi K_i}{\partial \sigma_l}\text{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)\Phi K_j + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi K_i \text{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)\frac{\partial K_j^T\Phi^T}{\partial \sigma_l} \quad (3.4.18)\end{aligned}$$

Now,

$$\frac{\partial \Phi K_i}{\partial \sigma_l} = \sum_{i=1}^{k}\frac{\partial \Phi}{\partial \sigma_l}K_i + \sum_{i=1}^{k}\Phi\frac{\partial K_i}{\partial \sigma_l} = \sum_{i=1}^{k}(-\Phi P_l \Phi)K_i + \sum_{i=1}^{k}\Phi\frac{\partial K_i}{\partial \sigma_l}. \quad (3.4.19)$$

and

$$\begin{aligned}\frac{\partial K_i}{\partial \sigma_l} &= \{X^T(\boldsymbol{\gamma})\frac{\partial}{\partial \sigma_l}(\frac{\partial \Sigma^{-1}}{\partial \gamma_i})X(\boldsymbol{\gamma}) + \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \gamma_i}\frac{\partial \Sigma^{-1}}{\partial \sigma_l}X(\boldsymbol{\gamma}) - X^T(\boldsymbol{\gamma})\frac{\partial \Sigma^{-1}}{\partial \sigma_l}\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}\\ &\quad -X^T(\boldsymbol{\gamma})\frac{\partial}{\partial \sigma_l}\big(\frac{\partial \Sigma^{-1}}{\partial \gamma_i}\big)X(\boldsymbol{\gamma}) - \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \gamma_i}\frac{\partial \Sigma^{-1}}{\partial \sigma_l}X(\boldsymbol{\gamma})\}\boldsymbol{\beta}\\ &= -X^T(\boldsymbol{\gamma})\frac{\partial \Sigma^{-1}}{\partial \sigma_l}\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}\boldsymbol{\beta}.\end{aligned}$$

Let us denote $M_{li} = \Phi X^T(\boldsymbol{\gamma})\partial\Sigma^{-1}/\partial\sigma_l \times \partial X(\boldsymbol{\gamma})/\partial\gamma_i \times \boldsymbol{\beta}$.

Therefore,

$$\begin{aligned}\frac{\partial \Phi K_i}{\partial \sigma_l} &= \sum_{i=1}^{k}(-\Phi P_l\Phi)K_i + \sum_{i=1}^{k}\Phi - X^T(\boldsymbol{\gamma})\frac{\partial \Sigma^{-1}}{\partial \sigma_l}\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}\boldsymbol{\beta}\\ &= \sum_{i=1}^{k}(-\Phi P_l\Phi)K_i - M_{li}. \quad (3.4.20)\end{aligned}$$

Using (3.4.19) and (3.4.20), (3.4.18) becomes,

$$\frac{\partial \Psi(\boldsymbol{\sigma})}{\partial \sigma_l} = \sum_{i=1}^{k}\sum_{i=1}^{k}\left\{(-\Phi P_l \Phi K_i - M_{li})\text{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j)K_j^T\Phi^T + \Phi K_i \text{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j)(-\Phi P_l \Phi K_j - M_{lj})^T\right\}.$$

Substituting the above in (3.4.17), we get

$$\begin{aligned}
&\Psi(\widehat{\boldsymbol{\sigma}}) \\
&= \Psi(\boldsymbol{\sigma}) + \sum_{l=1}^{r}\Bigg[\sum_{i=1}^{k}\left\{(-\Phi P_l \Phi K_i - M_{li})\text{var}(\widehat{\gamma}_i)K_i^T\Phi^T + \Phi K_i \text{var}(\widehat{\gamma}_i)(-\Phi P_l \Phi K_i - M_{li})^T\right\} \\
&+ \sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{i=1}^{k}\left\{(-\Phi P_l \Phi K_i - M_{li})\text{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j)K_j^T\Phi^T + \Phi K_i \text{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j)(-\Phi P_l \Phi K_j - M_{lj})^T\right\}\Bigg] \\
&+ \sum_{l=1}^{r}\sum_{m=1}^{r}\frac{\partial^2 \Psi}{\partial \sigma_l \partial \sigma_m}(\widehat{\sigma}_l - \sigma_l)(\widehat{\sigma}_m - \sigma_m) \\
&= \Psi(\sigma) + O_p(n^{-2}). \qquad (3.4.21)
\end{aligned}$$

Therefore, the adjusted covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $\widehat{\Phi}_A = \widehat{\Phi} + 2\widehat{\Lambda} - \widehat{R}^{\star} + \widehat{\Psi}$ and $E(\widehat{\Phi}_A) = \text{var}(\widehat{\boldsymbol{\beta}}) + O(n^{-2})$. This completes the proof.

3.4.3 Derivation of $\widetilde{E}$ and $\widetilde{V}$ used in Theorem 2

We begin by computing the expected value and variance of the F statistic

$$F = \frac{1}{l}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1}L^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

The expected value is approximated using the conditional expectation arguments.

$$\begin{aligned} E(F) &= E\{E(F|\widehat{\boldsymbol{\sigma}})\}, \\ E(F|\widehat{\boldsymbol{\sigma}}) &= E\{\frac{1}{l}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1}L^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})|\widehat{\boldsymbol{\sigma}}\} \\ &= \frac{1}{l}\Big[E\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T L\}(L^T\widehat{\Phi}_A L)^{-1}E\{L^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\} \\ &\quad +\text{trace}\big\{(L^T\widehat{\Phi}_A L)^{-1}\text{var}(L^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}))\big\}\Big]. \end{aligned}$$

Since $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$, therefore

$$lE(F|\widehat{\boldsymbol{\sigma}}) = \text{trace}\{(L^T\widehat{\Phi}_A L)^{-1}(L^T V L)\}, \text{where } V = \text{var}(\widehat{\boldsymbol{\beta}}) = \Phi + \Lambda + \Psi. \qquad (3.4.22)$$

Since $\widehat{\Phi}_A = \widehat{\Phi} + \widehat{A}^* + \widehat{\Psi}$, where $\widehat{A}^* = 2\widehat{\Phi}\big\{\sum_{l=1}^r\sum_{m=1}^r \text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)(\widehat{Q}_{lm} - \widehat{P}_l\widehat{\Phi}\widehat{P}_m - \frac{1}{4}\widehat{R}_{lm})\big\}\widehat{\Phi}$ (Alnosaier 2007), we get

$$\begin{aligned} (L^T\widehat{\Phi}_A L)^{-1}(L^T V L) &= \{(L^T\widehat{\Phi}L) + (L^T\widehat{A}^*L) + (L^T\widehat{\Psi}L)\}^{-1}(L^T V L) \\ &= \big[(L^T\widehat{\Phi}L)\big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}\big]^{-1}(L^T V L) \\ &= \big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}^{-1}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= (L^T\widehat{\Phi}L)^{-1}(L^T V L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T V L) - \\ &\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T V L). \end{aligned}$$

From Alnosaier (2007), we notice that

$$E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T V L)\}] = l + A_2 + 2A_3 + O_p(n^{-3/2}), \qquad (3.4.23)$$

$$E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^TVL)\}] = E\{\text{trace}(\Theta\widehat{A}^*)\}$$
$$= 2A_3 + O(n^{-3/2}), \quad (3.4.24)$$

where

$$\Theta = L(L^T\Phi L)^{-1}L^T,$$
$$A_2 = \sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\text{trace}(\Theta\Phi P_l\Phi\Theta\Phi P_m\Phi),$$
$$A_3 = \sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\text{trace}\big(\Theta\Phi(Q_{lm} - P_l\Phi P_m - \frac{1}{4}R_{lm})\big).$$

Next, we need to consider

$$(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^TVL) = (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L) +$$
$$(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Lambda L) +$$
$$(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Psi L) +$$
$$= (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L) + O(n^{-2}).$$

Using the Taylor series expansion for $(L^T\widehat{\Phi}L)^{-1}$ about $\boldsymbol{\sigma}$, we get

$$\begin{aligned}
& (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L) + O(n^{-2}) \\
= & \left\{(L^T\Phi L)^{-1} + \sum_{l=1}^{r}(\widehat{\sigma}_l - \sigma_l)\frac{\partial(L^T\Phi L)^{-1}}{\partial\sigma_l}\right. \\
& \left. + \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}(\widehat{\sigma}_l - \sigma_l)(\widehat{\sigma}_m - \sigma_m)\frac{\partial^2(L^T\Phi L)^{-1}}{\partial\sigma_l\partial\sigma_m}\right\}(L^T\widehat{\Psi}L)^{-1} \\
\times & \left\{(L^T\Phi L)^{-1} + \sum_{l=1}^{r}(\widehat{\sigma}_l - \sigma_l)\frac{\partial(L^T\Phi L)^{-1}}{\partial\sigma_l}\right. \\
& \left. + \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}(\widehat{\sigma}_l - \sigma_l)(\widehat{\sigma}_m - \sigma_m)\frac{\partial^2(L^T\Phi L)^{-1}}{\partial\sigma_l\partial\sigma_m}\right\}(L^T\Phi L) \\
= & (L^T\Phi L)^{-1}(L^T\widehat{\Psi}L) + O_p(n^{-3/2}).
\end{aligned}$$

As shown in Theorem 1, $E(\widehat{\Psi}) = \Psi + O_p(n^{-2})$, therefore

$$\begin{aligned}
E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^TVL)\}] &= E\{\text{trace}(\Theta\widehat{\Psi})\} \\
&= A_4 + O(n^{-3/2}), \qquad (3.4.25)
\end{aligned}$$

where

$$A_4 = \sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\text{trace}(\Theta\widehat{\Psi}).$$

