A Small-Sample Randomization-Based Approach to Semi-Parametric Estimation and Misspecification in Generalized Linear Mixed Models

Mohammad Zakir Hossain
School of Mathematical Sciences
Queen Mary University of London

A thesis submitted for the Degree of
Doctor of Philosophy

April 2017

Acknowledgements

I would like to thank my supervisor Dr Heiko Grossmann for his generous guidance, advice and continuous encouragement over the period of this work. I am thankful to my co-supervisors and assessors Professor Mark Jerrum, Dr Lawrence Pettit and Dr Barbara Bogacka for their valuable comments during my PhD study.

I am grateful to the School of Mathematical Sciences and Queen Mary University of London for funding my PhD studentship. I also would like to thank Queen Mary University of London for supporting me from the Postgraduate Research Fund to present a paper at The JSM (Joint Statistical Meetings) 2015 held on August 8-13, 2015, Seattle, Washington, USA.

I also gratefully acknowledge the comments by Professor Steve Gilmour, Dr Mizanur Khondoker and Dr Mahbub Latif on certain parts of the thesis. I would like to thank my colleagues and friends especially Nick, Shadeque, Reza, Peter, Lutfor, Shihan and Iftakhar for their constant support and inspiration. I am also grateful to the Department of Statistics at the University of Dhaka, Bangladesh for approving the study leave to pursue my PhD research.

Finally, I am deeply indebted to my parents and family members, especially to my wife Nilima, daughter Areeba and son Abrar whose determined efforts, unconditional sacrifices, love and moral support have helped me to achieve my goal.

Statement of Originality

I, Mohammad Zakir Hossain, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third partys copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:
Date:

Abstract

In a generalized linear mixed model (GLMM), the random effects are typically uncorrelated and assumed to follow a normal distribution. However, findings from recent studies on how the misspecification of the random effects distribution affects the estimated model parameters are inconclusive. In the thesis, we extend the randomization approach for deriving linear models to the GLMM framework. Based on this approach, we develop an algorithm for estimating the model parameters of the randomization-based GLMM (RB-GLMM) for the completely randomized design (CRD) which does not require normally distributed random effects. Instead, the discrete uniform distribution on the symmetric group of permutations is used for the random effects. Our simulation results suggest that the randomization-based algorithm may be an alternative when the assumption of normality is violated.

In the second part of the thesis, we consider an RB-GLMM for the randomized complete block design (RCBD) with random block effects. We investigate the effect of misspecification of the correlation structure and of the random effects distribution via simulation studies. In the simulation, we use the variance-covariance matrices derived from the randomization approach. The misspecified model with uncorrelated random effects is fitted to data generated from the model with correlated random effects. We also fit the model with normally distributed random effects to data simulated from models with different random effects distributions. The simulation results show that misspecification of both the correlation structure and of the random effects distribution has hardly any effect on the estimates of the fixed effects parameters. However, the estimated variance components are frequently severely biased and standard errors of these estimates are substantially higher.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Generalized linear models (GLMs) are an extension of classical linear models and are widely used for analyzing categorical data, such as binary responses or counts. Generalized linear mixed models (GLMMs) generalize GLMs further by introducing random effects. In GLMMs, the random effects are often used to accommodate correlation among the observations. Such data are common in applied fields such as biology, medical or biomedical studies. The random effects are also used in GLMMs to model count data with overdispersion [see Rabe-Hesketh and Skrondal (2012, p.706) and Stroup (2012, p.340)]. Overdispersion usually occurs in Poisson models when the variance of the responses is greater than the mean and is caused by positive correlation between counts or responses [Hilbe (2011, p.141) and Stroup (2012, p. 340)].

The random effects in a GLMM are usually assumed to have a normal distribution. This is mainly done for computational simplicity in the analysis for estimating the model parameters. However, inferences that are based on the normality assumption may be incorrect if the actual distribution of the random effects is not normal.

Recently it has been investigated how misspecification of the random effects distribution affects the estimates of the model parameters. The findings from misspecification studies however seem to be inconclusive. Some studies report

that misspecification has a strong effect (Litière et al., 2007, 2008) while others conclude that it has very little effect (Neuhaus et al., 2013, McCulloch and Neuhaus, 2011). This motivates us to develop an alternative approach where no parametric distribution for the random effects in a GLMM is assumed. Instead, we derive GLMMs, moments of random effects and likelihood functions by using the randomization of the underlying design as a starting point. In particular, we consider the completely randomized design (CRD) and the randomized complete block design (RCBD).

More precisely, we extend the randomization approach of Kempthorne (1955) for deriving linear models to GLMMs. We derive the moments of the random effects, and then construct the likelihood function of the derived GLMM from the randomization. Based on this approach, we develop an algorithm for estimating the model parameters and compare via simulation studies the results obtained from our algorithm with a standard GLMM which assumes normality of the random effects distribution. Our approach is semi-parametric in the sense that for the given realization of the random effects the conditional distribution of the responses given the random effects is parametric and a member of the exponential family. Assuming a parametric distribution for the random effects is however not required.

In the next section, we begin by describing some background concepts, in particular model equations of linear models, GLMs and GLMMs. These are required to extend the randomization ideas for deriving linear models to the GLMM framework.

1.2 GLMs and GLMMs

A classical linear model assumes that the $n \times 1$ vector $\boldsymbol{Y}$ of responses can be expressed as

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

where $\boldsymbol{X}$ is the $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random errors. The components of $\boldsymbol{\epsilon}$ are uncorrelated random variables with mean zero and constant variance σ^2. In addition, for inferential purposes it is usually assumed that the components of

$\boldsymbol{\epsilon}$, or equivalently the components of $\boldsymbol{Y}$, are normally distributed. The corresponding linear models are also referred to as normal linear models.

GLMs (McCullagh and Nelder, 1989, p.26-28) generalize the normal linear model in two ways:

(i) The response variables $Y_1, \ldots, Y_n$ can have any distribution from the exponential family. The probability density function (pdf) or probability mass function (pmf) of the response variable Y_i can then be expressed as:

$$f(y_i) = \exp\left[\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i, \phi)\right],$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are some known functions. Moreover, θ and ϕ denote the canonical and the dispersion parameters, respectively.

(ii) A sufficiently smooth link function $g : \mathbb{R} \to \mathbb{R}$ is used to transform the expected response $E(Y_i)$ for $i = 1, \ldots, n$ such that

$$g(E(Y_i)) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta},$$

where $\boldsymbol{x}_i$ denotes a $p \times 1$ column vector of (coded) covariates, η_i is called the linear predictor and $\boldsymbol{\beta}$ is again a vector of unknown model parameters. In matrix notation, this relationship can be expressed as

$$\boldsymbol{g(E(Y))} = \boldsymbol{\eta} = \boldsymbol{X\beta}, \tag{1.1}$$

where $\boldsymbol{E(Y)} = (E(Y_1), \ldots, E(Y_n))^\top$ is the expectation of the response vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$. Also $\boldsymbol{g(E(Y))}$ and $\boldsymbol{\eta}$ are vectors with components $g(E(Y_i))$ and η_i respectively.

GLMMs are an extension of GLMs where random effects are introduced into the linear predictor. The conditional distribution with conditional pdf $f(y_i|\boldsymbol{u})$, of the response variable Y_i given the $q \times 1$ vector of random effects $\boldsymbol{u}$ is considered and assumed to be a member of the exponential family. Similar to GLMs it is then assumed that (McCulloch et al., 2008, p.189-190)

$$\boldsymbol{g(E(Y|u))} = \boldsymbol{X\beta + Zu}, \tag{1.2}$$

where now $\boldsymbol{E(Y|u)} = (E(Y_1|\boldsymbol{u}), \ldots, E(Y_n|\boldsymbol{u}))^\top$ is the conditional expectation

and $\boldsymbol{g(E(Y|u))}$ has components $g(E(Y_i|\boldsymbol{u}))$. As before $g(\cdot)$ is the link function and $\boldsymbol{Z}$ is an $n \times q$ model matrix for the random effects. It is assumed that the expectation is equal to $\boldsymbol{E}(\boldsymbol{u}) = \boldsymbol{0}$ and that the variance-covariance matrix is equal to $\boldsymbol{V}(\boldsymbol{u}) = \boldsymbol{\Sigma}$. In the simplest case, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_q$, where $\boldsymbol{I}_q$ is the identity matrix of order q. Moreover, it is typically assumed that the vector $\boldsymbol{u}$ has a multivariate normal distribution, that is, $\boldsymbol{u} \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$.

1.3 Review of Literature

There are different aspects of misspecification in GLMMs. For instance, McCulloch et al. (2008) consider (in Chapter 12) departures from assumptions such as misspecification of the link function, omission of important covariates, misspecification of binary responses and misspecification of the random effects distribution. However, we focus in the thesis on two aspects in particular; one is misspecification of the distribution for the random effects only, e.g. normal distribution rather than misspecified link function in GLMMs, and the other is correlated random effects. In this section, we review articles particularly related to the misspecification of the random effects distribution in GLMMs. We also review a number of articles on correlated random effects in GLMMs, as will be the case in our randomization approach.

1.3.1 Misspecification in GLMMs

Recently there has been interest in the extent to which estimated GLMM parameters which are obtained under the normality assumption are, for example, biased or less precise if the distribution of the random effects is not normal. To this end, simulation studies have been performed which have lead to conflicting conclusions. First we summarize the studies which found that misspecification has an effect and then review those which claim the contrary.

Heagerty and Kurland (2001) considered a mixed effects logistic regression model and calculated the asymptotic bias of the maximum likelihood estimates (MLEs) of the parameters for different situations with non-normal random effects distributions. They concluded that the misspecification of the random effects distribution can lead to a considerable amount of bias in the maximum

likelihood estimates of the fixed effects parameters and variance components.

Likewise, Agresti et al. (2004) investigated some examples of misspecified random effects distributions for GLMMs by considering logistic regression models for proportions and log-odds ratios as well as a frailty model for survival analysis. For the logistic regression models the true random effects distributions considered in their simulation studies were normal, uniform or exponential, whereas for the frailty model gamma and uniform distributions were used. In order to investigate the effect of misspecification of the random effects distribution, the data were analyzed under the assumption that the random effects in the logistic regression models had a normal distribution while for the frailty model a gamma distribution for the random effects was assumed. They found that the MLEs of the fixed effects parameters were biased in both scenarios due to the misspecified random effects distributions.

Litière et al. (2007) used a logistic regression random-intercept model in simulation studies to investigate the type I and type II errors associated with the Wald test for the mean structure when the true random effects distributions are not correctly specified. In this framework, data were generated by considering the true random effects as having normal, power function, discrete or mixture distributions. The model was fitted by assuming normal distributions for the random effects. They found that the type I and type II errors were severely affected and, depending on the distribution of the random effects, the power of the tests was either increased or decreased.

Continuing this line of research, Litière et al. (2008) performed further simulation studies by considering normal, uniform, exponential, chi-square, log-normal, power function, discrete, symmetric and asymmetric mixtures of two normals as the true random effects distributions in order to examine the impact of misspecification on the estimation procedures and hypothesis tests in GLMMs. They found that the MLEs of the fixed effects parameters were not consistent when incorrect assumptions about the random effects distributions were made. Also, the estimates of variance components were found to be severely biased. Moreover, the power of tests and type I error rates associated with Wald tests were observed to be seriously affected and these effects could become more serious in the presence of more than one random effect in the

model.

Different conclusions emerge from some other misspecification studies. Neuhaus et al. (2011) considered a logistic regression random-intercept model where normal and power function were used as the true random effects distributions, whereas for the analysis normally distributed random effects were assumed. They found only slightly increased type II error rates of the Wald test in their simulation studies that could be attributed to the misspecification of the random effects distribution. They also used a logistic regression random-intercept and slope model by considering the multivariate normal and the symmetric mixture of two multivariate normals as the true random effects distributions, whereas the model was fitted by assuming the normal distribution for the random effects in the analysis. They found that most of the estimated fixed effects parameters are slightly biased due to misspecification of the random effects distribution. They also observed that the estimates of variance components may be biased.

McCulloch and Neuhaus (2011) used maximum likelihood estimation in GLMMs to investigate the effect of misspecification of the random effects distributions for logistic regression models. They considered normal and a mixture of bivariate-normals as the true random effects distributions for their simulation studies while the analysis was performed by assuming a bivariate normal distribution for the random effects. They found that the estimates of the fixed effects parameters were approximately unbiased. The findings contradict to much of the previous literature (e.g., Heagerty and Kurland (2001), Litière et al. (2008) and Agresti et al. (2004)). McCulloch and Neuhaus (2011) argued that the contradiction is due to the fact that results for the non-clustered data situation are incorrectly interpreted as relevant to the cluster data setting. Also Neuhaus et al. (2011) replicates the simulation study of Litière et al. (2007), but they got different results. They also observed that the estimate of the intercept may be biased when the distribution of the random intercept is not correctly specified.

Further simulation studies for generalized linear mixed models have been reported by Neuhaus et al. (2013). They generated data considering logistic and Poisson regression models with random intercepts and slopes. Some bivariate

distributions, namely, t, exponential, log–normal and Tukey, were considered as the true random effects distributions. The data were analyzed assuming a bivariate normal distribution for the random intercepts and slopes i.e., a misspecified random effects distribution. The authors observed little bias in the estimates of the fixed effects parameters. They also observed that a misspecified random effects distribution can bias estimates of variance components in both logistic and Poisson regression models. The key results of the misspecification studies are summarized in Table 1.1.

As a consequence of the misspecification of the random effects distribution, usually assumed to be a normal in a GLMM (see McCulloch et al. (2008, p.299-301)), a number of authors suggest using the nonparametric maximum likelihood approach for estimating the model parameters. To implement this approach several methods have been proposed (e.g., see Butler et al. (1997), Wang (2007), Laird (1978), Aitkin (1999) and Agresti et al. (2004)) where a normality assumption for the random effects distribution is not required. Butler et al. (1997) suggested nonparametric mixing distribution of the random effects for analyzing the repeated binary measures. Wang (2007) proposed an algorithm for the computation of the nonparametric maximum likelihood estimates considering a mixing distribution function of the random effects. Laird (1978) proposed also a mixing distribution function in the nonparametric maximum likelihood estimation for the incomplete data problems. Aitkin (1999) described an expectation maximization (EM) algorithm for nonparametric maximum likelihood estimation in GLMs with variance component structure. Agresti et al. (2004) applied this approach in their simulation study for estimating the model parameters in a GLMM. Moreover, Piepho and McCulloch (2004) proposed the Johnson family of distributions and Magder and Zeger (1996) considered mixtures of normal densities for the random effects distribution. However, these methods are not widely used in data analysis as the distributions are usually highly parameterized. As a result these methods have not been implemented in mainstream statistical packages.

1.3.2 Correlated Random Effects

In standard GLMMs, random effects are usually assumed to be uncorrelated. However, there is some literature where random effects are correlated. Work related to correlated random effects can be found in Chapter 8 of Lee et al.

Table 1.1: Summary of published misspecification studies

No.	Articles	Models	Random effects distributions		Important findings
			True	Assumed	
1	Heagerty and Kurland (2001)	• Logistic random intercept and slope	• Gamma (λ,1) • Normal	• Normal	• Biased MLEs of the fixed effects. • Bias in variance components estimates.
2	Agresti et al. (2004)	• Logistic: proportions and log-odds	• Normal, Uniform Exponential, Two-point mixture	• Normal • Nonparametric	• Biased MLEs of the fixed effects in both models. • Biased variance components estimates.
		• Frailty: Ratios	• Gamma, Uniform • Two-point mixture	• Gamma • Nonparametric	
3	Litière et al. (2007)	• Logistic random intercept	• Normal • Power function • Discrete • Asymmetric mixture	• Normal	• Type I, II errors and power of Wald test are severely affected.
4	Litière et al. (2008)	• Logistic random intercept	• Normal, Uniform • Exponential, Chi-square • Log-normal, Discrete • Power function • Symmetric and asymmetric mixtures of two normals	• Normal	• MLEs of fixed effects are inconsistent. • Variance components estimates are severely biased. • Power and type I error of Wald test are seriously affected and more severe for more than one random effect.
		• Logistic random intercept and slope	• Multivariate normal • Symmetric mixture of two multivariate normals.	• Normal	
5	Neuhaus et al. (2011)	• Logistic random intercept	• Normal • Power function	• Normal	• Type II error of Wald test is slightly increased.
		• Logistic random intercept and slope	• Multivariate normal • Symmetric mixtures of two multivariate normals	• Normal	• Most of the fixed effects estimates are slightly affected. • Variance components estimates may be biased.
6	McCulloch and Neuhaus (2011)	• Logistic random intercept	• Tukey	• Normal • Tukey	• MLEs of fixed effects are approximately unbiased and intercept may be biased.
		• Logistic random intercept and slope	• Normal • Mixture of bi-variate normals	• Bi-variate normal	• Small bias in fixed effect estimates.
7	Neuhaus et al. (2013)	• Logistic and Poisson: random intercept and slope	• Bi-variate: • Normal, t • Exponential • Log-normal, Tukey	• Bi-variate normal	• Little bias in fixed effect estimates. • Little effect on confidence interval. • Can yield biased estimates of intercept and variance components.

(2006). They consider hierarchical generalized linear models (HGLMs) with correlated random effects. They also present some examples of twin and family data in Section 8.7, where the random effects are correlated. In the thesis, we consider HGLMs because the existing software packages for HGLMs have better support for correlated random effects. In addition, Stroup (2012) describes the close linkage between design of experiments and statistical modelling in several chapters of his book. In particular, correlated random errors arising from repeated measures are described in Chapters 14 and 15.

Moreover, in the literature there are at least two applications of GLMMs, equivalently hierarchical generalized linear models (HGLMs), with correlated random effects by Lee and Lee (2012). They investigate the performance of the hierarchical likelihood (h-likelihood) method for HGLMs with correlated random effects via a simulation study and apply their method to real data on seed germination and lip cancer. They apply a binomial-beta HGLM to the seed germination data set where the response distribution is binomial and the correlated random effects are assumed to have a Beta distribution. They also apply a Poisson HGLM with correlated random effects to the lip cancer data, considering the response distribution to be Poisson and using a Markov random field model for the random effects.

There is also an application of HGLMs with correlated random effects in the field of genetics by Noh et al. (2006). They consider logistic regression mixed models with correlated random effects assumed to follow the normal distribution. They use eight explanatory variables and four random effects for the analysis of preeclampsia data.

Moreover, some recent studies focus on the linkage between the design of experiments and statistical modelling, in particular in GLMs and GLMMs. For example, Woods and Van de Ven (2011) considered GLMs (non-normal responses) and blocked designs for experiments, while Waite and Woods (2015) looked at GLMMs for blocked experiments with random block effects. Similar work related to GLMMs can also be found in Stroup (2012).

1.4 Randomization Approach

The randomization approach for deriving linear models from the design was introduced by Kempthorne (1955). This approach is also summarized in Chapters 6 and 7 of Hinkelmann and Kempthorne (2008) for the CRD and RCBD respectively. Kempthorne (1955) used the assumption of unit-treatment additivity where the responses depend on contributions from units and treatments to derive linear models. We adapt this idea to the linear predictor of a GLM. The randomization of units is then modelled in order to derive GLMMs. The randomization ideas for both the CRD and RCBD are described in the following sections.

1.4.1 Randomization in the CRD

A CRD is a design in which the selected treatments are assigned or allocated to the experimental units completely at random. This is the simplest design where we are interested in comparing the treatments in situations with homogeneous experimental units. Randomization is the process of allocating the treatments to the experimental units completely by a chance mechanism. Kempthorne (1955) described this randomization process by using random numbers and expressed it mathematically with design random variables (Hinkelmann and Kempthorne, 2008, p.154-155). More precisely, for t treatments, each of which is replicated r times, and $n = tr$ experimental units, a random sequence of the numbers $1, \ldots, n$ is generated. Then, the t treatments are allocated to the randomized units such that the first set of r units receive the first treatment and the second set of r units receive the second treatment, and so on.

An equivalent way to achieve this randomization is by randomly permuting the units and then writing down the assignment of treatments to the randomized units in a systematic way. For example, suppose there are six units u_1, u_2, u_3, u_4, u_5, u_6, and three equally replicated treatments so that $t = 3$ and $r = 2$. After permuting the units, suppose the order of the units is u_3, u_6, u_5, u_1, u_2, u_4. We then relabel the units as $\tilde{u}_1 = u_3$, $\tilde{u}_2 = u_6$, $\tilde{u}_3 = u_5$, $\tilde{u}_4 = u_1$, $\tilde{u}_5 = u_2$ and $\tilde{u}_6 = u_4$. In general, $\tilde{u}_k = u_{\pi(k)}$, where π is a randomly chosen permutation of the set $\{1, 2, 3, 4, 5, 6\}$. That is $\pi \in S_6$, where S_6 is the symmetric group of order 6. In our example, the permutation π maps the set $\{1, 2, 3, 4, 5, 6\}$ onto

itself as follows

$$1 \mapsto 3, \;\; 2 \mapsto 6, \;\; 3 \mapsto 5, \;\; 4 \mapsto 1, \;\; 5 \mapsto 2, \;\; 6 \mapsto 4$$

and the randomization of units is done by using a randomly selected element π of the symmetric group S_6. A description of the symmetric group S_n of n units is given in the appendix (see Section A.1.1). Treatments are then assigned to the randomized units $\tilde{u}_k$, for $k = 1, \ldots, 6$, systematically such that $\tilde{u}_1$ and $\tilde{u}_2$ receive treatment 1; $\tilde{u}_3$ and $\tilde{u}_4$ receive treatment 2; and so on.

Formally, for general values of r and t, and hence $n = tr$, the randomization of units in a CRD is modelled by randomly choosing a permutation π from the symmetric group S_n. As will be explained in Section 2.1, by assuming unit-treatment additivity on the scale of the linear predictor of a given GLM, it can then be shown that the randomization gives rise to a random effect in the linear predictor. The model thus becomes a GLMM. Conditionally, the linear predictor of this model can be represented exactly as in (1.2). This is the generalization of the randomization-based linear model of Hinkelmann and Kempthorne (2008, p.159) to a randomization-based GLMM (RB-GLMM). The moments of the random effects are also derived from the randomization and it is found that the random effects are correlated. More precisely, there is an exchangeable (i.e., compound symmetric) correlation structure among the random effects.

GLMMs for the CRD, in which the variance-covariance matrix of the random effects has the same form as in the RB-GLMM, but where the random effects are assumed to have a normal distribution, have been considered in some applications. Lee et al. (2006, p.256) use this model to analyze family data where the correlation comes from genetic and common family-environment effects. This model is also considered in Rabe-Hesketh and Skrondal (2012, p.706) and can be applied for modelling count data with overdispersion. A model with correlated random effects is applied in genetic epidemiology and animal breeding (Lee et al., 2006, p.251). For example, Wong (personal communication) obtained a real data set from the field of animal breeding and was interested in applying a GLMM with the same correlation structure for the random effects as in the randomization-based model (https://stat.ethz.ch/pipermail/r-sig-mixed-models/2010q4/004629.html).

We derive the likelihood function of the RB-GLMM by using general probability and measure theoretic concepts. The relevant background material on measure theory is described in Section A.2 of the appendix. In the derivation of the likelihood function we do not assume any parametric distribution for the random effects. Instead, we use the discrete uniform distribution on the symmetric group S_n. The derived likelihood function can be expressed in terms of the permanent (Section A.4) of a suitably chosen matrix, as described in Section 2.5. One needs to maximize the likelihood function or equivalently the permanent of the matrix to obtain the maximum likelihood estimates of the model parameters. However, it is well known that the direct computation of the permanent is a very difficult problem (see Section 2.6).

The direct maximization of the likelihood function is also complicated as the summation in the likelihood equation does not commute with taking natural logarithms. In order to handle this problem, an alternative approach is considered where one needs to maximize a minorization function rather than the log-likelihood (Lange, 2013, p.186-187). The advantage of this approach is that the derivative of the minorization function can be found more easily. We derive a minorization function for our likelihood function in Section 3.2. Based on this minorization function, we develop an estimation algorithm where we combine the iterative weighted least squares (IWLS) algorithm for standard GLMs, with best linear predictors (BLP) of the random effects.

In Section 3.7, we conduct simulation studies in order to compare the estimated model parameters obtained from the randomization-based algorithm with standard GLMM estimates where normality is assumed for the random effects distribution. The results show that for misspecified random effects distributions, the randomization-based algorithm gives more precise estimates of the model parameters in most cases than the standard GLMM with normally distributed random effects. However, because of its computational requirements, currently our algorithm can only be used for small samples. This is due to the fact that the algorithm considers all $n!$ permutations of the symmetric group S_n. This number increases rapidly with the increase of the sample size.

However, it is for small samples, that violations of the normality assumption for

the random effects may be expected to have the most serious effects and where the randomization-based approach may provide a useful alternative. This is a hypothesis at this stage rather than the results produced in the thesis. It may be possible to verify this claim in the future, after better computation has been developed to enable model fitting for larger sample sizes. Finally, we apply our model to real count data which exhibit overdispersion in Section 3.8.

1.4.2 Randomization in the RCBD

An RCBD is a design in which the whole set of experimental units is arranged in several blocks which are internally homogeneous. The treatments are allocated randomly to the experimental units within each block such that each treatment occurs exactly once in each block. For the RCBD, there are two types of randomization for deriving linear models; one for random block effects associated with Brien and Bailey (2006) and one for fixed block effects which is in line with the approach of Kempthorne (1955).

Extending the approach of Brien and Bailey (2006), we derive a RB-GLMM for the RCBD with random block effects (see Chapter 4). The corresponding randomization of b blocks and t units within each block is modelled by elements of the *wreath product* $S_t \wr S_b$ of two symmetric groups S_t and S_b (Section A.1.3). In this setup, the symmetric group S_t represents the randomization of units and S_b stands for the randomization of blocks. The variance-covariance matrices of the random block effects and errors are derived by using the randomization approach. We find that the random block effects and errors are correlated; the details are given in Chapter 4. We also derive the likelihood function for this model.

In addition to the models presented in the thesis, we have also considered Kempthorne's (1955) approach where the same randomization process is applied as for the CRD to each block. Generalizing the approach of Kempthorne (1955), we have derived a RB-GLMM for the RCBD with fixed block effects (see Appendix C). The corresponding randomization of units is modelled by using the *direct product* S_t^b of b instances of the symmetric group S_t, where S_t represents the randomization of t units within each fixed block (Section A.1.2). We also derived the likelihood function and moments of the random effects for

this RB-GLMM. However, in order to keep the thesis to a reasonable length, the details of these derivations are not presented here. Instead, we present a summary of the main results in Appendix C (see Section C.1). Likewise, the details of the derivation of the likelihood function for the RB-GLMM for the RCBD with random block effects are not presented in the thesis. The form of the derived likelihood function is however given in Section C.2 of the appendix.

We use simulation studies to investigate the impact of misspecification of the correlation structure and of the distribution of the random effects for data which are simulated from the GLMM for the RCBD with random block effects in Chapter 4. More precisely, we fit several GLMMs with standard assumptions to these data. This work is in line with previous studies where the misspecification of the random effects distribution has been investigated. Previous studies mostly considered uncorrelated random effects. However, in our case, as a consequence of the randomization, the random block effects and errors are correlated. The results show that there is hardly any effect on the estimates of fixed treatment effects parameters in terms of their biases and standard errors (SEs). However, this is not the case for the estimates of variance components and it is found that these are frequently severely biased. We also find that the SEs of the variance components estimates are substantially higher for the misspecified models than the corresponding estimates for the true models in most cases where misspecification of the random effects distribution was present. Similar results were also found by Neuhaus et al. (2011, 2013). An outline of the structure of the thesis is given in the next section.

1.5 Structure of the Thesis

The thesis has been organized into two parts. In Part I, we derive the randomization-based GLMM for the CRD and describe an algorithm for estimating the model parameters. In Part II, we investigate the impact of misspecification of the correlation structure and of the distribution of the random effects in a GLMM using variance-covariance matrices derived from the randomization approach for the RCBD with random block effects.

Part I consists of Chapters 2 and 3 while Part II consists of Chapters 4 and 5. The second chapter focuses on the derivation of the RB-GLMM for the CRD

and the likelihood function. The randomization-based estimation algorithm with a simulation study and an application to a real data set are described in Chapter 3. The derivation of the RB-GLMM for the RCBD with random block effects and of the moments of the random effects are described in Chapter 4. Chapter 5 presents a simulation study which, in the context of the RB-GLMM in Chapter 4, investigates misspecification of the random effects distribution and of the correlation structure. Finally, conclusion and future work are contained in Chapter 6. Background material and additional results are given in the appendix. Moreover, the *R*-code used in the thesis is given on a DVD (digital video disc) submitted with the thesis.

Part I

Semi-Parametric Estimation in a GLMM Based on Randomization

Chapter 2

Derivation of the Model and Likelihood Function

In this chapter we begin by deriving the RB-GLMM and the likelihood function for the CRD where the responses depend on contributions from units and treatments. In Section 2.1, the randomization approach for deriving linear models is adapted to the GLM framework. In order to do this, we consider a response variable whose distribution is a member of the exponential family. We assume that the linear predictor is the sum of a constant contribution from the unit on which the response is measured and of a constant contribution from the treatment which is applied to the unit. In other words, we assume that the principle of unit-treatment additivity holds on the scale of the linear predictor of a GLM.

The randomization of units is then modelled by using permutations from the symmetric group. A consequence of the randomization is that the contribution of the experimental units to the linear predictor becomes a random variable. The resulting model for the response is then a GLMM whose conditional expectation can be expressed as in equation (1.2).

Deriving the GLMM based on this randomization idea requires working out the conditional expectation in the framework of general probability theory, which in turn builds on measure theory. We derive some useful results using measure theoretic concepts which we use to derive the model and the likelihood function. In Section 2.1.5, we derive moments of the random effects based on the

randomization idea and show that the random effects are correlated due to the randomization. This is similar to the work of Hinkelmann and Kempthorne (2008, p.160) in the context of linear models.

In Section 2.2, the likelihood function for the RB-GLMM is derived by using general probability and measure theoretic concepts. In deriving the likelihood function, no parametric distribution is assumed for the random effects. More precisely, the distribution of the random effects is derived from the uniform distribution on the symmetric group of permutations.

The results for the GLMM which are derived from the randomization for the CRD are summarized in matrix notation and presented in Section 2.3. Moreover, we factorize the singular variance-covariance matrix of the random effects in order to facilitate the use of available software packages for fitting GLMMs. We describe the factorization of this matrix in Section 2.4. We show that for the CRD the likelihood function for the RB-GLMM is a multiple of the permanent (Section A.4) of a suitably chosen matrix (Section 2.5). In order to obtain the maximum likelihood estimates of the model parameters one needs to maximize the likelihood function or equivalently the permanent of the matrix. However, the direct computation of the permanent is well known to be a very hard problem. The complexity of calculating the permanent and its consequences for estimating the model parameters of the RB-GLMM are briefly described in Section 2.6.

2.1 Derivation of the Model

In this section, our main goal is to derive the RB-GLMM for the CRD. In order to do so, we first adapt Kempthorne's (1955) randomization approach for deriving linear models to GLMs. We then express the randomization-based model as a standard GLMM. As the standard form of GLMMs involves the conditional expectation of the responses given the vector of random effects, we derive the conditional expectation in accordance with its general measure-theoretic definition. In measure theory, the conditional expectation is defined with respect to a σ-field. The details of this derivation are given in Section 2.1.3. Then in Section 2.1.4, applying a link function to the conditional expectation we obtain the standard form (1.2) for the RB-GLMM for the CRD.

2.1.1 Adapting Kempthorne's Approach to GLMs

In order to derive the RB-GLMM for the CRD, we adapt the approach of Kempthorne (1955) for deriving linear models from the randomization to the link-transformed mean response in the framework of GLMs. Let $\{1, \ldots, n\}$ be a set of units and $J = \{1, \ldots, t\}$ be a set of t distinct treatments. Let $X_{i,j}$ be the response for unit i and treatment j whose distribution is assumed to be from the exponential family. The $X_{i,j}$ are potential outcomes which are not all simultaneously observable (Rubin, 2005). Formally $X_{i,j}$ is a real random variable on a probability space $(\Omega, \mathscr{F}, P)$ (see appendix Section A.2.1), where Ω is a non-empty set, $\mathscr{F}$ is a σ-field on Ω and P is a probability measure on $\mathscr{F}$. We assume that for every design $d = (j_1, \ldots, j_n) \in J^n$ the random variables $X_{1,j_1}, \ldots, X_{n,j_n}$ are independent.

Let $g : \mathbb{R} \to \mathbb{R}$ be a known link function such that the assumption of unit-treatment additivity (Kempthorne, 1955) holds for the transformed mean of $X_{i,j}$. Then we can write

$$g(E(X_{i,j})) = u_i + v_j \tag{2.1}$$

for every $i \in \{1, \ldots, n\}$ and every $j \in J$, where u_i and v_j denote constants which represent contributions from the i-th unit and j-th treatment, respectively. Here we assume that all u_i are different. Further letting $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$, $\bar{v} = \frac{1}{t}\sum_{j=1}^{t} v_j$, $\mu = \bar{u} + \bar{v}$, $\alpha_j = v_j - \bar{v}$ and $e_i = u_i - \bar{u}$ we can rewrite equation (2.1) as

$$g(E(X_{i,j})) = \mu + \alpha_j + e_i \tag{2.2}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, t$, where μ is the grand mean, α_j is the treatment effect and the unit errors e_i are all different. The form of equation (2.2) is similar to the specification of the linear predictor (1.1) in a GLM as there is no random effect in this setup. More precisely, GLMs model the expectation by applying a suitable link function g to the mean responses. It is now our next goal to derive a GLMM in standard form, similar to the specification in (1.2), by using the randomization to introduce random effects into (2.2).