Therefore, substituting (3.4.23)-(3.4.25) in (3.4.22), we obtain the approximation of the expected value of the F statistics (3.1.7).

$$E(F) = 1 + \frac{A_2}{l} - \frac{A_4}{l} + O(n^{-3/2}). \qquad (3.4.26)$$

Now, we need to compute the variance of the F statistic.

$$\text{var}(F) = E\{\text{var}(F|\widehat{\boldsymbol{\sigma}})\} + \text{var}\{E(F|\widehat{\boldsymbol{\sigma}})\}. \tag{3.4.27}$$

Consider

$$\begin{aligned} \text{var}(F|\widehat{\boldsymbol{\sigma}}) &= \frac{1}{l^2}\text{var}\big\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1}L^T(\widehat{\boldsymbol{\beta}})|\widehat{\boldsymbol{\sigma}}\big\} \\ &= \frac{2}{l^2}\text{trace}\big\{(L^T\widehat{\Phi}_A L)^{-1}(L^T V L)\big\}^2, \text{ since it is a quadratic form.} \end{aligned} \tag{3.4.28}$$

Using similar arguments as used in Alnosaier (2007), we approximate the following

$$\begin{aligned} (L^T\widehat{\Phi}_A L)^{-1}(L^T V L) &= \{(L^T\widehat{\Phi}L) + (L^T\widehat{A}^*L) + (L^T\widehat{\Psi}L)\}^{-1}(L^T V L) \\ &= \big\{(L^T\widehat{\Phi}L)\big(I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big)\big\}^{-1}(L^T V L) \\ &= \big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}^{-1}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= (L^T\widehat{\Phi}L)^{-1}(L^T V L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T V L) - \\ &\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T V L). \end{aligned}$$

Therefore

$$
\begin{aligned}
& \operatorname{trace}\big\{(L^T\widehat{\Phi}_A L)^{-1}(L^T V L)\big\}^2 \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T V L)\big\}^2 \\
& -2\operatorname{trace}\{(L^T\widehat{\Phi} L)^{-1}(L^T V L)(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{A}^* L)(L^T\widehat{\Phi} L)^{-1}(L^T V L)\} \\
& -2\operatorname{trace}\{(L^T\widehat{\Phi} L)^{-1}(L^T V L)(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{\Psi} L)(L^T\widehat{\Phi} L)^{-1}(L^T V L)\} + O_p(n^{-2}) \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} \\
& -2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{A}^* L) \\
& \cdot(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} - 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L) \\
& \cdot(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{\Psi} L)(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} + O_p(n^{-2}) \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)\big\} + \\
& \operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi} L)^{-1}(L^T\Lambda L)\big\} + \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi} L)^{-1}(L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{A}^* L)(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{\Psi} L)(L^T\widehat{\Phi} L)^{1}(L^T\Phi L)\big\} + O_p(n^{-2})
\end{aligned}
$$

Using the Taylor series expansion about $\boldsymbol{\sigma}$, we get

$$
\begin{aligned}
= \;& \operatorname{trace}\big\{I + 2\sum_{l=1}^{r}(\widehat{\sigma}_l - \sigma_l)\frac{\partial}{\partial\sigma_l}(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L) + \\
& \sum_{l=1}^{r}\sum_{m=1}^{r}\operatorname{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\frac{\partial^2}{\partial\sigma_l\partial\sigma_m}(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)\big\} + \\
& \operatorname{trace}\big\{\sum_{l=1}^{r}\sum_{m=1}^{r}\operatorname{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\frac{\partial}{\partial\sigma_l}(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)\frac{\partial}{\partial\sigma_m}(L^T\widehat{\Phi} L)^{-1}(L^T\Phi L)\big\} + \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Lambda L)\big\} + 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{A}^* L)\big\} - 2\operatorname{trace}\big\{(L^T\widehat{\Phi} L)^{-1}(L^T\widehat{\Psi} L)\big\} + O_p(n^{-3/2}).
\end{aligned}
$$

Using $E(\widehat{\Psi}) = \Psi + O(n^{-2})$ and (3.4.28), we approximate the expected value of the conditional variance as follows,

$$E\{\text{var}(F|\widehat{\boldsymbol{\sigma}})\} = \frac{2}{l^2}(l + 3A_2) + O(n^{-3/2}). \tag{3.4.29}$$

Also, using the expression of the mean i.e. (3.4.26), we obtain

$$\text{var}\{E(F|\widehat{\boldsymbol{\sigma}})\} = \frac{A_1}{l^2} + O(n^{-3/2}), \text{ where} \tag{3.4.30}$$

$$A_1 = \sum_{l=1}^{r}\sum_{m=1}^{r} \text{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\text{trace}(\Theta\Phi P_l\Phi)\text{trace}(\Theta\Phi P_m\Phi).$$

Using (3.4.29) and (3.4.30) in (3.4.27) we obtain

$$\text{var}(F) = \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2} + O(n^{-3/2}).$$

Therefore the approximate expected value and variance of the F statistic that will be used for matching moments of λF with those of $F(l, d)$ distribution are given as

$$\widetilde{E} = 1 + \frac{A_2}{l} - \frac{A_4}{l},$$
$$\widetilde{V} = \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2}.$$

This completes the proof.

3.4.4 Implementation of Simulation Study

We derive the required quantities for a block design experiment, where the fixed effect covariates are generated from a Gaussian distribution. Therefore,

$$\frac{\partial X^{\star}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1} = \begin{pmatrix} 0 & : X^{\star} = X \\ 1 & : X^{\star} = \boldsymbol{\gamma}_1 \end{pmatrix}$$

and

$$\frac{\partial X^{\star}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2} = 0.$$

Also, now we need to compute $\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}$ and $\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}$. In the following, $\Sigma = \sigma_e^2 D_2 + \sigma_b^2 D_1$, and for our case we take $D_2 = I$; therefore

$$\frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2} = D_2 = I$$
$$\frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2} = D_1.$$

The REML equations are given by:

$$2\frac{\partial l(\boldsymbol{\sigma})}{\partial \sigma_e^2} = -\text{trace}(G) + \mathbf{Y}^T G G \mathbf{Y} \text{ since} D_2 = I, \tag{3.4.31}$$
$$2\frac{\partial l(\boldsymbol{\sigma})}{\partial \sigma_b^2} = -\text{trace}(G D_1) + \mathbf{Y}^T G D_1 G \mathbf{Y}, \tag{3.4.32}$$
where $G = \Sigma^{-1} - \Sigma^{-1} X^{\star} (X^{\star} \Sigma^{-1} X^{\star})^{-1} X^{\star T} \Sigma^{-1}$.

Again, for notational convenience, we will replace $X^{\star}$ by X. To obtain MLE of σ_e^2 and

σ_b^2, we need to set the equations (3.4.31) and (3.4.32) to zero.

$$\begin{aligned}
&\left|\frac{\partial l}{\partial \sigma_e^2}\right|_{\sigma_e^2=\widehat{\sigma}_e^2,\sigma_b^2=\widehat{\sigma}_b^2} = 0,\\
\Longrightarrow\quad & \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) = \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\\
\Longrightarrow\quad & \frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) = \frac{\partial}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\}. \qquad (3.4.33)
\end{aligned}$$

and

$$\begin{aligned}
&\left|\frac{\partial l}{\partial \sigma_b^2}\right|_{\sigma_e^2=\widehat{\sigma}_e^2,\sigma_b^2=\widehat{\sigma}_b^2} = 0,\\
&\Longrightarrow\ \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1) = \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1 G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\\
&\Longrightarrow\ \frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1) = \frac{\partial}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1 G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\}. \qquad (3.4.34)
\end{aligned}$$

Let us consider the LHS (left hand side) of equation (3.4.31)

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}\{G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\}\\
=\ & [\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}) - \text{trace}\{\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}X)^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\}]\\
=\ & \{\frac{\partial}{\partial \widehat{\sigma}_e^2}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1})\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial}{\partial \widehat{\sigma}_b^2}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1})\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\} - \frac{\partial}{\partial \boldsymbol{\gamma}}(\text{trace}(M(\boldsymbol{\gamma}))), \text{where}\\
& M = \Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}X)^{-1})X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1},\\
=\ & \text{trace}(\frac{\partial \text{trace}\Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \Sigma}\frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2})\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \text{trace}(\frac{\partial \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1})}{\partial \Sigma}\frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2}) - \frac{\partial}{\partial \boldsymbol{\gamma}}(\text{trace}(M(\boldsymbol{\gamma})))\\
=\ & \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_2)\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_1)\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \text{trace}(\frac{\partial M(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}). \quad (3.4.35)
\end{aligned}$$

We need to explicitly solve one of the terms in (3.4.35),

$$\frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}}$$

$$= \frac{\partial \Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \boldsymbol{\gamma}} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} + \tag{3.4.36}$$

$$\Sigma(\hat{\sigma}(\gamma))^{-1}(\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} + \tag{3.4.37}$$

$$\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\frac{\partial (X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}}{\partial \boldsymbol{\gamma}} X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} + \tag{3.4.38}$$

$$\Sigma(\hat{\sigma}(\gamma))^{-1}X(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma(\hat{\sigma}(\gamma))^{-1} + \tag{3.4.39}$$

$$\Sigma(\hat{\sigma}(\gamma))^{-1}X(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\frac{\partial \Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \boldsymbol{\gamma}}. \tag{3.4.40}$$

Further, we need to solve all the terms (3.4.36)-(3.4.40) separately using standard matrix calculus. Therefore consider (3.4.36),

$$\left(\frac{\partial \Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \widehat{\sigma}_e^2}\cdot\frac{\partial \sigma_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial \Sigma(\hat{\sigma}(\gamma))^{-1}}{\partial \widehat{\sigma}_b^2}\cdot\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\right)$$

$$\cdot X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}$$

$$= \left\{(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}) - (\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}})\right\}$$

$$\cdot X(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}. \tag{3.4.41}$$

Solving (3.4.38) we get,

$$
\begin{aligned}
&-\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\Big\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\\
&-X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\\
&-X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}X\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\\
&+X^T\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Big\}(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}. \quad (3.4.42)
\end{aligned}
$$

Solving (3.4.40) we get,

$$
\begin{aligned}
&\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\times\\
&\Big\{-\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}-\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\Big\}. \quad (3.4.43)
\end{aligned}
$$

Substituting the expressions from (3.4.37), (3.4.39), (3.4.41), (3.4.42) and (3.4.43) we get,

$$
\begin{aligned}
& \frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} \\
= \;& \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ - \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
+ \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
+ \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} \Big\} \\
+ \;& \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ - \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
+ \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
- \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} \Big\} \\
+ \;& \Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
- \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
- \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
+ \;& \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1}. \qquad (3.4.44)
\end{aligned}
$$

Now substituting (3.4.44) in (3.4.35), we obtain the following,

$$
\begin{aligned}
&\frac{\partial \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))}{\partial \boldsymbol{\gamma}} \\
= \;& \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_2) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& - \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& \cdot D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1}) \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_1) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& - \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& \cdot D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1}) \Big\} \\
& - \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} (X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& - \text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1}). \qquad (3.4.45)
\end{aligned}
$$

Moving on, we compute the RHS (right hand side) of equation (3.4.33),

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}} (\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y}) \\
& = \mathbf{Y}^T \frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} + \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} \mathbf{Y} \qquad (3.4.46)
\end{aligned}
$$

Consider,

$$\begin{aligned}
&\frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}}\{\Sigma(\hat{\sigma}(\gamma))^{-1} - \Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\} \\
&= -\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} - \frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}}. \qquad (3.4.47)
\end{aligned}$$

Now using (3.4.44) in (3.4.47), we get,

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}}(\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}) \\
= \; & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\Big\{ -\Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& -\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& \cdot D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) (X'(\gamma) \Sigma(\hat{\sigma}(\gamma))^{-1} X(\gamma))^{-1} \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} \Big\} \\
& +\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\Big\{ -\Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& -\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& \cdot D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} \Big\} \\
& -\Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& +\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma(\hat{\sigma}(\gamma))^{-1} \\
& -\Sigma(\hat{\sigma}(\gamma))^{-1} X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma(\hat{\sigma}(\gamma))^{-1}. \qquad (3.4.48)
\end{aligned}
$$

For notational convenience let us define three new quantities U, U_1 and V as follows,

$$\begin{aligned} U &= \Big\{\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad -\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad \cdot D_2\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad +\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}\Big\}, \quad (3.4.49) \end{aligned}$$

$$\begin{aligned} U_1 &= \Big\{\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad -\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad \cdot D_1\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad +\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}\Big\}, \quad (3.4.50) \end{aligned}$$

and

$$\begin{aligned} V &= -\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad +\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad +\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1} \\ &\quad -\Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})(X'(\gamma)\Sigma(\hat{\sigma}(\gamma))^{-1}X(\gamma))^{-1}\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma(\hat{\sigma}(\gamma))^{-1}. \quad (3.4.51) \end{aligned}$$

Now, substituting the equations $(3.4.48), (3.4.49), (3.4.50), (3.4.51)$ in $(3.4.46)$ we get RHS

of (3.4.33) as

$$
\begin{aligned}
& \mathbf{Y}^T\Big\{\frac{\partial\widehat{\sigma}_e^2}{\partial\boldsymbol{\gamma}}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}+U)+\frac{\partial\widehat{\sigma}_b^2}{\partial\boldsymbol{\gamma}}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}+U_1) \\
& +V\Big\}G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}+\mathbf{Y}^TG(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\Big\{\frac{\partial\widehat{\sigma}_e^2}{\partial\boldsymbol{\gamma}}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}+U) \\
& +\frac{\partial\widehat{\sigma}_b^2}{\partial\boldsymbol{\gamma}}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}+U_1)+V\Big\}\mathbf{Y} \\
= \; & \frac{\partial\widehat{\sigma}_e^2}{\partial\boldsymbol{\gamma}}\Big\{\mathbf{Y}^TUG\mathbf{Y}+\mathbf{Y}^TGU\mathbf{Y}-\mathbf{Y}^T\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1} \\
& -\mathbf{Y}^TG\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y}\Big\} \\
& +\frac{\partial\widehat{\sigma}_b^2}{\partial\boldsymbol{\gamma}}\Big\{\mathbf{Y}^TU_1G\mathbf{Y}+\mathbf{Y}^TGU_1\mathbf{Y}-\mathbf{Y}^T\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1} \\
& -\mathbf{Y}^TG\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}\mathbf{Y}\Big\}+\mathbf{Y}^TVG\mathbf{Y}+\mathbf{Y}^TGV\mathbf{Y}. \qquad (3.4.52)
\end{aligned}
$$

Equating the RHS (3.4.52) and LHS (3.4.45) of (3.4.31) we get,

$$
\begin{aligned}
& \frac{\partial\widehat{\sigma}_e^2}{\partial\boldsymbol{\gamma}}\Big\{\text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_2)+\text{trace}(U)-\mathbf{Y}^TUG\mathbf{Y}-\mathbf{Y}^TGU\mathbf{Y} \\
& +\mathbf{Y}^T\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\Sigma(\hat{\sigma}(\gamma))^{-1}G\mathbf{Y}+\mathbf{Y}^TG\Sigma(\hat{\sigma}(\gamma))^{-1}D_2\mathbf{Y}\Big\} \\
& +\frac{\partial\widehat{\sigma}_b^2}{\partial\boldsymbol{\gamma}}\Big\{\text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_1)+\text{trace}(U_1)-\mathbf{Y}^TU_1G\mathbf{Y}-\mathbf{Y}^TGU_1\mathbf{Y} \\
& +\mathbf{Y}^T\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1}G\mathbf{Y}+\mathbf{Y}^TG\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\mathbf{Y}\Big\} \\
= \; & \mathbf{Y}^TVG\mathbf{Y}+\mathbf{Y}^TGV\mathbf{Y}-\text{trace}(V). \qquad (3.4.53)
\end{aligned}
$$

Similarly, we need to solve both sides of the equation (3.4.32). Consider the LHS

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}}\{\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}D_1 - \Sigma(\hat{\sigma}(\gamma))^{-1}X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma(\hat{\sigma}(\gamma))^{-1}D_1)\} \\
= & \frac{\partial}{\partial \widehat{\sigma}_e^2}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}D_1)\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial}{\partial \widehat{\sigma}_b^2}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}D_1)\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \frac{\partial}{\partial \boldsymbol{\gamma}}(\text{trace}(M(\boldsymbol{\gamma})D_1)) \\
= & \text{trace}\{\frac{\partial}{\partial \Sigma}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}D_1)\frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2}\} + \text{trace}\{\frac{\partial}{\partial \Sigma}\text{trace}(\Sigma(\hat{\sigma}(\gamma))^{-1}D_1)\frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2}\}\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \\
& -\frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(M(\boldsymbol{\gamma})D_1). \qquad (3.4.54)
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(M(\boldsymbol{\gamma})D_1) &= \text{trace}(\frac{\partial MD_1}{\boldsymbol{\gamma}}) \\
&= \text{trace}(\frac{\partial M}{\partial \boldsymbol{\gamma}} \cdot D_1) \\
&= \text{trace}\{(-\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}U - \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}U_1 - V) \cdot D_1\} \\
&= \text{trace}(-UD_1)\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} - \text{trace}(U_1D_1)\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \text{trace}(VD_1). \qquad (3.4.55)
\end{aligned}$$

Using (3.4.55) in (3.4.54), we get

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(GD_1) &= \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\{-\text{trace}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1})^TD_2 + \text{trace}(UD_1)\} \\
&\quad + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\{-\text{trace}(-\Sigma(\hat{\sigma}(\gamma))^{-1}D_1\Sigma(\hat{\sigma}(\gamma))^{-1})^TD_1 + \text{trace}(U_1D_1)\} \\
&\quad + \text{trace}(VD_1). \qquad (3.4.56)
\end{aligned}$$

Solving RHS of (3.4.34) we get,

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}}(\mathbf{Y}^T G D_1 G \mathbf{Y}) \\
= \; & \mathbf{Y}^T (\frac{\partial G}{\partial \boldsymbol{\gamma}}) D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \frac{\partial G}{\partial \boldsymbol{\gamma}} \mathbf{Y} \\
= \; & \mathbf{Y}^T \Big\{ \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} (U - \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} (U_1 - \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1}) + V \Big\} D_1 G \mathbf{Y} \\
& + \mathbf{Y}^T G D_1 \Big\{ \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} (U - \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1}) \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} (U_1 - \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1}) + V \Big\} \mathbf{Y} \\
= \; & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ \mathbf{Y}^T U D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 U \mathbf{Y} - \mathbf{Y}^T \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} \mathbf{Y} \Big\} + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ \mathbf{Y}^T U_1 D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 U_1 \mathbf{Y} \\
& - \mathbf{Y}^T \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 G \mathbf{Y} - \mathbf{Y}^T G D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} \mathbf{Y} \Big\} \\
& + \mathbf{Y}^T V D_1 G \mathbf{Y} - \mathbf{Y}^T G D_1 V \mathbf{Y}. \qquad (3.4.57)
\end{aligned}
$$

Therefore, equating LHS (3.4.56) = RHS (3.4.57) we get,

$$
\begin{aligned}
& \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ - \text{trace}((\Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1})^T D_2) + \text{trace}(U D_1) - \mathbf{Y}^T U D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 U \mathbf{Y} + \mathbf{Y}^T \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_2 \Sigma(\hat{\sigma}(\gamma))^{-1} \mathbf{Y} \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ - \text{trace}((\Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1})^T D_1) + \text{trace}(U_1 D_1) - \mathbf{Y}^T U_1 D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 U_1 \mathbf{Y} + \mathbf{Y}^T \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} D_1 \Sigma(\hat{\sigma}(\gamma))^{-1} \mathbf{Y} \Big\} \\
& = \mathbf{Y}^T V D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 V \mathbf{Y} - \text{trace}(V D_1). \qquad (3.4.58)
\end{aligned}
$$

Finally, we can solve the equations (3.4.53) and (3.4.58) simultaneously to obtain $\partial \widehat{\sigma}_e^2 / \partial \boldsymbol{\gamma}$

and $\partial\widehat{\sigma}_b^2/\partial\boldsymbol{\gamma}$.