2.1.2 Modelling the Randomization

We continue to consider the random variables $X_{i,j}$ on $(\Omega, \mathscr{F}, P)$. In order to represent the randomization of units by using permutations, let S_n be the symmetric group of permutations of the set $\{1, \ldots, n\}$ and U_n be the uniform distribution on S_n. The probability of picking one element π from the symmetric group S_n is $U_n(\{\pi\}) = 1/n!$ for every $\pi \in S_n$. For $i = 1, \ldots, n$ define $\tilde{\epsilon}_i : S_n \to \mathbb{R}$ by

$$\tilde{\epsilon}_i(\pi) = \sum_{\sigma \in S_n} e_{\sigma(i)} \, 1_{\{\sigma\}}(\pi) = e_{\pi(i)} \quad \text{for every} \quad \pi \in S_n, \tag{2.3}$$

where the function $1_{\{\sigma\}}(\pi)$ is an indicator function which is equal to one for $\sigma = \pi$ and zero otherwise. Here $\tilde{\epsilon}_i$ is a random variable on the probability space $(S_n, \mathscr{P}(S_n), U_n)$, where $\mathscr{P}(S_n)$ is the power set of S_n. For later use let $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n)^\top$ be the vector of random variables $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n$. Further define $\epsilon_i : S_n \times \Omega \to \mathbb{R}$ by

$$\epsilon_i(\pi, \omega) = \tilde{\epsilon}_i(\pi) \quad \text{for every} \quad (\pi, \omega) \in S_n \times \Omega \tag{2.4}$$

on the probability space $(S_n \times \Omega, \mathscr{P}(S_n) \otimes \mathscr{F}, U_n \otimes P)$, where $S_n \times \Omega$ is the Cartesian product, $\mathscr{P}(S_n) \otimes \mathscr{F}$ is the product σ-field and $U_n \otimes P$ is the product measure.

Let $d = (j_1, \ldots, j_n) \in J^n$ be a fixed design which uses treatment j_i for the i-th run of the experiment. For every $i = 1, \ldots, n$ we define a random variable $Y_{i,j_i} : S_n \times \Omega \to \mathbb{R}$ by

$$Y_{i,j_i}(\pi, \omega) = X_{\pi(i),j_i}(\omega) \quad \text{for every} \quad (\pi, \omega) \in S_n \times \Omega \tag{2.5}$$

on the probability space $(S_n \times \Omega, \mathscr{P}(S_n) \otimes \mathscr{F}, U_n \otimes P)$ which represents the response for the i-th randomized unit and treatment j_i. Table 2.1 illustrates the relationship between response variables defined in (2.5) and the original variable for the situation with six units and three equally replicated treatments that was considered in Section 1.4.1.

Table 2.1: Responses under randomization of units for a fixed design in a CRD.

Fixed systematic design		Non-randomized	Random	Randomized
observation	treatment	responses	permutation	responses
i	j_i	X_{i,j_i}	$\pi(i)$	$Y_{i,j_i} = X_{\pi(i),j_i}$
1	1	$X_{1,1}$	3	$Y_{1,1} = X_{3,1}$
2	1	$X_{2,1}$	6	$Y_{2,1} = X_{6,1}$
3	2	$X_{3,2}$	5	$Y_{3,2} = X_{5,2}$
4	2	$X_{4,2}$	1	$Y_{4,2} = X_{1,2}$
5	3	$X_{5,3}$	2	$Y_{5,3} = X_{2,3}$
6	3	$X_{6,3}$	4	$Y_{6,3} = X_{4,3}$

2.1.3 Conditional Expectation

We want to derive a GLMM from (2.2) and the randomization, in which the vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ plays the role of the vector $\boldsymbol{u}$ of random effects in equation (1.2). In order to obtain an equation similar to (1.2) we need to find $g(E(Y_{i,j_i}|\boldsymbol{\epsilon}))$ where the conditional expectation $E(Y_{i,j_i}|\boldsymbol{\epsilon})$ of Y_{i,j_i} given $\boldsymbol{\epsilon}$ is defined as $E(Y_{i,j_i}|\mathscr{F}_0)$ and $\mathscr{F}_0 = \mathscr{F}(\boldsymbol{\epsilon})$ is the σ-field generated by the vector $\boldsymbol{\epsilon}$ (Billingsley, 1985, p.466). For every component ϵ_i of $\boldsymbol{\epsilon}$ there exists the σ-field $\mathscr{F}(\epsilon_i)$ generated by ϵ_i which consists of the sets $\epsilon_i^{-1}(B) = \{(\pi, \omega) \in S_n \times \Omega : \epsilon_i(\pi, \omega) \in B\}$, where B is a Borel set. So $\mathscr{F}(\epsilon_i) = \{\epsilon_i^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}$, where $\mathscr{B}(\mathbb{R})$ is the Borel σ-field (Section A.2.1). In addition, there is the σ-field generated by the vector $\boldsymbol{\epsilon}$ of random variables $\epsilon_1, \ldots, \epsilon_n$ for which we write $\mathscr{F}(\boldsymbol{\epsilon})$. Equivalently, $\mathscr{F}(\boldsymbol{\epsilon})$ is the smallest σ-field which contains the union $\bigcup_{i=1}^{n} \mathscr{F}(\epsilon_i)$ (Section A.2.4). Now in order to find $E(Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon}))$ we want to apply Theorem 2 from the appendix (see Section A.2.6). This theorem gives a simple way to calculate the conditional expectation, when the σ-field with respect to which we want to calculate the conditional expectation is generated by a partition of the space on which the random variable is defined.

Lemma 1 and Corollary 1 below enable us to prove that $\mathscr{F}(\boldsymbol{\epsilon})$ is the σ-field generated by the partition $\{\{\pi\} \times \Omega : \pi \in S_n\}$ of $S_n \times \Omega$. This result enables us to apply Theorem 2 in order to find the conditional expectation $E(Y_{i,j_i}|\boldsymbol{\epsilon})$.

Lemma 1. *Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ be the random vector with components ϵ_i defined by (2.4). The σ-field $\tilde{\mathscr{F}} = \{A \times \Omega : A \subseteq S_n\}$ generated by the set*

$\{\{\pi\} \times \Omega : \pi \in S_n\}$ *is a subset of* $\mathscr{F}(\boldsymbol{\epsilon})$ *i.e.,* $\{A \times \Omega : A \subseteq S_n\} \subseteq \mathscr{F}(\boldsymbol{\epsilon})$.

Proof. That $\tilde{\mathscr{F}}$ is the σ-field generated by $\{\{\pi\} \times \Omega : \pi \in S_n\}$ follows from part (*iii*) of Lemma 19 (see appendix Section A.3) with $S = S_n$ and $\mathscr{F} = \mathscr{P}(S_n)$. Let $\pi \in S_n$ be fixed. Then we have $\epsilon_i(\pi, \omega) = \tilde{\epsilon}_i(\pi) = e_{\pi(i)}$ using (2.4) and (2.3). It follows that $\epsilon_i^{-1}(\{e_{\pi(i)}\}) = \left\{(\sigma, \omega) \in S_n \times \Omega : \epsilon_i(\sigma, \omega) = e_{\pi(i)}\right\} \in \mathscr{F}(\boldsymbol{\epsilon})$. It also follows by using the intersection property of the σ-field in equation (A.1) of the appendix that $\bigcap_{i=1}^{n} \epsilon_i^{-1}\left(\{e_{\pi(i)}\}\right) \in \mathscr{F}(\boldsymbol{\epsilon})$. Now

$$
\begin{aligned}
& \bigcap_{i=1}^{n} \epsilon_i^{-1}\left(\{e_{\pi(i)}\}\right) = \bigcap_{i=1}^{n} \left\{(\sigma, \omega) \in S_n \times \Omega : \epsilon_i(\sigma, \omega) = e_{\pi(i)}\right\} \\
= & \bigcap_{i=1}^{n} \left\{(\sigma, \omega) \in S_n \times \Omega : \tilde{\epsilon}_i(\sigma) = e_{\pi(i)}\right\} = \bigcap_{i=1}^{n} \left(\left\{\sigma \in S_n : \tilde{\epsilon}_i(\sigma) = e_{\pi(i)}\right\} \times \Omega\right) \\
= & \left(\bigcap_{i=1}^{n} \left\{\sigma \in S_n : \tilde{\epsilon}_i(\sigma) = e_{\pi(i)}\right\}\right) \times \Omega = \left(\bigcap_{i=1}^{n} \left\{\sigma \in S_n : e_{\sigma(i)} = e_{\pi(i)}\right\}\right) \times \Omega \\
= & \{\pi\} \times \Omega
\end{aligned}
$$

since all e_i are assumed to be different. So for every $\pi \in S_n$ the set $\{\pi\} \times \Omega \in \mathscr{F}(\boldsymbol{\epsilon})$. Therefore $\{\{\pi\} \times \Omega : \pi \in S_n\} \subseteq \mathscr{F}(\boldsymbol{\epsilon})$. Thus the σ-field generated by the set $\{\{\pi\} \times \Omega : \pi \in S_n\}$ is a subset of $\mathscr{F}(\boldsymbol{\epsilon})$. Hence $\tilde{\mathscr{F}} \subseteq \mathscr{F}(\boldsymbol{\epsilon})$. ■

By applying part (*iv*) of Lemma 19 to $\boldsymbol{V} = \boldsymbol{\epsilon}$ it follows that

$$\mathscr{F}(\boldsymbol{\epsilon}) \subseteq \{A \times \Omega : A \subseteq S_n\}. \tag{2.6}$$

Thus from (2.6) and Lemma 1 we obtain the following result.

Corollary 1. *The* σ*-field generated by the vector* $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ *with components* ϵ_i *defined by (2.4) is* $\mathscr{F}(\boldsymbol{\epsilon}) = \{A \times \Omega : A \subseteq S_n\}$.

The elements of $\{\{\pi\} \times \Omega : \pi \in S_n\}$ are mutually disjoint. Furthermore, we have $S_n \times \Omega = \bigcup_{\pi \in S_n} (\{\pi\} \times \Omega)$ and so $\{\{\pi\} \times \Omega : \pi \in S_n\}$ is a partition of $S_n \times \Omega$. Also $\mathscr{F}(\boldsymbol{\epsilon})$ is the σ-field generated by $\{Z_l : l = 1, \ldots, n!\} = \{\{\pi_l\} \times \Omega : l = 1, \ldots, n!\}$. Now to find $E(Y_{i,j_i}|\boldsymbol{\epsilon}) = E(Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon}))$ we can apply Theorem 2 (Section A.2.6) to $S_n \times \Omega$ and $Z_l = \{\pi_l\} \times \Omega$ for $l = 1, \ldots, n!$, where $\pi_1, \ldots, \pi_{n!}$ are the distinct elements of S_n.

Lemma 2. *The conditional expectation of* Y_{i,j_i} *given* $\boldsymbol{\epsilon}$ *is*

$$E(Y_{i,j_i}|\boldsymbol{\epsilon}) = \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega} \int_\Omega X_{\sigma(i),j_i}(\tilde{\omega})\, P(d\tilde{\omega}). \tag{2.7}$$

Proof. By using Theorem 2 from Section A.2.6 we can write

$$\begin{aligned}
E(Y_{i,j_i}|\boldsymbol{\epsilon}) &= E(Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon})) \\
&= \sum_{l=1}^{n!} \frac{1}{(U_n\otimes P)(\{\pi_l\}\times\Omega)} \int_{\{\pi_l\}\times\Omega} Y_{i,j_i}(\pi,\tilde{\omega})(U_n\otimes P)d(\pi,\tilde{\omega}) 1_{\{\pi_l\}\times\Omega} \\
&= \sum_{\sigma\in S_n} \frac{1}{(U_n\otimes P)(\{\sigma\}\times\Omega)} \int_{\{\sigma\}\times\Omega} Y_{i,j_i}(\pi,\tilde{\omega})(U_n\otimes P)\, d(\pi,\tilde{\omega})\, 1_{\{\sigma\}\times\Omega} \\
&= \sum_{\sigma\in S_n} \frac{1}{U_n(\{\sigma\})\, P(\Omega)} \int_{\{\sigma\}} \int_\Omega Y_{i,j_i}(\pi,\tilde{\omega})\, P(d\tilde{\omega})\, U_n(d\pi)\, 1_{\{\sigma\}\times\Omega} \\
&= \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega} \int_\Omega Y_{i,j_i}(\sigma,\tilde{\omega}) P(d\tilde{\omega}) = \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega} \int_\Omega X_{\sigma(i),j_i}(\tilde{\omega}) P(d\tilde{\omega}).
\end{aligned}$$

■

Note that $E(Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon}))$ is a random variable. For every $(\pi,\omega)\in S_n\times\Omega$ the realization of the conditional expectation is equal to

$$\begin{aligned}
E(Y_{i,j_i}|\boldsymbol{\epsilon})(\pi,\omega) &= E(Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon}))(\pi,\omega) = \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega}(\pi,\omega) \int_\Omega X_{\sigma(i),j_i}(\tilde{\omega}) P(d\tilde{\omega}) \\
&= \sum_{\sigma\in S_n} 1_{\{\sigma\}}(\pi) \int_\Omega X_{\sigma(i),j_i}(\tilde{\omega}) P(d\tilde{\omega}) = \sum_{\sigma\in S_n} 1_{\{\sigma\}}(\pi) E(X_{\sigma(i),j_i}) \\
&= E(X_{\pi(i),j_i})
\end{aligned} \tag{2.8}$$

which depends only on π but not ω. Therefore, the conditional expectation $E(Y_{i,j_i}|\boldsymbol{\epsilon})$ can be written as $E(Y_{i,j_i}|\boldsymbol{\epsilon}) = \left(\sum_{\sigma\in S_n} 1_{\{\sigma\}} \int_\Omega X_{\sigma(i),j_i}(\tilde{\omega}) P(d\tilde{\omega})\right) \circ \tilde{p}_1$, where $\tilde{p}_1 : S_n\times\Omega \to S_n$ with $\tilde{p}_1(\pi,\omega) = \pi$ is a projection or coordinate mapping.

Now consider the functions $f = E(Y_{i,j_i}|\boldsymbol{\epsilon})$ and $g = \boldsymbol{\epsilon}$ which are both defined on $S_n\times\Omega$. By the definition of the conditional expectation f is $\mathscr{F}(\boldsymbol{\epsilon})$-$\mathscr{B}(\mathbb{R})$-measurable. The well known factorization Lemma 18 (Section A.2.7) then implies the existence of a $\mathscr{B}(\mathbb{R}^n)$-$\mathscr{B}(\mathbb{R})$-measurable function h such that $f =$

$h \circ g$. By letting $E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet) = h$ we can write

$$E(Y_{i,j_i}|\boldsymbol{\epsilon}) = E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet) \circ \boldsymbol{\epsilon}, \tag{2.9}$$

where $E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet) : \mathbb{R}^n \to \mathbb{R}$ is defined on the image set of $\boldsymbol{\epsilon}$ i.e., $\text{Im}(\boldsymbol{\epsilon}) = \{(e_{\pi(1)}, \ldots, e_{\pi(n)}) : \pi \in S_n\}$ with $E(Y_{i,j_i}|\boldsymbol{\epsilon} = (e_{\pi(1)}, \ldots, e_{\pi(n)})) = E(X_{\pi(i),j_i})$. In the following section, we introduce a link function to the conditional expectation $E(Y_{i,j_i}|\boldsymbol{\epsilon})$ in order to obtain a standard form of GLMMs.

2.1.4 Incorporating the Link Function

Let g be the link function for which (2.2) holds. It follows from (2.9) that

$$\begin{aligned} g\left(E(Y_{i,j_i}|\boldsymbol{\epsilon})\right) &= g \circ E(Y_{i,j_i}|\boldsymbol{\epsilon}) = g \circ (E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet) \circ \boldsymbol{\epsilon}) \\ &= (g \circ E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet)) \circ \boldsymbol{\epsilon} \\ &= g\left(E(Y_{i,j_i}|\boldsymbol{\epsilon} = \bullet)\right) \circ \boldsymbol{\epsilon}. \end{aligned} \tag{2.10}$$

This can be verified by applying the functions to arbitrary elements $(\pi, \omega) \in S_n \times \Omega$. Now from equation (2.10) using (2.8) and (2.2) we have

$$\begin{aligned} g\left(E(Y_{i,j_i}|\boldsymbol{\epsilon})\right)(\pi, \omega) &= g\left(E(X_{\pi(i),j_i})\right) = \mu + \alpha_{j_i} + e_{\pi(i)} \\ &= \mu + \alpha_{j_i} + \epsilon_i(\pi, \omega), \end{aligned} \tag{2.11}$$

where $\epsilon_i(\pi, \omega) = \tilde{\epsilon}_i(\pi) = \sum_{\sigma \in S_n} e_{\sigma(i)}\ 1_{\{\sigma\}}(\pi) = e_{\pi(i)}$. Hence from equation (2.11) and our previous derivations, we obtain

$$\text{Model 1} : g(E(Y_{i,j_i}|\boldsymbol{\epsilon})) = \mu + \alpha_{j_i} + \epsilon_i \tag{2.12}$$

for every $i = 1 \ldots, n$ and $j = 1, \ldots, t$. One can see that the only random variable on the right hand side of (2.12) is ϵ_i and the treatment j_i is fixed by the design d. Here the random variable ϵ_i is treated as the random effect in the GLMM framework for the CRD. The form in (2.12) is our generalization of the derived linear model of Hinkelmann and Kempthorne (2008, p.159) to a GLMM.

2.1.5 Moments of Random Effects

In matrix notation, equation (2.12) can be written in the standard GLMM form (1.2) with $\boldsymbol{Z}$ being the identity matrix and $\boldsymbol{u} = \boldsymbol{\epsilon}$. Below we use the notation u_i and e_i as introduced in Section 2.1.1. Usually in a GLMM each of the random effects $\boldsymbol{u}$ is assumed to be normally and independently distributed with mean zero and constant variance. In our project, we do not assume any parametric distribution for the random effects. By using the randomization distribution of $\boldsymbol{\epsilon}$ we derive the mean, variance, covariance and correlation of the random effects below. To this end we introduce some notation. Set $Q_{i,i^*} = \{\pi \in S_n : \pi(i) = i^*\}$ for every i and i^*. Then $S_n = \bigcup_{i^*=1}^{n} Q_{i,i^*}$. Clearly, the number of elements of Q_{i,i^*} is $|Q_{i,i^*}| = (n-1)!$. Moreover, for $k \neq l$ we have $S_n = \bigcup_{k^* \neq l^*}^{n} (Q_{k,k^*} \bigcap Q_{l,l^*})$ and the number of elements of $Q_{k,k^*} \bigcap Q_{l,l^*}$ for $k \neq l$ and $k^* \neq l^*$ is $|Q_{k,k^*} \bigcap Q_{l,l^*}| = (n-2)!$.

2.1.5.1 Expectation and Variance

In this section we derive the mean and the variance of the random effects from the randomization. By using the measure theoretic definition of the expectation (Section A.2.5) we can compute the following results.

Lemma 3. *The expectation and variance of an element ϵ_i of the vector of random errors $\boldsymbol{\epsilon}$ are*

(i) $E(\epsilon_i) = 0$;

(ii) $Var(\epsilon_i) = \sigma^2$, *where* $\sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2$.

Proof. *(i)* For every unit i the expectation of ϵ_i is given by

$$\begin{aligned} E(\epsilon_i) &= \int_{S_n}\int_{\Omega} \epsilon_i(\pi,\omega)P(d\omega)U_n(d\pi) = \frac{1}{n!}\sum_{\pi\in S_n}\int_{\Omega}\epsilon_i(\pi,\omega)P(d\omega) \\ &= \frac{1}{n!}\sum_{\pi\in S_n}\int_{\Omega}\tilde{\epsilon}_i(\pi)P(d\omega) = \frac{1}{n!}\sum_{\pi\in S_n}\tilde{\epsilon}_i(\pi) = \frac{1}{n!}\sum_{\pi\in S_n} e_{\pi(i)} \\ &= \frac{1}{n!}\sum_{i^*=1}^{n}\sum_{\pi\in Q_{i,i^*}} e_{i^*} = \frac{(n-1)!}{n!}\sum_{i^*=1}^{n} e_{i^*} = \frac{1}{n}\sum_{i^*=1}^{n}(u_{i^*} - \bar{u}) = 0. \end{aligned}$$

(*ii*) As we obtain $E(\epsilon_i) = 0$ in part (*i*), the variance of ϵ_i is equal to

$$\begin{aligned} Var(\epsilon_i) &= E(\epsilon_i^2) = \int_{S_n} \int_{\Omega} \epsilon_i^2(\pi, \omega) P(d\omega) U_n(d\pi) \\ &= \frac{1}{n!} \sum_{\pi \in S_n} \int_{\Omega} \epsilon_i^2(\pi, \omega) P(d\omega) = \frac{1}{n!} \sum_{\pi \in S_n} \tilde{\epsilon}_i^2(\pi) = \frac{1}{n!} \sum_{\pi \in S_n} e_{\pi(i)}^2 \\ &= \frac{1}{n!} \sum_{i^*=1}^{n} \sum_{\pi \in Q_{i,i^*}} e_{i^*}^2 = \frac{(n-1)!}{n!} \sum_{i^*=1}^{n} e_{i^*}^2 = \frac{1}{n} \sum_{i^*=1}^{n} (u_{i^*} - \bar{u})^2 = \sigma^2. \end{aligned}$$

■

From the above expressions one can see that neither the expectation nor the variance of ϵ_i depends on i which means that the components of the vector of random effects $\boldsymbol{\epsilon}$ have mean zero and common variance σ^2. We use these results in predicting the random effects and for estimating the variance component in Chapter 3. Next we compute the covariance and correlation between components of the vector of random effects.

2.1.5.2 Covariance and Correlation

In Lemma 4, we use the results for the expectation and variance from Lemma 3 to calculate the covariance and correlation for the random effects.

Lemma 4. *The covariance and correlation between two components ϵ_k and ϵ_l of the vector of random errors $\boldsymbol{\epsilon}$ with $k \neq l$ are*

(*i*) $Cov(\epsilon_k, \epsilon_l) = -\dfrac{1}{n-1}\sigma^2$;

(*ii*) $Corr(\epsilon_k, \epsilon_l) = -1/(n-1)$,

where σ^2 is the common variance of the random errors defined in Lemma 3.

Proof. (i) For $k \neq l$ the covariance of ϵ_k and ϵ_l can be computed as

$$\begin{aligned} Cov(\epsilon_k, \epsilon_l) &= E(\epsilon_k \epsilon_l) = \frac{1}{n!} \sum_{\pi \in S_n} e_{\pi(k)} e_{\pi(l)} = \frac{1}{n!} \sum_{k^* \neq l^*}^{n} \sum_{\pi \in Q_{k,k^*} \bigcap Q_{l,l^*}} e_{k^*}\ e_{l^*} \\ &= \frac{(n-2)!}{n!} \sum_{k^* \neq l^*}^{n} e_{k^*}\ e_{l^*} = \frac{1}{n(n-1)} \sum_{k^*=1}^{n} e_{k^*} \sum_{\substack{l^*=1 \\ l^* \neq k^*}}^{n} e_{l^*} \\ &= \frac{1}{n(n-1)} \sum_{k^*=1}^{n} e_{k^*} \left(\sum_{l^*=1}^{n} e_{l^*} - e_{k^*} \right) = -\frac{1}{n(n-1)} \sum_{k^*=1}^{n} e_{k^*}^2 \\ &= -\frac{1}{n(n-1)} \sum_{k^*=1}^{n} (u_{k^*} - \bar{u})^2 = -\frac{1}{n-1} \sigma^2. \end{aligned}$$

(ii) By using Lemma 3 (ii) and applying part (i) we calculate the correlation between ϵ_k and ϵ_l for $k \neq l$ and as

$$Corr(\epsilon_k, \epsilon_l) = \frac{Cov(\epsilon_k, \epsilon_l)}{\sqrt{Var(\epsilon_k) Var(\epsilon_l)}} = -\frac{1}{n-1}.$$

■

The results in Lemma 4 show that the random errors are correlated due to the randomization. A similar result was found by Hinkelmann and Kempthorne (2008, p.160) for the derived linear model associated with the CRD.

2.2 Derivation of the Likelihood Function

For standard GLMMs, the method of maximum likelihood estimation is used for estimating the model parameters, in particular, the fixed effect parameters. The likelihood function is obtained by integrating out the random effects assuming that the vector of random effects has a multivariate normal distribution. The likelihood function is therefore often complicated as it generally involves a high-dimensional integral with respect to the random effects distribution (McCulloch et al., 2008, p.193).

In the RB-GLMM, we use the same general idea. However, since the distribution of the random effects is derived from the uniform distribution on the symmetric group, the integral is a general measure theoretic analogue which

takes the form of a sum over all the elements of the symmetric group. The construction of the likelihood function is based on the randomization of units for the CRD using the symmetric group S_n. Then by using the concept of conditional probability, joint probability and the law of total probability we derive the likelihood function for the RB-GLMM (2.12). The likelihood function is derived by using some auxiliary results which are stated and proved below.

2.2.1 Conditional Independence and Joint Probability

In order to derive the joint distribution of $Y_{1,j_1}, \ldots, Y_{n,j_n}$ we show that these are conditionally independent (McCulloch et al., 2008, p.189) given the vector of random errors in Lemma 5.

Lemma 5. *The random variables $Y_{1,j_1}, \ldots, Y_{n,j_n}$ are conditionally independent for every realization of the vector $\boldsymbol{\epsilon}$ of random errors.*

Proof. Let $A_1, \ldots, A_n$ be elements of the Borel σ-field $\mathscr{B}(\mathbb{R})$. For every $\pi \in S_n$ we can write the conditional probability of $Y_{1,j_1}, \ldots, Y_{n,j_n}$ given the vector of random errors $\boldsymbol{\epsilon} = \left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)$ as

$$\begin{aligned} & (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \boldsymbol{\epsilon} = \left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\right) \\ = \; & (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \boldsymbol{\epsilon}^{-1}\left(\{\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\}\right)\right) \\ = \; & (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \tilde{\boldsymbol{\epsilon}}^{-1}\left(\{\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\}\right) \times \Omega\right). \end{aligned} \tag{2.13}$$

As we assume that the values $e_1, \ldots, e_n$ are all different, we have

$$\tilde{\boldsymbol{\epsilon}}^{-1}\left(\{\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\}\right) = \{\pi\}. \tag{2.14}$$

This is true because if $\sigma \in \tilde{\boldsymbol{\epsilon}}^{-1}\left(\{(e_{\pi(1)}, \ldots, e_{\pi(n)})\}\right) = \{\tau \in S_n : \tilde{\boldsymbol{\epsilon}}(\tau) \in \{(e_{\pi(1)}, \ldots, e_{\pi(n)})\}\}$ then $\tilde{\boldsymbol{\epsilon}}(\sigma) = (e_{\pi(1)}, \ldots, e_{\pi(n)})$. Hence $e_{\sigma(i)} = e_{\pi(i)}$ for every $i = 1, \ldots, n$. It follows that the two elements σ and π in S_n are equal, i.e., $\sigma = \pi$ as $e_1, \ldots, e_n$ are all different. Then using (2.14), the conditional probability

expression (2.13) can be rewritten as

$$\begin{aligned}
& (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \boldsymbol{\epsilon} = \left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\right) \\
= \; & (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \{\pi\} \times \Omega\right) \\
= \; & \frac{(U_n \otimes P)\left(\{Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n\} \cap (\{\pi\} \times \Omega)\right)}{(U_n \otimes P)\left(\{\pi\} \times \Omega\right)}. \qquad (2.15)
\end{aligned}$$

Now in order to simplify the above expression the intersection in the numerator of (2.15) can be written as

$$\begin{aligned}
& \{Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n\} \cap (\{\pi\} \times \Omega) \\
= \; & \{(\sigma, \omega) \in S_n \times \Omega : Y_{1,j_1}(\sigma, \omega) \in A_1, \ldots, Y_{n,j_n}(\sigma, \omega) \in A_n\} \cap (\{\pi\} \times \Omega) \\
= \; & \{(\pi, \omega) : \omega \in \Omega \text{ and } Y_{1,j_1}(\pi, \omega) \in A_1, \ldots, Y_{n,j_n}(\pi, \omega) \in A_n\} \\
= \; & \{\pi\} \times \{\omega \in \Omega : X_{\pi(1),j_1}(\omega) \in A_1, \ldots, X_{\pi(n),j_n}(\omega) \in A_n\} \\
= \; & \{\pi\} \times \{X_{\pi(1),j_1} \in A_1, \ldots, X_{\pi(n),j_n} \in A_n\}. \qquad (2.16)
\end{aligned}$$

Therefore, using (2.16) in (2.15) and that X_{i,j_i} and $X_{i',j_{i'}}$ are independent for $i \neq i'$ (see Section 2.1.1), we obtain the conditional probability of $Y_{1,j_1}, \ldots, Y_{n,j_n}$ given $\boldsymbol{\epsilon} = \left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)$ as

$$\begin{aligned}
& (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \boldsymbol{\epsilon} = \left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)\right) \\
= \; & \frac{(U_n \otimes P)\left(\{\pi\} \times \{X_{\pi(1),j_1} \in A_1, \ldots, X_{\pi(n),j_n} \in A_n\}\right)}{(U_n \otimes P)\left(\{\pi\} \times \Omega\right)} \\
= \; & \frac{U_n(\{\pi\}) P\left(\{X_{\pi(1),j_1} \in A_1, \ldots, X_{\pi(n),j_n} \in A_n\}\right)}{U_n(\{\pi\}) P(\Omega)} \\
= \; & P\left(X_{\pi(1),j_1} \in A_1\right) \ldots P\left(X_{\pi(n),j_n} \in A_n\right) = \prod_{i=1}^{n} P\left(X_{\pi(i),j_i} \in A_i\right) \qquad (2.17) \\
= \; & \prod_{i=1}^{n} P\left(\{\omega \in \Omega : X_{\pi(i),j_i}(\omega) \in A_i\}\right) = \prod_{i=1}^{n} P\left(\{\omega \in \Omega : Y_{i,j_i}(\pi, \omega) \in A_i\}\right) \\
= \; & \prod_{i=1}^{n} \frac{U_n(\{\pi\}) P\left(\{\omega \in \Omega : Y_{i,j_i}(\pi, \omega) \in A_i\}\right)}{U_n(\{\pi\}) P(\Omega)} \\
= \; & \prod_{i=1}^{n} \frac{(U_n \otimes P)\left(\{(\sigma, \omega) \in S_n \times \Omega : Y_{i,j_i}(\sigma, \omega) \in A_i\} \cap (\{\pi\} \times \Omega)\right)}{(U_n \otimes P)\left(\{\pi\} \times \Omega\right)} \\
= \; & \prod_{i=1}^{n} \frac{(U_n \otimes P)(\{Y_{i,j_i} \in A_i\} \cap (\{\pi\} \times \Omega))}{(U_n \otimes P)\left(\{\pi\} \times \Omega\right)} = \prod_{i=1}^{n} (U_n \otimes P)(Y_{i,j_i} \in A_i | \{\pi\} \times \Omega) \\
= \; & \prod_{i=1}^{n} (U_n \otimes P)\left(Y_{i,j_i} \in A_i | \boldsymbol{\epsilon} = (e_{\pi(1)}, \ldots, e_{\pi(n)})\right)
\end{aligned}$$

which shows that the random variables $Y_{1,j_1}, \ldots, Y_{n,j_n}$ are conditionally independent for the given realization of $\boldsymbol{\epsilon}$. ■

Next, in addition to the results in Lemma 5, we derive the joint probability by using the law of total probability. We consider the joint probability $(U_n \otimes P)(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n)$ where the response variables $Y_{1,j_1}, \ldots, Y_{n,j_n}$ take values in Borel sets $A_1, \ldots, A_n$. These probabilities uniquely determine the joint distribution of the variables $Y_{1,j_1}, \ldots, Y_{n,j_n}$ (Billingsley, 1985, p.265) and we present the derived result in Lemma 6.

Lemma 6. *Let $A_1, \ldots, A_n$ be elements of $\mathscr{B}(\mathbb{R})$. Then*

$$(U_n \otimes P)(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n) = \int_{S_n} \prod_{i=1}^{n} P\left(X_{\pi(i),j_i} \in A_i\right) U_n(d\pi). \tag{2.18}$$

Proof. Let $\mathrm{Im}(\boldsymbol{\epsilon}) = \{\boldsymbol{\epsilon}(\pi, \omega) : (\pi, \omega) \in S_n \times \Omega\}$ be the image of $\boldsymbol{\epsilon}$. By using the law of total probability it follows that

$$\begin{aligned}
&(U_n \otimes P)(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n) \\
=& \sum_{\boldsymbol{c} \in \mathrm{Im}(\boldsymbol{\epsilon})} (U_n \otimes P)(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \boldsymbol{\epsilon} = \boldsymbol{c})(U_n \otimes P)(\boldsymbol{\epsilon} = \boldsymbol{c}) \\
=& \sum_{\pi \in S_n} (U_n \otimes P)(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \{\pi\} \times \Omega)(U_n \otimes P)(\{\pi\} \times \Omega)
\end{aligned} \tag{2.19}$$

which is true since for every $\boldsymbol{c} \in \mathrm{Im}(\boldsymbol{\epsilon})$ we can write

$$\begin{aligned}
\{\boldsymbol{\epsilon} = \boldsymbol{c}\} &= \{(\pi, \omega) \in S_n \times \Omega : \boldsymbol{\epsilon}(\pi, \omega) = \boldsymbol{c}\} = \{(\pi, \omega) \in S_n \times \Omega : \tilde{\boldsymbol{\epsilon}}(\pi) = \boldsymbol{c}\} \\
&= \{\pi \in S_n : \tilde{\boldsymbol{\epsilon}}(\pi) = \boldsymbol{c}\} \times \Omega = \{\pi \in S_n : (e_{\pi(1)}, \ldots, e_{\pi(n)}) = \boldsymbol{c}\} \times \Omega \\
&= \{\pi\} \times \Omega,
\end{aligned}$$

where we use again that $e_1, \ldots, e_n$ are all distinct. It follows from equation

(2.19) and by using the intermediate result (2.17) that

$$\begin{aligned}
& (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n\right) \\
= & \sum_{\pi \in S_n} (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \{\pi\} \times \Omega\right) U_n(\{\pi\})\ P(\Omega) \\
= & \frac{1}{n!} \sum_{\pi \in S_n} (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \{\pi\} \times \Omega\right) \\
= & \int_{S_n} (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n | \{\pi\} \times \Omega\right) U_n(d\pi) \\
= & \int_{S_n} \prod_{i=1}^{n} P\left(X_{\pi(i),j_i} \in A_i\right) U_n\ (d\pi).
\end{aligned}$$

■

Now using the derived result for the joint probability in Lemma 6 we derive the joint density of the response random variables in Theorem 1. We then obtain the expression for the likelihood function from the joint density in the following section.