Remark: Here we have derived the quantities to implement the simulation for a study with one random effect. In the studies, which have more than one random effect, similar method can be used to derive the required quantities.

Chapter 4

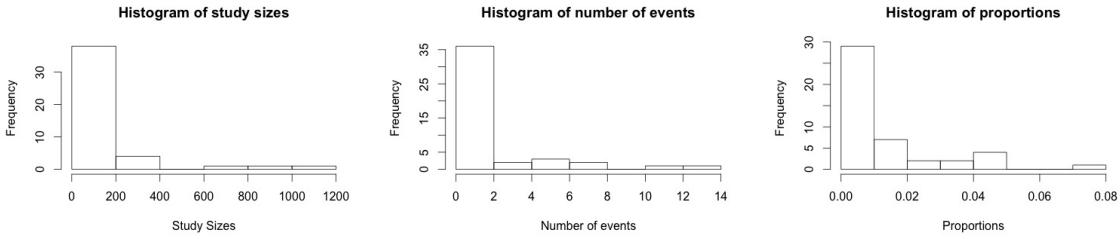
Application: Meta-analysis using random-effects models

Meta-analysis provides a way of combining outcomes across different studies. The use of meta-analysis is very frequent in clinical trials and many other applications. The International Conference on Harmonization E9 guideline states that meta-analyses should be prospectively planned with the clinical trials program in the development of a new treatment. In clinical trials, meta-analysis is primarily used to quantify the effect size and its uncertainty of the primary or the key secondary endpoints. The Cochrane Collaboration provides the world's largest repository of systematic review and meta-analysis frequently used in clinical trials. The underlying assumption of the commonly used statistical methods is that all key endpoints have reasonable number of outcomes/effects across different studies. However, there are instances (mainly in safety studies) when the primary interests of the outcomes from different studies are rare (low incidences) and meta-analysis is required to capture the effect size. For example, consider a recently investigated meta-analysis study of three types of needles used in the EUS-FNA[1] procedure for gastrointestinal lesions (for the confidentiality reason, we are not disclosing the name of the sponsor here). This procedure is used to obtain biopsies before performing any pancreatic surgery. The three most commonly used needles in this procedure are 19-gauge (19G), 22-gauge (22G) and 25-gauge (25G) with

[1]EUS-FNA: endoscopic ultrasound fine needle aspiration

the 19G needle having the largest diameter. Given the larger size, the 19G needle can yield sufficient tissue for diagnosis but is relatively difficult to maneuver and is suspected to lead to more complications like excessive bleeding (Trindade & Berzin, 2013) especially with the less-experienced endosonographers. The focus of the recently conducted meta-analysis was to compare the overall complication rates due to pancreatitis, bleeding due to the needle, perforation, infection, fever, pain or elevated pancreatic enzyme. Since the usage of 19G is limited, only six studies are available in the literature reporting overall complications which are as follows 0/38, 0/72, 2/18, 1/44, 0/30 and 0/13, where the denominator represents the total number of subjects in the study and numerator represents the number of subjects reporting complications. On the other hand, 45 studies are available reporting the usage of 22G or 25G needles of which 27 studies reported zero overall complications and 18 studies reported some amount of complications mostly between 1 to 5. A summary of the data is provided in Figure 1.

Figure 4.1: Summary of the 22/25G data



The studies included in the above investigation are based on a similar patient population. The paucity of the events from different studies and non-randomization nature of these studies poses a serious threat to the statistical comparison and exacerbates the use of all statistical methods that are not designed for these situations.

In the following, section 4.1 and section 4.2 present our extension of the selected CPR

methods and the new proposed methods incorporating zero-inflated models, section 4.3 presents the implementation of the methods on real world data, section 4.4 presents a simulation study where the comparisons are made to the MH method, 1959 and the Peto method, 1985 and finally section 4.5 draws the conclusion.

4.1 Likelihood-based approach due to CPR

Suppose we observe data from n different studies, out of which m are from treatment arm 1 and $n-m$ are from treatment arm 2. Let Y_i be the number of adverse events observed among n_i subjects in the i^{th} study. Let X_i be the indicator covariate indicating treatment arm 1. In this study, we are interested in detecting the difference between the two treatment groups. Since the response variable is discrete, we begin by assuming a Poisson model on Y_i. The methods suggested in CPR were used only for randomized studies where every study observed subjects from both treatment arms (treatment arm 1 and 2). However, in our setup, we want to draw a comparison between two treatments where some or all studies do not have observed subjects in both treatment arms. We derive the methods in CPR for our setup. In the following models, we write down the explicit likelihood and maximize it to get the estimates of the parameters of interest.

4.1.1 Poisson model with fixed relative risk: POIF

Following a similar model as in CPR, Y_i follows a Poisson distribution with rate $n_i\ \lambda_i$ where,

$$\lambda_i = \xi_i e^{\tau X_i}. \tag{4.1.1}$$

Following similar notation as in CPR, the parameter ξ_i represents the baseline event rate while τ represents the log relative risk associated with the treatment. Here the baseline event rates vary from study to study, but the relative risk (e^{τ}) is constant. A gamma distribution is assumed on the ξ_i i.e. $\xi_i \sim \text{Gamma}(\alpha, \beta)$. The likelihood of the model can be written as follows,

$$L_n(\theta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}\Gamma(y_i+\alpha)}{\Gamma(\alpha)(\beta+n_i e^{\tau X_i})^{(y_i+\alpha)}} \frac{(n_i e^{\tau X_i})^{y_i}}{y_i!} \tag{4.1.2}$$

where, $\boldsymbol{\theta} = (\alpha, \beta, \tau)'$ represents vector of model parameters. The above likelihood equation can be maximised to obtain the estimated parameters.

4.1.2 Poisson model with random relative risk: POIR

Model (4.1.1) is extended by assuming log relative risk to be random, hence, in this case both baseline risk and relative risks are varied from study to study. Therefore, the model is

$$\lambda_i = \xi_i e^{(\tau_i X_i)}, \tag{4.1.3}$$

where the baseline risk is $\xi_i \sim$ Gamma (α, β) and the log relative risk is $\tau_i \sim N(\mu, \sigma^2)$. The resulting likelihood function is given by,

$$\prod_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\beta^{\alpha}\Gamma(y_i+\alpha)}{\Gamma(\alpha)(\beta+n_i e^{\tau X_i})^{(y_i+\alpha)}} \frac{(n_i e^{\tau X_i})^{y_i}}{y_i!} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\tau-\mu)^2/2\sigma^2} d\tau. \tag{4.1.4}$$

We approximate the above likelihood function using gaussian quadrature and then obtain the estimate of $\boldsymbol{\theta} = (\alpha, \beta, \mu, \sigma)'$.

4.2 Proposed methods for meta-analysis based on zero inflated Poisson models

On further examination of the EUS-guided FNA data described in the introduction, we observe a majority of responses as zeros. We suspect that there are more zeros in the data than would be predicted by a standard Poisson model. Since the data is overdispersed, we propose zero inflated models where the mean and variance do not have to be equal. This leads us to propose the following models.

4.2.1 Zero inflated Poisson model with fixed relative risk: ZIPF

We assume that Y_i follows a Poisson distribution with mean $n_i\lambda_i$ and inflation parameter π_i, i.e.

$$Y_i \sim \left\{ \begin{array}{ccc} P(n_i\lambda_i) & \text{w.p.} & \pi_i \\ 0 & \text{w.p.} & 1-\pi_i \end{array} \right\}.$$

As assumed in previous models, $\lambda_i = \xi_i e^{\tau X_i}$ where $\xi_i \sim Gamma(\alpha, \beta)$. Further, we use a beta distribution to model the inflation parameter i.e. $\pi_i \sim Beta(c, d)$. The full likelihood $L(\alpha, \beta, \tau, c, d)$ has been derived in the appendix. We can obtain the estimates of $\boldsymbol{\theta} = (\alpha, \beta, \tau, c, d)'$ by maximizing the likelihood.

4.2.2 Zero inflated Poisson model with random relative risk: ZIPR

In method (4.2.1), we accounted for the study variation ξ_i. Now we would like to introduce variation in τ. Therefore, we extend model 4.2.1 to $\lambda_i = \xi_i e^{\tau_i X_i}$ to allow for pos-

sibility of study to study heterogeneity with respect to relative risk. Both the baseline parameters and the relative risk are assumed to be random. Here, $\xi_i \sim Gamma(\alpha, \beta)$ and $\tau_i \sim Normal(\mu, \sigma^2)$. Further the inflation parameter is assumed to follow a beta distribution i.e. $\pi_i \sim Beta(c, d)$. To be able to get the estimates of the parameters $(\alpha, \beta, \mu, \sigma, \pi)$, we need to compute the likelihood of the specified model, which is approximated using numerical methods like the Gaussian quadrature and Laplace approximations. The method is detailed in the appendix.

4.3 Numerical studies

This section explicates the numerical studies using the methods discussed in the previous section. First, we discuss the analysis using the EUS-FNA study discussed in the introduction and show how the proposed methods can be applied in a non RCT setup. Next, we explicate the randomized meta-analysis by revisiting a recently published controversial meta-analysis study published in the New England Journal of Medicine journal (Nissen & Wolski, 2007, 2011).

4.3.1 Example 1: Overall complications due to needles in EUS-FNA

As stated in section 1, the objective of the study is to compare the overall complication rates due to the 19G and the 22G/25G needles in EUS guided FNA procedure. Only 6 studies are found in the literature reporting the overall surgical complications with 19G; whereas 45 studies are found using either the 22G needle or the 25G needle reporting the same. The traditional meta-analysis examines each treatment group separately by computing the pooled

estimates and the corresponding confidence intervals of the overall complication rates across all studies. Since many of the observations are zeros either in 19G group or in 22G/25G group, almost all statistical software to our knowledge became computationally vulnerable; therefore, to bypass the computational errors, a continuity correction 0.5 is added to the studies with zero counts. The naive estimates for the two groups are shown in Table 4.1.

Table 4.1 Initial proportion estimates of the treatment groups with and without continuity correction (CC)

	Treatment 1: 19G		Treatment 2: 22/25G		Relative Risk	
	Rate	95% C.I.	Rate	95% C.I.	Estimate	95% C.I.
With 0.5 CC	0.0036	[0,0.0180]	0.0063	[0.0039,0.0086]	0.571	[0.204,1.602]
Without any CC	0.0139	[0,0.0295]	0.0124	[0.0096,0.0152]	1.121	[0.262,4.801]

Based on the results in Table 4.1, by comparing the estimates, we notice that the overall complication rates of 22/25G (0.63%) is higher than that of 19G (0.36%). However, the 95% confidence intervals of the complication rates of 19G and 22/25G group overlap. Notice that the interval for the 19G group is much wider as compared to the 22/25G group, as the number of studies and the corresponding sample sizes in 19G group is much smaller than the 22/25G group. However, in this analysis, we have added a continuity correction of 0.5 to the studies with zero counts which can lead to incorrect results. When we base our calculations on raw data (without any continuity correction); we find that 3 out of 215 people had complications in the 19G group producing the risk to be 0.0139 as opposed to 74 out of 5957 (risk=0.0124) in the 22G/25G group. Hence, we obtain a relative risk of 1.121 suggesting that the overall complication rate is about 12% higher in the 19G group as compared to the 22G/25G group. Based on the traditional methods, no conclusion can be

drawn and further investigation is required.

For further analysis, we use the models described in sections 4.1 and 4.2. Since the data for each group contains excessive zeros, it is relevant to test if a significant number of zeros are coming from a zero state as opposed to the Poisson distribution. For this purpose, we conduct a likelihood ratio test to test the following hypothesis. $H_0 : \pi_i = 1 \ \forall i = 1, .., n$ in the model $Y_i \sim Poisson(n_i \xi_i e^{\tau X_i})$ with the inflation parameter as π_i, where $\xi_i \sim Gamma(\alpha, \beta)$, $\tau_i \sim N(\mu, \sigma^2)$ and π_i is the probability that the data follows a Poisson distribution without any zero inflation. The likelihood ratio test statistic is given by

$$-2log\left(\frac{\hat{L}(\hat{\theta}_{H_0})}{\hat{L}(\hat{\theta})}\right) \tag{4.3.5}$$

where, $\hat{\theta}_{H_0} = argmax_{\theta, H_0} \hat{L}(\theta)$. Since we are testing for $\pi_i = 1$, which is on the boundary of the parameter space, the likelihood ratio test statistics follows a mixture of χ^2 distribution with $50 : 50$ of χ^2_0 and χ^2_1 as discussed in Self and Liang, 1987. The value of the test statistic is 3.298 and the p-value is 0.03. Based on the test results we reject the null hypothesis thereby indicating that a zero inflated Poisson model would be more appropriate for this data set.

Estimation of relative risk

The methods in CPR were only derived for a randomized-control study. It was not tested for non randomized studies where the number of studies in the two arms are unequal; like our proposed data. However, we were able to extend the first two methods based on Poisson models from CPR for our set up. Hence, we use all models to fit the data for comparison. For models POIF and ZIPF, where τ is fixed, estimation of relative risk is straightforward. We can calculate the asymptotic confidence interval of τ by using the information matrix. But for

models POIR and ZIPR, where τ is random, we need to compute the posterior distribution of τ_i i.e. $f(\tau_i|\mathbf{y})$. Subsequently, the posterior mean can be calculated. Assuming τ_i, ξ_i and π_i to be mutually independent, the posterior distribution of τ_i when $X_i = 1$ can be obtained by

$$\begin{aligned} f(\tau_i|\mathbf{y}) &= \frac{f(\mathbf{y}|\tau_i).f(\tau_i)}{\int f(\mathbf{y}|\tau_i).f(\tau_i)d\tau_i}, \\ &= \frac{f(y_i|\tau_i).f(\tau_i)}{\int f(y_i|\tau_i).f(\tau_i)d\tau_i}, \text{ since } \tau_i \text{ and } y_j \text{ are independent for } i \neq j \end{aligned}$$

where $f(y_i|\tau_i) = \int_{-\infty}^{\infty} f(y_i|\tau_i, \boldsymbol{\theta_i}) f(\boldsymbol{\theta_i}) d\boldsymbol{\theta_i}$.

Here, $\boldsymbol{\theta_i} = \xi_i$ for POIR and $\boldsymbol{\theta_i} = (\xi_i, \pi_i)$ for ZIPR. The estimate of log relative risk for the i^{th} study when $\{X_i = 1\}$ is denoted by $\hat{\tau}_i = E(\tau_i|\mathbf{y})$ can be approximated using numerical methods such as gaussian quadrature. This was subsequently used to obtain the estimate of the log relative risk. We used parametric bootstrap to compute the standard error of the estimate. The results for the four models are given in Table 4.2. Here, for models POIR and ZIPR, we report the average of $\hat{\tau}_i' s$ as an estimate of the log relative risk.

Table 4.2 Estimates of log relative risk for EUS-FNA study

		log relative risk (τ)		C.I. of relative risk
Model	Log likelihood	Estimate	SE	95% C.I.
POIF	-67.886	0.36	0.77	(0.32, 6.48)
POIR	-67.836	0.13	0.83	(0.22, 5.79)
ZIPF	-66.237	0.47	0.77	(0.35, 7.23)
ZIPR	-66.179	0.26	0.64	(0.37, 4.55)

From the results, the mean of the baseline event rate is approximately 0.01 (not shown in

the table) for all methods, indicating little between study variations. For models ZIPF and ZIPR we observe that about 60% (mean of inflation parameter) of the observations come from the Poisson distribution while the remaining 40% come from a zero state. Since we are interested in comparing the overall complications in the 2 treatment groups, our parameter of interest in the above models is the log relative risk given by τ. From model ZIPR, the point estimate of the relative risk (e^{τ}) is 1.297, which suggests that the subjects in treatment group 1 (19G) have about 30.0% higher risk of overall complications as compared to those in treatment group 2 (22/25G).

The estimates given in Table 4.2 can be used to draw inferences about relative risk. The 95% confidence interval for the relative risk for model ZIPR is $(0.37, 4.55)$ which suggests that there is a 95% chance that the relative risk of overall complications in the 19G group compared to 22/25G is between 0.37 and 4.55. All models indicate higher complication rates in the 19G group, however, the estimates from none of the models are statistically significant; the confidence intervals are very wide. Although the relative risk estimate is not statistically different from one, it is worth noticing that the standard error of τ is smallest for ZIPR thereby resulting in tighter confidence interval as compared to the other three models. The table also gives the value of the log likelihood for the four models which is maximum for the model ZIPR.

The estimates obtained from all four models are different from those obtained in Table 4.1 which were based on naive calculations where a 0.5 correction was made to studies with zero events. No information is lost from the original data; they do not make any corrections or delete any part of the data. Also, the excessive zeros can be accommodated when they are observed by using a zero inflated distribution. Further, we will show by simulation that the zero inflated models (ZIPF and ZIPR) are a better fit for the data.

Simulation study

We need to find the model that best fits the data. We use a parametric bootstrap approach as described in Efron & Tibshirani, 1993 and Allcroft & Glasbey, 2003. The idea is to simulate data from each model and fit all the models to the same data set and compute and compare the goodness of fit statistics which in our case is the maximized log likelihood estimate. By simulating from each model in turn, we determine whether any model is consistent with the data. The best model is expected to behave in the same manner as the original data set . The formal step by step procedure is given as follows,

1. Fit the 4 models to the original observed data and estimate the model parameters by optimizing the likelihood. Record the values of the maximized log likelihood in the vector $\mathbf{u} = (u_1, u_2, u_3, u_4)'$.

2. From each of the four fitted models in step 1, replicate the data 200 times using the model parameters obtained in Step 1, generating 800 (200×4) data sets in all.

3. Refit all the four models to each replicated data set and record the maximized log likelihood estimates. So, for each data set we have four maximized log likelihood estimates corresponding to the four models say $\mathbf{l} = (l_1, l_2, l_3, l_4)'$. There are 200 such vectors when the data is replicated from the i^{th} model. Let $\bar{\mathbf{u}}_\mathbf{i}$ denote the mean of the these 200 vectors and S_i denote the sample covariance matrix for these 200 vectors.

4. Compare the maximized log likelihoods of the original observed data to that of the replicated data for each of the 4 models.