2.2.2 Joint Density and Likelihood Function

For $i = 1, \ldots, n$ let the random variable X_{i,j_i} have a probability density function $f_{i,j_i}(x_i; \boldsymbol{\theta})$ which depends on the parameter vector $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ depends on the fixed effects parameter vector $\boldsymbol{\beta}$ and the unknown vector $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$. We then use Lemma 6 to derive the joint probability density function of $Y_{1,j_1}, \ldots, Y_{n,j_n}$.

Theorem 1. *Assume that the random variable X_{i,j_i} has a density $f_{i,j_i}(x_i, \boldsymbol{\theta})$ for every $i = 1, \ldots, n$. Then the joint probability density function of Y_{1,j_1}, …, Y_{n,j_n} is*

$$f_{Y_{1,j_1},\ldots,Y_{n,j_n}}(y_1, \ldots, y_n; \boldsymbol{\theta}) = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i; \boldsymbol{\theta}). \tag{2.20}$$

Proof. Let $A_1, \ldots, A_n$ be a collection of sets in the Borel σ-field $\mathscr{B}(\mathbb{R})$. It follows that

$$P\left(X_{\pi(i),j_i} \in A_i\right) = \int_{A_i} f_{\pi(i),j_i}(x_i; \boldsymbol{\theta})\ dx_i. \tag{2.21}$$

Substituting (2.21) in the joint probability equation (2.18) (see Lemma 6), we

can write

$$\begin{aligned}
& (U_n \otimes P)\left(Y_{1,j_1} \in A_1, \ldots, Y_{n,j_n} \in A_n\right) \\
= & \int_{S_n} \prod_{i=1}^{n} \int_{A_i} f_{\pi(i),j_i}(x_i;\boldsymbol{\theta}) dx_i U_n(d\pi) \\
= & \int_{S_n} \int_{A_1} \ldots \int_{A_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(x_i;\boldsymbol{\theta}) dx_n \ldots dx_1 U_n(d\pi) \\
= & \frac{1}{n!} \sum_{\pi \in S_n} \int_{A_1} \ldots \int_{A_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(x_i;\boldsymbol{\theta}) dx_n \ldots dx_1 \\
= & \int_{A_1} \ldots \int_{A_n} \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(x_i;\boldsymbol{\theta}) dx_n \ldots dx_1. \qquad (2.22)
\end{aligned}$$

Equation (2.22) shows that the mapping $f_{Y_{1,j_1},\ldots,Y_{n,j_n}}$ defined by

$$(y_1, \ldots, y_n) \mapsto \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i;\boldsymbol{\theta}) \qquad (2.23)$$

is the joint probability density function of the distribution of the random vector $(Y_{1,j_1}, \ldots, Y_{n,j_n})$ on the probability space $(S_n \times \Omega, \mathscr{P}(S_n) \otimes \mathscr{F}, U_n \otimes P)$. ■

Finally we obtain the likelihood function $L(\boldsymbol{\theta})$ from the randomization using the joint density (2.20).

Corollary 2. *Under the assumption of Theorem 1 the likelihood function for the RB-GLMM for the CRD is equal to*

$$L(\boldsymbol{\theta}) = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i;\boldsymbol{\theta}), \qquad (2.24)$$

where $f_{\pi(i),j_i}(y_i;\boldsymbol{\theta})$ is the conditional density function of the random variable Y_{i,j_i} given π and $\boldsymbol{\theta}$ is the vector of parameters of interest.

$L(\boldsymbol{\theta})$ in (2.24) depends on the unknown vector $\boldsymbol{e}$ in addition to $\boldsymbol{\beta}$. The elements of $\boldsymbol{e}$ are estimated by using the best linear predictor in Section 3.5. The likelihood function for a standard GLMM with normally distributed random effects (McCulloch et al., 2008, p.193) is given by

$$L(\boldsymbol{\theta}) = \int \prod_{i=1}^{n} f_{Y_i|\boldsymbol{u}}(y_i;\boldsymbol{\theta}|\boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u}, \qquad (2.25)$$

where $f_{\boldsymbol{U}}(\boldsymbol{u})$ is the density function of a multivariate normal distribution with

mean zero and variance-covariance matrix $\boldsymbol{V}(\boldsymbol{u}) = \sigma^2\boldsymbol{I}$, and $f_{Y_i|\boldsymbol{u}}(y_i;\boldsymbol{\theta}|\boldsymbol{u})$ is the conditional density of the response Y_i given $\boldsymbol{u}$. However, in our case the likelihood function derived from the randomization (2.24) may be regarded as a discrete version of (2.25) in which the multivariate normal distribution is replaced by the uniform distribution on the symmetric group S_n.

2.3 Summary of RB-GLMM for the CRD

In this section, we summarize the derived results of the RB-GLMM for the CRD. We refer to the model equation (2.12) as the RB-GLMM for the CRD where the conditional distribution of Y_{i,j_i} given $\boldsymbol{\epsilon}$ is from the exponential family. Here the random variables $X_{i,j_i}, \ldots, X_{n,j_n}$ are assumed to be independent for a fixed design $d = (j_1, \ldots, j_n)$ and whose distribution is a member of the exponential family as before (see Section 2.1.1). The relationship between Y_{i,j_i} and X_{i,j_i} is described for a fixed specific design d in Table 2.1. The likelihood function for this model is given in (2.24) where the density $f_{\pi(i),j_i}(y_i;\boldsymbol{\theta})$ is a member of the exponential family. More precisely, the random variables $Y_{1,j_1}, \ldots, Y_{n,j_n}$ are conditionally independent for the given realization of the vector $\boldsymbol{\epsilon}$ of random errors (see Lemma 5).

To summarize the derived model in matrix notation, let $\boldsymbol{Y}$ be the $n \times 1$ vector of responses, $\boldsymbol{X}$ be the $n \times (t+1)$ design matrix for fixed effects. Then the derived GLMM (Model (2.12)) from the randomization is

$$\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.26}$$

where $\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})$ is the conditional expectation, $\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right)$ is a vector with components $g\left(E(Y_{i,j_i}|\boldsymbol{\epsilon})\right)$, $g(\cdot)$ is the link function as before, $\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_t)^\top$ is a $(t+1) \times 1$ vector of treatment effect parameters fixed by the specific design d and $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of random errors. Note that the design matrix $\boldsymbol{X}$ in (2.26) is overparameterized. In Section 3.7.2, we use the overparameterized design matrix to simulate the data. However, for estimating the model parameters we use the effects-coded parameters and the corresponding design matrix. For example, $\tilde{\boldsymbol{\beta}} = (\mu, \alpha_1, \ldots, \alpha_{t-1})^\top$ is the effects-coded fixed effects parameter vector where $\alpha_t = -\alpha_1 - \ldots - \alpha_{t-1}$ for an equi-replicated design.

Moreover, from Lemma 3 and 4 we have $\boldsymbol{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and the variance-covariance

matrix of $\boldsymbol{\epsilon}$ is $\boldsymbol{V}(\boldsymbol{\epsilon}) = \dfrac{\sigma^2}{n-1}\left(n\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^\top\right) = \sigma^2\boldsymbol{\mathcal{P}}$, where $\boldsymbol{\mathcal{P}}$ is the correlation matrix and equal to

$$\boldsymbol{\mathcal{P}} = \frac{1}{n-1}\left(n\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^\top\right) = \begin{pmatrix} 1 & -\frac{1}{n-1} & \dots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 & \dots & -\frac{1}{n-1} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n-1} & -\frac{1}{n-1} & \dots & 1 \end{pmatrix}. \quad (2.27)$$

The symmetric $n \times n$ matrix $\boldsymbol{\mathcal{P}}$ is singular as can be verified by noting that its columns sum up to the zero vector.

2.4 Factorization of the Singular Correlation Matrix

The matrix $\boldsymbol{\mathcal{P}}$ in (2.27) does not have an inverse. This causes problems in applying standard estimation and inferential procedures for the model parameters. For example, software for estimating the parameters of a GLMM can usually not handle multivariate normal random effects $\boldsymbol{u}$ with variance-covariance matrix equal to $\boldsymbol{V}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{\mathcal{P}}$. In addition, the density of the multivariate normal distribution with the singular variance-covariance matrix does not exist. To overcome this problem we factorize the matrix $\boldsymbol{\mathcal{P}}$ as $\boldsymbol{\mathcal{P}} = \boldsymbol{L}\boldsymbol{L}^\top$ and replace $\boldsymbol{\epsilon}$ with $\boldsymbol{L}\tilde{\boldsymbol{\epsilon}}^*$, where $\boldsymbol{L}$ is an $n \times (n-1)$ matrix and $\tilde{\boldsymbol{\epsilon}}^*$ is an $(n-1) \times 1$ random vector with $\boldsymbol{E}(\tilde{\boldsymbol{\epsilon}}^*) = \mathbf{0}$ and $\boldsymbol{V}(\tilde{\boldsymbol{\epsilon}}^*) = \sigma^2\boldsymbol{I}_{n-1}$. In other words, $\tilde{\boldsymbol{\epsilon}}^*$ satisfies the assumptions of the standard GLMM (1.2). The vector $\boldsymbol{L}\tilde{\boldsymbol{\epsilon}}^*$ in the GLMM with $\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\tilde{\boldsymbol{\epsilon}}^*)\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{L}\tilde{\boldsymbol{\epsilon}}^*$ has the same mean and variance-covariance matrix as the random vector $\boldsymbol{\epsilon}$ in the GLMM with $\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ derived from the randomization. This can be seen by noting that $\boldsymbol{V}(\boldsymbol{L}\tilde{\boldsymbol{\epsilon}}^*) = \boldsymbol{L}\boldsymbol{V}(\tilde{\boldsymbol{\epsilon}}^*)\boldsymbol{L}^\top = \sigma^2\boldsymbol{L}\boldsymbol{L}^\top = \sigma^2\boldsymbol{\mathcal{P}} = \boldsymbol{V}(\boldsymbol{\epsilon})$. Below we present a lemma and its constructive proof of the factorization $\boldsymbol{\mathcal{P}} = \boldsymbol{L}\boldsymbol{L}^\top$.

Lemma 7. *Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ be an orthonormal basis of $\mathbb{R}^n$, where $\boldsymbol{u}_1 = \dfrac{1}{\sqrt{n}}\mathbf{1}_n$. Then the $n \times n$ singular matrix $\boldsymbol{\mathcal{P}}$ can be factorized as $\boldsymbol{\mathcal{P}} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L} = \sqrt{\lambda_2}\tilde{\boldsymbol{U}}$, $\tilde{\boldsymbol{U}} = (\boldsymbol{u}_2, \dots, \boldsymbol{u}_n)$ and $\lambda_2 = \dfrac{n}{n-1}$ is the non-zero positive eigenvalue of $\boldsymbol{\mathcal{P}}$ with multiplicity $n-1$.*

Proof. We know from Rao (1973, p.67) that the matrix $\boldsymbol{\mathcal{P}}$ has only two dis-

tinct eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \dfrac{n}{n-1}$ with multiplicity 1 and $n-1$, respectively. Let $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ be the orthogonal matrix corresponding to the basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. Then by Theorem B.19 of Christensen (1987, p.332) we have

$$\boldsymbol{U}^\top \boldsymbol{\mathcal{P}} \boldsymbol{U} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_2), \tag{2.28}$$

where $n-1$ diagonal elements of the matrix on the right hand side of (2.28) are equal to λ_2. Letting $d_1 = \lambda_1 = 0$ and $d_i = \lambda_2$ for $i = 2, \ldots, n$ it follows that

$$\begin{aligned} \boldsymbol{\mathcal{P}} &= \boldsymbol{U}\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_2)\boldsymbol{U}^\top = \boldsymbol{U}\text{diag}(d_1, d_2, \ldots, d_n)\boldsymbol{U}^\top \\ &= \sum_{i=1}^{n} d_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \sum_{i=2}^{n} d_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \lambda_2 \sum_{i=2}^{n} \boldsymbol{u}_i \boldsymbol{u}_i^\top = \lambda_2 \tilde{\boldsymbol{U}} \tilde{\boldsymbol{U}}^\top = \boldsymbol{L}\boldsymbol{L}^\top. \end{aligned}$$

■

The factorization in Lemma 7 will be used in fitting HGLMs with correlated random effects by setting $\boldsymbol{L}$ as the design matrix for the random effects for estimating the model parameters in the estimation software in Chapter 3. In the next section, we express the likelihood function (2.24), derived from the randomization, as the permanent of a suitably chosen square matrix.

2.5 Likelihood Function as Permanent

The permanent of a matrix is similar to the determinant defined in Section A.4. The calculation of the determinant of a matrix is easy, however, this is not the case for the permanent. Using the definition of the permanent of the square matrix (A.11), the derived likelihood function can be expressed in terms of the multiple of a square matrix where each element is the probability density function of the random variables Y_{i,j_i}.

Corollary 3. *Under the assumption of Theorem 1 the likelihood function $L(\boldsymbol{\theta})$ can be expressed as the permanent of a square matrix.*

Proof. Apart from the factor $1/n!$ the right-hand side of equation (2.24) is

the permanent (Section A.4) of the $n \times n$ square matrix $\boldsymbol{D}$, where

$$\boldsymbol{D} = \begin{pmatrix} f_{1,j_1}(y_1;\boldsymbol{\theta}) & f_{1,j_2}(y_2;\boldsymbol{\theta}) & \dots & f_{1,j_n}(y_n;\boldsymbol{\theta}) \\ f_{2,j_1}(y_1;\boldsymbol{\theta}) & f_{2,j_2}(y_2;\boldsymbol{\theta}) & \dots & f_{2,j_n}(y_n;\boldsymbol{\theta}) \\ \vdots & \vdots & & \vdots \\ f_{n,j_1}(y_1;\boldsymbol{\theta}) & f_{n,j_2}(y_2;\boldsymbol{\theta}) & \dots & f_{n,j_n}(y_n;\boldsymbol{\theta}) \end{pmatrix}. \tag{2.29}$$

More precisely, the permanent of $\boldsymbol{D}$ is computed as

$$\begin{aligned} \text{per}\boldsymbol{D} &= \sum_{\pi \in S_n} f_{\pi(1),j_1}(y_1;\boldsymbol{\theta})\, f_{\pi(2),j_2}(y_2;\boldsymbol{\theta}) \dots f_{\pi(n),j_n}(y_n;\boldsymbol{\theta}) \\ &= \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i;\boldsymbol{\theta}). \end{aligned} \tag{2.30}$$

Therefore, using (2.30) in (2.24) we can write the likelihood function in terms of the permanent of the square matrix $\boldsymbol{D}$ as

$$L(\boldsymbol{\theta}) = \frac{1}{n!}\text{per}\boldsymbol{D} = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i;\boldsymbol{\theta}).$$

■

To compute the likelihood function it is necessary to calculate the permanent of the $n \times n$ square matrix $\boldsymbol{D}$ in equation (2.29). It is well known that the exact calculation of the permanent is very difficult because it involves the summation over all permutations of the symmetric group S_n and hence computationally time consuming. In the next section, we briefly describe the computational complexity of calculating the permanent, equivalently the likelihood function, and its consequences for estimating the model parameters of interest.

2.6 Complexity of Calculating the Likelihood Function

Complexity of an algorithm refers to the time and/or space required to compute all the necessary steps. Usually the order of the complexity of an algorithm is expressed in terms of the "big-O" notation (Arora and Barak, 2009, p.3). For example, $O(1)$ denotes constant running time, $O(n)$ denotes linear, $O(n^2)$

quadratic, $O(n^c)$ for polynomial time of order c and $O(c^n)$ for non-polynomial or exponential time, where c ia a constant and n is the input size of an algorithm. The notation refers to the upper bound of the complexity of an algorithm.

There are various types of complexity classes such as P, NP and #P (Jerrum, 2003). The class P is the set of problems that can be solved in polynomial time. Problems in the NP class can be solved in non-polynomial time or exponential time. The class #P is for problems that can be computed in non-deterministic polynomial time. In the complexity class #P, the most difficult problems are #P-complete. It has been shown that computing the permanent of a matrix is a #P-complete problem (Valiant, 1979). Also it appears that Ryser's method (Minc, 1984) is the most efficient for the exact computation of the permanent in terms of complexity. This method is evaluated using $O(2^n n^2)$ arithmetic operations.

However, a well known approximation algorithm for calculating the permanent of a matrix with non-negative entries has been developed by Jerrum et al. (2004) with run time $O(n^{10}(\log n)^3)$. Bezáková et al. (2008) improved this algorithm with reduced run time $O(n^7(\log n)^4)$. Also Huber and Law (2008) developed an algorithm for approximation of the permanent by using a different approach with expected run time $O(n^4 \log n)$.

The above computational complexity immediately tells us that it requires a huge number of operations and a huge amount of time to run an algorithm based on the randomization for calculating and maximizing the likelihood function in order to estimate the model parameters.

As a consequence of the above complexity results, we only consider the model for the simple underlying design CRD in order to estimate the model parameters. At this stage, the computational complexity does not allow us to estimate the model parameters of more complex designs. Even for the derived GLMM for the CRD we can only consider small examples. Although we do not use the permanent directly for the purpose of estimation in the randomization approach, it requires a high performance parallel computing cluster even for small examples. Considering the computational complexity for cal-

culating the permanent, equivalently the likelihood function, we develop the randomization-based estimation algorithm only for the derived model for the CRD in the next chapter.

Chapter 3

Estimation

We describe the estimation algorithm for the RB-GLMM for the CRD (Model given in (2.26)). The direct maximization of the likelihood function (2.24) is complicated because the summation in the likelihood does not commute with taking natural logarithms. To overcome this difficulty, we consider an alternative approach in which we maximize a minorization function rather than the likelihood function (Lange, 2013, p.186-187). The derivative of the minorization function can be found more easily. In general, we are applying a minorization-maximization (MM) algorithm to maximize the likelihood function (see Chapter 8, Lange (2013)). This type of algorithm includes a well-known expectation-maximization (EM) algorithm which is commonly used for latent variable models including GLMMs as a special case (e.g. McCulloch (1997)). The definition of the minorization function and background concepts are described in Section 3.1. We derive the minorization function for the likelihood function in Section 3.2. We then investigate the concavity of the minorization function and of the log-likelihood in Section 3.3.

In order to implement the algorithm for estimating the model parameters, we derive iterative weighted least squares (IWLS) equations in Section 3.4. We also derive the best linear predictors (BLP) of the random effects in Section 3.5. In Section 3.6, we give detailed steps and a flowchart of the algorithm. We conduct a simulation study for estimating the model parameters using the randomization-based algorithm with some examples in Section 3.7. In Section 3.8, we present an application of the RB-GLMM to a real data set. Section 3.9 briefly discusses computational limitations and future plans for generalizing the algorithm to larger samples.

3.1 Minorization Function

We introduce the minorization function as a surrogate function of the log-likelihood for estimating the RB-GLMM parameters for the CRD due to the fact that the minorization function is concave, and hence has a unique global maximum. Let $\boldsymbol{\theta} \in \mathbb{R}^p$ be the vector of p unknown parameters. A function $m : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is a minorization function for $\log L(\boldsymbol{\theta})$ if it satisfies the following properties (see Lange (2013, p.186-187), Lange et al. (2000), Hunter and Lange (2004), Wu et al. (2010)) :

(a) For all $\boldsymbol{\gamma}, \boldsymbol{\tau} \in \mathbb{R}^p$ it holds

$$m(\boldsymbol{\gamma}, \boldsymbol{\tau}) \leq \log L(\boldsymbol{\gamma}) \tag{3.1}$$

(b) For all $\boldsymbol{\gamma} \in \mathbb{R}^p$

$$m(\boldsymbol{\gamma}, \boldsymbol{\gamma}) = \log L(\boldsymbol{\gamma}). \tag{3.2}$$

Instead of maximizing $\log L(\boldsymbol{\theta})$ we consider the maximization of the minorization function $m(\boldsymbol{\gamma}, \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_0$ be some initial estimate of $\boldsymbol{\theta}$. For every $k = 0, 1, 2, \ldots$ let $\hat{\boldsymbol{\theta}}_{k+1}$ be a vector that maximizes the minorization function $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ when it is only regarded as a function of $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\theta}}_k$ is held fixed. It then follows by using property (3.1) that

$$\log L(\hat{\boldsymbol{\theta}}_{k+1}) \geq m(\hat{\boldsymbol{\theta}}_{k+1}, \hat{\boldsymbol{\theta}}_k) \geq m(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\theta}}_k). \tag{3.3}$$

Using the second property (3.2) we can write

$$m(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\theta}}_k) = \log L(\hat{\boldsymbol{\theta}}_k). \tag{3.4}$$

Therefore finally using (3.3) and (3.4) we obtain

$$\log L(\hat{\boldsymbol{\theta}}_{k+1}) \geq \log L(\hat{\boldsymbol{\theta}}_k) \tag{3.5}$$

which ensures that the log-likelihood increases or stays constant from iteration to iteration. This is illustrated later in Section 3.3 numerically with an example (see Table 3.1).

3.2 Derivation of the Minorization Function

In this section, we derive a minorization function for the likelihood function $L(\boldsymbol{\theta})$ in (2.24) for the RB-GLMM for the CRD. We recall that

$$L(\boldsymbol{\theta}) = \frac{1}{n!} \sum_{\pi \in S_n} \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i; \boldsymbol{\theta}),$$

where $f_{\pi(i),j_i}(y_i; \boldsymbol{\theta})$ is the density of a distribution from the exponential family. In order to derive the minorization function for the likelihood function, we first introduce some notation and definitions. Let $h : S_n \times \mathbb{R}^p \to \mathbb{R}$ be the function defined by

$$h(\pi, \boldsymbol{\gamma}) = \prod_{i=1}^{n} f_{\pi(i),j_i}(y_i; \boldsymbol{\gamma}) \tag{3.6}$$

for every $\pi \in S_n$ and $\boldsymbol{\gamma} \in \mathbb{R}^p$. Furthermore, we define $\tilde{h} : S_n \times \mathbb{R}^p \to \mathbb{R}$ by

$$\tilde{h}(\pi, \boldsymbol{\tau}) = \frac{h(\pi, \boldsymbol{\tau})}{\int_{S_n} h(\sigma, \boldsymbol{\tau}) U_n(d\sigma)}, \tag{3.7}$$

where, as before, U_n denotes the uniform distribution on the symmetric group S_n. The function $\tilde{h}$, defined in (3.7), is non-negative and satisfies

$$\int_{S_n} \tilde{h}(\sigma, \boldsymbol{\tau}) U_n(d\sigma) = 1.$$

We now define a measure D with respect to another measure by using $\tilde{h}$ as a density, as described in Section A.2.8. Using equation (A.8), substituting $M = D$, $f = \tilde{h}$ and $N = U_n$, we obtain

$$D(A) = \int_A \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) \tag{3.8}$$

for every $A \in \mathscr{P}(S_n)$. Let $X : S_n \to \mathbb{R}$ be the random variable defined by $X(\pi) = \dfrac{h(\pi, \boldsymbol{\gamma})}{h(\pi, \boldsymbol{\tau})}$ on the probability space $(S_n, \mathscr{P}(S_n), D)$. It is worth noting that the random variable X is non-negative by definition. Then the expectation of X with respect to D, denoted by $E_D(X)$, (Section A.2.5) is defined as

$$E_D(X) = \int_{S_n} X(\pi) D(d\pi). \tag{3.9}$$

Equation (3.9) can be rewritten by using the equation (A.9) in the appendix

with $g = X$, $M = D$, $f = \tilde{h}$ and $N = U_n$. It follows that

$$E_D(X) = \int_{S_n} X(\pi)\tilde{h}(\pi, \boldsymbol{\tau})U_n(d\pi), \tag{3.10}$$

since the probability measure D has density $\tilde{h}$ with respect to U_n. By using the above definitions and results we derive the following lemma.

Lemma 8. *The function $m(\boldsymbol{\gamma}, \boldsymbol{\tau})$ defined by*

$$m(\boldsymbol{\gamma}, \boldsymbol{\tau}) = \log L(\boldsymbol{\tau}) + \frac{1}{n!}\sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\gamma}) - \frac{1}{n!}\sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\tau}) \tag{3.11}$$

is a minorization function for the log-likelihood function $\log L(\boldsymbol{\theta})$ of the RB-GLMM (2.26) for the CRD.

Proof. We begin by considering the difference of log-likelihood functions as

$$\begin{aligned}
\log L(\boldsymbol{\gamma}) - \log L(\boldsymbol{\tau}) &= \log \frac{L(\boldsymbol{\gamma})}{L(\boldsymbol{\tau})} = \log \frac{(1/n!)\sum_{\pi \in S_n}\prod_{i=1}^n f_{\pi(i),j_i}(y_i; \boldsymbol{\gamma})}{(1/n!)\sum_{\pi \in S_n}\prod_{i=1}^n f_{\pi(i),j_i}(y_i; \boldsymbol{\tau})} \\
&= \log \frac{\int_{S_n}\prod_{i=1}^n f_{\pi(i),j_i}(y_i; \boldsymbol{\gamma})U_n(d\pi)}{\int_{S_n}\prod_{i=1}^n f_{\pi(i),j_i}(y_i; \boldsymbol{\tau})U_n(d\pi)} \\
&= \log \frac{\int_{S_n} h(\pi, \boldsymbol{\gamma})U_n(d\pi)}{\int_{S_n} h(\pi, \boldsymbol{\tau})U_n(d\pi)} \\
&= \log \int_{S_n} \frac{h(\pi, \boldsymbol{\gamma})}{h(\pi, \boldsymbol{\tau})} \frac{h(\pi, \boldsymbol{\tau})}{\int_{S_n} h(\sigma, \boldsymbol{\tau})U_n(d\sigma)} U_n(d\pi) \\
&= \log \int_{S_n} \frac{h(\pi, \boldsymbol{\gamma})}{h(\pi, \boldsymbol{\tau})} \tilde{h}(\pi, \boldsymbol{\tau})U_n(d\pi) \\
&= \log \int_{S_n} X(\pi)\tilde{h}(\pi, \boldsymbol{\tau})U_n(d\pi) \\
&= \log E_D(X)
\end{aligned} \tag{3.12}$$

which is obtained by using equation (3.10). Jensen's inequality (Billingsley, 1985, p.75, 283) shows that for every concave function

$$\log E(X) \geq E(\log X). \tag{3.13}$$

Therefore, by using (3.13) we can see that

$$\log L(\boldsymbol{\gamma}) - \log L(\boldsymbol{\tau}) = \log E_D(X) \geq E_D(\log X). \tag{3.14}$$

Moreover, the expectation $E_D(\log X)$ can be written as

$$\begin{aligned} E_D(\log X) &= \int_{S_n} [\log X(\pi)]\, \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) = \int_{S_n} \left[\log \frac{h(\pi, \boldsymbol{\gamma})}{h(\pi, \boldsymbol{\tau})}\right] \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) \\ &= \int_{S_n} [\log h(\pi, \boldsymbol{\gamma})]\, \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) - \int_{S_n} [\log h(\pi, \boldsymbol{\tau})]\, \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi). \end{aligned} \tag{3.15}$$

Using equation (3.15) in (3.14) it follows that

$$\begin{aligned} \log L(\boldsymbol{\gamma}) &\geq \log L(\boldsymbol{\tau}) + \int_{S_n} [\log h(\pi, \boldsymbol{\gamma})]\, \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) \\ &\quad - \int_{S_n} [\log h(\pi, \boldsymbol{\tau})]\, \tilde{h}(\pi, \boldsymbol{\tau}) U_n(d\pi) \\ &= \log L(\boldsymbol{\tau}) + \frac{1}{n!} \sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\gamma}) \\ &\quad - \frac{1}{n!} \sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\tau}). \end{aligned}$$

Defining the function m by

$$m(\boldsymbol{\gamma}, \boldsymbol{\tau}) = \log L(\boldsymbol{\tau}) + \frac{1}{n!} \sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\gamma}) - \frac{1}{n!} \sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\tau}) \tag{3.16}$$

we can see that $m(\boldsymbol{\gamma}, \boldsymbol{\tau}) \leq \log L(\boldsymbol{\gamma})$ and $m(\boldsymbol{\gamma}, \boldsymbol{\gamma}) = \log L(\boldsymbol{\gamma})$. That is, m satisfies the properties (3.1) and (3.2) and is hence a minorization function for $\log L(\boldsymbol{\theta})$. ■

3.3 Concavity and Monotonicity of the Minorization Function

In this section we demonstrate the concavity of the minorization function m in (3.16). Using (3.6) we can write

$$\log h(\pi, \boldsymbol{\gamma}) = \log \prod_{i=1}^{n} f_{\pi(i), j_i}(y_i; \boldsymbol{\gamma}) \tag{3.17}$$

and it can be shown that as a function of $\boldsymbol{\gamma}$ only, $\log h(\pi, \boldsymbol{\gamma})$ is concave (Bickel and Doksum, 2001, p.59-61). They have shown that $\log h(\pi, \boldsymbol{\gamma})$ is concave

for the full exponential family where $f_{\pi(i),j_i}(y_i;\boldsymbol{\gamma})$ is a member of this family. Hence it follows that for fixed $\boldsymbol{\tau}$, the function T defined by

$$T_{\boldsymbol{\tau}}(\boldsymbol{\gamma}) = \sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\gamma})$$

is also concave since a linear combination of concave functions with positive coefficients is concave. It follows that the minorization function m defined in Lemma 8 is concave. The log-likelihood function $\log L(\boldsymbol{\theta})$ may not be a concave function while the minorization function m is always concave. The concavity of the minorization function m is shown graphically in Figure 3.1.

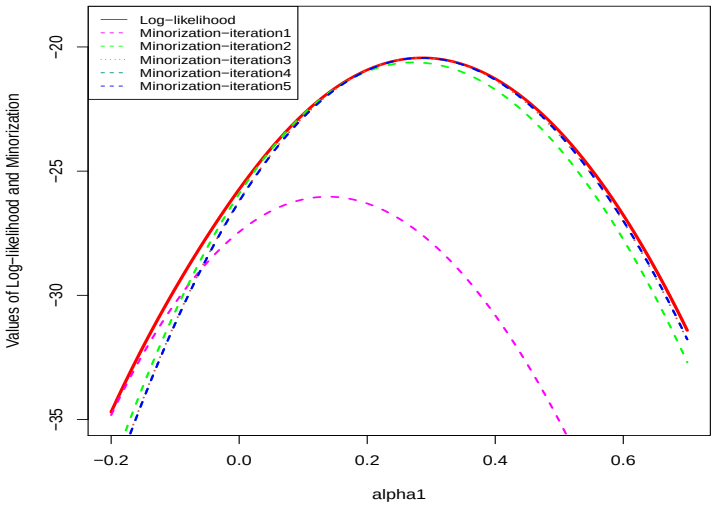

Figure 3.1: Log-likelihood and different minorization functions

Figure 3.1 shows minorization functions $m(\cdot, \hat{\theta}_k)$, $k = 1, \ldots, 5$ at different iterations and the log-likelihood. The maximization of the minorization functions is performed iteratively until a maximum of the log-likelihood is reached. The figure shows that the maxima of the minorization functions at different iterations gradually approach the maximum of the log-likelihood function.

Moreover, we consider a numerical example to illustrate that the log-likelihood increases or stays constant from iteration to iteration. In this example, we consider the number of treatments $t = 2$ and replications $r = 3$. We use the log-likelihood of the RB-GLMM for the CRD of α_1 assuming that the values of μ and the random effects ϵ_i are known. The results are given in Table 3.1.

Table 3.1: Example of increasing log-likelihood from iteration to iteration when true $\alpha_1 = 0.25$.

Iteration	$\hat{\alpha}_1$	Log-likelihood	Minorization
1	0.14071	−19.98390	−20.15749
2	0.17086	−19.89267	−19.90640
3	0.18183	−19.87107	−19.87294
4	0.18594	−19.86453	−19.86480
5	0.18749	−19.86229	−19.86233
6	0.18808	−19.86147	−19.86148
7	0.18830	−19.86117	−19.86117
8	0.18838	−19.86106	−19.86106

From Table 3.1 it can be seen that the log-likelihood increases from iteration 1 to 2, 2 to 3 and so on. By iteration 8, the log-likelihood is constant up to three decimal points i.e., the value of the log-likelihood is the same as the value of the previous iteration. Moreover, this is also the case for the minorization function. The log-likelihood and the minorization functions are the same at iteration 8.

3.4 Iterative Weighted Least Squares

Of the terms defining the minorization function m in Lemma 8, only the middle term depends on $\boldsymbol{\gamma}$. Therefore maximizing the minorization function $m(\boldsymbol{\gamma}, \boldsymbol{\tau})$ for fixed $\boldsymbol{\tau}$ it is equivalent to maximizing the function $\sum_{\pi \in S_n} \tilde{h}(\pi, \boldsymbol{\tau}) \log h(\pi, \boldsymbol{\gamma})$. More precisely, for a given estimate $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}$, the maximization of $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ as a function of $\boldsymbol{\gamma}$ is equivalent to maximizing

$$\begin{aligned} T_{\hat{\boldsymbol{\theta}}_k}(\boldsymbol{\gamma}) &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \log h(\pi, \boldsymbol{\gamma}) \\ &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \log \left(\prod_{i=1}^{n} f_{\pi(i), j_i}(y_i; \boldsymbol{\gamma}) \right) \\ &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} \log f_{\pi(i), j_i}(y_i; \boldsymbol{\gamma}). \end{aligned} \tag{3.18}$$

We will show that for a given estimate $\hat{\boldsymbol{\theta}}_k$, a maximum of $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ can be

computed using the familiar iterative weighted least squares (IWLS) algorithm for GLMs. In equation (3.18), $f_{\pi(i),j_i}(y_i;\boldsymbol{\gamma})$ is the density of a distribution from the exponential family. We now write $f_{\pi(i),j_i}(y_i;\boldsymbol{\gamma})$ in the form of the exponential family with some known functions $a(\cdot), b(\cdot), c(\cdot)$ and dispersion parameter ϕ (McCullagh and Nelder, 1989, p.28) as

$$f_{\pi(i),j_i}(y_i;\boldsymbol{\gamma}) = \exp\left((y_i\theta_{\pi(i),j_i,\boldsymbol{\gamma}} - b(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}))/a(\phi) + c(y_i,\phi)\right), \tag{3.19}$$

where $\theta_{\pi(i),j_i,\boldsymbol{\gamma}}$ is the (canonical) parameter related to the vector $\boldsymbol{\gamma}$ and π in S_n. More precisely, from the properties of the standard GLMs, it follows that

$$E(Y_{i,j_i}|\boldsymbol{\epsilon}=\boldsymbol{e}) = \mu_{\pi(i),j_i,\boldsymbol{\gamma}} = b'\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right) \tag{3.20}$$

and

$$\mathrm{var}(Y_{i,j_i}|\boldsymbol{\epsilon}=\boldsymbol{e}) = a(\phi)b''\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right). \tag{3.21}$$

We recall the RB-GLMM for the CRD (2.26) that

$$\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

The $n \times (t+1)$ design matrix $\boldsymbol{X}$ for the fixed treatments with number of parameters $t+1$ is defined by

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_i^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1s} & \dots & x_{1(t+1)} \\ x_{21} & x_{22} & \dots & x_{2s} & \dots & x_{2(t+1)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{is} & \dots & x_{i(t+1)} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{ns} & \dots & x_{n(t+1)} \end{pmatrix}. \tag{3.22}$$

Note that the elements of $\boldsymbol{X}$ are either 1 or 0. Also the $n \times n$ model matrix $\boldsymbol{Z}_\pi$ for the random effects is defined as

$$\boldsymbol{Z}_\pi = \begin{pmatrix} \boldsymbol{z}_{\pi(1)}^\top \\ \boldsymbol{z}_{\pi(2)}^\top \\ \vdots \\ \boldsymbol{z}_{\pi(i)}^\top \\ \vdots \\ \boldsymbol{z}_{\pi(n)}^\top \end{pmatrix}, \tag{3.23}$$

where the unit vector $\boldsymbol{z}_{\pi(i)}^{\top}$, $i = 1, \ldots, n$, is obtained by permuting the i-th row of the identity matrix $\boldsymbol{I}_n$ of order n.