Now, since we use negative log likelihood as the goodness of fit criteria their joint distribution is multivariate normal. Therefore, we can use Mahalanobis squared distance to compare the

models (Mardia *et al.*, 1979) which is given by,

$$D_i^2 = (u - \bar{u}_i)' S_i^{-1} (u - \bar{u}_i). \tag{4.3.6}$$

This is the standardized squared distance between the point obtained from the original data and the mean of the i^{th} model. We can test the hypothesis $H_0 : i^{th}$ model is correct, using the statistic D_i^2. Under the null hypothesis D_i^2 is distributed as $4 \times F_{4,199}$. Therefore the p-values can be obtained.

Table 4.3 Mahalanobis distances (EUS-FNA study)

Model	D^2	P-value
POIF	30.35	$\ll$0.001
POIR	2.66	0.617
ZIPF	2.60	0.626
ZIPR	1.928	0.748

By looking at the results in Table 4.3, we can conclude that POIF is not appropriate to fit this data. Also, close comparison shows that the value of Mahalanobis distance for the ZIPR model is the least among all four methods. This justifies the use of the zero inflated Poisson model as opposed to a Poisson model.

Discussion of results

19G needles are the largest in diameter among the three needle sizes. Therefore, it is suspected to yield the largest amount of tissue sample, but it can increase the chances of complications as mentioned in Gerke, 2009 and Trindade & Berzin, 2013. From the results of all four models, the point estimate of relative risk suggests that the rate of overall compli-

cations would be higher in the 19G group as compared to the 22/25G group which complies with this reasoning. However, the results from all models are not statistically significant and we believe that this is because we have limited data available on the usage of the 19G needle. Out of the four models, ZIPR provides us with the tightest confidence interval and is shown to fit the model best via the simulation study.

4.3.2 Example 2: Rosiglitazone study

We deploy our models to another dataset that is also discussed in CPR. The data used in this study is the myocardial infarction (MI) related death data from the rosiglitazone study. Rosiglitazone is used to treat patients with diabetes and researchers are interested in assessing the effect of this drug on myocardial infarctions (MI) and cardiovascular (CV) mortality. A key contribution towards this study was made by Nissen and Wolski, 2007. The data consisted of the number of deaths and the number of myocardial infarctions in 48 studies in the treatment group (group administered with the rosiglitazone drug) and 48 studies in the control group. They used 42 out of the 48 studies since there were no events in 6 of the studies in both groups. Their analysis concluded that rosiglitazone increases the risk of MI and CV mortality significantly. However, this paper was considered very controversial and many parallel analyses were performed. The Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes (RECORD) by Home *et al.*, 2009) concluded no significant increase of MI risk and a nonsignificant decrease in CV mortality risk due to rosiglitazone which contradicted with the previous results of Nissen and Wolski. Further analysis using larger data sets was done by Graham *et al.*, 2010 and Nissen and Wolski, 2011 which reconfirmed the negative impact of rosiglitazone on MI risk and CV mortality, due to which there is limited access to this drug in some countries. However, the effects of

rosiglitazone still remain controversial. In this study, we conduct the analysis using the 48 studies as discussed in Nissen and Wolski, 2007 and CPR. As done in the previous example, we conduct a likelihood ratio test to test $H_0 : \pi_i = 1 \; \forall i = 1, .., n$ and fail to reject the null hypothesis. Hence, we cannot establish the appropriateness of one model over another for this data. Therefore, we compare the results obtained from all four models. Table 4.4 gives estimates of the mean and standard error (SE) of the relative risk based on different methods.

Table 4.4 Estimates of log relative risk for Rosiglitazone study

		log relative risk (τ)		C.I. of relative risk
Model	Log likelihood	Mean	SE	95% C.I.
POIF	-124.41	0.25	0.17	(0.92, 1.79)
POIR	-129.35	0.25	0.16	(0.94, 1.75)
ZIPF	-129.44	0.25	0.28	(0.74, 2.22)
ZIPR	-129.38	0.24	0.15	(0.95, 1.70)

In this study, only about 10% of the zeros were from a zero state. The results obtained from all four models are similar. We conclude that the risk of MI is approximately 30% higher in the treatment group. By looking at the confidence intervals, all four models indicate marginal significance. Although the log likelihood values do not favor the zero inflated models over the Poisson models, the ZIPR model provides the tightest confidence interval. We will also show by simulation that the proposed models using zero inflated Poisson distribution are a better fit for the data.

Simulation study

We conduct a similar simulation study as described in section (4.3.1) for the rosiglitazone data and obtain the results in Table 4.5.

Table 4.5 Mahalanobis distances (Rosiglitazone study)

Model	D^2	P-value
POIF	1.640	0.80
POIR	1.111	0.89
ZIPF	1.592	0.81
ZIPR	0.669	0.95

From the results, we observe that all the models fit quite satisfactorily to the data but close comparison shows that the values of Mahalanobis distance (D^2) for the model ZIPR is the least among all four methods. Among POIF and ZIPF where τ is assumed to be non-random, ZIPF performs better than POIF in terms of goodness of fit.

Discussion of results

From model ZIPR, we obtain the 95% C.I. for relative risk as (0.95, 1.70) and a p-value of 0.10 for testing $\tau = 0$ indicating marginal significance. We hence confirm the results obtained by Nissen and Wolski, 2007, 2011 and Cai, Parast & Ryan, 2010; that rosiglitazone causes a marginally significant increase in the risk of myocardial infarctions.

4.4 Simulation II

In this section we numerically analyze the performance of our models under different setups. We also compare their performance to the Mantel-Haenszel (MH) method and the Peto method. Since MH and Peto methods cannot compute the estimate of relative risk when the studies are non-randomized, therefore no comparisons could be made for data similar to example 1 (EUS-FNA study). Hence, we used the study sizes from example 2 (rosiglitazone

study) to simulate our data. All simulations were done in R, and the estimates for MH and Peto method were computed using the package 'meta' (Guido Schwarzer, 2014).

Simulation setup: In this study, the number of studies in treatment and control arm is chosen as 48. The study sizes are chosen to be same as those in rosiglitazone study. We generate the number of adverse events using a binomial distribution with different probabilities of adverse events to simulate extremely rare event data to moderately rare event data. The probabilities were chosen as 0.001 (such that about 75% of the observations are zeros) and 0.005 (such that about 45% of the observations are zeros). For simulation purposes, the value of relative risk is taken as one i.e. there is no difference between the two groups. For each setup, 500 data sets are generated. We compute the estimates of relative risks followed by the mean squared error (MSE). We also compute the size of the test for each method to test the following hypothesis $H_0 : \tau = 0$, where τ is the log relative risk at 0.05 significance level. Further comparisons include the power of the tests where the true value of the log-relative risk τ is assumed to be 0.5. The results from the simulation study are summarized in Table 4.6.

Table 4.6 Comparison of methods on the basis of MSE, Size and Power of test

Model	MSE	Size	Power
Extremely rare events			
(75% zeros)			
MH	0.173	0.032	0.98
Peto	0.183	0.036	0.99
POIF	0.183	0.130	0.88
ZIPF	0.107	0.050	0.89
Moderately rare events			
(45% zeros)			
MH	0.031	0.040	0.96
Peto	0.032	0.056	0.98
POIF	0.030	0.054	0.94
ZIPF	0.031	0.054	0.94

We also computed the estimates for the random-effects models (POIR and ZIPR) and obtained similar results as in the fixed effects models (POIF and ZIPF respectively). Hence, to avoid repetition and redundancy, here we only include the estimates from fixed effect models in the Table. As shown in Table 4.6, for extremely rare event data, the zero-inflated Poisson model provides the least value of mean squared error while preserving the size of the test. For moderately rare event data the Poisson and zero-inflated Poisson model perform equally well since the Poisson distribution is able to accommodate for almost all the zeros in the data. In both cases, even though the MH and the Peto methods have reasonably low mean squared errors, but they are unable to preserve the size of the test. MH and Peto

methods provide better power but we believe this is because they are unable to preserve the size.

4.5 Discussion

In this chapter, we proposed the use of zero inflated Poisson model with random effects for a clinical trial when the event rate is rare. Earlier work has been done under a randomized control setup where the incidence rate is low. The Poisson random effects models explained in CPR do not account for the unequal number of studies in the two treatment groups. We extend the models in CPR to a non randomized trial with rare events to draw inferences on two treatments. In our methods, we are able to write the explicit likelihood for each study in each arm separately which allows us to use our methods for a non treatment-control type study. The modeling function assumes that the baseline event rate, log relative risk and the zero inflation rates are independent of each other. Although, we haven't seen any concrete examples suggesting otherwise, this may not always be true.

In this chapter, the motivational example of the EUS-FNA study has only 6 studies in treatment group 1 as opposed to 45 in treatment group 2. To address some of the concerns regarding reliability of results for such a setup, we conducted a simulation study for a variety of cases with varying number of studies in each group where we compute the estimates and mean squared error (MSE) of the log-relative risk. MSE is used as a measure of reliability. We concluded that in all cases, the log-relative risk τ is estimated accurately, however, as the number of available studies increases, MSE reduces significantly approximately at a rate of $\sqrt{n}$.

On the basis of data analysis and simulation results, the zero inflated Poisson models

are evidently superior to the Poisson models when the event rate is extremely rare and the ratio of the number of studies in the two arms is highly uneven (EUS-FNA study). However, the number of parameters to be estimated increases when we use a zero inflated Poisson model. Also, it is worth pointing out that, in general, Poisson model with random effects is computationally difficult to handle. CPR adopted a bayesian approach using MCMC methods to estimate the relative risk. Zero-inflated Poisson model with random effects being a more complicated model requires the approximation and maximization of a non-convex likelihood function over six parameters. The estimates obtained from standard optimization routines like *nlminb* and *optim* in R, are heavily dependent on the choice of initial values. To deal with this issue, we used a Differential Evolution optimization algorithm implemented in the package DEoptim (Mullen et al., 2011) in R which does not require the function to be either continuous or differentiable or convex. It optimizes the function over a given range of parameter values and the computing time is reasonably fast.