Suppose that the vector $\boldsymbol{\gamma}$ is partitioned as $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\top}, \boldsymbol{\gamma}_2^{\top})^{\top}$ where $\boldsymbol{\gamma}_1$ is a $(t+1) \times 1$ vector of unknown fixed effects parameters and $\boldsymbol{\gamma}_2$ is the $n \times 1$ vector of the unit errors e_i, $i = 1, \ldots, n$, defined by equation (2.2). We write the vector of linear predictors, $\boldsymbol{\eta}_{\pi,\boldsymbol{\gamma}}$, by incorporating the random effects with fixed treatment effects in GLMMs setup as $\boldsymbol{\eta}_{\pi,\boldsymbol{\gamma}} = \boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{Z}_{\pi}\boldsymbol{\gamma}_2$. We then write the ith element of $\boldsymbol{\eta}_{\pi,\boldsymbol{\gamma}}$, denoted by $\eta_{\pi(i),j_i,\boldsymbol{\gamma}}$, for the fixed design $d = (j_1, \ldots, j_n)$ as

$$\eta_{\pi(i),j_i,\boldsymbol{\gamma}} = \boldsymbol{x}_i^{\top}\boldsymbol{\gamma}_1 + \boldsymbol{z}_{\pi(i)}^{\top}\boldsymbol{\gamma}_2. \tag{3.24}$$

Because of (3.19) we can rewrite equation (3.18) as

$$T_{\hat{\boldsymbol{\theta}}_k}(\boldsymbol{\gamma}) = \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \left(\sum_{i=1}^{n} \left(y_i \theta_{\pi(i),j_i,\boldsymbol{\gamma}} - b\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right)\right) / a(\phi) + \sum_{i=1}^{n} c\left(y_i, \phi\right) \right). \tag{3.25}$$

Thus maximizing (3.25) as a function of $\boldsymbol{\gamma}$ is equivalent to maximizing

$$\begin{aligned} \tilde{T}_{\hat{\boldsymbol{\theta}}_k}(\boldsymbol{\gamma}) &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \left(\sum_{i=1}^{n} \left(y_i \theta_{\pi(i),j_i,\boldsymbol{\gamma}} - b\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right)\right) / a(\phi) \right) \\ &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} T_{i,\boldsymbol{\gamma}}^{\pi}, \end{aligned} \tag{3.26}$$

where

$$T_{i,\boldsymbol{\gamma}}^{\pi} = \left(y_i \theta_{\pi(i),j_i,\boldsymbol{\gamma}} - b\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right)\right) / a(\phi). \tag{3.27}$$

We now compute the partial derivatives of equation (3.26) with respect to the component $\gamma_{1,s}$, $s = 1, \ldots, t+1$, of $\boldsymbol{\gamma}_1$ by applying the chain rule of differentiation

$$\begin{aligned} \frac{\partial \tilde{T}_{\hat{\boldsymbol{\theta}}_k}(\boldsymbol{\gamma})}{\partial \gamma_{1,s}} = U_s &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} \frac{\partial T_{i,\boldsymbol{\gamma}}^{\pi}}{\partial \gamma_{1,s}} \\ &= \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} \left(\frac{\partial T_{i,\boldsymbol{\gamma}}^{\pi}}{\partial \theta_{\pi(i),j_i,\boldsymbol{\gamma}}} . \frac{\partial \theta_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}} . \frac{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \gamma_{1,s}} \right) \end{aligned} \tag{3.28}$$

We compute each partial derivative of (3.28) separately. First using (3.27) and

(3.20) we write

$$\begin{aligned}\frac{\partial T^{\pi}_{i,\boldsymbol{\gamma}}}{\partial \theta_{\pi(i),j_i,\boldsymbol{\gamma}}} &= \left(y_i - b'\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right)\right)/a(\phi) \\ &= \left(y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}}\right)/a(\phi). \end{aligned} \tag{3.29}$$

Furthermore, using (3.21) we compute

$$\frac{\partial \theta_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}} = \frac{1}{\dfrac{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \theta_{\pi(i),j_i,\boldsymbol{\gamma}}}} = \frac{1}{b''\left(\theta_{\pi(i),j_i,\boldsymbol{\gamma}}\right)} = \frac{a(\phi)}{\mathrm{var}(Y_{i,j_i}|\boldsymbol{\epsilon}=\boldsymbol{e})}. \tag{3.30}$$

Moreover, using (3.24) and assuming the vector $\boldsymbol{\gamma}_2$ is known we then calculate

$$\begin{aligned}\frac{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \gamma_{1,s}} &= \frac{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \eta_{\pi(i),j_i,\boldsymbol{\gamma}}} . \frac{\partial \eta_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \gamma_{1,s}} \\ &= \frac{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \eta_{\pi(i),j_i,\boldsymbol{\gamma}}} \frac{\partial (\boldsymbol{x}_i^{\top}\boldsymbol{\gamma}_1 + \boldsymbol{z}_{\pi(i)}^{\top}\boldsymbol{\gamma}_2)}{\partial \gamma_{1,s}} \\ &= \frac{1}{g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}})} x_{is} \end{aligned} \tag{3.31}$$

with $g(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) = \eta_{\pi(i),j_i,\boldsymbol{\gamma}}$ and $g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) = \dfrac{\partial \eta_{\pi(i),j_i,\boldsymbol{\gamma}}}{\partial \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}$ where g is the link function as before. Finally substituting (3.29), (3.30) and (3.31) into (3.28), we can write

$$\begin{aligned} U_s =& \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} \left(\frac{y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}}}{a(\phi)} \frac{a(\phi)}{\mathrm{var}(Y_{i,j_i}|\boldsymbol{\epsilon}=\boldsymbol{e})} \frac{1}{g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}})} x_{is} \right) \\ =& \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \sum_{i=1}^{n} \left((y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}}) \frac{1}{\mathrm{var}(Y_{i,j_i}|\boldsymbol{\epsilon}=\boldsymbol{e})} \frac{1}{g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}})} x_{is} \right) \end{aligned} \tag{3.32}$$

Gathering partial derivatives with respect to $\gamma_{1,s}$, $s = 1, \ldots, t+1$ into a $(t+1) \times 1$ column vector we get

$$\frac{\partial \tilde{T}_{\hat{\boldsymbol{\theta}}_k}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1} = \boldsymbol{U}_{\pi,\boldsymbol{\gamma}} = \sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_k) \boldsymbol{X}^{\top} \boldsymbol{W}_{\pi,\boldsymbol{\gamma}} \boldsymbol{D}_{\pi,\boldsymbol{\gamma}} (\boldsymbol{y} - \boldsymbol{\mu}_{\pi,\boldsymbol{\gamma}}). \tag{3.33}$$

The matrices and vectors in (3.33) are defined as follows:

$$\boldsymbol{W}_{\pi,\boldsymbol{\gamma}} = \text{diag}\left(\frac{1}{\text{var}(Y_{1,j_1}|\boldsymbol{\epsilon}=\boldsymbol{e})(g'(\mu_{\pi(1),j_1,\boldsymbol{\gamma}}))^2}, \ldots, \frac{1}{\text{var}(Y_{1,j_1}|\boldsymbol{\epsilon}=\boldsymbol{e})(g'(\mu_{\pi(n),j_n,\boldsymbol{\gamma}}))^2}\right), \tag{3.34}$$

$$\boldsymbol{D}_{\pi,\boldsymbol{\gamma}} = \text{diag}\left(g'(\mu_{\pi(1),j_1,\boldsymbol{\gamma}}), \ldots, g'(\mu_{\pi(n),j_n,\boldsymbol{\gamma}})\right) \tag{3.35}$$

and $\boldsymbol{\mu}_{\pi,\boldsymbol{\gamma}} = (\mu_{\pi(1),j_1,\boldsymbol{\gamma}}, \ldots, \mu_{\pi(n),j_n,\boldsymbol{\gamma}})^\top$. Now, using the first-order Taylor approximation to the link function g at $\mu_{\pi(i),j_i,\boldsymbol{\gamma}}$ (McCullagh and Nelder, 1989, p.40) we have

$$\begin{aligned} g(y_i) &\simeq g(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) + (y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}})g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) \\ &= \boldsymbol{x}_i^\top\boldsymbol{\gamma}_1 + \boldsymbol{z}_{\pi(i)}^\top\boldsymbol{\gamma}_2 + (y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}})g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) \end{aligned}$$

for every i. Setting the new working variable $\tilde{z}_{\pi(i),j_i,\boldsymbol{\gamma}} = g(\mu_{\pi(i),j_i,\boldsymbol{\gamma}}) + (y_i - \mu_{\pi(i),j_i,\boldsymbol{\gamma}})g'(\mu_{\pi(i),j_i,\boldsymbol{\gamma}})$ equation (3.33) can be approximated by

$$\sum_{\pi\in S_n} \tilde{h}(\pi,\hat{\boldsymbol{\theta}}_k)\boldsymbol{X}^\top\boldsymbol{W}_{\pi,\boldsymbol{\gamma}}(\tilde{\boldsymbol{z}}_{\pi,\boldsymbol{\gamma}} - \boldsymbol{X}\boldsymbol{\gamma}_1 - \boldsymbol{Z}_\pi\boldsymbol{\gamma}_2) \tag{3.36}$$

where $\tilde{\boldsymbol{z}}_{\pi,\boldsymbol{\gamma}} = (\tilde{z}_{\pi(1),j_1,\boldsymbol{\gamma}}, \ldots, \tilde{z}_{\pi(n),j_n,\boldsymbol{\gamma}})^\top$. Equating (3.36) to $\mathbf{0}$ yields the form of the IWLS equation as

$$\boldsymbol{X}^\top\left(\sum_{\pi\in S_n}\tilde{h}(\pi,\hat{\boldsymbol{\theta}}_k)\boldsymbol{W}_{\pi,\boldsymbol{\gamma}}\right)\boldsymbol{X}\boldsymbol{\gamma}_1 = \boldsymbol{X}^\top\sum_{\pi\in S_n}\tilde{h}(\pi,\hat{\boldsymbol{\theta}}_k)\boldsymbol{W}_{\pi,\boldsymbol{\gamma}}(\tilde{\boldsymbol{z}}_{\pi,\boldsymbol{\gamma}} - \boldsymbol{Z}_\pi\boldsymbol{\gamma}_2). \tag{3.37}$$

3.5 Best Linear Predictor of Random Effects

Implementing the algorithm for estimating the vector $\boldsymbol{\beta}$ of fixed effect parameters in the RB-GLMM (2.26), one needs to predict the vector $\boldsymbol{\epsilon}$ of random errors. This is due to the fact that the estimation of $\boldsymbol{\beta}$ depends on $\boldsymbol{\epsilon}$, however, the vector $\boldsymbol{\epsilon}$ is an unobserved random variable. In order to do so we use the best linear predictor (BLP) of the vector of random errors $\boldsymbol{\epsilon}$. The formula is given by McCulloch et al. (2008, p.307) and Searle et al. (1992, p.266) as

$$\text{BLP}(\boldsymbol{\epsilon}) = \hat{\boldsymbol{\epsilon}} = \boldsymbol{\mu}_{\boldsymbol{\epsilon}} + \boldsymbol{C}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}}), \tag{3.38}$$

where $\boldsymbol{\mu}_{\boldsymbol{\epsilon}} = E(\boldsymbol{\epsilon})$ is the expectation of $\boldsymbol{\epsilon}$, $\boldsymbol{\mu}_{\boldsymbol{Y}} = E(\boldsymbol{Y})$ is the mean for the vector of responses $\boldsymbol{Y}$, $\boldsymbol{C} = Cov(\boldsymbol{\epsilon}, \boldsymbol{Y})$ is the variance-covariance matrix be-

tween $\boldsymbol{\epsilon}$ and $\boldsymbol{Y}$, and $\boldsymbol{\Sigma} = Cov(\boldsymbol{Y})$ is also the variance-covariance matrix of $\boldsymbol{Y}$. In the RB-GLMM (2.26) for a fixed design $d = (j_1, \ldots, j_n)$, we recall that $\boldsymbol{Y} = (Y_{1,j_1}, \ldots, Y_{n,j_n})^\top$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$. The formula given by (3.38) is based on the first and second moments of $\boldsymbol{\epsilon}$ and $\boldsymbol{Y}$ but without any assumption of normality.

The derivation of the BLP formula is solely based on the moments of the random effects using the minimized criterion $E((\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon})^\top(\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon}))$, which is known as the mean square error of prediction (MSEP) (see Searle et al. (1992, p.261, 267)). It follows that the linear predictor $\hat{\boldsymbol{\epsilon}}$ computed according to the BLP formula (3.38) is the best among all other linear predictors in the sense that

$$E((\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon})^\top(\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon})) \leq E((\hat{\boldsymbol{\epsilon}}^* - \boldsymbol{\epsilon})^\top(\hat{\boldsymbol{\epsilon}}^* - \boldsymbol{\epsilon})), \tag{3.39}$$

where $\hat{\boldsymbol{\epsilon}}^*$ is any other linear predictor of $\boldsymbol{\epsilon}$. For the RB-GLMM the expected values are computed on the product space $S_n \times \Omega$.

Equation (3.38) is applicable to all forms of $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}$ and therefore we can apply (3.38) for predicting the vector of random errors $\boldsymbol{\epsilon}$ in the RB-GLMM (2.26) for the CRD. In order to apply (3.38) into the estimation algorithm we now derive some auxiliary results. The detailed derivation of these results is given below. We first compute the expectation and variance of an element Y_{i,j_i} and also the covariance between two elements Y_{k,j_k} and Y_{l,j_l} for $k \neq l$ of $\boldsymbol{Y}$.

Lemma 9. *Let X_{i,j_i} be the response whose distribution is assumed to be from the exponential family. Let Y_{i,j_i} be the response for the i-th randomized experimental unit and treatment j_i which is an element of the vector of responses $\boldsymbol{Y}$. Then the expectation and variance of Y_{i,j_i}, and the covariance between Y_{k,j_k} and Y_{l,j_l} for $k \neq l$ are*

(i) $E(Y_{i,j_i}) = \dfrac{1}{n}\sum_{i^*=1}^{n} E(X_{i^*,j_i}) = m_{j_i}$,

(ii) $Var(Y_{i,j_i}) = \dfrac{1}{n}\sum_{i^*=1}^{n}\left(Var(X_{i^*,j_i}) + (E(X_{i^*,j_i}))^2\right) - m_{j_i}^2$,

(iii) $Cov(Y_{k,j_k}, Y_{l,j_l}) = \dfrac{1}{n-1} m_{j_k} m_{j_l} - \dfrac{1}{n(n-1)}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l})$.

Proof. (i) The expectation of Y_{i,j_i} for every i and j_i is given by

$$\begin{aligned} E(Y_{i,j_i}) &= \int_{S_n}\int_{\Omega} Y_{i,j_i}(\pi,\omega)\, P\,(d\omega)\, U_n\,(d\pi) = \frac{1}{n!}\sum_{\pi\in S_n}\int_{\Omega} X_{\pi(i),j_i}(\omega) P\,(d\omega) \\ &= \frac{1}{n!}\sum_{\pi\in S_n} E(X_{\pi(i),j_i}) = \frac{1}{n!}\sum_{i^*=1}^{n}\sum_{\pi\in Q_{i,i^*}} E(X_{i^*,j_i}) \\ &= \frac{(n-1)!}{n!}\sum_{i^*=1}^{n} E(X_{i^*,j_i}) = \frac{1}{n}\sum_{i^*=1}^{n} E(X_{i^*,j_i}) = m_{j_i}. \end{aligned}$$

(ii) Using the result in part (i), the variance of Y_{i,j_i} is equal to

$$\begin{aligned} Var(Y_{i,j_i}) &= E(Y_{i,j_i}^2) - (E(Y_{i,j_i}))^2 \\ &= \int_{S_n}\int_{\Omega} Y_{i,j_i}^2(\pi,\omega)\, P\,(d\omega)\, U_n\,(d\pi) - m_{j_i}^2 \\ &= \frac{1}{n!}\sum_{\pi\in S_n}\int_{\Omega} X_{\pi(i),j_i}^2(\omega) P\,(d\omega) - m_{j_i}^2 \\ &= \frac{1}{n}\sum_{i^*=1}^{n} E(X_{i^*,j_i}^2) - m_{j_i}^2 \\ &= \frac{1}{n}\sum_{i^*=1}^{n} \left(Var(X_{i^*,j_i}) + (E(X_{i^*,j_i}))^2\right) - m_{j_i}^2. \end{aligned}$$

(iii) Again by using the result in part (i), we can write the covariance of Y_{k,j_k} and Y_{l,j_l} for $k \neq l$ as

$$\begin{aligned} Cov(Y_{k,j_k}, Y_{l,j_l}) &= E(Y_{k,j_k}Y_{l,j_l}) - E(Y_{k,j_k})E(Y_{l,j_l}) \\ &= E(Y_{k,j_k}Y_{l,j_l}) - m_{j_k}m_{j_l}. \end{aligned} \tag{3.40}$$

Now $E(Y_{k,j_k}Y_{l,j_l})$ in (3.40) is equal to

$$
\begin{aligned}
& E(Y_{k,j_k}Y_{l,j_l}) = \int_{S_n}\int_{\Omega} Y_{k,j_k}(\pi,\omega)\; Y_{l,j_l}(\pi,\omega)P(d\omega)U_n(d\pi) \\
= \;& \frac{1}{n!}\sum_{\pi\in S_n}\int_{\Omega} X_{\pi(k),j_k}(\omega)X_{\pi(l),j_l}(\omega)P(d\omega) = \frac{1}{n!}\sum_{\pi\in S_n} E(X_{\pi(k),j_k}X_{\pi(l),j_l}) \\
= \;& \frac{1}{n!}\sum_{k^*\neq l^*}^{n}\sum_{\pi\in Q_{k,k^*}\bigcap Q_{l,l^*}} E(X_{k^*,j_k}X_{l^*,j_l}) = \frac{(n-2)!}{n!}\sum_{k^*\neq l^*}^{n} E(X_{k^*,j_k}X_{l^*,j_l}) \\
= \;& \frac{(n-2)!}{n!}\sum_{k^*\neq l^*}^{n} E(X_{k^*,j_k}X_{l^*,j_l}) = \frac{1}{n(n-1)}\sum_{k^*=1}^{n}\sum_{\substack{l^*=1\\ l^*\neq k^*}}^{n} E(X_{k^*,j_k})E(X_{l^*,j_l}) \\
= \;& \frac{1}{n(n-1)}\sum_{k^*=1}^{n}\left(\sum_{l^*=1}^{n} E(X_{k^*,j_k})E(X_{l^*,j_l}) - E(X_{k^*,j_k})E(X_{k^*,j_l})\right) \\
= \;& \frac{1}{n(n-1)}\left(\sum_{k^*=1}^{n} E(X_{k^*,j_k})\sum_{l^*=1}^{n} E(X_{l^*,j_l}) - \sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l})\right) \\
= \;& \frac{1}{n(n-1)}\left(n^2 m_{j_k}m_{j_l} - \sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l})\right) \\
= \;& \frac{n}{n-1}m_{j_k}m_{j_l} - \frac{1}{n(n-1)}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l}). \qquad (3.41)
\end{aligned}
$$

Using (3.41) in (3.40) we can rewrite

$$
\begin{aligned}
Cov(Y_{k,j_k},Y_{l,j_l}) &= \frac{n}{n-1}m_{j_k}m_{j_l} - \frac{1}{n(n-1)}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l}) \\
&\quad -m_{j_k}m_{j_l} \\
&= \frac{1}{n-1}m_{j_k}m_{j_l} - \frac{1}{n(n-1)}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l}).
\end{aligned}
$$

■

The results in Lemma 9 are similar to those in Section 7.3 of McCulloch et al. (2008). Next in order to derive the covariance matrix $\boldsymbol{C}$ between $\boldsymbol{\epsilon}$ and $\boldsymbol{Y}$ we compute the following results.

Lemma 10. *The covariances between ϵ_i and Y_{i,j_i}, and ϵ_k and Y_{l,j_l} for $k \neq l$ are*

(i) $Cov(\epsilon_i, Y_{i,j_i}) = \frac{1}{n}\sum_{i^*=1}^{n} e_{i^*}E(X_{i^*,j_i})$,

(ii) $Cov(\epsilon_k, Y_{l,j_l}) = -\dfrac{1}{(n-1)n}\sum_{k^*=1}^{n} e_{k^*}E(X_{k^*,j_l})$.

Proof. *(i)* We know from part (i) of Lemma 3 that $E(\epsilon_i) = 0$. The covariance between ϵ_i and Y_{i,j_i} can be written as

$$\begin{aligned} Cov(\epsilon_i, Y_{i,j_i}) &= E(\epsilon_i Y_{i,j_i}) - E(\epsilon_i)E(Y_{i,j_i}) = E(\epsilon_i Y_{i,j_i}) \\ &= E(E(\epsilon_i Y_{i,j_i}|\boldsymbol{\epsilon})). \end{aligned} \tag{3.42}$$

Now we consider the inner expectation in (3.42),

$$\begin{aligned} &E(\epsilon_i Y_{i,j_i}|\boldsymbol{\epsilon}) = E(\epsilon_i Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon})) \\ &= \sum_{l=1}^{n!} \frac{1}{(U_n \otimes P)(\{\pi_l\} \times \Omega)} \int_{\{\pi_l\}\times\Omega} \epsilon_i(\pi,\tilde{\omega}) Y_{i,j_i}(\pi,\tilde{\omega})\,(U_n \otimes P)\,d(\pi,\tilde{\omega})\, 1_{\{\pi_l\}\times\Omega} \\ &= \sum_{\sigma\in S_n} \frac{1}{(U_n \otimes P)(\{\sigma\} \times \Omega)} \int_{\{\sigma\}\times\Omega} \epsilon_i(\pi,\tilde{\omega}) Y_{i,j_i}(\pi,\tilde{\omega})\,(U_n \otimes P)\,d(\pi,\tilde{\omega})\, 1_{\{\sigma\}\times\Omega} \\ &= \sum_{\sigma\in S_n} \frac{1}{U_n(\{\sigma\})\,P(\Omega)} \int_{\{\sigma\}} \int_{\Omega} \epsilon_i(\pi,\tilde{\omega}) Y_{i,j_i}(\pi,\tilde{\omega}) P(d\tilde{\omega})\,U_n(d\pi)\;\; 1_{\{\sigma\}\times\Omega} \\ &= \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega} \int_{\Omega} \tilde{\epsilon}_i(\pi) Y_{i,j_i}(\sigma,\tilde{\omega})\,P(d\tilde{\omega}) \\ &= \sum_{\sigma\in S_n} 1_{\{\sigma\}\times\Omega} \int_{\Omega} e_{\pi(i)} X_{\sigma(i),j_i}(\tilde{\omega})\,P(d\tilde{\omega}) \\ &= \sum_{\sigma\in S_n} e_{\pi(i)} 1_{\{\sigma\}\times\Omega} \int_{\Omega} X_{\sigma(i),j_i}(\tilde{\omega})\,P(d\tilde{\omega}). \end{aligned} \tag{3.43}$$

We can rewrite (3.43) for every $(\pi,\omega) \in S_n \times \Omega$ as

$$\begin{aligned} E(\epsilon_i Y_{i,j_i}|\mathscr{F}(\boldsymbol{\epsilon}))(\pi,\omega) &= \sum_{\sigma\in S_n} e_{\pi(i)} 1_{\{\sigma\}\times\Omega}(\pi,\omega) \int_{\Omega} X_{\sigma(i),j_i}(\tilde{\omega})\,P(d\tilde{\omega}) \\ &= \sum_{\sigma\in S_n} e_{\pi(i)} 1_{\{\sigma\}}(\pi) \int_{\Omega} X_{\sigma(i),j_i}(\tilde{\omega})\,P(d\tilde{\omega}) \\ &= \sum_{\sigma\in S_n} e_{\pi(i)} 1_{\{\sigma\}}(\pi) E(X_{\sigma(i),j_i}) \\ &= e_{\pi(i)} E(X_{\pi(i),j_i}). \end{aligned} \tag{3.44}$$

Using (3.44) in (3.42) the covariance between ϵ_i and Y_{i,j_i} can be rewritten

as

$$\begin{aligned} Cov(\epsilon_i, Y_{i,j_i}) &= E(e_{\pi(i)}E(X_{\pi(i),j_i})) = \int_{S_n} e_{\pi(i)}E(X_{\pi(i),j_i})U_n(d\pi) \\ &= \frac{1}{n!}\sum_{\pi\in S_n} e_{\pi(i)}E(X_{\pi(i),j_i}) = \frac{1}{n!}\sum_{i^*=1}^{n}\sum_{\pi\in Q_{i,i^*}} e_{i^*}E(X_{i^*,j_i}) \\ &= \frac{(n-1)!}{n!}\sum_{i^*=1}^{n} e_{i^*}E(X_{i^*,j_i}) = \frac{1}{n}\sum_{i^*=1}^{n} e_{i^*}E(X_{i^*,j_i}). \end{aligned}$$

(ii) Similar to part (i) and using (3.44), the covariance between ϵ_k and Y_{l,j_l} for $k \neq l$ is equal to

$$\begin{aligned} Cov(\epsilon_k, Y_{l,j_l}) &= E(\epsilon_k Y_{l,j_l}) = E(E(\epsilon_k Y_{l,j_l}|\boldsymbol{\epsilon})) = E\left(e_{\pi(k)}E(X_{\pi(l),j_l})\right) \\ &= \int_{S_n} e_{\pi(k)}E(X_{\pi(l),j_l})U_n(d\pi) = \frac{1}{n!}\sum_{\pi\in S_n} e_{\pi(k)}E(X_{\pi(l),j_l}) \\ &= \frac{1}{n!}\sum_{k^*\neq l^*}^{n}\sum_{\pi\in Q_{k,k^*}\bigcap Q_{l,l^*}} e_{k^*}E(X_{l^*,j_l}) \\ &= \frac{(n-2)!}{n!}\sum_{k^*\neq l^*}^{n} e_{k^*}E(X_{l^*,j_l}) \\ &= \frac{1}{n(n-1)}\sum_{k^*=1}^{n} e_{k^*}\sum_{\substack{l^*=1 \\ l^*\neq k^*}}^{n} E(X_{l^*,j_l}) \\ &= \frac{1}{n(n-1)}\sum_{k^*=1}^{n} e_{k^*}\left(\sum_{l^*=1}^{n} E(X_{l^*,j_l}) - E(X_{k^*,j_l})\right) \\ &= \frac{1}{n(n-1)}\left(\sum_{k^*=1}^{n} e_{k^*}\sum_{l^*=1}^{n} E(X_{l^*,j_l}) - \sum_{k^*=1}^{n} e_{k^*}E(X_{k^*,j_l})\right) \\ &= -\frac{1}{(n-1)n}\sum_{k^*=1}^{n} e_{k^*}E(X_{k^*,j_l}). \end{aligned}$$

■

The results in Lemmas 9 and 10 will allow us to explicitly express $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{C}$ in the BLP equation (3.38) in matrix notation. Let $\boldsymbol{M}$ be the $n \times n$ matrix

$$\boldsymbol{M} = \begin{pmatrix} E(X_{1,j_1}) & \cdots & E(X_{1,j_n}) \\ \vdots & & \vdots \\ E(X_{n,j_1}) & \cdots & E(X_{n,j_n}) \end{pmatrix} \tag{3.45}$$

and $\boldsymbol{V}$ be the $n \times n$ matrix

$$\boldsymbol{V} = \begin{pmatrix} Var(X_{1,j_1}) & \cdots & Var(X_{1,j_n}) \\ \vdots & & \vdots \\ Var(X_{n,j_1}) & \cdots & Var(X_{n,j_n}) \end{pmatrix}. \tag{3.46}$$

The elements of $\boldsymbol{M}$ and $\boldsymbol{V}$ are functions of the vector $\boldsymbol{\beta}$ of fixed effects parameters and of the vector $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$ where e_i's are defined in Section 2.1.1. Now from Lemma 9(i) we can write the mean vector $\boldsymbol{\mu_Y}$ of responses as

$$\boldsymbol{\mu_Y} = E(\boldsymbol{Y}) = E(Y_{1,j_1}, \ldots, Y_{n,j_n}) = \frac{1}{n}\boldsymbol{M}^\top \mathbf{1}_n. \tag{3.47}$$

For every vector $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$ let $\mathrm{diag}(\boldsymbol{a})$ be the diagonal matrix with elements $a_1, \ldots, a_n$ on the main diagonal. Also $\mathrm{diag}^2(\boldsymbol{a}) = \mathrm{diag}(\boldsymbol{a})\ \mathrm{diag}(\boldsymbol{a})$. Now we can write the results of Lemma 9 and 10 in matrix notation below.

Lemma 11. *Let $\boldsymbol{a_M}$ be the vector containing the diagonal elements of $\boldsymbol{M}^\top\boldsymbol{M}$. Then the variance-covariance matrix $\boldsymbol{\Sigma}$ of the vector $\boldsymbol{Y}$ of responses is*

$$\begin{aligned} \boldsymbol{\Sigma} \;=\; & \frac{1}{(n-1)n}\Big((n-1)\,\mathrm{diag}(\boldsymbol{V}^\top \mathbf{1}_n) + n\,\mathrm{diag}(\boldsymbol{a_M}) - \mathrm{diag}^2(\boldsymbol{M}^\top \mathbf{1}_n) \\ & -\boldsymbol{M}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{M}\Big). \end{aligned} \tag{3.48}$$

Lemma 12. *The covariance matrix $\boldsymbol{C}$ between $\boldsymbol{\epsilon}$ and $\boldsymbol{Y}$ is equal to*

$$\boldsymbol{C} = \frac{1}{(n-1)n}(n\,\mathrm{diag}(\boldsymbol{e}^\top\boldsymbol{M}) - \mathbf{1}_n\boldsymbol{e}^\top\boldsymbol{M}). \tag{3.49}$$

One can show that the diagonal elements of $\boldsymbol{\Sigma}$ in (3.48) have the same form in part (ii) of Lemma 9 and that the off-diagonal elements are exactly the same as in Lemma 9(iii). Similarly, the diagonal and off-diagonal elements of $\boldsymbol{C}$ in (3.49) are the same as in Lemma 10(i) and (ii) respectively. The proof of Lemma 11 and 12 is given in Appendix B.

Finally, for the RB-GLMM (2.26) using $\boldsymbol{\mu_\epsilon} = \mathbf{0}$ (Section 2.1.6), the BLP equation (3.38) reduces to

$$\mathrm{BLP}(\boldsymbol{\epsilon}) = \hat{\boldsymbol{\epsilon}} = \pi(\hat{\boldsymbol{e}}) = \boldsymbol{C}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu_Y}), \tag{3.50}$$

where $\boldsymbol{\mu_Y}$, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ are given in equations (3.47), (3.48) and (3.49) respectively. The right hand side of (3.50) depends on $\boldsymbol{e}$, and so this will be used as an update equation in an iterative procedure. Note that $\boldsymbol{\Sigma}$ and $\boldsymbol{C}$ depend only on $\boldsymbol{M}$ and $\boldsymbol{V}$. Therefore, during the algorithm one can also easily update $\boldsymbol{\Sigma}$ and $\boldsymbol{C}$ to get $\boldsymbol{\Sigma}^{(l+1)}$ and $\boldsymbol{C}^{(l+1)}$, say from, $\boldsymbol{M}^{(l+1)}$ and $\boldsymbol{V}^{(l+1)}$.