4.6 Proofs of Chapter 4

Derivation of model 4.2.1

Assume Y_i follows a Poisson distribution with mean $n_i\lambda_i$ and inflation parameter π_i, i.e.

$$Y_i \sim \left\{ \begin{array}{ccc} P(n_i\lambda_i) & \text{w.p.} & \pi_i \\ 0 & \text{w.p.} & 1-\pi_i \end{array} \right\}.$$

Here $\lambda_i = \xi_i e^{\tau X_i}$ and $\xi_i \sim Gamma(\alpha, \beta)$. Define

$$a_i = \left\{ \begin{array}{ll} 0 & : Y_i = 0 \\ & \\ 1 & : Y_i > 0 \end{array} \right\}.$$

The conditional likelihood of Y given ξ is written as

$$\prod_{i=1}^{n} \quad \left(1 - \pi_i + \pi_i e^{-n_i \xi_i e^{\tau X_i}}\right)^{1-a_i} \left(\pi_i e^{-n_i \xi_i e^{\tau X_i}} \frac{(n_i \xi_i e^{\tau X_i})^{y_i}}{y_i!}\right)^{a_i}. \tag{4.6.7}$$

Since $\xi_i \sim Gamma(\alpha, \beta)$ we can integrate out ξ_i from (4.6.7), to obtain the marginal likelihood.

Consider the density of Y_i for some i,

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(1 - \pi_i + \pi_i e^{-n_i \xi_i e^{\tau X_i}}\right)^{1-a_i} \left(\pi_i e^{-n_i \xi_i e^{\tau X_i}} \frac{(n_i \xi_i e^{\tau X_i})^{y_i}}{y_i!}\right)^{a_i} \xi_i^{\alpha-1} e^{-\xi_i \beta} d\xi_i \tag{4.6.8}$$

Now, if $a_i = 0$, (4.6.8) becomes

$$\begin{aligned} &\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty (1 - \pi_i + \pi_i e^{-n_i \xi_i e^{\tau X_i}}) \xi_i^{\alpha-1} e^{\xi_i \beta} d\xi_i \\ &\qquad = 1 - \pi_i + \pi_i \frac{\beta^\alpha}{\Gamma(\alpha)} \int \xi_i^{\alpha-1} e^{-\xi_i (\beta + n_{i1} e^{\tau X_i})} d\xi_i = 1 - \pi_i + \frac{\pi_i \beta^\alpha}{(\beta + n_i e^{\tau X_i})^\alpha}. \end{aligned} \tag{4.6.9}$$

Similarly, if $a_i = 1$, (4.6.8) becomes

$$\begin{aligned} &\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \pi_i e^{-n_i \xi_i e^{\tau X_i}} \frac{(n_i \xi_i e^{\tau X_i})^{y_i}}{y_i!} \xi_i^{\alpha-1} e^{-\beta \xi_i} d\xi_i \\ &= \frac{\pi_i (n_i e^{\tau X_i})^{y_i}}{y_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \int \xi_i^{y_i + \alpha - 1} e^{-\xi_i (\beta + n_i e^{\tau X_i})} d\xi_i \\ &= \frac{\pi_i (n_i e^{\tau X_i})^{y_i}}{y_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(y_i + \alpha)}{(\beta + n_i e^{\tau X_i})^{y_i + \alpha}}. \end{aligned} \tag{4.6.10}$$

Therefore, using (4.6.9) and (4.6.10) the density of Y_i i.e. (4.6.8) is given by

$$\left(1-\pi_i+\frac{\pi_i\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha}\right)^{1-a_i}\left(\frac{\pi_i\beta^\alpha}{(\beta+n_ie^{\tau X_i})^{y_i+\alpha}}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{(n_ie^{\tau X_i})^{y_i}}{y_i!}\right)^{a_i}. \quad (4.6.11)$$

Further, we assume a Beta distribution on the inflation parameter i.e. $\pi_i \sim Beta(c,d)$. Therefore we need to express the likelihood in terms of c and d. Again, we consider the density of Y_i. If $a_i = 0$, when integrated over π_i, (4.6.11) can be written as

$$\begin{aligned}
& \frac{1}{\beta(c,d)}\int_0^1(1-\pi_i+\frac{\pi_i\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha})\pi_i^{c-1}(1-\pi_i)^{d-1}d\pi_i \\
= & \frac{1}{\beta(c,d)}\int_0^1 1-\pi_i(1-\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha})\pi_i^{c-1}(1-\pi_i)^{d-1}d\pi_i \\
= & \frac{1}{\beta(c,d)}\int_0^1\left((1-\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha})(1-\pi_i)+\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha}\right)\pi_i^{c-1}(1-\pi_i)^{d-1}d\pi_i \\
= & \frac{1}{\beta(c,d)}(1-\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha})\int_0^1\pi_i^{c-1}(1-\pi_i)^{d-1}d\pi_i+\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha} \\
= & \frac{\beta(c,d+1)}{\beta(c,d)}(1-\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha})+\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha}. \qquad (4.6.12)
\end{aligned}$$

If $a_i = 1$, when integrated over π_i, (4.6.11) can be written as

$$\begin{aligned}
& \frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^{y_i+\alpha}}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{(n_ie^{\tau X_i})^{y_i}}{y_i!}\frac{1}{\beta(c,d)}\int_0^1\pi_i\pi_i^{c-1}(1-\pi_i)^{d-1}d\pi_i \\
= & \frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^{y_i+\alpha}}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{(n_ie^{\tau X_i})^{y_i}}{y_i!}\frac{\beta(c+1,d)}{\beta(c,d)}. \qquad (4.6.13)
\end{aligned}$$

Hence using (4.6.12) and (4.6.13), the resulting likelihood becomes

$$\begin{aligned}
L(\alpha,\beta,\tau,c,d) = & \prod_{i=1}^n\left(\frac{\beta(c,d+1)}{\beta(c,d)}\big(1-\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha}\big)+\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^\alpha}\right)^{1-a_i} \\
& \left(\frac{\beta(c+1,d)}{\beta(c,d)}\frac{\beta^\alpha}{(\beta+n_ie^{\tau X_i})^{y_i+\alpha}}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{(n_ie^{\tau X_i})^{y_i}}{y_i!}\right)^{a_i}
\end{aligned}$$

We maximize the above likelihood and obtain the estimate of $\boldsymbol{\theta} = (\alpha, \beta, \tau, c, d)'$.

Derivation of model 4.2.2

Consider the density of Y_i for some $i \in \{\text{treatment arm 1}\}$. Since $\tau_i \sim N(\mu, \sigma^2)$, for $a_i = 0$, (4.6.9) becomes

$$
\begin{aligned}
& \int_{-\infty}^{\infty} \mathbb{1}_{\{X_i=1\}}(1-\pi_i+\frac{\pi_i\beta^\alpha}{(\beta+n_ie^\tau)^\alpha})\frac{1}{\sqrt{(2\pi)}\sigma}e^{-\frac{1}{2}\frac{(\tau-\mu)^2}{\sigma^2}}d\tau \\
= \;& \mathbb{1}_{\{X_i=1\}}\left(1-\pi_i+\frac{\pi_i\beta^\alpha}{\sqrt{(2\pi)}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(\tau-\mu)^2}{\sigma^2}}/(\beta+n_ie^\tau)^\alpha d\tau\right) \\
= \;& \mathbb{1}_{\{X_i=1\}}\left(1-\pi_i+\frac{\pi_i\beta^\alpha}{\sqrt{(\pi)}\sigma}\int_{-\infty}^{\infty} e^{-t^2}/(\beta+n_ie^{\sqrt{2}\sigma t+\mu})^\alpha dt, \text{ where } \frac{\tau-\mu}{\sqrt{2}\sigma}=t\right) \\
= \;& \mathbb{1}_{\{X_i=1\}}\left(1-\pi_i+\frac{\pi_i\beta^\alpha}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} f(t)w(t)dt\right), \text{ where} \\
& f(t)=\frac{1}{(\beta+n_ie^{\sqrt{2}\sigma t+\mu})^\alpha} \qquad \text{and} \qquad w(t)=e^{-t^2}.
\end{aligned}
$$

We can integrate the above expression using Gauss-Hermite Quadrature.