We now consider the situation, in particular, that the random variables $X_{i,j}$ follow the Poisson distribution and we call the corresponding RB-GLMM the Poisson RB-GLMM. We show the calculation of the BLP equation (3.50) for the Poisson RB-GLMM for the CRD (2.26). Recall that for the given fixed design $d = (j_1, \ldots, j_n)$ we have $\boldsymbol{\mu_Y} = E(Y_{1,j_1}, \ldots, Y_{n,j_n}) = \frac{1}{n}\boldsymbol{M}^\top \mathbf{1}_n$. For the Poisson responses and the CRD, from (2.2) we have $E(X_{i,j}) = \exp(\mu + \alpha_j + e_i)$. Let $\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_{t-1})^\top \in \mathbb{R}^t$ be the effects-coded fixed effects parameter vector and $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$ be defined as before. Define the $t \times t$ matrix

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{1}_{t-1} & \boldsymbol{I}_{t-1} \\ 1 & -\mathbf{1}_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 1 \\ 1 & -1 & \cdots & -1 \end{pmatrix}. \tag{3.51}$$

Then $\boldsymbol{A\beta} = (\mu + \alpha_1, \ldots, \mu + \alpha_t)^\top$ where $\alpha_t = -\alpha_1 - \ldots - \alpha_{t-1}$. Let $\boldsymbol{g}_i \in \mathbb{R}^t$ be the i-th unit (column) vector in $\mathbb{R}^t$ with 1 in row i and 0 otherwise. Then we can write $\boldsymbol{g}_i^\top \boldsymbol{A\beta} = \mu + \alpha_i$. For the given design $d = (j_1, \ldots, j_n)$ define the $n \times t$ matrix

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{g}_{j_1}^\top \\ \vdots \\ \boldsymbol{g}_{j_n}^\top \end{pmatrix}. \tag{3.52}$$

Note that in the special case of an equi-replicated design d with t treatments and replication r where the treatments are in the order $j_1 = \ldots = j_r = 1, j_{r+1} = \ldots = j_{2r} = 2, \ldots, j_{(t-1)r+1} = \ldots = j_{tr} = t$ then the matrix $\boldsymbol{G}$ simplifies to the Kronecker product $\boldsymbol{G} = \boldsymbol{I}_t \otimes \mathbf{1}_r$. It follows that for Poisson responses, in general, the matrix $\boldsymbol{M}$ in (3.45) can be computed as

$$\boldsymbol{M} = \exp(\mathbf{1}_n \mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{GA\beta}) + \boldsymbol{e}\mathbf{1}_n^\top), \tag{3.53}$$

where the exponential function is applied to every element of the $n \times n$ matrix

$$\mathbf{1}_n \mathbf{1}_n^\top \text{diag}(\boldsymbol{GA\beta}) + \boldsymbol{e}\mathbf{1}_n^\top. \tag{3.54}$$

For the Poisson responses the mean is equal to the variance. It follows that $\boldsymbol{M} = \boldsymbol{V}$. Now from (3.48) we can write the variance-covariance matrix for the Poisson responses ($\boldsymbol{\Sigma}_P$) as

$$\begin{aligned} \boldsymbol{\Sigma}_P &= \frac{1}{(n-1)n}\Big((n-1)\,\text{diag}(\boldsymbol{M}^\top \mathbf{1}_n) + n\,\text{diag}(\boldsymbol{a_M}) - \text{diag}^2(\boldsymbol{M}^\top \mathbf{1}_n) \\ &\quad -\boldsymbol{M}^\top (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top)\boldsymbol{M} \Big), \end{aligned} \tag{3.55}$$

where $\boldsymbol{M}$ is given by equation (3.53). Using $\boldsymbol{\Sigma}_P$ we can write the BLP equation for the Poisson RB-GLMM as

$$\hat{\boldsymbol{\epsilon}} = \boldsymbol{C}\boldsymbol{\Sigma}_P^{-1}(\boldsymbol{y} - \boldsymbol{\mu_Y}). \tag{3.56}$$

3.6 Algorithm

In this section, we describe the randomization-based algorithm for implementing the estimation procedure for the RB-GLMM (2.26) for the CRD. In the algorithm, the systematic part of the model for the fixed treatment effects represented by $\boldsymbol{X\beta}$. However, for the effects due to the units we restrict ourselves to the case where $\boldsymbol{Z}_\pi$ is a matrix obtained by permuting the rows of the identity matrix $\boldsymbol{I}_n$ of order n and a vector $\boldsymbol{\epsilon} = \boldsymbol{e}$ of errors whose components are a fixed permutation (corresponding to the actual randomization used in the experiment) of the errors $e_i, i = 1, ..., n$, defined by equation (2.2).

The parameters of interest to be estimated are the components of the vector $\boldsymbol{\beta}$. In order to be able to compute an estimate we also need to know the vector $\boldsymbol{\epsilon}$ which is however not observable. Whenever $\boldsymbol{\epsilon}$ is needed in a computation we use equation (3.50). In what follows we denote the vector obtained by combining $\boldsymbol{\epsilon}$ and $\boldsymbol{\beta}$ by $\boldsymbol{\theta}$. In other words $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\epsilon}^\top)^\top$. The algorithm performs an outer iteration in which a new minorization function $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ based on the estimate $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}$ from the k-th loop of that outer iteration is computed. Note that given $\hat{\boldsymbol{\theta}}_k$ the function $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ depends only on $\boldsymbol{\gamma}$. Having determined $m(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}_k)$ an inner iteration is invoked to find a vector $\boldsymbol{\gamma}^*$ that maximizes this

function. To this end, $\boldsymbol{\gamma}$ is partitioned in the same way as $\boldsymbol{\theta}$ as $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top$ where the dimension of $\boldsymbol{\gamma}_1$ is the same as that of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_2$ has n components.

From the estimate $\hat{\boldsymbol{\theta}}_k$ we obtain initial estimates $\boldsymbol{\gamma}_1^{(0)}$ and $\boldsymbol{\gamma}_2^{(0)}$ for starting the inner iteration (by using the sub-vectors of $\hat{\boldsymbol{\theta}}_k$ corresponding to $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$, respectively). During the $(l+1)$th iteration of the inner iteration first an estimate $\boldsymbol{\gamma}_1^{(l+1)}$ of $\boldsymbol{\gamma}_1$ is computed by solving (3.37) for $\boldsymbol{\gamma}_1$. In this step, we set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(l)} = \left((\boldsymbol{\gamma}_1^{(l)})^\top, (\boldsymbol{\gamma}_2^{(l)})^\top\right)^\top$ and $\boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_2^{(l)}$ in (3.37), where $\boldsymbol{\gamma}_1^{(l)}$ and $\boldsymbol{\gamma}_2^{(l)}$ are estimates from the lth inner iteration. Subsequently, an update $\boldsymbol{\gamma}_2^{(l+1)}$ is calculated by using $\left((\boldsymbol{\gamma}_1^{(l)})^\top, (\boldsymbol{\gamma}_2^{(l)})^\top\right)^\top$ as a surrogate for $\boldsymbol{\theta}$ in the BLP formula (3.50). When the inner iteration has converged after, say, l^* iterations, an update $\hat{\boldsymbol{\theta}}_{k+1}$ of the estimate in the outer loop is given by $\hat{\boldsymbol{\theta}}_{k+1} = \left((\boldsymbol{\gamma}_1^{(l^*)})^\top, (\boldsymbol{\gamma}_2^{(l^*)})^\top\right)^\top$. After convergence of the outer loop we obtain an estimate of $\boldsymbol{\beta}$ and a prediction of $\boldsymbol{\epsilon}$. The detailed steps of the algorithm are now outlined as follows:

1. Input: Responses $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, design matrix $\boldsymbol{X}$.
2. Initial estimate $\hat{\boldsymbol{\theta}}_0$ for outer iteration:
   (a) Calculate estimate $\hat{\boldsymbol{\beta}}_0$ of $\boldsymbol{\beta}$ from the GLM without random effects and linear predictor $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$.
   (b) Set $\hat{\boldsymbol{\epsilon}}_0 = (\boldsymbol{y} - \hat{\boldsymbol{\mu}})\dfrac{\partial \hat{\boldsymbol{\eta}}}{\partial \hat{\boldsymbol{\mu}}}$ (McCullagh and Nelder, 1989, p.40) which are residuals obtained from the GLM fit.
   (c) Set $\hat{\boldsymbol{\theta}}_0 = (\hat{\boldsymbol{\beta}}_0^\top, \hat{\boldsymbol{\epsilon}}_0^\top)^\top$.
3. Start outer iteration: Set $k = 1$
4. Computation of $\tilde{h}$: Calculate $\tilde{h}(\pi, \hat{\boldsymbol{\theta}}_{k-1})$ for all $\pi \in S_n$.
5. Initial estimate $\hat{\boldsymbol{\gamma}}^{(0)} = ((\hat{\boldsymbol{\gamma}}_1^{(0)})^\top, (\hat{\boldsymbol{\gamma}}_2^{(0)})^\top)^\top$ for inner iteration: Set $\hat{\boldsymbol{\gamma}}^{(0)} = \hat{\boldsymbol{\theta}}_{k-1}$.
6. Start inner iteration: Set $l = 1$.
7. Preliminary calculations for WLS (3.37): Calculate

$$\sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_{k-1}) \boldsymbol{W}_{\pi, \boldsymbol{\gamma}}$$

and

$$\sum_{\pi \in S_n} \tilde{h}(\pi, \hat{\boldsymbol{\theta}}_{k-1}) \boldsymbol{W}_{\pi,\boldsymbol{\gamma}} (\boldsymbol{z}_{\pi,\boldsymbol{\gamma}} - \boldsymbol{Z}_{\pi} \boldsymbol{\gamma}_2)$$

using $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}^{(l-1)}$ and $\hat{\boldsymbol{\gamma}}_2 = \hat{\boldsymbol{\gamma}}_2^{(l-1)}$.

8. Update estimate of $\boldsymbol{\gamma}_1$: Solve (3.37) for $\boldsymbol{\gamma}_1$ and use the solution as the new estimate $\hat{\boldsymbol{\gamma}}_1^{(l)}$.
9. Update estimate of $\boldsymbol{\gamma}_2$: Calculate $\hat{\boldsymbol{\gamma}}_2^{(l)}$ using (3.50) substituting $\boldsymbol{\gamma}_2$ for $\boldsymbol{\epsilon}$.
10. New estimate from inner iteration: Set $\hat{\boldsymbol{\gamma}}^{(l)} = ((\hat{\boldsymbol{\gamma}}_1^{(l)})^\top, (\hat{\boldsymbol{\gamma}}_2^{(l)})^\top)^\top$.
11. Test of convergence of $\hat{\boldsymbol{\gamma}}_1$ for inner iteration: If $||\hat{\boldsymbol{\gamma}}_1^{(l)} - \hat{\boldsymbol{\gamma}}_1^{(l-1)}|| < \tilde{\delta}$ where $\tilde{\delta}$ is a very small positive quantity, terminate inner iteration. Otherwise set $l = l + 1$ and go to step 7.
12. New estimate from outer iteration: Set $\hat{\boldsymbol{\theta}}_k = (\hat{\boldsymbol{\beta}}_k^\top, \hat{\boldsymbol{\epsilon}}_k^\top)^\top = \hat{\boldsymbol{\gamma}}^{(l)}$.
13. Test of convergence of $\hat{\boldsymbol{\beta}}$ for outer iteration: If $||\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1}|| < \tilde{\delta}$, terminate outer iteration. Otherwise set $k = k + 1$ and go to step 4.
14. Output: The estimate $\hat{\boldsymbol{\theta}}_k = (\hat{\boldsymbol{\beta}}_k^\top, \hat{\boldsymbol{\epsilon}}_k^\top)^\top$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\epsilon}^\top)^\top$.

The steps of the algorithm are given in Figure 3.2 via a flowchart.

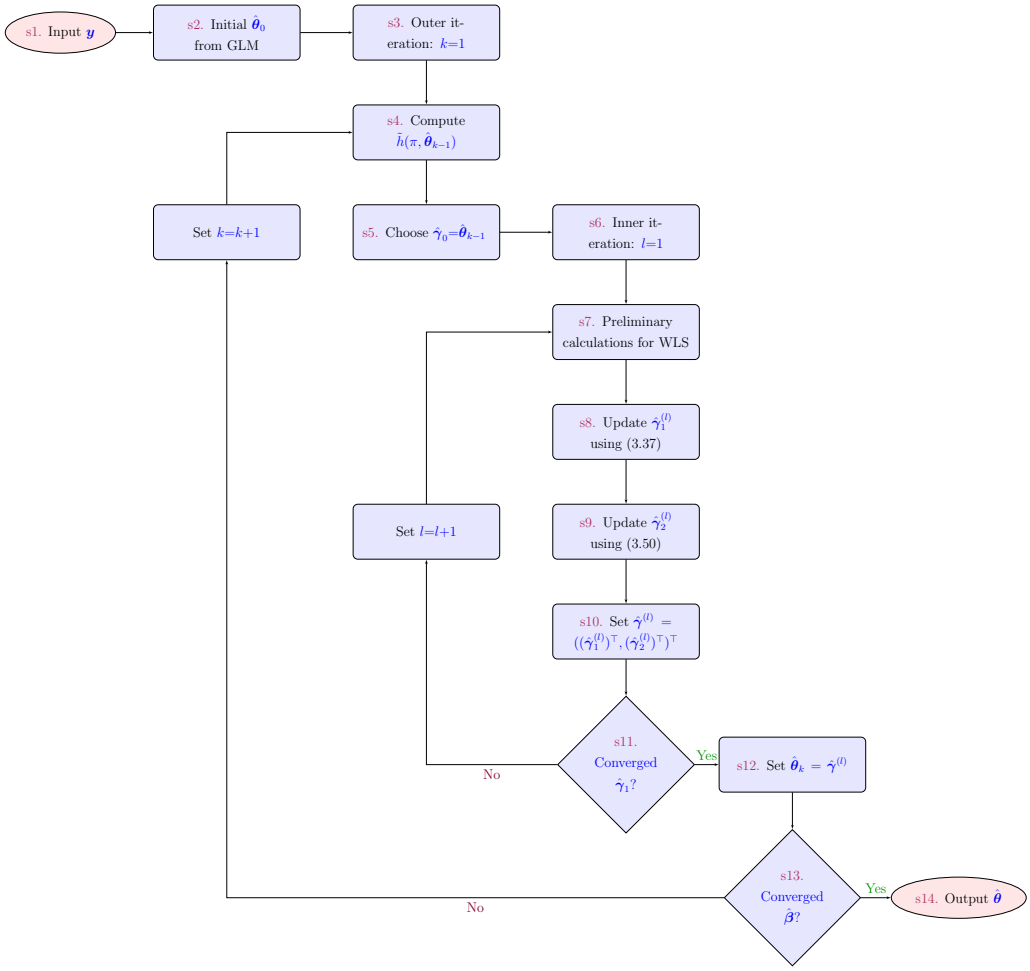

Figure 3.2: Flowchart of the randomization-based algorithm.

3.7 Simulation Study

We assess the estimated parameters for the RB-GLMM (2.26) (equivalently model (2.12)) for the CRD via simulation studies by using the randomization-based algorithm. To this end, we generated Poisson responses from the RB-GLMM and run the algorithm for estimating the model parameters on the simulated data. We then compare the results obtained from the randomization-based algorithm with results obtained from fitting a Poisson GLM and a HGLM (hierarchical generalized linear model) to the same set of simulated data. We use the R-package **hglm** for fitting the HGLM which is based on the h-likelihood method of estimation (Lee and Nelder, 1996, Lee and Lee, 2012) and suitable for correlated random effects in GLMMs (Rönnegård et al., 2010). When fitting the HGLM, we assume correlated and normally distributed ran-

dom effects. In Section 3.7.1, we describe the simulation settings for different examples for the CRD including unit contributions, treatment contributions and true parameter values. Section 3.7.2 describes the generation of the data from the RB-GLMM for the CRD. We present and compare the simulation results in Section 3.7.3. Moreover, Section 3.7.4 summarizes the findings of the simulation studies.

3.7.1 Simulation Settings

We consider several examples for simulating data from the randomization-based model (2.12). We first specify the number of treatments (t) and the number of replications of each treatment (r) allocated to the experimental units according to a CRD. It follows that in total there are $n = rt$ experimental units. We then specify the i-th unit contributions (u_i) and the j-th treatment contributions (v_j) in a fixed design (see model (2.1)).

For Examples 1-3, we set $t = 2$, $v_1 = 1.75$ and $v_2 = 1.25$ where the values of the v_i's have been chosen arbitrarily. We then randomly generated unit contributions u_i, for $i = 1, 2, \ldots, n$, from the t-distribution with three degrees of freedom (t_3) for Examples 1 and 2, and from the standard normal distribution $N(0, 1)$ for Example 3. Following the notation in equation (2.2), we recode the u_i and v_j to specify true values for the overall mean μ, the fixed treatment effects parameters α_j, for $j = 1, 2 \ldots, t$, and the unit errors e_i, for $i = 1, 2 \ldots, n$, with variance component σ^2. The u_i and e_i values for the different examples are given in Table 3.2.

Table 3.2: Unit contributions used in examples.

Examples	t	r	u_i and e_i values for $i = 1, 2 \ldots, n$	Distribution
Example 1	2	3	u_i: 0.790, -2.064, -1.345, 0.247, 2.987, 0.075	t_3
			e_i: 0.675, -2.179, -1.460, 0.132, 2.872, -0.040	-
Example 2			u_i: 1.176, 0.429, -0.058, 0.836, 1.261, 2.843	t_3
			e_i: 0.095, -0.652, -1.140, -0.245, 0.180, 1.762	-
Example 3			u_i: -0.626, 0.184, -0.836, 1.595, 0.330, -0.820	$N(0, 1)$
			e_i: -0.597, 0.213, -0.807, 1.624, 0.359, -0.791	-

For estimating the model parameters we use the constraint $\sum_{j=1}^{t} \alpha_j = 0$ and estimate $(t - 1)$ fixed treatment effect parameters $\alpha_1, \ldots, \alpha_{t-1}$ and μ. True

parameter values for different examples are given in Table 3.3.

3.7.2 Data Generation

Recall that the design matrix $\boldsymbol{X}$ is overparameterized and we use the same $\boldsymbol{X}$ and $\boldsymbol{\beta}$ for simulating the data which are defined in Section 2.3. We simulated 1000 Poisson data sets of size $n = 6$ from the RB-GLMM for the CRD (2.26). As can be seen from Table 3.2, the u_i and e_i values are the same in 1000 simulations. This is because we generated u_i values only once in the simulation. We used the linear predictor $\boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where, in each simulation, $\boldsymbol{X\beta}$ is fixed while $\boldsymbol{\epsilon}$ is formed by randomly permuting the e_i values given in Table 3.2. More precisely, in each simulation $\boldsymbol{\epsilon} = (e_{\pi(1)}, \ldots, e_{\pi(6)})^\top$ for some randomly chosen $\pi \in S_6$. As a result, the data are generated from the Poisson distribution using many different linear predictors based on the randomization approach.

3.7.3 Simulation Results

In this section, we present simulation results in order to assess and compare the estimates and their standard errors. For implementing the model estimation, we run the randomization-based algorithm where we do not assume any distribution for the random effects. In addition, we fitted a Poisson GLM and a HGLM to the same set of simulated data. The GLM and HGLM are considered as misspecified models in the sense that there are no random effects in the GLM framework, and the random effects of the HGLM fitting are assumed to follow the normal distribution. For each combination of the true parameters, the data are generated from the RB-GLMM for the CRD.

We summarize the results by averaging over all data sets where the estimation algorithms converged, except the HGLM in Example 2. In Example 2, we excluded the outputs for the data set 319 before summarizing the results. This is because we observed an unusual outlier for the estimated variance component $\hat{\sigma}^2$ ($\hat{\sigma}^2 = 1160.5370066$), when the true value of σ^2 was 0.822. In order to compare the estimated model parameters, we present a summary of the simulation results in Table 3.3.

From Table 3.3, it can be seen that the GLM and the randomization-based

algorithm converged in 100% of all cases for each example. The convergence for the HGLM estimation algorithm varies slightly between the examples. The HGLM estimation algorithm converged for 981 (98.1%), 988 (98.8%) and 991 (99.1%) out of 1000 simulated data sets for Examples 1-3 respectively. In these examples, the convergence is higher for the randomization-based algorithm than for the HGLM estimation algorithm.

Table 3.3: True parameters, estimates, empirical standard errors (SE), mean square error of prediction (MSEP) and convergence under different models.

Example 1: True $\mu = 1.615$, $\alpha_1 = 0.25$ and $\sigma^2 = 2.60$.								
Models	$\hat{\mu}$	$SE(\hat{\mu})$	$\hat{\alpha}_1$	$SE(\hat{\alpha}_1)$	$\hat{\sigma}^2$	$SE(\hat{\sigma}^2)$	$MSEP$	Converged
GLM	2.429	0.229	0.236	1.087	–	–	–	100%
HGLM	1.454	0.339	0.271	0.851	4.313	3.374	41.763	98.1%
RB-GLMM	1.970	**0.210**	**0.247**	**0.428**	**1.246**	**0.515**	**5.275**	100%
Example 2: True $\mu = 2.581$, $\alpha_1 = 0.25$ and $\sigma^2 = 0.822$.								
Models	$\hat{\mu}$	$SE(\hat{\mu})$	$\hat{\alpha}_1$	$SE(\hat{\alpha}_1)$	$\hat{\sigma}^2$	$SE(\hat{\sigma}^2)$	$MSEP$	Converged
GLM	2.891	**0.125**	**0.263**	0.610	–	–	–	100%
HGLM	2.529	0.155	0.270	0.460	1.292	0.955	12.806	98.8%
RB-GLMM	2.666	**0.125**	0.224	**0.250**	**0.562**	**0.199**	**0.905**	100%
Example 3: True $\mu = 1.971$, $\alpha_1 = 0.25$ and $\sigma^2 = 0.741$.								
Models	$\hat{\mu}$	$SE(\hat{\mu})$	$\hat{\alpha}_1$	$SE(\hat{\alpha}_1)$	$\hat{\sigma}^2$	$SE(\hat{\sigma}^2)$	$MSEP$	Converged
GLM	2.241	0.170	**0.285**	0.574	–	–	–	100%
HGLM	1.887	0.222	0.291	0.482	1.192	0.973	11.473	99.1%
RB-GLMM	2.075	**0.166**	0.205	**0.320**	**0.436**	**0.225**	**1.109**	100%

From Table 3.3, it can also be seen that the randomization-based estimated overall mean effect ($\hat{\mu}$) is less biased than the corresponding GLM estimates in each example. In each example, both the randomization-based and the GLM estimates of μ are overestimated whereas the HGLM estimates are underestimated. The empirical standard errors (SE) of $\hat{\mu}$ are substantially smaller for the randomization-based estimates than for the GLM and for the HGLM estimates in each example, except for the GLM in Example 2. In Example 2, the SE of $\hat{\mu}$ (0.125) for the GLM is the same as for the RB-GLMM.

In Example 1, the randomization-based estimated treatment effect parameter $\hat{\alpha}_1$ is very close to the true α_1 i.e., less biased ($\alpha_1 = 0.25$ and $\hat{\alpha}_1 = 0.247$) than the corresponding estimates obtained from the GLM ($\hat{\alpha}_1 = 0.236$) and the HGLM ($\hat{\alpha}_1 = 0.271$) estimation algorithm. However, in Examples 2 and 3, the randomization-based $\hat{\alpha}_1$ is underestimated (0.224 and 0.205) whereas the

corresponding estimates for the GLM (0.263 and 0.285) and the HGLM (0.270 and 0.291) are overestimated. Moreover, the randomization-based $\hat{\alpha}_1$ is also more precise as its SE is substantially smaller than the GLM and the HGLM estimates in each example.

The estimated variance component $\hat{\sigma}^2$ obtained from the randomization-based algorithm is also less biased in each example ($\sigma^2 = 2.60$, $\hat{\sigma}^2 = 1.246$ for Example 1; $\sigma^2 = 0.822$, $\hat{\sigma}^2 = 0.562$ for Example 2; and $\sigma^2 = 0.741$, $\hat{\sigma}^2 = 0.436$ for Example 3) than the corresponding HGLM estimates ($\hat{\sigma}^2 = 4.313$ for Example 1, $\hat{\sigma}^2 = 1.292$ for Example 2 and $\hat{\sigma}^2 = 1.192$ for Example 3). Moreover, the randomization-based estimate of σ^2 is underestimated whereas the corresponding HGLM estimate is overestimated in each example. It is also the case that the empirical standard errors of $\hat{\sigma}^2$ are substantially smaller for the randomization-based algorithm than for the HGLM estimation algorithm in each example.

Table 3.3 also shows that the MSEP values for predicting the random effects, calculated by (3.39), are substantially smaller for the RB-GLMM (5.275, 0.905 and 1.109 for Examples 1-3 respectively) than the corresponding values for the HGLM (41.775, 12.806 and 11.473 for Examples 1-3 respectively).

3.7.4 Conclusion

From the simulation results it can be concluded that, for the examples considered, the randomization-based estimated fixed effects parameters ($\hat{\mu}$ and $\hat{\alpha}_1$) are less biased and more precise in the Poisson GLMM than the corresponding GLM estimates in most cases. This is due to the fact that there are no random effects in the GLM framework, whereas the data used in the simulation were generated with random effects in addition to fixed effects from the RB-GLMM.

The estimated fixed effects parameters obtained from both the randomization-based and the HGLM estimation algorithm in Poisson GLMMs are very close to true values in most cases. However, the randomization-based estimated fixed effects parameters are more precise than the HGLM estimates. Moreover, the randomization-based algorithm gives better and more precise estimates of the variance component in Poisson GLMMs than the HGLM estimation algorithm

when there was misspecification of the random effects distribution. A possible explanation is that the randomization-based algorithm does not make any assumption about the random effects distribution, while normality is assumed in the HGLM estimation algorithm.

3.8 An Example with Real Data

The Poisson GLMM with individual level random effects (number of observations is equal to the number of elements of the vector of the random effects), equivalently the RB-GLMM for the CRD (2.26), can be applied to model overdispersed count data (Rabe-Hesketh and Skrondal, 2012, p.706). For details on overdispersion see Chapter 7 of Hilbe (2011). It has shown in Chapters 8 and 9 by Hilbe (2011) that a negative binomial (NB) regression can also be applied as an alternative to a Poisson regression for analyzing the overdispersed count data. The IWLS estimation algorithm is also used in the negative binomial GLM as in standard GLMs.

Overdispersion can be identified by using the value of the Pearson χ^2-statistic divided by the degrees of freedom (df). We call this value the *dispersion*. If this is greater than 1, then the model is overdispersed (Hilbe, 2011, p.142). The Pearson χ^2-statistic and residuals [Hilbe (2011, p.62) and Stroup (2012, p. 343)] are defined by

$$\text{Pearson } \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i^*)^2}{V(\mu_i^*)} \tag{3.57}$$

and

$$\text{residuals} = \frac{y_i - \mu_i^*}{\sqrt{V(\mu_i^*)}} \tag{3.58}$$

respectively, where μ_i^* is the expected counts and V is the variance function. The function V is equal to μ_i^* and $\mu_i^* + \kappa\mu_i^{*2}$ for the Poisson and NB regression models respectively where κ is the dispersion parameter (Hilbe, 2011, p.63). When $\kappa \to 0$ the negative binomial model becomes Poisson (Hilbe, 2011, p.221). We consider an overdispersed count data set for fitting the different models, the details of which are described in the next sections.

3.8.1 Data and Variables

We use the Medpar data which stands for the US national Medicare inpatient hospital database. This data set is organized yearly from the hospital records and prepared for each state. The details can be found in Hilbe (2011, p.100) and the R-package **COUNT**. There are 115 variables in the full Medpar data set and 14 million records, with one record for each hospitalization.

In the *medpar* file, the data come from 1991 Medicare files for the state of Arizona and only one diagnostic group (DRG 112) is considered. The data have been selected at random from the original data file and there are 1495 observations in the *medpar* data file with 10 variables. The data were downloaded from the website $http://www.cambridge.org/9780521857727$ of Cambridge University Press.

The count response variable is the *los*: Length-of-Stay (Days in the Hospital) and we only consider the covariate *hmo* (Health Maintenance Organization). The variable *hmo* is binary and coded as $hmo = 1$ if the patient belongs to a Health Maintenance Organization, 0 otherwise.

3.8.2 Analysis

In this section, we first fitted a Poisson GLM to the full *medpar* data set consisting of 1495 observations from 54 hospitals in order to investigate whether overdispersion is present or not. The summary of the results is given in Table 3.4.

Table 3.4: Results obtained by fitting Poisson GLM to the full *medpar* data with $n = 1495$.

Variable	*Estimate*	*SE*	*p-value*	*Dispersion*
Constant	2.310	0.009	0.000	7.821
hmo	-0.150	0.024	0.000	

From Table 3.4 it can be seen that the Pearson dispersion value is 7.821 which shows overdispersion in the data. Moreover, the *hmo* is found to be highly significant (p-value < 0.01). Similar results were also found by Hilbe (2011) in

Chapter 6. To analyze this data set, Hilbe (2011) again fitted the NB regression and showed in Chapter 8 that NB regression fits better than the Poisson GLM in the presence of overdispersion.

However, we now consider a subset of the *medpar* data in order to fit the Poisson RB-GLMM. In the full *medpar* data set, of the 54 hospitals, there were 40 hospitals that each had more than 6 observations. We selected one hospital from these 40 hospitals at random and found 74 observations in the selected hospital. We then selected 6 observations at random from this hospital and fitted different models to the selected subset of *medpar* data (*medpar_small*). The summary of results obtained from fitting different models to the *medpar_small* data set is given in Table 3.5.

Table 3.5 shows that the Pearson dispersion value for the Poisson GLM is 2.457 which shows the presence of overdispersion in the selected subset of the *medpar* data. We then fitted the NB regression model to the same subset of the data and observed that the Pearson dispersion is 1.351. It can be concluded that there is still some overdispersion as the dispersion value is greater than 1.

Table 3.5: Results obtained from fitting different models to *medpar_small* data with $n = 6$.

Poisson GLM				
Variable	*Estimate*	*SE*	*p-value*	*Dispersion*
Constant	1.833	0.200	0.000	2.457
hmo	-0.734	0.455	0.106	
Negative Binomial GLM				
Variable	*Estimate*	*SE*	*p-value*	*Dispersion*
Constant	1.833	0.282	0.000	1.351
hmo	-0.734	0.571	0.198	
Poisson HGLM				
Variable	*Estimate*	*SE*	*p-value*	*Dispersion*
Constant	1.705	0.282	0.104	0.387
hmo	-0.743	0.731	0.495	
Poisson RB-GLMM				
Variable	*Estimate*	*SE*	*p-value*	*Dispersion*
Constant	1.623	0.250	0.000	0.677
hmo	-0.234	0.856	0.785	

In addition, we fitted a Poisson HGLM with correlated and normally distributed random effects, and the RB-GLMM where the assumption of normality is not required. In the RB-GLMM, we compute the Hessian (second derivative of the log-likelihood) matrix by using the R-package **numDeriv** and hence obtain the variance-covariance matrix of the estimated parameters in order to calculate standard errors of the estimates. The estimated variance component $\hat{\sigma}^2$ is calculated by Lemma 3(ii) in Section 2.1.5 and found to be 0.328 and 0.148 for the HGLM and the RB-GLMM respectively.

It has shown that standard errors of the estimates may be underestimated because of the overdispersion in the data (Hilbe, 2011, p.141). From Table 3.5 it can be observed that this is the case in the Poisson GLM compared with the other three models.

The Pearson dispersion values are found to be less than 1 in both the HGLM (0.387) and RB-GLMM (0.677) and suggest that the Poisson GLMM with individual level random effects may be a good choice for analyzing the overdispersed count data. However, the main goal is to show an application of the RB-GLMM to real data set. Moreover, we compare the observed counts with

predicted values under different models in Table 3.6 and Figure 3.3.

Table 3.6: Observed vs Predicted days in Hospital under different models.

No.	Observed	Predicted			
		Poisson-GLM	NB-GLM	HGLM	RB-GLMM
1	10	6	6	9	9
2	1	3	3	2	3
3	9	6	6	8	8
4	2	6	6	3	3
5	5	3	3	4	3
6	4	6	6	4	4

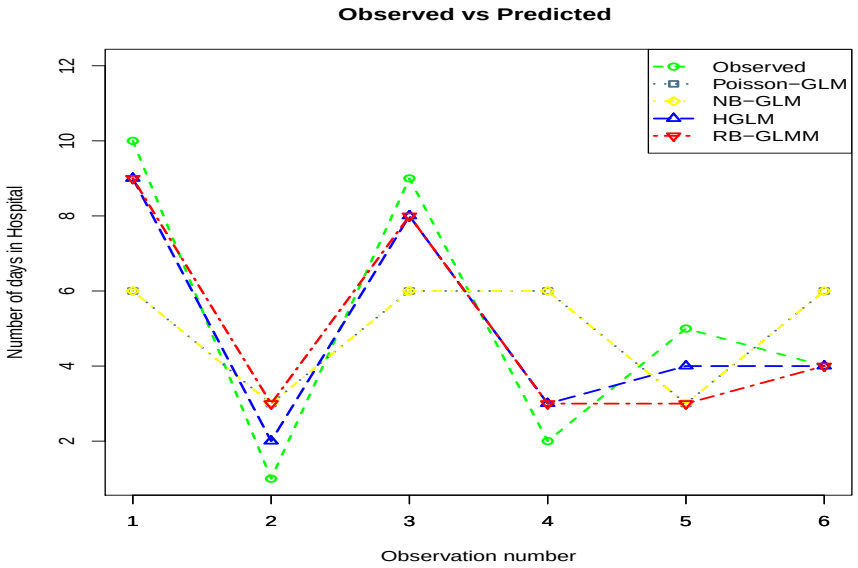

Figure 3.3: Observed vs Predicted days in Hospital

It can be seen from Table 3.6 and Figure 3.3 that the predicted number of days in the hospital are very close to the corresponding observed days for both the HGLM and the RB-GLMM. However, this is not the case in the Poisson and NB regressions.

3.9 Computational Limitations

The results obtained from the randomization-based algorithm in Sections 3.7 and 3.8 are promising. However, there are some computational limitations for

applicability of the randomization-based algorithm. Here we discuss some of the challenges for implementing the algorithm. The number of all possible permutations ($n!$) in the symmetric group S_n increases rapidly with the increase in sample size n and so does both the memory and computational time needed by the randomization-based estimation algorithm. The algorithm then becomes computationally infeasible for $n > 10$ on standard personal computers (PCs) such as 8-core PCs.

To overcome this difficulty we deal with the required computations by using batch-wise permutations instead of a single permutation. We split all the permutations into batches of suitable size (based on the number of available cores) so that each one of them can be stored and processed into the available memory. The results are then aggregated when all batch-wise computations are completed. However, another challenge in handling the large number of batches is that it sometimes exceeds the available memory due to garbage collection failure. We enforce the memory deallocation prior to invoking another batch and avoid exceeding the memory limit. We are then able to handle the memory problem by deleting the collected garbage stored from previous computations.

The next challenge to deal with is the required computation time. We approach the problem by processing a number of batches in parallel on different cores. Using the High Performance Computing (HPC) services at Queen Mary University of London, we managed to implement the randomization-based estimation algorithm for up to $n = 12$ ($12! = 479001600$ permutations) with 48 cores (12 CPUs-central processing units with 4 cores each). However, in the HPC facility, it took approximately 32 hours to finish the processing of the algorithm for a single data set of size $n = 12$.

As a result, to implement the algorithm for larger samples, we plan to use a random subset of permutations instead of all possible permutations from S_n. We will then compute all the relevant steps of the algorithm for the randomly selected subset of permutations and summarize the results based on this subset of permutations.

Part II

Misspecification of Correlation and Distribution of Random Effects in GLMMs

Chapter 4

Derivation of the RB-GLMM for the RCBD with Random Block Effects

In the second part of my project, we investigate the impact of misspecification of the correlation structure and of the random effects distribution via an extensive simulation study (see Chapter 5). To this end, we first derive a RB-GLMM for the RCBD with random block effects. A model with random block effects can be derived by randomizing the blocks in addition to the units (Section 1.4.2). The symmetric groups S_b and S_t are used for the randomization of b blocks and t treatments within each block, respectively. In the context of linear models Brien and Bailey (2006) represent the corresponding randomization by elements of the *wreath product* $S_t \wr S_b$ of S_t and S_b (Section A.1.3). Here we generalize this approach to GLMMs. The generalization follows a pattern similar to the derivation of the model for the CRD in Section 2.1.

We begin in Section 4.1 by formulating a model within the GLM framework. In Section 4.2, we introduce some notation and random variables which are used in the derivation of the model. To obtain the GLMM for the RCBD, we derive the conditional expectation of the responses given the vector of the random effects. This is the content of Section 4.3. This derivation is similar to the derivation of the conditional expectation in the model for the CRD in Section 2.1.3. However, in the RCBD setting we have two random effects, one for random blocks and the other is associated with random errors for the

experimental units. In Section 4.4, we apply the link function to the derived conditional expectation and finally obtain the RB-GLMM for the RCBD with random block effects, which we call Model 2.

We then derive moments of the random effects, random blocks and random errors, in Section 4.5. It is shown that the random effects are correlated and that the random blocks and random unit errors are uncorrelated. In Section 4.6, we summarize the derived results and present Model 2 in matrix notation with variance-covariance matrices for random blocks and unit errors derived from the randomization. The derivations are similar to those in Section 2.1 but are complicated by the more complex type of design.

4.1 GLM for the RCBD

For $i = 1, \ldots, t$, $j = 1, \ldots, t$ and $k = 1, \ldots, b$ let the random variable $X_{i,j,k}$ represent the response for the i-th unit, j-th treatment and k-th block whose distribution is assumed to be from the exponential family. Formally $X_{i,j,k}$ is a real random variable as before on a probability space $(\Omega, \mathscr{F}, P)$ (see appendix Section A.2.1). We adapt the assumption of unit-treatment additivity (Cox, 2009, Kempthorne, 1955) for $g(E(X_{i,j,k}))$ where $E(X_{i,j,k})$ is the expectation of $X_{i,j,k}$ and g is the link function as before. More precisely, adapting the approach of Kempthorne (1955) we assume that

$$g(E(X_{i,j,k})) = u_{i,k} + v_j, \tag{4.1}$$

where $u_{i,k}$ and v_j denote fixed contributions from the i-th unit of the k-th block and the j-th treatment, respectively. We also assume that for every k the values $u_{i,k}$ are all different. Write $\bar{u} = \frac{1}{bt}\sum_{i=1}^{t}\sum_{k=1}^{b} u_{i,k}$, $\bar{v} = \frac{1}{t}\sum_{j=1}^{t} v_j$ and $\bar{u}_k = \frac{1}{t}\sum_{i=1}^{t} u_{i,k}$. Letting $\mu = \bar{u} + \bar{v}$, $\alpha_j = v_j - \bar{v}$, $\beta_k = \bar{u}_k - \bar{u}$ and $e_{i,k} = u_{i,k} - \bar{u}_k$, equation (4.1) can be rewritten as

$$g(E(X_{i,j,k})) = \mu + \alpha_j + \beta_k + e_{i,k}, \tag{4.2}$$

where μ is the grand mean, α_j is the j-th treatment effect, β_k is the k-th block effect and the unit errors $e_{1,k}, \ldots, e_{t,k}$ are assumed to be all different for every $k = 1, \ldots, b$. We consider the treatment j as being fixed for every block k by

the specific design $d = (j_1, \ldots, j_t)$ and assume that

(i) for every block k the random variables $X_{1,j_1,k}, \ldots, X_{t,j_t,k}$ are independent, in particular, $X_{i,j_i,k}$ and $X_{i',j_{i'},k}$ are independent for $i \neq i'$ which means that in the same block outcomes for the different units are independent;

(ii) for $k \neq k'$ the random variables $X_{i,j_i,k}$ and $X_{i',j_{i'},k'}$ are also independent i.e., outcomes for units in different blocks are independent.

4.2 Notation and Definitions

Let U be the uniform distribution on $S_t \wr S_b$ and $a = t!^b b!$ is the number elements in the wreath product $S_t \wr S_b$. The probability of randomly selecting an element $(\boldsymbol{\pi}, \delta) = (\pi_1, \ldots, \pi_b, \delta)$ from $S_t \wr S_b$ is $U(\{(\boldsymbol{\pi}, \delta)\}) = 1/a$. For $i = 1, \ldots, t$ and $k = 1, \ldots, b$ we define $\tilde{\epsilon}_{i,k} : S_t \wr S_b \to \mathbb{R}$ by

$$\tilde{\epsilon}_{i,k}(\boldsymbol{\pi}, \delta) = \sum_{(\boldsymbol{\sigma},\tau) \in S_t \wr S_b} e_{\sigma_{\tau(k)}(i),\tau(k)} \; 1_{\{(\boldsymbol{\sigma},\tau)\}}(\boldsymbol{\pi}, \delta) = e_{\pi_{\delta(k)}(i),\delta(k)} \tag{4.3}$$

and $\tilde{B}_k : S_t \wr S_b \to \mathbb{R}$ by

$$\tilde{B}_k(\boldsymbol{\pi}, \delta) = \sum_{\tau \in S_b} \beta_{\tau(k)} 1_{\{\tau\}}(\delta) = \beta_{\delta(k)} \tag{4.4}$$

for every $(\boldsymbol{\pi}, \delta) \in S_t \wr S_b$. Here $\tilde{\epsilon}_{i,k}$ and $\tilde{B}_k$ are random variables on the probability space $(S_t \wr S_b, \mathscr{P}(S_t \wr S_b), U)$, where $\mathscr{P}(S_t \wr S_b)$ is the power set of $S_t \wr S_b$. Note that $\tilde{\boldsymbol{B}} = \left(\tilde{B}_1, \ldots, \tilde{B}_b\right)^\top$ is the vector of random variables $\tilde{B}_1, \ldots, \tilde{B}_b$. Further define $\epsilon_{i,k} : S_t \wr S_b \times \Omega \to \mathbb{R}$ by

$$\epsilon_{i,k}((\boldsymbol{\pi}, \delta), \omega) = \tilde{\epsilon}_{i,k}(\boldsymbol{\pi}, \delta) \tag{4.5}$$

and $B_k : S_t \wr S_b \times \Omega \to \mathbb{R}$ with

$$B_k((\boldsymbol{\pi}, \delta), \omega) = \tilde{B}_k(\boldsymbol{\pi}, \delta) \tag{4.6}$$

for every $((\boldsymbol{\pi}, \delta), \omega) \in S_t \wr S_b \times \Omega$. Therefore $\epsilon_{i,k}$ and B_k are random variables on the probability space $(S_t \wr S_b \times \Omega, \mathscr{P}(S_t \wr S_b) \otimes \mathscr{F}, U \otimes P)$, where $S_t \wr S_b \times \Omega$ is the Cartesian product, $\mathscr{P}(S_t \wr S_b) \otimes \mathscr{F}$ is the product σ-field and $U \otimes P$ is

the product measure on $\mathscr{P}(S_t \wr S_b) \otimes \mathscr{F}$. For the fixed design $d = (j_1, \ldots, j_t)$ in each block, and for every $i = 1, \ldots, t$, and $k = 1, \ldots, b$, we define $Y_{i,j_i,k} : S_t \wr S_b \times \Omega \to \mathbb{R}$ by

$$Y_{i,j_i,k}((\boldsymbol{\pi}, \delta), \omega) = X_{\pi_{\delta(k)}(i),j_i,\delta(k)}(\omega) \tag{4.7}$$

for every $((\boldsymbol{\pi}, \delta), \omega) \in S_t \wr S_b \times \Omega$. This is a random variable on the probability space $(S_t \wr S_b \times \Omega, \mathscr{P}(S_t \wr S_b) \otimes \mathscr{F}, U \otimes P)$ which represents the response after the randomization of blocks and units to treatments within block. That is $Y_{i,j_i,k}$ is the response for the unit with label i and treatment j_i in block k.

4.3 Conditional Expectation

Our main goal is to derive from the randomization and equation (4.2) a GLMM in standard form by incorporating two random effects associated with the RCBD. This requires deriving a form similar to the one in equation (1.2). In order to do this, we first derive the conditional expectation of the responses $Y_{i,j_i,k}$ given the vector of two random effects, random blocks ($\boldsymbol{B}$) and random errors ($\boldsymbol{\epsilon}$), by using the same approach as in Section 2.1.3. More precisely, we compute $E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) = E(Y_{i,j_i,k}|\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B}))$, where $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$ is the σ-field generated by the vector $(\boldsymbol{\epsilon}, \boldsymbol{B}) = (\epsilon_{1,1}, \ldots, \epsilon_{t,1}, \ldots, \epsilon_{1,b}, \ldots, \epsilon_{t,b}, B_1, \ldots, B_b)^\top$ corresponding to the vector of random errors $\boldsymbol{\epsilon} = (\epsilon_{1,1}, \ldots, \epsilon_{t,1}, \ldots, \epsilon_{1,b}, \ldots, \epsilon_{t,b})^\top$ and the vector of random blocks $\boldsymbol{B} = (B_1, \ldots, B_b)^\top$.

For every component $\epsilon_{i,k}$ and B_k of $(\boldsymbol{\epsilon}, \boldsymbol{B})$ there exist the σ-fields generated by $\epsilon_{i,k}$ and B_k which we call $\mathscr{F}(\epsilon_{i,k})$ and $\mathscr{F}(B_k)$, respectively. Moreover, $\mathscr{F}(\epsilon_{i,k})$ and $\mathscr{F}(B_k)$ consist of the sets

$$\epsilon_{i,k}^{-1}(B) = \{((\boldsymbol{\pi}, \delta), \omega) \in S_t \wr S_b \times \Omega : \epsilon_{i,k}((\boldsymbol{\pi}, \delta), \omega) \in B\}$$

and

$$B_k^{-1}(B) = \{((\boldsymbol{\pi}, \delta), \omega) \in S_t \wr S_b \times \Omega : B_k((\boldsymbol{\pi}, \delta), \omega) \in B\},$$

respectively, where B is a Borel set. It follows that $\mathscr{F}(\epsilon_{i,k}) = \{\epsilon_{i,k}^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}$ and $\mathscr{F}(B_k) = \{B_k^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}$, where $\mathscr{B}(\mathbb{R})$ is the Borel σ-field. In addition, there is the σ-field generated by the vector $(\boldsymbol{\epsilon}, \boldsymbol{B})$ of random vari-

ables, for which we write $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$, and which is defined as the σ-field generated by the union of $\mathscr{F}(\epsilon_{1,1}), \ldots \mathscr{F}(\epsilon_{t,1}), \ldots \mathscr{F}(\epsilon_{1,b}), \ldots, \mathscr{F}(\epsilon_{t,b}), \mathscr{F}(B_1), \ldots, \mathscr{F}(B_b)$ i.e., $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$ is the σ-field generated by $\left(\bigcup_{k=1}^{b}\bigcup_{i=1}^{t}\mathscr{F}(\epsilon_{i,k})\right)\bigcup\left(\bigcup_{k=1}^{b}\mathscr{F}(B_k)\right)$. Equivalently, $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$ is the smallest σ-field (Section A.2.4) which contains the union $\left(\bigcup_{k=1}^{b}\bigcup_{i=1}^{t}\mathscr{F}(\epsilon_{i,k})\right)\bigcup\left(\bigcup_{k=1}^{b}\mathscr{F}(B_k)\right)$.

Now we wish to find the conditional expectation $E(Y_{i,j_i,k}|\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B}))$. The following results, Lemma 13 and Corollary 4, will help us to derive that $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$ is the σ-field generated by the partition $\{\{(\boldsymbol{\pi}, \delta)\}\times\Omega : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\}$. This allows us to use Theorem 2 (Section A.2.6) to derive the conditional expectation $E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B})$. The derivations of these results are similar to the derivations in Section 2.1.3. The details of these derivations are given below.

Lemma 13. *Let $\epsilon_{i,k}$ and B_k be the components of the random vector $(\boldsymbol{\epsilon}, \boldsymbol{B})$ defined by (4.5) and (4.6), respectively. The σ-field $\{A \times \Omega : A \subseteq S_t \wr S_b\}$ generated by the set $\{\{(\boldsymbol{\pi}, \delta)\} \times \Omega : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\}$ is a subset of $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$ i.e., $\{A \times \Omega : A \subseteq S_t \wr S_b\} \subseteq \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$.*

Proof. We need to show that for every $(\boldsymbol{\pi}, \delta) \in S_t \wr S_b$ the set $\{(\boldsymbol{\pi}, \delta)\} \times \Omega \in \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$. If we can show this, then we can write $\{\{(\boldsymbol{\pi}, \delta)\} \times \Omega : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\} \subseteq \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$. Therefore, the σ-field generated by the set $\{\{(\boldsymbol{\pi}, \delta)\} \times \Omega : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\}$ is a subset of $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$. Applying part (iii) of Lemma 19 of the appendix with $S = S_t \wr S_b$ and $s = (\boldsymbol{\pi}, \delta)$ it can be seen that $\{A \times \Omega : A \subseteq S_t \wr S_b\}$ is the σ-field generated by the set $\{\{(\boldsymbol{\pi}, \delta)\} \times \Omega : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\}$ and it follows that $\{A \times \Omega : A \subseteq S_t \wr S_b\} \subseteq \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$. Now let $(\boldsymbol{\pi}, \delta) \in S_t \wr S_b$ be fixed. Then by definition and using (4.5) and (4.3) we have $\epsilon_{i,k}((\boldsymbol{\pi}, \delta), \omega) = \tilde{\epsilon}_{i,k}((\boldsymbol{\pi}, \delta)) = e_{\pi_{\delta(k)}(i),\delta(k)}$. Moreover, by using (4.6) and (4.4) we can write $B_k((\boldsymbol{\pi}, \delta), \omega) = \tilde{B}_k((\boldsymbol{\pi}, \delta)) = \beta_{\delta(k)}$. It follows that

$$\begin{aligned} & \epsilon_{i,k}^{-1}(\{e_{\pi_{\delta(k)}(i),\delta(k)}\}) \\ = \; & \left\{((\boldsymbol{\sigma}, \tau), \omega) \in S_t \wr S_b \times \Omega : \epsilon_{i,k}((\boldsymbol{\sigma}, \tau), \omega) = e_{\pi_{\delta(k)}(i),\delta(k)}\right\} \in \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B}) \end{aligned}$$

and

$$B_k^{-1}(\{\beta_{\delta(k)}\}) = \left\{((\boldsymbol{\sigma}, \tau), \omega) \in S_t \wr S_b \times \Omega : B_k((\boldsymbol{\sigma}, \tau), \omega) = \beta_{\delta(k)}\right\} \in \mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$$

since $\epsilon_{i,k}$ and B_k are measurable. It also follows that by using the intersection

property (A.1) of the σ-field

$$\bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\epsilon_{i,k}^{-1}\left(\left\{e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\right)\in\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B}) \tag{4.8}$$

and

$$\bigcap_{k=1}^{b}B_k^{-1}(\{\beta_{\delta(k)}\})\in\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B}). \tag{4.9}$$

Under the assumption that $\beta_1,\ldots,\beta_b$ are all different we can write

$$\begin{aligned}
\bigcap_{k=1}^{b}B_k^{-1}\left(\{\beta_{\delta(k)}\}\right) &= \bigcap_{k=1}^{b}\left\{\left((\boldsymbol{\sigma},\tau),\omega\right)\in S_t\wr S_b\times\Omega : B_k\left((\boldsymbol{\sigma},\tau),\omega\right)=\beta_{\delta(k)}\right\}\\
&= \bigcap_{k=1}^{b}\left\{\left((\boldsymbol{\sigma},\tau),\omega\right)\in S_t\wr S_b\times\Omega : \tilde{B}_k\left(\boldsymbol{\sigma},\tau\right)=\beta_{\delta(k)}\right\}\\
&= \bigcap_{k=1}^{b}\left(\left\{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b : \tilde{B}_k\left(\boldsymbol{\sigma},\tau\right)=\beta_{\delta(k)}\right\}\times\Omega\right)\\
&= \left(\bigcap_{k=1}^{b}\left\{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b : \tilde{B}_k\left(\boldsymbol{\sigma},\tau\right)=\beta_{\delta(k)}\right\}\right)\times\Omega\\
&= \left(\bigcap_{k=1}^{b}\left\{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b : \beta_{\tau(k)}=\beta_{\delta(k)}\right\}\right)\times\Omega.
\end{aligned} \tag{4.10}$$

Since we assume that for every $k=1,\ldots,b$ the values $e_{1,k},\ldots,e_{t,k}$ are all different so again we can write

$$\begin{aligned}
&\bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\epsilon_{i,k}^{-1}\left(\left\{e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\right)\\
&= \bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\left\{\left((\boldsymbol{\sigma},\tau),\omega\right)\in S_t\wr S_b\times\Omega : \epsilon_{i,k}\left((\boldsymbol{\sigma},\tau),\omega\right)=e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\\
&= \bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\left\{\left((\boldsymbol{\sigma},\tau),\omega\right)\in S_t\wr S_b\times\Omega : \tilde{\epsilon}_{i,k}\left(\boldsymbol{\sigma},\tau\right)=e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\\
&= \bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\left(\left\{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b : \tilde{\epsilon}_{i,k}\left(\boldsymbol{\sigma},\tau\right)=e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\times\Omega\right)\\
&= \left(\bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\left\{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b : \tilde{\epsilon}_{i,k}\left(\boldsymbol{\sigma},\tau\right)=e_{\pi_{\delta(k)}(i),\delta(k)}\right\}\right)\times\Omega.
\end{aligned} \tag{4.11}$$

Now by using (4.10) and (4.11) we see that

$$\left(\bigcap_{k=1}^{b}\bigcap_{i=1}^{t}\epsilon_{i,k}^{-1}\left(\{e_{\pi_{\delta(k)}(i),\delta(k)}\}\right)\right)\bigcap\left(\bigcap_{k=1}^{b}B_k^{-1}\left(\{\beta_{\delta(k)}\}\right)\right)=\{(\boldsymbol{\pi},\delta)\}\times\Omega.$$

Thus, using (4.8) and (4.9), for every $(\boldsymbol{\pi},\delta)\in S_t\wr S_b$ the set $\{(\boldsymbol{\pi},\delta)\}\times\Omega\in\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})$. ■

Now by applying part (iv) of Lemma 19 to $\boldsymbol{V}=(\boldsymbol{\epsilon},\boldsymbol{B})$ with

$$(\boldsymbol{\epsilon},\boldsymbol{B})=(\epsilon_{1,1},\ldots,\epsilon_{t,1},\ldots,\epsilon_{1,b},\ldots,\epsilon_{t,b},B_1,\ldots,B_b)^{\top} \tag{4.12}$$

we can write

$$\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})\subseteq\{A\times\Omega:A\subseteq S_t\wr S_b\}. \tag{4.13}$$

Thus from equation (4.13) and Lemma 13 we obtain the following result:
Corollary 4. *The σ-field $\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})$ generated by the vector $(\boldsymbol{\epsilon},\boldsymbol{B})$ defined in (4.12) is $\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})=\{A\times\Omega:A\subseteq S_t\wr S_b\}$.*

Furthermore, we have $S_t\wr S_b\times\Omega=\bigcup_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}(\{(\boldsymbol{\pi},\delta)\}\times\Omega)$. Since $\{A\times\Omega:A\subseteq S_t\wr S_b\}$ is the σ-field generated by the set $\{\{(\boldsymbol{\pi},\delta)\}\times\Omega:(\boldsymbol{\pi},\delta)\in S_t\wr S_b\}$ it follows from Corollary 4 that $\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})$ is the σ-field generated by $\{\{(\boldsymbol{\pi},\delta)\}\times\Omega:(\boldsymbol{\pi},\delta)\in S_t\wr S_b\}$.

In order to find $E(Y_{i,j_i,k}|\boldsymbol{\epsilon},\boldsymbol{B})=E(Y_{i,j_i,k}|\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B}))$ we can apply Theorem 2 in the appendix to $\tilde{\Omega}=S_t\wr S_b\times\Omega$ and $Z_l=\left\{\left(\boldsymbol{\pi}^{(l)},\delta_l\right)\right\}\times\Omega=\left\{\left(\pi_1^{(l)},\ldots,\pi_b^{(l)},\delta_l\right)\right\}\times\Omega$ for $l=1,\ldots,t!^b b!=a$ (say), where $\left(\boldsymbol{\pi}^{(1)},\delta_1\right)=\left(\pi_1^{(1)},\ldots,\pi_b^{(1)},\delta_1\right),\ldots,\left(\boldsymbol{\pi}^{(a)},\delta_a\right)=\left(\pi_1^{(a)},\ldots,\pi_b^{(a)},\delta_a\right)$ are the different elements of $S_t\wr S_b$. So s in Theorem 2 is equal to the number of elements in $S_t\wr S_b$ i.e., $s=t!^b b!=a$. Also $\mathscr{G}$ in Theorem 2 is the σ-field generated by $\{Z_l:l=1,\ldots,a\}=\left\{\left\{\left(\boldsymbol{\pi}^{(l)},\delta_l\right)\right\}\times\Omega:l=1,\ldots,a\right\}$ so $\mathscr{G}=\mathscr{F}(\boldsymbol{\epsilon},\boldsymbol{B})$.

Next, in Lemma 14 we derive the conditional expectation $E(Y_{i,j_i,k}|\boldsymbol{\epsilon},\boldsymbol{B})$ by substituting the random variable $Y_{i,j_i,k}:S_t\wr S_b\times\Omega\to\mathbb{R}$ defined on the probability space $(\Omega,\mathscr{F},P)=(S_t\wr S_b\times\Omega,\mathscr{P}(S_t\wr S_b)\otimes\mathscr{F},U\otimes P)$ for the random variable X in Theorem 2.

Lemma 14. *The conditional expectation of* $Y_{i,j_i,k}$ *given* $(\boldsymbol{\epsilon}, \boldsymbol{B})$ *is*

$$E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) = \sum_{(\boldsymbol{\sigma},\tau)\in S_t \wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} \int_\Omega X_{\sigma_{\tau(k)}(i),j,\tau(k)}(\tilde{\omega}) P(d\tilde{\omega}). \tag{4.14}$$

Proof. According to Theorem 2 in the appendix and by using equation (4.7) the conditional expectation is equal to

$$\begin{aligned}
& E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) = E(Y_{i,j_i,k}|\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})) \\
= & \sum_{l=1}^{a} \frac{1}{(U\otimes P)(\{(\boldsymbol{\pi}^{(l)}, \delta_l)\}\times\Omega)} \int_{\left\{(\boldsymbol{\pi}^{(l)},\delta_l)\right\}\times\Omega} Y_{i,j_i,k}((\boldsymbol{\pi},\delta),\tilde{\omega})(U\otimes P) \\
& \qquad d\left((\boldsymbol{\pi},\delta),\tilde{\omega}\right) 1_{\left\{(\boldsymbol{\pi}^{(l)},\delta_l)\right\}\times\Omega} \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} \frac{1}{\left(U\otimes P\right)\left(\{(\boldsymbol{\sigma},\tau)\}\times\Omega\right)} \int_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} Y_{i,j_i,k}((\boldsymbol{\pi},\delta),\tilde{\omega})(U\otimes P) \\
& \qquad d((\boldsymbol{\pi},\delta),\tilde{\omega}) 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} \frac{1}{U(\{(\boldsymbol{\sigma},\tau)\})P(\Omega)} \int_{\{(\boldsymbol{\sigma},\tau)\}} \int_\Omega Y_{i,j_i,k}((\boldsymbol{\pi},\delta),\tilde{\omega}) P(d\tilde{\omega}) U(d(\boldsymbol{\pi},\delta)) \\
& \qquad 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} \int_\Omega Y_{i,j_i,k}((\boldsymbol{\sigma},\tau),\tilde{\omega}) P(d\tilde{\omega}) \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega} \int_\Omega X_{\sigma_{\tau(k)}(i),j,\tau(k)}(\tilde{\omega}) P(d\tilde{\omega})
\end{aligned}$$

■

The conditional expectation $E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B})$ is a random variable. For every $((\boldsymbol{\pi},\delta),\omega) \in S_t \wr S_b \times \Omega$ the corresponding realization of the conditional expectation is equal to

$$\begin{aligned}
& E(Y_{i,j_i,k}|\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B}))((\boldsymbol{\pi},\delta),\omega) \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}\times\Omega}((\boldsymbol{\pi},\delta),\omega) \int_\Omega X_{\sigma_{\tau(k)}(i),j_i,\tau(k)}\left(\tilde{\omega}\right) P\left(d\tilde{\omega}\right) \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}}\left(\boldsymbol{\pi},\delta\right) \int_\Omega X_{\sigma_{\tau(k)}(i),j_i,\tau(k)}\left(\tilde{\omega}\right) P(d\tilde{\omega}) \\
= & \sum_{(\boldsymbol{\sigma},\tau)\in S_t\wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}}(\boldsymbol{\pi},\delta) E(X_{\sigma_{\tau(k)}(i),j_i,\tau(k)}) = E(X_{\pi_{\delta(k)}(i),j_i,\delta(k)}).
\end{aligned} \tag{4.15}$$

Equation (4.15) does not depend on $\omega \in \Omega$, rather only on the element $(\boldsymbol{\pi},\delta) \in$

$S_t \wr S_b$. Therefore we can write

$$E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) = \left(\sum_{(\boldsymbol{\sigma},\tau)\in S_t \wr S_b} 1_{\{(\boldsymbol{\sigma},\tau)\}} \int_{\Omega} X_{\sigma_{\tau(k)}(i),j_i,\tau(k)}\left(\tilde{\omega}\right) P\left(d\tilde{\omega}\right) \right) \circ \hat{p}_1,$$

where $\hat{p}_1 : S_t \wr S_b \times \Omega \to S_t \wr S_b$ with $\hat{p}_1\left(\left(\boldsymbol{\pi}, \delta\right), \omega\right) = \left(\boldsymbol{\pi}, \delta\right)$.

Similar to Section 2.1.3 it is useful to factorize the conditional expectation. In order to apply the factorization lemma (Lemma 18 in Section A.2.7) we substitute the functions $f = E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B})$ and $g = (\boldsymbol{\epsilon}, \boldsymbol{B})$ which are both defined on $S_t \wr S_b \times \Omega$. It follows that the function f is $\mathscr{F}(\boldsymbol{\epsilon}, \boldsymbol{B})$-$\mathscr{B}(\mathbb{R})$-measurable by the definition of the conditional expectation. Then, by Lemma 18, there exists a function $h = E(Y_{i,j_i,k}|\left(\boldsymbol{\epsilon}, \boldsymbol{B}\right) = \bullet)$ which is $\mathscr{B}(\mathbb{R}^n)$-$\mathscr{B}(\mathbb{R})$-measurable with $n = b(t+1)$ such that $f = h \circ g$. Therefore we can write

$$E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) = E(Y_{i,j_i,k}|(\boldsymbol{\epsilon}, \boldsymbol{B}) = \bullet) \circ (\boldsymbol{\epsilon}, \boldsymbol{B}), \tag{4.16}$$

where $E(Y_{i,j_i,k}|\left(\boldsymbol{\epsilon}, \boldsymbol{B}\right) = \bullet)$ is defined on the image set of $(\boldsymbol{\epsilon}, \boldsymbol{B})$ i.e., $\mathrm{Im}(\boldsymbol{\epsilon}, \boldsymbol{B}) = \{(e_{\pi_{\delta(1)}(1),\delta(1)}, \ldots, e_{\pi_{\delta(1)}(t),\delta(1)}, \ldots, e_{\pi_{\delta(b)}(1),\delta(b)}, \ldots e_{\pi_{\delta(b)}(t),\delta(b)}, \beta_{\delta(1)}, \ldots \beta_{\delta(b)}) : (\boldsymbol{\pi}, \delta) \in S_t \wr S_b\}$. It follows that $E(Y_{i,j_i,k}|(\boldsymbol{\epsilon}, \boldsymbol{B}) = (e_{\pi_{\delta(1)}(1),\delta(1)}, \ldots, e_{\pi_{\delta(1)}(t),\delta(1)}, \ldots, e_{\pi_{\delta(b)}(1),\delta(b)}, \ldots e_{\pi_{\delta(b)}(t),\delta(b)}, \beta_{\delta(1)}, \ldots \beta_{\delta(b)})) = E(X_{\pi_{\delta(k)}(i),j_i,\delta(k)})$.

4.4 Link Function to Conditional Expectation

In GLMMs, we model the conditional expectation of the responses given the vector of random effects by using a suitable link function. Now, in order to derive the RB-GLMM for the RCBD we introduce a link function g as before. It follows from (4.16) that

$$\begin{aligned} g\left(E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B})\right) &= g \circ E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}) \\ &= g \circ (E(Y_{i,j_i,k}|\left(\boldsymbol{\epsilon}, \boldsymbol{B}\right) = \bullet) \circ (\boldsymbol{\epsilon}, \boldsymbol{B})) \\ &= (g \circ E(Y_{i,j_i,k}|\left(\boldsymbol{\epsilon}, \boldsymbol{B}\right) = \bullet)) \circ (\boldsymbol{\epsilon}, \boldsymbol{B}) \\ &= g\left(E(Y_{i,j_i,k}|\left(\boldsymbol{\epsilon}, \boldsymbol{B}\right) = \bullet)\right) \circ (\boldsymbol{\epsilon}, \boldsymbol{B}). \end{aligned} \tag{4.17}$$

For every element $((\boldsymbol{\pi}, \delta), \omega) \in S_t \wr S_b \times \Omega$, we use (4.17), (4.15) and (4.2) to write $g(E(Y_{i,j_i,k}|\boldsymbol{\epsilon}, \boldsymbol{B}))$ as

$$
\begin{aligned}
g(E(Y_{i,j_i,k}|\boldsymbol{\epsilon},\boldsymbol{B}))((\boldsymbol{\pi},\delta),\omega) &= g(E(X_{\pi_{\delta(k)}(i),j_i,\delta(k)})) \\
&= \mu + \alpha_{j_i} + \beta_{\delta(k)} + e_{\pi_{\delta(k)}(i),\delta(k)} \\
&= \mu + \alpha_{j_i} + B_k((\boldsymbol{\pi},\delta),\omega) + \epsilon_{i,k}((\boldsymbol{\pi},\delta),\omega),
\end{aligned} \tag{4.18}
$$

where $\epsilon_{i,k}\left((\boldsymbol{\pi},\delta),\omega\right) = \tilde{\epsilon}_{i,k}(\boldsymbol{\pi},\delta) = e_{\pi_{\delta(k)}(i),\delta(k)}$ and $B_k\left((\boldsymbol{\pi},\delta),\omega\right) = \tilde{B}_k(\boldsymbol{\pi},\delta) = \beta_{\delta(k)}$. Finally, equation (4.18) and our previous results enable us to write

$$
\text{Model 2} : g\left(E(Y_{i,j_i,k}|\boldsymbol{\epsilon},\boldsymbol{B})\right) = \mu + \alpha_{j_i} + B_k + \epsilon_{i,k} \tag{4.19}
$$

for every $i = 1\ldots,t$, $j = 1,\ldots,t$ and $k = 1,\ldots,b$. There are two random variables, B_k and $\epsilon_{i,k}$, on the right hand side of (4.19) and the treatment j_i is fixed by the specific design d as before. More precisely, μ is the grand mean, α_{j_i} is the fixed treatment effect and B_k and $\epsilon_{i,k}$ are two random effects due to random blocks and random unit errors respectively.

4.5 Moments of Random Effects

In this section, we derive expectations, variances and covariances for the components of both vectors of random effects, that is, for the random blocks $\boldsymbol{B} = (B_1,\ldots,B_b)^\top$ and for the random errors $\boldsymbol{\epsilon} = (\epsilon_{1,1},\ldots,\epsilon_{t,1},\ldots,\epsilon_{1,b},\ldots,\epsilon_{t,b})^\top$ from the randomization associated with Model 2 in (4.19). The derivations are similar to those for Model 1 in Section 2.1.5 but are more complicated due to the underlying more complex design.

4.5.1 Random Errors

We derive the expectations, variances and covariances for the components of $\boldsymbol{\epsilon}$. In order to derive these results, we first set $Q^*_{k,k^*} = \{(\boldsymbol{\pi},\delta) \in S_t \wr S_b : \delta(k) = k^*\}$. Then, the wreath product of S_t and S_b is $S_t \wr S_b = \bigcup_{k^*=1}^{b} Q^*_{k,k^*}$. Furthermore, set $Q^*_{k,k^*,i,i^*} = \{(\boldsymbol{\pi},\delta) \in S_t \wr S_b : \delta(k) = k^*, \pi_{k^*}(i) = i^*\}$. Then $Q^*_{k,k^*} = \bigcup_{i^*=1}^{t} Q^*_{k,k^*,i,i^*}$. It follows that $S_t \wr S_b = \bigcup_{k^*=1}^{b}\bigcup_{i^*=1}^{t} Q^*_{k,k^*,i,i^*}$. The number of elements of Q^*_{k,k^*,i,i^*}

is $|Q^*_{k,k^*,i,i^*}| = (b-1)!(t-1)!t!^{(b-1)}$. Moreover, for $i \neq i^{'}$ we can write $S_t \wr S_b$ $= \bigcup_{k^*=1}^{b} \bigcup_{i^* \neq i'^*}^{t} (Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k,k^*,i',i'^*})$ and $|Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k,k^*,i',i'^*}| = (b-1)!\ (t-2)!\ t!^{(b-1)}$. Also, for $k \neq k^{'}$ we have $S_t \wr S_b = \bigcup_{k^* \neq k'^*}^{b} \left(Q^*_{k,k^*} \bigcap Q^*_{k',k'^*}\right) = \bigcup_{k^* \neq k'^*}^{b} \bigcup_{i^*=1}^{t} \bigcup_{i'^*=1}^{t} \left(Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k',k'^*,i',i'^*}\right)$ and $|Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k',k'^*,i',i'^*}| = (b-2)!\ (t-1)!^2\ t!^{(b-2)}$. Recall $u_{i,k}$ values and their mean $\bar{u}$ have already been defined in Section 4.1. With these preliminaries we can prove the following results.