If $a_i = 1$, the density of Y_i for some $i \in \{\text{treatment arm 1}\}$ is given by

$$
\begin{aligned}
& \mathbb{1}_{\{X_i=1\}}\frac{\beta^\alpha\pi_i}{\sqrt{2\pi}\sigma}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{n_i^{y_i}}{y_i!}\int_{-\infty}^{\infty}\frac{e^{\tau y_i}}{(\beta+n_ie^\tau)^{y_i+\alpha}}e^{-\frac{1}{2}\frac{(\tau-\mu)^2}{\sigma^2}}d\tau \\
= \;& \mathbb{1}_{\{X_i=1\}}\frac{\beta^\alpha\pi_i}{\sqrt{2\pi}\sigma}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{n_i^{y_i}}{y_i!}\sigma e^{-\frac{1}{2}(\frac{\mu^2}{\sigma^2}-(\frac{\mu}{\sigma}+y_i\sigma)^2)}\int\frac{e^{-\frac{t^2}{2}}}{(\beta+n_ie^{\sigma t+\mu+y_i\sigma^2})^{y_i+\alpha}}dt, \text{ where} \\
& \frac{\tau}{\sigma}-(\frac{\mu}{\sigma}+y_i\sigma)=t \\
= \;& \mathbb{1}_{\{X_i=1\}}\frac{\beta^\alpha\pi_i}{\sqrt{\pi}}\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\frac{n_i^{y_i}}{y_i!}e^{-\frac{1}{2}(\frac{\mu^2}{\sigma^2}-(\frac{\mu}{\sigma}+y_i\sigma)^2)}\int_{-\infty}^{\infty} f_1(t)w_1(t)dt, \text{ where} \\
& f_1(t)=\frac{1}{(\beta+n_ie^{\sqrt{2}\sigma t+\mu+y_i\sigma^2})^{y_i+\alpha}} \qquad \text{and} \qquad w_1(t)=e^{-t^2}.
\end{aligned}
$$

We can use Gauss-Hermite Quadrature to approximate the density of Y_i. Next, we assume

a beta distribution on π and following similar arguments the resulting likelihood is given by

$$
\begin{aligned}
L(\alpha,\beta,\mu,\sigma,c,d) \;=\; & \prod_{i=1}^{n} \mathbb{1}_{\{X_i=1\}} \left(\frac{\beta(c,d+1)}{\beta(c,d)}(1-\frac{\beta^\alpha h_{1i}}{\sqrt{\pi}}) + \frac{\beta^\alpha h_{1i}}{\sqrt{\pi}} \right)^{1-a_i} \\
& \left(\frac{\beta(c+1,d)}{\beta(c,d)} \frac{\beta^\alpha}{\sqrt{\pi}} \frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha) y_i!} n_i^{y_i} e^{-\frac{1}{2}(\frac{\mu^2}{\sigma^2} - (\frac{\mu}{\sigma}+y_i\sigma)^2)} h_{2i} \right)^{a_i} \times \\
& \prod_i \mathbb{1}_{\{X_i=0\}} \left(\frac{\beta(c,d+1)}{\beta(c,d)}(1-\frac{\beta^\alpha}{(\beta+n_j)^\alpha}) + \frac{\beta^\alpha}{(\beta+n_j)^\alpha} \right)^{1-a_i} \\
& \left(\frac{\beta(c+1,d)}{\beta(c,d)} \frac{\beta^\alpha}{(\beta+n_j)^{y_j+\alpha}} \frac{\Gamma(y_j+\alpha)}{\Gamma(\alpha) y_j!} n_j^{y_j} \right)^{a_i}, \text{where}
\end{aligned}
$$

$$
h_{1i} = \int_{-\infty}^{\infty} e^{-t^2}/(\beta + n_i e^{\sqrt{2}\sigma t+\mu})^\alpha dt \quad \text{and} \quad h_{2i} = \int_{-\infty}^{\infty} e^{-t^2}/(\beta + n_i e^{\sqrt{2}\sigma t+\mu+y_i\sigma^2})^{y_i+\alpha} dt.
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Allcroft DJ, Glasbey CA, A simulation-based method for model evaluation. *Statistical Modelling, An International Journal* 2003; **3**(1): 1-13.

[2] Alnosaier WS. *Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models.* PhD thesis, Oregon State University, USA, 2007.

[3] Barndorff-Nielsen OE and Hall P. On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* 1988; **75**: 374-378.

[4] Bartlett MS. Properties of sufficiency and statistical test. *Proc R Soc London, Ser A* 1937; **160**: 268-282.

[5] Bradburn MJ, Deeks JJ, Berlin JA, Localio AR: Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2007; **26**: 53-77.

[6] Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Statistics in Medicine* 2010; **29**: 2078-2089.

[7] Chen Q, Ibrahim JG, Chen MH, Senchaudhuri P, Theory and Inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis* 2008; **99**(6): 1302-1331.

[8] Cox DR and Hinkley DV. *Theoretical Statistics.* Chapman and Hall, 1990.

[9] Cox DR and Reid N. Parameter orthogonality and approximate conditional inference. *J R Stat Soc B* 1987; **49**: 1-39.

[10] Daniels MJ, Gatsonis C. Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American Statistical Association* 1999; **94**: 29-30.

[11] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**: 177-188.

[12] DiCiccio TJ and Stern SE. Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J R Stat Soc B* 1994; **56**: 397-408.

[13] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. *Chapman and Hall, New York* 1993.

[14] Fitzmaurice G, Laird N, Ware J, Applied Longitudinal Analysis. *John Wiley & Sons: New York* 2004.

[15] Gerke H. EUS-guided FNA: better samples with smaller needles? *Gastrointestinal Endoscopy* 2009; **70**(6): 1098-1100.

[16] Goldstein H. *Design and Analysis of Longitudinal Studies.* London: Academic Press, 1979.

[17] Graham DJ, Ouellet-Hellstrom R, MaCurdy TE et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. *Journal of the American Medical Association* 2010; **304**(4): 411-418.

[18] Gregory KB. *A Comparison of Denominator Degrees of Freedom Approximation Methods in the Unbalanced Two-Way Factorial Mixed Model.* Masters paper, Texas A&M University, USA, 2011.

[19] Guido Schwarzer. meta: Meta-Analysis with R. *R package version 3.7-0.* 2014.

[20] Harville DA, Maximum likelihood approaches to variance component estimation and related problems. *Journal of American Statistical Association* 1977; **72**: 320-340.

[21] Harville DA, Decomposition of prediction error. *Journal of the American Statistical Association* 1985; **80**: 132-138.

[22] Harville DA, Jeske DR, Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* 1992; **87**: 724-731.

[23] Henry K, Erice A et al., A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. AIDS Clinical Trial Group 193A Study Team. *Journal of acquired immune deficiency syndromes and human retrovirology.* 1998; **19**(4): 339-349.

[24] Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons Ltd, England.

[25] Home PD, Pocock SJ, Beck-Nielsen H et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for Type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet* 2009; **373**(9681): 2125-2135.

[26] Ibrahim JG, Incomplete data in generalized linear models. *Journal of the American Statistical Association* 1990; **85**(411): 765-769.

[27] Ibrahim JG, Chen MH, Lipsitz SR, Missing responses in generalized linear mixed models when the missing data mechanism is non ignorable. *Biometrika* 2001; **88**(2): 551-564.

[28] Ibrahim JG, Chen MH, Lipsitz SR, Herring AH, Missing-Data methods for generalized linear models. *Journal of the American Statistical Association* 2005; **100**(469): 332-346.

[29] International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Statistical priniciples for clinical trials: E9

[30] Kackar AN and Harville DA. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J Am Stat Assoc* 1984; **79**: 853-862.

[31] Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**: 983-997.

[32] Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963-974.

[33] Lane PW. Meta-analysis of incidence of rare events. *Statistical Methods in Medical Research* 2013; **22**: 117-132.

[34] Leeper JD, Chang SW, Comparison of general multivariate and random-effects models for growth curve analysis: incomplete data small sample situations. *Journal of Statistical Computation and Simulation* 1992; **44**(1-2): 93-104.

[35] Little RJA and Rubin DB. *Statistical Analysis with Missing Data.* New York: John Viley & Sons, 1987.

[36] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease.*Journal of the National Cancer Institute* 1959; **22**(4): 719 -748.

[37] Mardia KV, Kent JT, Bibby JM. *Multivariate Analysis.* Academic Press, London, 1979.

[38] McKeon JJ, F approximations to the distribution of Hotelling's T_0^2. *Biometrika* 1974; **61**(2): 381-383.

[39] Melo TFN, Ferrari SLP and Cribari-Neto Fransisco. Improved testing inference in mixed linear models. *Comput Stat Data Anal* 2009; **53**(7): 2573-2582.

[40] Mullen K, Ardia D, Gil D, Windover D, Cline J. DEoptim: An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software* 2011; **40**(6): 1-26.

[41] Nissen S, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine* 2007; **356**(24): 2467-2471.

[42] Nissen S, Wolski K. An updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Archives of Internal Medicine* 2010; **170**(14): 1191-1201.

[43] Patterson HD and Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika* 1971; **58**: 545-554.

[44] R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2014.

[45] Satterthwaite FF. Synthesis of variance. *Psychometrika* 1941; **6**: 309-316.

[46] Self SG, Liang KY. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association* 1987; **82**(398): 605-610.

[47] Stubbendick AL, Ibrahim JG, Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* 2003; **59**(4): 1140-1150.

[48] The Cochrane Collaboration. The Cochrane Policy Manual

[49] Thompson R. Maximum likelihood estimation of variance components. *Mathematische Operationsforschung und Statistik, Series Statistics* 1980; **11**: 545-561.

[50] Tian L, Cai T, Pfeffer M, Piankov N, Cremieux P-Y, Wei LJ. Exact and efficient inference procedure for meta analysis its application to the analysis of independent 2x2 tables with all available data but without artificial continuity correction. *Biostatistics* 2009; **10**: 275-281.

[51] Trindade AJ, Berzin TM. Clinical controversies in endoscopic ultrasound. *Gastroenterology Report 1* 2013; 33-41.

[52] Verbeke G and Molenberghs G. *Linear mixed models in practice: A SAS oriented approach.* New York: Springer-Verlag, 1997.

[53] Verbeke G and Molenberghs G. *Linear mixed models for longitudinal data.* New York: Springer, 2000.

[54] Weiss RE, Modeling Longitudinal Data. *Springer: New York* 2005.

[55] Welham SJ and Thompson R. Likelihood ratio tests for fixed model terms using residual maximum likelihood. *J R Stat Soc B* 1997; **59**(3): 701-714.

[56] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases* 1985; **27**(5): 335 -371.

[57] Zucker DM, Lieberman O and Manor O. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *J R Stat Soc B* 2000; **62**(4): 827-838.