Lemma 15. *Let* $\boldsymbol{\epsilon} = (\epsilon_{1,1}, \ldots, \epsilon_{t,1}, \ldots, \epsilon_{1,b}, \ldots, \epsilon_{t,b})^\top$ *be the vector of random unit errors in Model 2 (4.19). Then*

(i) $E(\epsilon_{i,k}) = 0$,

(ii) $Var(\epsilon_{i,k}) = \sigma^2_W$,

(iii) $Cov(\epsilon_{i,k}, \epsilon_{i',k}) = -\dfrac{1}{t-1}\sigma^2_W$ *for every* k *and* $i \neq i^{'}$,

(iv) $Cov(\epsilon_{i,k}, \epsilon_{i',k'}) = 0$ *for* $k \neq k^{'}$,

where $\sigma^2_W = \dfrac{1}{bt}\sum_{k^*=1}^{b}\sum_{i^*=1}^{t}(u_{i^*,k^*} - \bar{u}_{k^*})^2$.

Proof. (i) The expectation of $\epsilon_{i,k}$ is given by

$$\begin{aligned}
E(\epsilon_{i,k}) &= \int_{S_t \wr S_b}\int_{\Omega} \epsilon_{i,k}((\boldsymbol{\pi},\delta),\omega)P(d\omega)U(d(\boldsymbol{\pi},\delta)) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\int_{\Omega}\epsilon_{i,k}((\boldsymbol{\pi},\delta),\omega)P(d\omega) = \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\tilde{\epsilon}_{i,k}(\boldsymbol{\pi},\delta) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b} e_{\pi_{\delta(k)}(i),\delta(k)} = \frac{1}{a}\sum_{k^*=1}^{b}\sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*}} e_{\pi_{k^*}(i),k^*} \\
&= \frac{1}{a}\sum_{k^*=1}^{b}\sum_{i^*=1}^{t}\sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*,i,i^*}} e_{i^*,k^*} = \frac{(b-1)!(t-1)!t!^{(b-1)}}{t!^b b!}\sum_{k^*=1}^{b}\sum_{i^*=1}^{t} e_{i^*,k^*} \\
&= \frac{1}{bt}\sum_{k^*=1}^{b}\sum_{i^*=1}^{t} e_{i^*,k^*} = \frac{1}{bt}\sum_{k^*=1}^{b}\sum_{i^*=1}^{t}(u_{i^*,k^*} - \bar{u}_{k^*}) = 0.
\end{aligned}$$

(ii) Now by using $E(\epsilon_{i,k})$=0 from part (i), the variance of $\epsilon_{i,k}$ is equal to

$$\begin{aligned}
Var(\epsilon_{i,k}) &= E(\epsilon_{i,k}^2) = \int_{S_t \wr S_b} \int_{\Omega} \epsilon_{i,k}^2((\boldsymbol{\pi},\delta),\omega)\, P\,(d\omega)\, U\,(d(\boldsymbol{\pi},\delta)) \\
&= \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} \int_{\Omega} \epsilon_{i,k}^2((\boldsymbol{\pi},\delta),\omega) P\,(d\omega) \\
&= \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} \tilde{\epsilon}_{i,k}^2(\boldsymbol{\pi},\delta) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} e^2_{\pi_{\delta(k)}(i),\delta(k)} \\
&= \frac{1}{a} \sum_{k^*=1}^{b} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*}} e^2_{\pi_{k^*}(i),k^*} = \frac{1}{a} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*,i,i^*}} e^2_{i^*,k^*} \\
&= \frac{(b-1)!(t-1)!t!^{(b-1)}}{t!^b b!} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} e^2_{i^*,k^*} \\
&= \frac{1}{bt} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} e^2_{i^*,k^*} = \frac{1}{bt} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} (u_{i^*,k^*} - \bar{u}_{k^*})^2 = \sigma_W^2.
\end{aligned}$$

(iii) For every k and $i \neq i^{'}$, we calculate the covariance of $\epsilon_{i,k}$ and $\epsilon_{i^{'},k}$ by parts (i) and (ii) as

$$\begin{aligned}
Cov(\epsilon_{i,k},\epsilon_{i',k}) &= E(\epsilon_{i,k}\epsilon_{i',k}) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} e_{\pi_{\delta(k)}(i),\delta(k)}\ e_{\pi_{\delta(k)}(i'),\delta(k)} \\
&= \frac{1}{a} \sum_{k^*=1}^{b} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*}} e_{\pi_{k^*}(i),k^*}\ e_{\pi_{k^*}(i'),k^*} \\
&= \frac{1}{a} \sum_{k^*=1}^{b} \sum_{i^*\neq i'^*}^{t} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k,k^*,i',i'^*}} e_{i^*,k^*}\ e_{i'^*,k^*} \\
&= \frac{(b-1)!(t-2)!t!^{(b-1)}}{t!^b b!} \sum_{k^*=1}^{b} \sum_{i^*\neq i'^*}^{t} e_{i^*,k^*}\ e_{i'^*,k^*} \\
&= \frac{1}{bt(t-1)} \sum_{k^*=1}^{b} \sum_{i^*\neq i'^*}^{t} e_{i^*,k^*}\ e_{i'^*,k^*} = \frac{1}{bt(t-1)} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} \sum_{\substack{i'^*=1 \\ i'^*\neq i^*}}^{t} e_{i^*,k^*}\ e_{i'^*,k^*} \\
&= \frac{1}{bt(t-1)} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} e_{i^*,k^*} \left(\sum_{i'^*=1}^{t} e_{i'^*,k^*} - e_{i^*,k^*} \right) \\
&= -\frac{1}{bt(t-1)} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} e^2_{i^*,k^*} = -\frac{1}{bt(t-1)} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} (u_{i^*,k^*} - \bar{u}_{k^*})^2 \\
&= -\frac{1}{t-1}\sigma_W^2.
\end{aligned}$$

(*iv*) Similarly, for $k \neq k^{'}$ and every i and $i^{'}$ we derive the covariance of $\epsilon_{i,k}$ and $\epsilon_{i^{'},k^{'}}$ as

$$\begin{aligned}
Cov(\epsilon_{i,k}, \epsilon_{i^{'},k^{'}}) &= E(\epsilon_{i,k}\epsilon_{i^{'},k^{'}}) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} e_{\pi_{\delta(k)}(i),\delta(k)} \; e_{\pi_{\delta(k^{'})}(i^{'}),\delta(k^{'})} \\
&= \frac{1}{a} \sum_{k^*\neq k^{'*}}^{b} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*} \bigcap Q^*_{k^{'},k^{'*}}} e_{\pi_{k^*}(i),k^*} \; e_{\pi_{k^{'*}}(i^{'}),k^{'*}} \\
&= \frac{1}{a} \sum_{k^*\neq k^{'*}}^{b} \sum_{i^*=1}^{t} \sum_{i^{'*}=1}^{t} \sum_{(\boldsymbol{\pi},\delta)\in Q^*_{k,k^*,i,i^*} \bigcap Q^*_{k^{'},k^{'*},i^{'},i^{'*}}} e_{i^*,k^*} \; e_{i^{'*},k^{'*}} \\
&= \frac{(b-2)!\;(t-1)!^2\;t!^{(b-2)}}{t!^b\;b!} \sum_{k^*\neq k^{'*}}^{b} \sum_{i^*=1}^{t} \sum_{i^{'*}=1}^{t} e_{i^*,k^*} \; e_{i^{'*},k^{'*}} \\
&= \frac{1}{b(b-1)t^2} \sum_{k^*\neq k^{'*}}^{b} \sum_{i^*=1}^{t} \left(e_{i^*,k^*} \sum_{i^{'*}=1}^{t} e_{i^{'*},k^{'*}} \right) \\
&= \frac{1}{b(b-1)t^2} \sum_{k^*\neq k^{'*}}^{b} \left(\sum_{i^*=1}^{t} e_{i^*,k^*} \right) \left(\sum_{i^{'*}=1}^{t} e_{i^{'*},k^{'*}} \right) = 0.
\end{aligned}$$

■

The result in part (*iii*) of Lemma 15 shows that the random errors are correlated when in the same block due to randomization, whereas part (*iv*) of Lemma 15 shows that they are uncorrelated when in different blocks.

4.5.2 Random Blocks

In this section we derive the expectations, variances and covariances for the components of $\boldsymbol{B}$ and we show that the random blocks are correlated. Recall that B_k has already been defined in Section 4.2. Set $Q_{k,k^*} = \{\delta \in S_b : \delta(k) = k^*\}$. Then $S_b = \bigcup_{k^*=1}^{b} Q_{k,k^*}$ and the number of elements of Q_{k,k^*} is $(b-1)!$. We further set $S_b = \bigcup_{k^*\neq l^*} (Q_{k,k^*} \bigcap Q_{l,l^*})$ which we use to derive the covariance between two different blocks.

Lemma 16. *The expectation and variance of an element B_k of the vector of random blocks $\boldsymbol{B}$ and the covariance of two components B_k and $B_{k'}$ are*

(i) $E(B_k) = 0$,

(ii) $Var(B_k) = \sigma_B^2$,

(iii) $Cov(B_k, B_{k'}) = -\dfrac{1}{b-1}\sigma_B^2$ *for* $k \neq k'$,

where $\sigma_B^2 = \dfrac{1}{b}\sum\limits_{k^*=1}^{b}(\bar{u}_{k^*} - \bar{u})^2$.

Proof. (i) The expectation of B_k is given by

$$\begin{aligned}
E(B_k) &= \int_{S_t \wr S_b}\int_{\Omega} B_k((\boldsymbol{\pi},\delta),\omega)P(d\omega)U(d(\boldsymbol{\pi},\delta)) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\int_{\Omega} B_k((\boldsymbol{\pi},\delta),\omega)P(d\omega) = \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\tilde{B}_k(\boldsymbol{\pi},\delta) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\beta_{\delta(k)} = \frac{t!^b}{t!^b b!}\sum_{\delta\in S_b}\beta_{\delta(k)} = \frac{1}{b!}\sum_{k^*=1}^{b}\sum_{\delta\in Q_{k,k^*}}\beta_{k^*} \\
&= \frac{(b-1)!}{b!}\sum_{k^*=1}^{b}\beta_{k^*} = \frac{1}{b}\sum_{k^*=1}^{b}\beta_{k^*} = \frac{1}{b}\sum_{k^*=1}^{b}(\bar{u}_{k^*} - \bar{u}) = 0.
\end{aligned}$$

(ii) The variance of B_k can be computed using $E(B_k) = 0$ as

$$\begin{aligned}
Var(B_k) &= E(B_k^2) = \int_{S_t\wr S_b}\int_{\Omega} B_k^2((\boldsymbol{\pi},\delta),\omega)P(d\omega)U(d(\boldsymbol{\pi},\delta)) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\int_{\Omega} B_k^2((\boldsymbol{\pi},\delta),\omega)P(d\omega) = \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\tilde{B}_k^2(\boldsymbol{\pi},\delta) \\
&= \frac{1}{a}\sum_{(\boldsymbol{\pi},\delta)\in S_t\wr S_b}\beta_{\delta(k)}^2 = \frac{t!^b}{t!^b\ b!}\sum_{\delta\in S_b}\beta_{\delta(k)}^2 = \frac{1}{b!}\sum_{k^*=1}^{b}\sum_{\delta\in Q_{k,k^*}}\beta_{k^*}^2 \\
&= \frac{(b-1)!}{b!}\sum_{k^*=1}^{b}\beta_{k^*}^2 = \frac{1}{b}\sum_{k^*=1}^{b}\beta_{k^*}^2 = \frac{1}{b}\sum_{k^*=1}^{b}(\bar{u}_{k^*} - \bar{u})^2 = \sigma_B^2.
\end{aligned}$$

(*iii*) Similarly for $k \neq l$ the covariance of B_k and B_l is

$$\begin{aligned}
Cov(B_k, B_l) &= E(B_k B_l) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b} \beta_{\delta(k)}\beta_{\delta(l)} = \frac{t!^b}{t!^b b!} \sum_{\delta \in S_b} \beta_{\delta(k)}\beta_{\delta(l)} \\
&= \frac{1}{b!} \sum_{k^* \neq l^*}^{b} \sum_{\delta \in Q_{k,k^*} \bigcap Q_{l,l^*}} \beta_{k^*}\beta_{l^*} = \frac{(b-2)!}{b!} \sum_{k^* \neq l^*}^{b} \beta_{k^*}\beta_{l^*} \\
&= \frac{1}{b(b-1)} \sum_{k^* \neq l^*}^{b} \beta_{k^*}\beta_{l^*} = \frac{1}{b(b-1)} \sum_{k^*=1}^{b} \sum_{\substack{l^*=1 \\ l^* \neq k^*}}^{b} \beta_{k^*}\beta_{l^*} \\
&= \frac{1}{b(b-1)} \sum_{k^*=1}^{b} \beta_{k^*} \left(\sum_{\substack{l^*=1 \\ l^* \neq k^*}}^{b} \beta_{l^*} \right) \\
&= \frac{1}{b(b-1)} \sum_{k^*=1}^{b} \beta_{k^*} \left(\sum_{l^*=1}^{b} \beta_{l^*} - \beta_{k^*} \right) = -\frac{1}{b(b-1)} \sum_{k^*=1}^{b} \beta_{k^*}^2 \\
&= -\frac{1}{b(b-1)} \sum_{k^*=1}^{b} \left(\bar{u}_{k^*} - \bar{u}\right)^2 = -\frac{1}{b-1}\sigma_B^2.
\end{aligned}$$

■

4.5.3 Covariance between Random Blocks and Errors

We now derive the covariance between random blocks and errors, and show that the random blocks and unit errors are uncorrelated.

Lemma 17. *Let* $\boldsymbol{B} = (B_1, \ldots, B_b)^\top$ *and* $\boldsymbol{\epsilon} = (\epsilon_{1,1}, \ldots, \epsilon_{t,1}, \ldots, \epsilon_{1,b}, \ldots, \epsilon_{t,b})^\top$ *be the vectors of random blocks and unit errors respectively in Model (4.19). Then*

(i) $Cov(B_k, \epsilon_{i,k}) = 0$,

(ii) $Cov(B_k, \epsilon_{i,k'}) = 0$ *for* $k \neq k'$.

Proof. (*i*) Using $E(\epsilon_{i,k}) = 0$ and $E(B_k) = 0$ from part (*i*) of Lemma 15 and 16 respectively, we compute the covariance of B_k and $\epsilon_{i,k}$ as

$$
\begin{aligned}
Cov(B_k, \epsilon_{i,k}) &= E(B_k \epsilon_{i,k}) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta) \in S_t \wr S_b} \beta_{\delta(k)} e_{\pi_{\delta(k)}(i),\delta(k)} \\
&= \frac{1}{a} \sum_{k^*=1}^{b} \sum_{(\boldsymbol{\pi},\delta) \in Q^*_{k,k^*}} \beta_{k^*} e_{\pi_{k^*}(i),k^*} \\
&= \frac{1}{a} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} \sum_{(\boldsymbol{\pi},\delta) \in Q^*_{k,k^*,i,i^*}} \beta_{k^*} e_{i^*,k^*} \\
&= \frac{(b-1)!(t-1)!t!^{(b-1)}}{t!^b b!} \sum_{k^*=1}^{b} \sum_{i^*=1}^{t} \beta_{k^*} e_{i^*,k^*} \\
&= \frac{1}{bt} \sum_{k^*=1}^{b} \beta_{k^*} \left(\sum_{i^*=1}^{t} e_{i^*,k^*} \right) = 0.
\end{aligned}
$$

(ii) Similarly for $k \neq k^{'}$ the covariance of B_k and $\epsilon_{i,k'}$ is

$$
\begin{aligned}
Cov(B_k, \epsilon_{i,k'}) &= E(B_k \epsilon_{i,k'}) = \frac{1}{a} \sum_{(\boldsymbol{\pi},\delta) \in S_t \wr S_b} \beta_{\delta(k)} e_{\pi_{\delta(k')}(i),\delta(k')} \\
&= \frac{1}{a} \sum_{k^* \neq k'^*}^{b} \sum_{(\boldsymbol{\pi},\delta) \in Q^*_{k,k^*} \bigcap Q^*_{k',k'^*}} \beta_{k^*} e_{\pi_{k'^*}(i),k'^*} \\
&= \frac{1}{a} \sum_{k^* \neq k'^*}^{b} \sum_{i^*=1}^{t} \sum_{(\boldsymbol{\pi},\delta) \in Q^*_{k,k^*} \bigcap Q^*_{k',k'^*,i,i^*}} \beta_{k^*} e_{i^*,k'^*} \\
&= \frac{(b-2)!(t-1)!^2 t!^{(b-2)}}{t!^b b!} \sum_{k^* \neq k'^*}^{b} \sum_{i^*=1}^{t} \beta_{k^*} e_{i^*,k'^*} \\
&= \frac{1}{b(b-1)t^2} \sum_{k^* \neq k'^*}^{b} \beta_{k^*} \left(\sum_{i^*=1}^{t} e_{i^*,k'^*} \right) = 0.
\end{aligned}
$$

■

4.6 Summary of Results

In order to summarize the derived results for Model 2 (4.19) we let $\boldsymbol{Y}$ be the $bt \times 1$ vector of responses and $\boldsymbol{X}$ be the $bt \times (t+1)$ design matrix for the fixed

treatment effects. We then rewrite Model 2 in matrix notation as

$$\boldsymbol{g}(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon},\boldsymbol{B})) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{B} + \boldsymbol{\epsilon}, \tag{4.20}$$

where $\boldsymbol{g}(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon},\boldsymbol{B}))$ is a vector of conditional expectations with components $g(E(Y_{i,j_i,k}|\epsilon,B))$ and $\boldsymbol{\beta} = (\mu,\alpha_1,\ldots,\alpha_t)^\top$ is the $(t+1)\times 1$ vector of fixed treatment effect parameters, $\boldsymbol{Z}$ is the model matrix of order $bt \times b$ corresponding to random block effects, $\boldsymbol{B}$ is the $b \times 1$ vector of random block effects, and $\boldsymbol{\epsilon}$ is the $bt \times 1$ vector of random errors. We call (4.20) the RB-GLMM for the RCBD with random block effects where the conditional distribution of $Y_{i,j_i,k}$ given $(\boldsymbol{\epsilon},\boldsymbol{B})$ is a member of the exponential family.

We can now summarize the results of Lemma 15 in matrix notation. The expectation can be written as $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and the variance-covariance matrix can be written as $V(\boldsymbol{\epsilon}) = \text{diag}(\boldsymbol{Q},\ldots,\boldsymbol{Q})$ with $\boldsymbol{Q} = \dfrac{\sigma_W^2}{t-1}(t\boldsymbol{I}_t - \boldsymbol{1}_t\boldsymbol{1}_t^\top) = \sigma_W^2\boldsymbol{\mathcal{P}}_t$, where $\boldsymbol{\mathcal{P}}_t = \dfrac{1}{t-1}(t\boldsymbol{I}_t - \boldsymbol{1}_t\boldsymbol{1}_t^\top)$ is a matrix of order $t \times t$. Moreover, from Lemma 16 we can write $E(\boldsymbol{B}) = \boldsymbol{0}$ and $V(\boldsymbol{B}) = \dfrac{\sigma_B^2}{b-1}\left(b\boldsymbol{I}_b - \boldsymbol{1}_b\boldsymbol{1}_b^\top\right) = \sigma_B^2\boldsymbol{\mathcal{P}}_b$ with the $b \times b$ matrix $\boldsymbol{\mathcal{P}}_b = \dfrac{1}{b-1}\left(b\boldsymbol{I}_b - \boldsymbol{1}_b\boldsymbol{1}_b^\top\right)$. It also follows from Lemma 17 that $Cov(\boldsymbol{B},\boldsymbol{\epsilon}) = \boldsymbol{0}$, i.e., the vectors for random blocks and unit errors are uncorrelated.

The matrices $\boldsymbol{\mathcal{P}}_b$ and $\boldsymbol{\mathcal{P}}_t$ are singular and hence do not have inverses. This creates a problem for fitting GLMMs in most of the available software packages. To handle this problem we factorize $\boldsymbol{\mathcal{P}}_b$ and $\boldsymbol{\mathcal{P}}_t$, as we did for Model 1 associated with the CRD in Section 2.4, so that the original variance-covariance matrices remain unchanged.

Similar to Section 2.4 we replace $\boldsymbol{B}$ with $\boldsymbol{L}\tilde{\boldsymbol{B}}^*$ such that $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{\mathcal{P}}_b$. Here $\boldsymbol{L}$ is a $b \times (b-1)$ matrix and $\tilde{\boldsymbol{B}}^*$ is a $(b-1)\times 1$ vector of random blocks with $E(\tilde{\boldsymbol{B}}^*) = \boldsymbol{0}$ and $\boldsymbol{V}(\tilde{\boldsymbol{B}}^*) = \sigma_B^2\boldsymbol{I}_{b-1}$. We also replace $\boldsymbol{\epsilon}$ with $\tilde{\boldsymbol{L}}\tilde{\boldsymbol{\epsilon}}^*$ where

$$\tilde{\boldsymbol{L}} = \begin{pmatrix} \boldsymbol{L}^* & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{L}^* & \ldots & \boldsymbol{0} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{L}^* \end{pmatrix}$$

is a matrix of order $bt \times b(t-1)$, $\mathbf{0}$ is the $t \times (t-1)$ matrix of zeros and $\boldsymbol{L}^*$ is a $t \times (t-1)$ matrix such that $\boldsymbol{L}^*\boldsymbol{L}^{*\top} = \boldsymbol{\mathcal{P}}_t$. Note that $\boldsymbol{L}^*$ is not unique.

In this case $\tilde{\boldsymbol{\epsilon}}^*$ is a $b(t-1) \times 1$ vector of random errors with $E(\tilde{\boldsymbol{\epsilon}}^*) = \mathbf{0}$ and $\boldsymbol{V}(\tilde{\boldsymbol{\epsilon}}^*) = \text{diag}(\tilde{\boldsymbol{Q}}, \ldots, \tilde{\boldsymbol{Q}})$ where $\tilde{\boldsymbol{Q}} = \sigma_W^2 \boldsymbol{I}_{t-1}$. More precisely, the vectors $\tilde{\boldsymbol{B}}^*$ and $\tilde{\boldsymbol{\epsilon}}^*$ satisfy the usual assumptions of the standard GLMM (1.2) and $Cov(\tilde{\boldsymbol{B}}^*, \tilde{\boldsymbol{\epsilon}}^*) = \mathbf{0}$. Now the vectors $\boldsymbol{L}\tilde{\boldsymbol{B}}^*$ and $\tilde{\boldsymbol{L}}\tilde{\boldsymbol{\epsilon}}^*$ in the GLMM with

$$\boldsymbol{g}(\boldsymbol{E}(\boldsymbol{Y}|\tilde{\boldsymbol{\epsilon}}, \tilde{\boldsymbol{B}})) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{L}\tilde{\boldsymbol{B}}^* + \tilde{\boldsymbol{L}}\tilde{\boldsymbol{\epsilon}}^* \tag{4.21}$$

have the same means and variance-covariance matrices as the random vectors $\boldsymbol{B}$ and $\boldsymbol{\epsilon}$ in the GLMM with

$$\boldsymbol{g}(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon}, \boldsymbol{B})) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{B} + \boldsymbol{\epsilon}$$

derived from the randomization for the RCBD with random block effects (Model (4.20)).

In Chapter 5, we investigate the impact of misspecification of the correlation structure and of the random effects distribution considering Model 2 (4.20) and the variance-covariance matrices derived from the randomization via simulation studies. The above factorization results allow us to calculate inverses of variance-covariance matrices for both vectors of random blocks $\boldsymbol{B}$ and random unit errors $\boldsymbol{\epsilon}$, which in turn allows us to apply HGLMs with correlated random effects to the simulated data.

Chapter 5

Simulation Study

We investigate the effect of misspecification of the correlation structure and of the distribution of the random effects on the estimated fixed effect parameters and variance components in GLMMs via simulation studies. To this end we simulated Poisson data from the GLMM (4.20) for the RCBD with random block effects using the variance-covariance matrices of the random blocks and unit errors that were derived from the randomization in Chapter 4. Different parametric models (see Section 5.1) are fitted to these data and then compared with the simulation results. In Section 5.2, we describe the simulation settings including true parameter values and specifications of the distributions of the random effects. The key simulation results are presented in Section 5.3 and some further results are given in the Appendices D to F. Moreover, we present results obtained from the additional comparisons of the models in Section 5.4. Section 5.5 summarizes the findings of the simulation studies.

5.1 Models

We consider five models ($M2$, $M2^*$, $M3$, $M4$, $M5$) in total; these models differ with respect to their linear predictors. We are interested in comparing these models in order to investigate the impact of model misspecification on the estimated model parameters. We distinguish a model in which the random effects, e.g., those for blocks and unit errors, are correlated, with the variance covariance matrices being derived from the randomization, from the model in which the random effects are uncorrelated ($M2$ versus $M2^*$). We distinguish models that in the linear predictor contain random effects for both blocks and

unit errors from models that contain only random effects for blocks ($M2$ versus $M3$) and only for unit errors ($M2$ versus $M4$). We recall the RB-GLMM for the RCBD with random block effects (model (4.20)). The vector of linear predictors $\boldsymbol{\eta}$, corresponding to model (4.20) with correlated random effects ($M2$) is defined as

$$M2 : \boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{ZB} + \boldsymbol{\epsilon}. \tag{5.1}$$

The detailed description of this model and the variance-covariance matrices for the random blocks and unit errors are given in Section 4.6.

We now consider a standard GLMM ($M2^*$) with uncorrelated random effects and the vector of linear predictors $\boldsymbol{\eta}^*$ is

$$M2^* : \boldsymbol{\eta}^* = \boldsymbol{X\beta} + \boldsymbol{ZB}^* + \boldsymbol{\epsilon}^*, \tag{5.2}$$

where $\boldsymbol{B}^*$ is the $b\times 1$ vector of uncorrelated random blocks with the expectation $E(\boldsymbol{B}^*) = \boldsymbol{0}$ and the variance-covariance matrix $V(\boldsymbol{B}^*) = \sigma_B^2 \boldsymbol{I}_b$. Moreover, $\boldsymbol{\epsilon}^*$ is the $bt \times 1$ vector of uncorrelated random errors with $E(\boldsymbol{\epsilon^*}) = \boldsymbol{0}$ and $V(\boldsymbol{\epsilon}^*) = \text{diag}(\boldsymbol{G}^*, \ldots, \boldsymbol{G}^*)$ where $\boldsymbol{G}^* = \sigma_W^2 \boldsymbol{I}_t$. The components between random blocks and errors are assumed to be uncorrelated i.e., $Cov(\boldsymbol{B^*, \epsilon^*}) = \boldsymbol{0}$.

Moreover, we consider two additional GLMMs ($M3$, $M4$) with the vector of linear predictors $\boldsymbol{\eta}$, given by

$$M3 : \boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{ZB} \tag{5.3}$$

and

$$M4 : \boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}. \tag{5.4}$$

In addition, we consider a GLM ($M5$) for fixed treatment effects with the linear predictor as

$$M5 : \boldsymbol{\eta} = \boldsymbol{X\beta}. \tag{5.5}$$

5.2 Simulation Settings

Following the simulation settings of previous studies (e.g., Agresti et al. (2004), Neuhaus et al. (2013) and McCulloch and Neuhaus (2011)) and in order to

generate reasonable Poisson data from the GLMM for the RCBD with random block effects, we consider the following settings for our simulation studies. The main goal is to investigate the impact of misspecification of the correlation structure and of the distribution of the random effects in GLMMs on the estimated model parameters.

We specify the number of treatments $t = 2, 3, 4$ and the number of blocks $b = 5t, 10t, 20t$. It follows that the number of blocks depends on the number of treatments. There are nine different combinations of a number of treatments and a number of blocks (t, b): (2,10), (2,20), (2,40), (3,15), (3,30), (3,60), (4,20), (4,40) and (4,80). We choose the error variances as $\sigma_W^2 = 0.01, 0.04, 0.09, 0.16$ and, hence, the corresponding standard deviations are $\sigma_W = 0.1, 0.2, 0.3, 0.4$. We also choose the block variances as $\sigma_B^2 = (1\sigma_W)^2$, $(2\sigma_W)^2$, $(3\sigma_W)^2$, $(4\sigma_W)^2$, $(5\sigma_W)^2$. For every combination of t, b and σ_W^2 we compute

$$t_{0.90,(t-1)(b-1)} \times \sqrt{\frac{2\sigma_W^2}{b}} = d_1,$$

$$t_{0.95,(t-1)(b-1)} \times \sqrt{\frac{2\sigma_W^2}{b}} = d_2$$

and

$$t_{0.99,(t-1)(b-1)} \times \sqrt{\frac{2\sigma_W^2}{b}} = d_3,$$

where $t_{0.90,(t-1)(b-1)}$, $t_{0.95,(t-1)(b-1)}$ and $t_{0.99,(t-1)(b-1)}$ are the upper 10%, 5% and 1% points of the t-distribution with $(t-1)(b-1)$ degrees of freedom. We then choose values for the fixed effect parameters $\tilde{\alpha}_j$ and μ for $j = 1, \ldots, t$ using d_r, $r = 1, 2, 3$ as

$$\tilde{\alpha}_j^{(r)} = \frac{3}{t-1} j d_r$$

and

$$\mu^{(r)} = \frac{1}{t} \sum_{j=1}^{t} \tilde{\alpha}_j^{(r)} = \frac{3(t+1)}{2(t-1)} d_r,$$

where the superscript on $\tilde{\alpha}_j^{(r)}$ and $\mu^{(r)}$ stands for the different 'least significant difference' (LSD) values. We note that the $\tilde{\alpha}_j$ parameters for the fixed effects are on a straight line. Moreover, the fixed effect parameters are chosen such that the difference $\tilde{\alpha}_t - \tilde{\alpha}_1$ is equal to three times the least significant difference for testing $H_0 : \tilde{\alpha}_j = \tilde{\alpha}_{j'}$ at significance levels 0.10, 0.05 and 0.01. The LSD is derived from the corresponding linear model and should be regarded as a proxy to the actual LSD. For the purpose of comparison with the simulation results

Table 5.2 does not give the original $\tilde{\alpha}_j$ values but the differences $\alpha_j = \tilde{\alpha}_j - \bar{\tilde{\alpha}}$.

We consider three different random effects distributions to generate random effects for blocks and errors:

(i) block effects are generated from $N(0, \sigma_B^2)$ and errors from $N(0, \sigma_W^2)$,

(ii) block effects are generated from $U(-a_B, a_B)$ and errors from $U(-a_W, a_W)$, where $a_B = \sigma_B\sqrt{3}$ and $a_W = \sigma_W\sqrt{3}$,

(ii) block effects are generated from $\frac{\sigma_B}{\sqrt{3}}t_3$ and errors from $\frac{\sigma_W}{\sqrt{3}}t_3$, where t_3 is the t-distribution with 3 degrees of freedom.

We note that, in all three cases, random effects for blocks and errors have mean zero and the same variance. There are 540 different possible combinations of the true parameter values. For each combination, we generated 1000 data sets from the Poisson distribution and then fitted different parametric models to the same set of simulated data. In the thesis, we focus on 36 of these combinations in particular those with relatively large and small values of true variance components σ_B^2 and σ_W^2.

5.3 Simulation Results

We present results for four different scenarios of combinations of variance components (σ_W^2, σ_B^2) and fixed treatments effect parameters (α_1, α_2, α_3). The true values for σ_W^2 and σ_B^2 under different scenarios are given in Table 5.1, each scenario consists of nine different treatment-block combinations and of the true parameter values. In Scenario-I and II we choose relatively large values for σ_W^2 and σ_B^2, compared to those chosen in Scenario-III and IV. The variance component σ_B^2 has the same relationship with σ_W (σ_B^2=$(3\sigma_W)^2$) in Scenario-I and Scenario-III, it also has the same relationship (σ_B^2=$(2\sigma_W)^2$) in Scenario-II and Scenario-IV.

Table 5.1: True values of σ_W^2 and σ_B^2 under different scenarios.

Scenario-I	**Scenario-II**
σ_W^2=0.16 σ_B^2=$(3\sigma_W)^2$=1.44	σ_W^2=0.16 σ_B^2=$(2\sigma_W)^2$=0.64
Scenario-III	**Scenario-IV**
σ_W^2=0.04 σ_B^2=$(3\sigma_W)^2$=0.36	σ_W^2=0.04 σ_B^2=$(2\sigma_W)^2$=0.16

True values for α_1, α_2 and α_3 are the same in Scenario-I and II, and also the same in Scenario-III and IV for all treatment-block combinations. These values can be seen in Table 5.2.

Table 5.2: True values of α_1, α_2 and α_3 under different scenarios.

	Para-	$t=2$			$t=3$			$t=4$		
Scenarios	meters	$b=10$	$b=20$	$b=40$	$b=15$	$b=30$	$b=60$	$b=20$	$b=40$	$b=80$
I and **II**	α_1	-0.371	-0.482	-0.325	-0.541	-0.371	-0.258	-0.454	-0.316	-0.222
	α_2	-	-	-	0.000	0.000	0.000	-0.151	-0.105	-0.074
	α_3	-	-	-	-	-	-	0.151	0.105	0.074
III and **IV**	α_1	-0.186	-0.241	-0.163	-0.270	-0.185	-0.129	-0.227	-0.158	-0.111
	α_2	-	-	-	0.000	0.000	0.000	-0.076	-0.053	-0.037
	α_3	-	-	-	-	-	-	0.076	0.053	0.037

We implement the model estimation using the R-package **hglm** (hierarchical generalized linear model) which is suitable for modelling correlated random effects in GLMMs (Rönnegård et al., 2010) based on the h-likelihood method of estimation (Lee and Nelder, 1996, Lee and Lee, 2012). However, we note that the **hglm** package gives the standard errors (SEs) of the variance components estimates ($\hat{\sigma}_B^2$, $\hat{\sigma}_W^2$) for the random effects on a log scale. We approximate SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ on the original scale by the *Delta*-method (Ver Hoef, 2012) using the formula

$$\mathrm{SE}(\hat{\theta}) = \hat{\theta}\,\mathrm{SE}(\log\hat{\theta}), \tag{5.6}$$

where $\hat{\theta} = \hat{\sigma}_B^2$ or $\hat{\sigma}_W^2$.

We compare results of the simulation studies for fixed effects and variance components estimates in terms of their biases and SEs. We calculate bias by subtracting median estimates from true values of the parameters, as done by Litière et al. (2008), McCulloch and Neuhaus (2011) and Neuhaus et al.

(2011, 2013) in their simulation studies. We present relative biases and SEs of the estimated fixed effect parameters and variance components obtained by fitting the true model and misspecified models due to misspecification of the correlation structure and of the random effects distribution for Scenario-I. Further simulation results are found to be similar for Scenario-II to IV and given in the Appendices D to F respectively.

5.3.1 Results without Misspecification

In this section, we assess the parameter estimates of the Poisson GLMM (model (4.20), equivalently $M2$, with the linear predictor (5.1)) under no model misspecification. Poisson data are simulated from $M2$ assuming correlated and normally distributed random effects (blocks and errors). We then fit the same model $M2$ assuming correlated and normally distributed random effects. More precisely, we consider the same assumptions that are used for generating the data and for the model fitting i.e., no model misspecification is assumed. We report the results by averaging the estimated parameters over all simulated data sets. The summary of results are presented in Table 5.3 including median estimates of the parameters as reported by Litière et al. (2008) and Neuhaus et al. (2011).

It can be seen from Table 5.3 that the estimated parameter values are close to the true values. The model-based ($M2$) standard errors (SE) and empirical standard errors (standard deviation of the estimates) are also very close to each other. As the h-likelihood estimation algorithm (Lee and Nelder, 1996, Lee and Lee, 2012) applied to our model is able to reproduce the true values of the parameters, this method appears suitable for fitting GLMMs with correlated random block effects and errors. The convergence of this estimation algorithm on the simulated data is very high (98.8 % to 100%) for all the nine treatment-block combinations.

Table 5.3: Estimates, standard errors and convergence obtained from fitting Poisson GLMM with linear predictor $M2$: $\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{ZB} + \boldsymbol{\epsilon}$ (5.1) under Scenario-I (σ^2_W=0.16, σ^2_B=1.44).

t	b	True values	Estimates		Standard Errors		Convergence (%)
			Mean	Median	Model-based	Empirical	
2	10	$\alpha_1 = -0.371$	-0.378	-0.381	0.148	0.154	98.8
		σ^2_B	1.530	1.372	0.743	0.828	
		σ^2_W	0.161	0.139	0.092	0.111	
	20	$\alpha_1 = -0.482$	-0.480	-0.479	0.101	0.101	100
		σ^2_B	1.459	1.403	0.482	0.503	
		σ^2_W	0.165	0.157	0.061	0.068	
	40	$\alpha_1 = -0.325$	-0.325	-0.320	0.080	0.077	100
		σ^2_B	1.498	1.461	0.353	0.404	
		σ^2_W	0.167	0.164	0.047	0.056	
3	15	$\alpha_1 = -0.541$	-0.549	-0.548	0.137	0.139	100
		$\alpha_2 = 0.000$	0.004	0.001	0.131	0.137	
		σ^2_B	1.547	1.427	0.598	0.646	
		σ^2_W	0.165	0.158	0.056	0.067	
	30	$\alpha_1 = -0.371$	-0.377	-0.375	0.108	0.110	100
		$\alpha_2 = 0.000$	0.004	0.003	0.104	0.105	
		σ^2_B	1.545	1.503	0.422	0.463	
		σ^2_W	0.166	0.162	0.043	0.056	
	60	$\alpha_1 = -0.258$	-0.260	-0.259	0.084	0.081	100
		$\alpha_2 = 0.000$	-0.001	0.000	0.081	0.080	
		σ^2_B	1.550	1.514	0.302	0.340	
		σ^2_W	0.169	0.166	0.034	0.045	
4	20	$\alpha_1 = -0.454$	-0.461	-0.457	0.138	0.138	99.9
		$\alpha_2 = -0.151$	-0.145	-0.142	0.131	0.133	
		$\alpha_3 = 0.151$	0.155	0.157	0.126	0.130	
		σ^2_B	1.528	1.481	0.511	0.538	
		σ^2_W	0.165	0.162	0.044	0.060	
	40	$\alpha_1 = -0.316$	-0.323	-0.322	0.107	0.105	100
		$\alpha_2 = -0.105$	-0.106	-0.103	0.103	0.100	
		$\alpha_3 = 0.105$	0.111	0.114	0.100	0.098	
		σ^2_B	1.531	1.490	0.362	0.406	
		σ^2_W	0.168	0.164	0.034	0.048	
	80	$\alpha_1 = -0.222$	-0.223	-0.223	0.081	0.080	100
		$\alpha_2 = -0.074$	-0.074	-0.072	0.079	0.079	
		$\alpha_3 = 0.074$	0.072	0.072	0.077	0.071	
		σ^2_B	1.521	1.501	0.256	0.279	
		σ^2_W	0.170	0.168	0.026	0.037	

5.3.2 Results with Misspecified Correlation Structure

We simulated Poisson data from $M2$ with correlated and normally distributed random effects (i.e., the same data as in Section 5.3.1). We fitted the misspecified model $M2^*$ with uncorrelated random effects (5.2) and the true model $M2$ with correlated random effects (5.1) to the same set of simulated data. Model $M2^*$ is misspecified in the sense that it assumes uncorrelated random effects. We then compare biases and empirical SEs of the estimated model parameters obtained from fitting $M2$ and $M2^*$.

5.3.2.1 Bias and SEs of Fixed Effects Estimates

In this section, we present relative biases and empirical SEs of the fixed effect estimates ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$) in Figure 5.1 and 5.2 respectively. It can be seen from Figure 5.1 that the percent relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are very close for the true model ($M2$) and the misspecified model ($M2^*$). Figure 5.2 also shows that SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ are almost the same in all nine treatment-block combinations for $M2$ and $M2^*$.

It follows that the misspecification of the correlation structure of the random effects has hardly any effect on the estimated fixed treatment effect parameters. That is, fixed effect estimates are not affected in terms of their biases and SEs due to misspecification of the correlation structure of the random effects distribution.

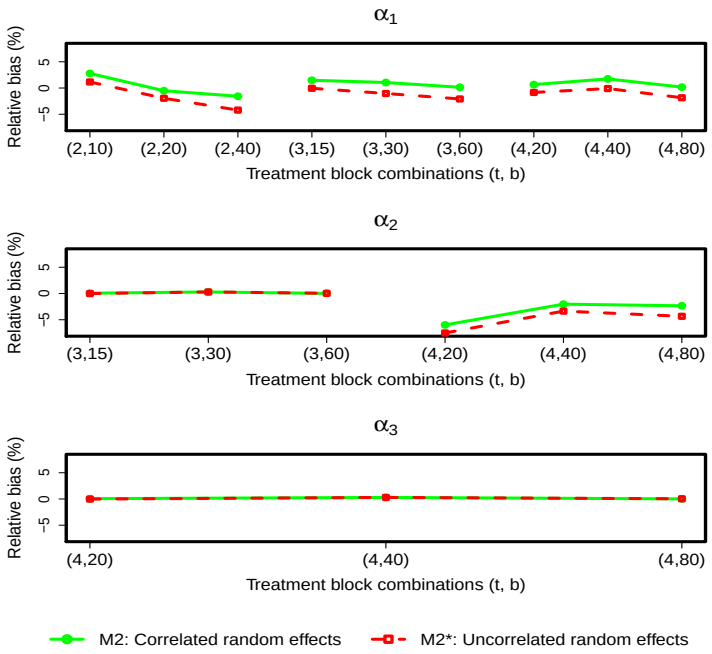

Figure 5.1: Comparison of relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-I between $M2$ (true model) and $M2^*$(misspecified model).

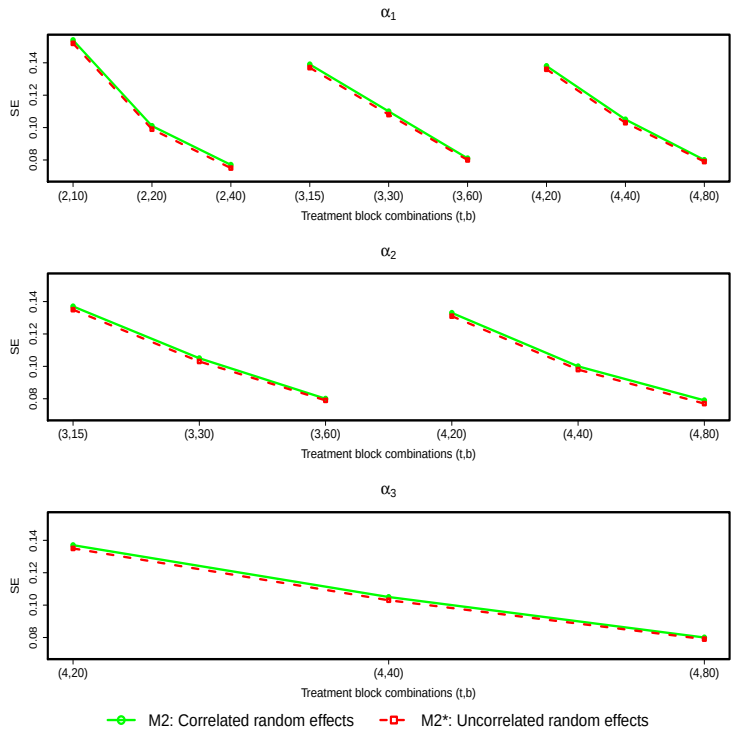

Figure 5.2: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-I between $M2$ (true model) and $M2^*$(misspecified model).

5.3.2.2 Bias and SEs of Variance Components Estimates

In this section, we compare relative biases (Figure 5.3) and empirical SEs (Figure 5.4) of the variance components estimates $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ between the true model ($M2$) and the misspecified model ($M2^*$). The top panel of Figure 5.3 represents relative biases of $\hat{\sigma}_B^2$ and the bottom panel shows relative biases for $\hat{\sigma}_W^2$. It can be seen from the top panel that the percent relative bias of $\hat{\sigma}_B^2$ for the true model is negligible (ranges from -5 % to 5 %) in all nine treatment-block combinations. However, $\hat{\sigma}_B^2$ seems to be consistently underestimated in all nine treatment-block combinations for the misspecified model $M2^*$ due to misspecification of the correlation structure for the random effects. The amount of bias is substantially higher (-15% to -5%) in the misspecified model compared to the true model.

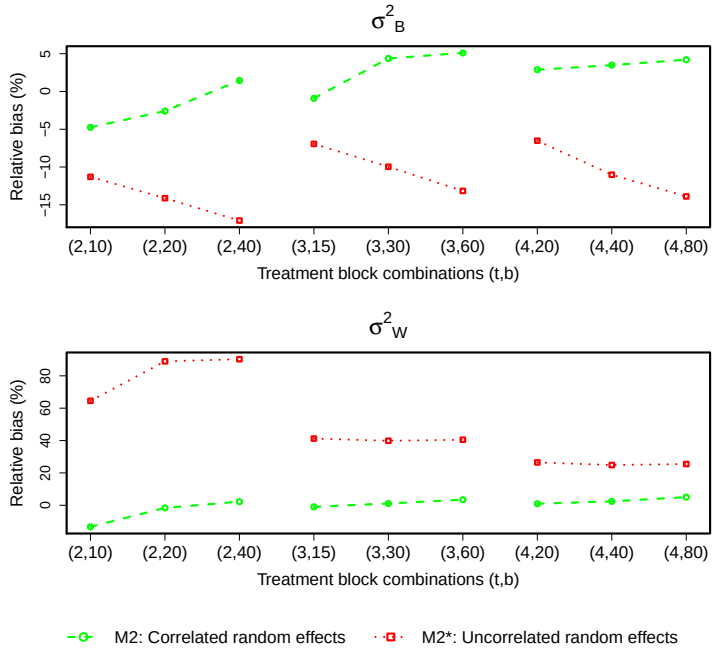

Figure 5.3: Comparison of relative biases of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-I between $M2$ (true model) and $M2^*$(misspecified model).

It can also be seen from the bottom panel of Figure 5.3 that the relative bias of $\hat{\sigma}_W^2$ is also negligible for the true model ($M2$) whereas it is substantially higher (20% to 90%) for the misspecified model ($M2^*$). Thus, both variance components estimates $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ are found to be severely biased due to misspecification of the correlation structure of the random blocks and errors.

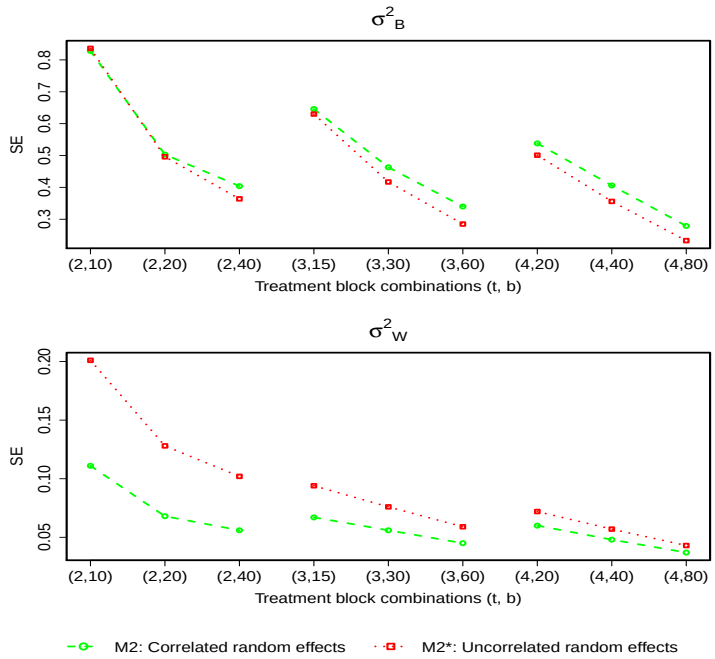

Figure 5.4: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-I between $M2$ (true model) and $M2^*$(misspecified model).

Moreover, from the top panel of Figure 5.4 it can be seen that there is not much difference in estimated SEs of $\hat{\sigma}_B^2$ between the true and misspecified models for all nine treatment-block combinations. However, there are some differences in the estimated SEs of $\hat{\sigma}_W^2$ between the true and misspecified models. In all nine treatment-block combinations, SEs of $\hat{\sigma}_W^2$ are found to be higher for the misspecified model $M2^*$ than the true model $M2$ due to misspecification of the correlation structure of the random effects. In summary, it appears that model misspecification in terms of correlation structure of the random effects has a substantial effect on the estimated variance components for both random blocks and errors.

5.3.3 Results with Misspecified Random Effects Distributions

We simulated Poisson data from $M2$ with correlated random effects (block effects and errors) those are generated from normal, uniform and transformed t_3 with the same mean and variance but different shapes. We fitted $M2$ under the assumption that random effects are normally distributed to the same set

of simulated data. That is, misspecification occurred when we fitted models assuming normally distributed random effects whereas the data are generated assuming random effects come from uniform and t_3 distributions.

5.3.3.1 Bias and SEs of Fixed Effects Estimates

Figure 5.5 and 5.6 show the percent relative biases and empirical SEs of the estimated fixed treatment effect parameters ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$) respectively. From Figure 5.5 it can be seen that there is no substantial differences in the relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ for all nine treatment-block combinations under three different random effects distributions (normal, uniform, t_3).

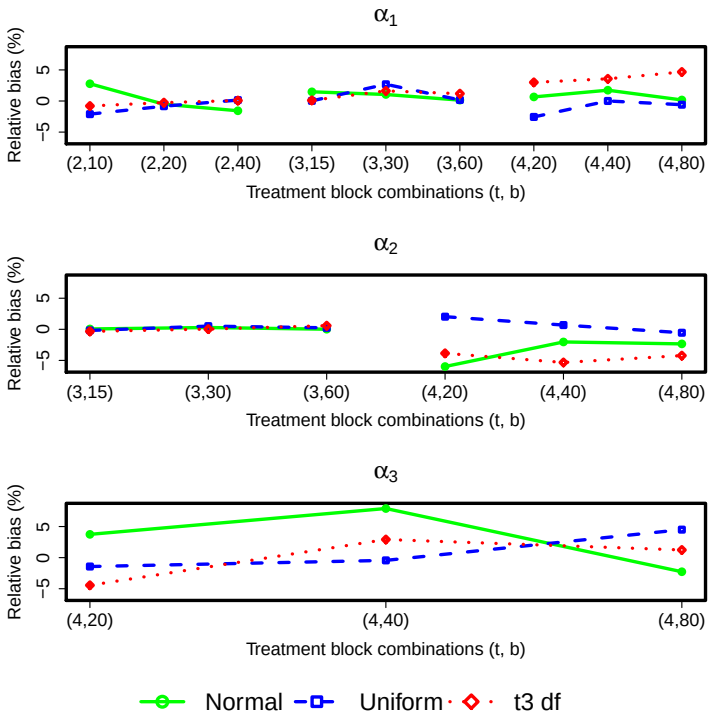

Figure 5.5: Comparison of percent relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-I among normal, uniform and t_3 random effects distributions.

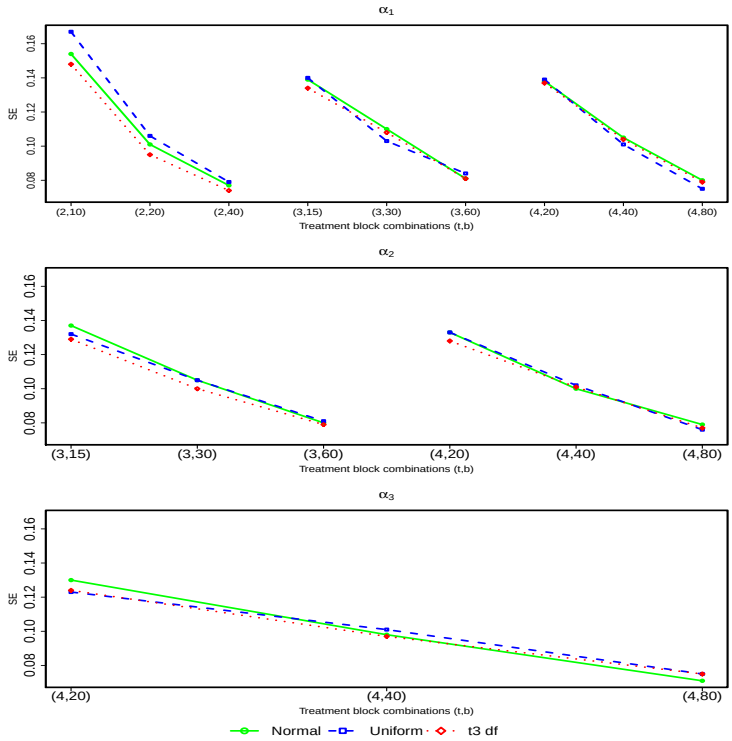

Figure 5.6: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-I among normal, uniform and t_3 random effects distributions.

For normal, uniform and t_3 random effects distributions, it can be seen from Figure 5.6 that the empirical SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ do not differ in all nine treatment-block combinations. We conclude that the misspecification of the random effects distribution does not have any effect in biases and standard errors on the estimates of the fixed treatment effect parameters.

5.3.3.2 Bias and SEs of Variance Components Estimates

We compare relative biases and SEs of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ under different random effects distributions in Figure 5.7 and 5.8 respectively. Relative biases of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ are presented at the top and bottom panel of Figure 5.7 respectively. From both panels it can be seen that the percent relative biases of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ are very close to each other in the case of normal and uniform random effects distributions. However, it seems that $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ are severely biased and underestimated in all treatment-block combinations when random effects are generated from t_3 (t-distribution with 3 degrees of freedom). The amount of bias is substantially higher (-40% to -10%) for t_3 than for normal and uniform

random effects distributions.

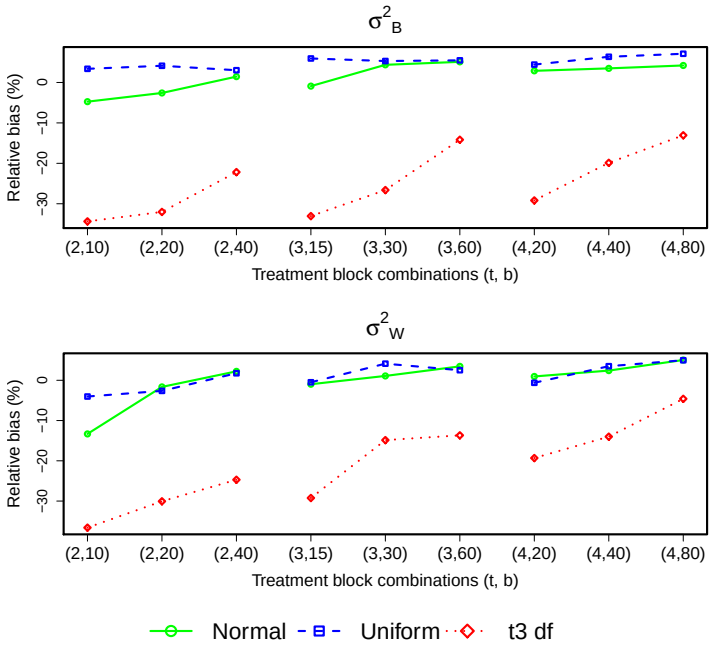

Figure 5.7: Comparison of percent relative biases of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-I among normal, uniform and t_3 random effects distributions.

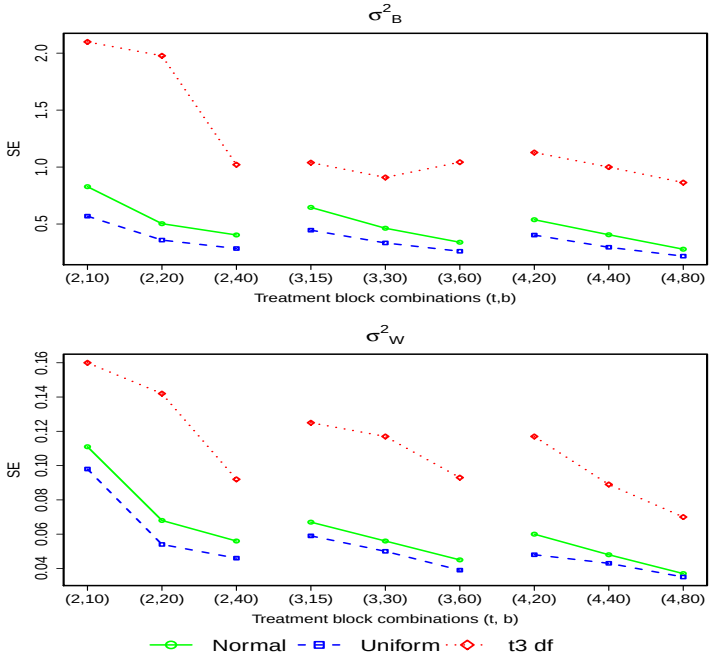

Figure 5.8: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-I among normal, uniform and t_3 random effects distributions.

Moreover, the top panel in Figure 5.8 represents empirical SEs of $\hat{\sigma}_B^2$ while the bottom panel represents empirical SEs of $\hat{\sigma}_W^2$. From both panels it can be seen that SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ are very close when random block effects and errors are generated from normal and uniform distributions. However, SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ seem to be substantially higher when random effects are simulated from t_3 than when simulated from normal and uniform distributions. In summary, we conclude that relative biases and SEs in variance components estimates $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ are higher when random effects are simulated from the heavy-tailed distribution t_3 than normal and uniform distributions.

5.4 Additional Model Comparisons

In this section, we investigate the effect of misspecification in GLMMs on the estimated model parameters in terms of their biases and empirical SEs by comparing different models:

(*a*) $M2$ vs $M3$: models with unit errors and without errors

(*b*) $M2$ vs $M4$: models with blocks and without blocks

(*c*) $M2$ vs $M5$: fitting a GLMM compared to a GLM

We generated Poisson data from model $M2$ assuming correlated and normally distributed random block effects and errors. All models are fitted to the same set of simulated data under the assumption of normally distributed random effects. The summary of results obtained by fitting the models is presented in Table 5.4.

Comparison (*a*) investigates if not modelling the random errors affects the estimates of fixed treatment effect parameters ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$) and the variance component estimate for block effects ($\hat{\sigma}_B^2$). From Table 5.4, it can be seen that the estimated values for $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are very close to their true parameter values and hence the relative biases in percentage of these estimates are also very close to each other for both $M2$ (true model) and $M3$ (misspecified). However, empirical SEs of these estimates are substantially higher for the misspecified model ($M3$) than for the true model ($M2$) in all nine treatment-block combinations. It can also be seen that biases and SEs of the variance component estimate $\hat{\sigma}_B^2$ are not too different for models $M2$ and $M3$.

Table 5.4: Estimates, relative biases (R.Bias) and empirical SEs of Poisson GLMMs ($M2$ to $M4$) and GLM ($M5$) with $\sigma_B^2 = 1.44$ and $\sigma_W^2 = 0.16$.

			$\boldsymbol{M2 : \eta = X\beta + ZB + \epsilon}$			$\boldsymbol{M3 : \eta = X\beta + ZB}$			$\boldsymbol{M4 : \eta = X\beta + \epsilon}$			$\boldsymbol{M5 : \eta = X\beta}$		
t	b	**True**	**Est.**	**R.Bias(%)**	**SE**	**Est.**	**R.Bias(%)**	**SE**	**Est.**	**R.Bias(%)**	**SE**	**Est**	**R.Bias(%)**	**SE**
2	10	$\alpha_1 = -0.371$	-0.378	1.977	0.154	-0.383	3.103	0.199	-0.354	-4.740	0.169	-0.383	3.127	0.199
		σ_B^2	1.530	6.221	0.828	1.531	6.353	0.813	-	-	-	-	-	-
		σ_W^2	0.161	0.680	0.111	-	-	-	0.340	112.647	0.247	-	-	-
	20	$\alpha_1 = -0.482$	-0.480	-0.434	0.101	-0.477	-1.092	0.152	-0.431	-10.627	0.121	-0.477	-1.092	0.152
		σ_B^2	1.459	1.348	0.503	1.477	2.537	0.504	-	-	-	-	-	-
		σ_W^2	0.165	3.431	0.068	-	-	-	0.371	131.710	0.174	-	-	-
	40	$\alpha_1 = -0.325$	-0.325	-0.011	0.077	-0.324	-0.539	0.121	-0.291	-10.528	0.092	-0.324	-0.539	0.121
		σ_B^2	1.498	4.029	0.404	1.496	3.884	0.399	-	-	-	-	-	-
		σ_W^2	0.167	4.104	0.056	-	-	-	0.326	103.697	0.124	-	-	-
3	15	$\alpha_1 = -0.541$	-0.549	1.554	0.139	-0.546	0.975	0.183	-0.550	1.831	0.198	-0.546	0.970	0.183
		$\alpha_2 = 0.000$	0.004	0.355	0.137	0.004	0.383	0.183	0.040	4.031	0.166	0.004	0.384	0.183
		σ_B^2	1.547	7.411	0.646	1.534	6.503	0.636	-	-	-	-	-	-
		σ_W^2	0.165	3.185	0.067	-	-	-	0.521	225.710	0.305	-	-	-
	30	$\alpha_1 = -0.371$	-0.377	1.667	0.110	-0.373	0.670	0.143	-0.372	0.410	0.139	-0.373	0.670	0.143
		$\alpha_2 = 0.000$	0.004	0.375	0.105	0.004	0.400	0.143	0.022	2.228	0.122	0.004	0.400	0.143
		σ_B^2	1.545	7.309	0.463	1.512	5.025	0.452	-	-	-	-	-	-
		σ_W^2	0.166	3.688	0.056	-	-	-	0.438	173.640	0.204	-	-	-
	60	$\alpha_1 = -0.258$	-0.260	0.747	0.081	-0.255	-1.102	0.109	-0.254	-1.633	0.100	-0.255	-1.102	0.109
		$\alpha_2 = 0.000$	-0.001	-0.095	0.080	-0.007	-0.726	0.106	0.004	0.425	0.094	-0.007	-0.726	0.106
		σ_B^2	1.550	7.672	0.340	1.503	4.349	0.330	-	-	-	-	-	-
		σ_W^2	0.169	5.410	0.045	-	-	-	0.406	153.702	0.144	-	-	-
4	20	$\alpha_1 = -0.454$	-0.461	1.541	0.138	-0.459	1.056	0.168	-0.487	7.333	0.207	-0.459	1.056	0.168
		$\alpha_2 = -0.151$	-0.145	-3.942	0.133	-0.149	-1.767	0.168	-0.118	-22.245	0.172	-0.149	-1.767	0.168
		$\alpha_3 = 0.151$	0.155	2.573	0.130	0.156	2.775	0.171	0.170	12.532	0.149	0.156	2.775	0.171
		σ_B^2	1.528	6.139	0.538	1.489	3.415	0.525	-	-	-	-	-	-
		σ_W^2	0.165	3.358	0.060	-	-	-	0.549	243.077	0.319	-	-	-
	40	$\alpha_1 = -0.316$	-0.323	2.135	0.105	-0.318	0.459	0.134	-0.333	5.145	0.142	-0.318	0.459	0.134
		$\alpha_2 = -0.105$	-0.106	0.444	0.100	-0.108	2.180	0.129	-0.094	-10.785	0.125	-0.108	2.180	0.129
		$\alpha_3 = 0.105$	0.111	4.772	0.098	0.109	3.151	0.126	0.119	12.720	0.110	0.109	3.151	0.126
		σ_B^2	1.531	6.293	0.406	1.477	2.550	0.394	-	-	-	-	-	-
		σ_W^2	0.168	4.786	0.048	-	-	-	0.492	207.350	0.417	-	-	-
	80	$\alpha_1 = -0.222$	-0.223	0.281	0.080	-0.224	0.756	0.101	-0.227	2.240	0.100	-0.224	0.756	0.101
		$\alpha_2 = -0.074$	-0.074	-0.623	0.079	-0.076	2.917	0.099	-0.068	-7.596	0.095	-0.076	2.917	0.099
		$\alpha_3 = 0.074$	0.072	-2.783	0.071	0.073	-1.955	0.091	0.075	1.370	0.080	0.073	-1.955	0.091
		σ_B^2	1.521	5.605	0.279	1.455	1.042	0.269	-	-	-	-	-	-
		σ_W^2	0.170	6.141	0.037	-	-	-	0.451	181.797	0.158	-	-	-

In (b), we compare models with blocks $(M2)$ and without blocks $(M4)$ to investigate the impact of block effects on the estimated treatment effect parameters $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$, and on the estimate $\hat{\sigma}^2_W$ of the error variance component. Table 5.4 shows that biases and empirical SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are slightly higher for $M4$ (misspecified model) than for $M2$ (true model) in most of the treatment-block combinations. However, the estimate $\hat{\sigma}^2_B$ seems to be severely biased and empirical SEs of this estimate are substantially higher for $M4$ than for $M2$ in all nine treatment-block combinations.

In (c), we investigate the impact of both random effects (blocks and errors) on the estimated fixed treatment effect parameters by comparing the models $M2$ and $M5$. It can be seen from Table 5.4 that the estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are not affected in terms of their biases when we fit the Poisson GLM $(M5)$ with fixed treatment effects to the data generated from the Poisson GLMM $(M2)$ with correlated random effects (blocks and errors). However, empirical SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are substantially higher for $M5$ (misspecified model) than for $M2$ (true model) in all nine treatment-block combinations.

5.5 Conclusion

We have investigated the impact of models misspecification in GLMMs setup via simulation studies on the estimated model parameters in terms of their biases and empirical SEs. We studied the misspecification of the random effects distribution, as done by most previous studies. However, unlike these studies, we also looked at misspecification of the correlation structure of the random effects as being derived from the randomization for the RB-GLMM with random block effects.

From the results in Section 5.3 we conclude that misspecification of both the correlation structure and the distribution of the random effects has hardly any effect on estimates of fixed treatment effect parameters ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$) in terms of their biases and empirical SEs. Similar results were reported by Neuhaus et al. (2013) and McCulloch and Neuhaus (2011). However, it has shown by Litière et al. (2008) that maximum likelihood estimates of fixed effect parameters are

inconsistent due to misspecification of the random effects distribution which is not in agreement with our findings.

The estimates of variance components ($\hat{\sigma}^2_B$, $\hat{\sigma}^2_W$) are frequently severely biased and their SEs are mostly higher due to misspecification of both random effects distribution and correlation structure. The estimated variance components were also found to be biased by Litière et al. (2008) and Neuhaus et al. (2011, 2013) when the misspecification of the random effects distribution was present.

The additional comparisons in Section 5.4 show that the fixed treatment effect estimates are affected when we do not consider the random errors. This is because SEs of these estimates are found to be substantially higher for the misspecified model than for the true model. However, the variance component estimate $\hat{\sigma}^2_B$ is not affected in terms of bias and SEs. Moreover, we conclude that the estimated fixed treatment effect parameters and the error variance component $\hat{\sigma}^2_W$ are found to be biased and SEs of these estimates are significantly higher when we ignore the block effects in the model fitting. The estimated fixed effect parameters are not affected in terms of their biases when we fit the GLM ($M5$) to the data generated from the GLMM ($M2$). However, empirical SEs of these estimates are found to be substantially higher for the misspecified model ($M5$) than for the true model ($M2$). To the best of our knowledge, such types of additional comparisons of models misspecification have not been considered by any previous studies.

Chapter 6

Conclusion and Future Work

6.1 Discussion and Conclusion

GLMMs are commonly used where random effects are incorporated for modelling the correlation or overdispersion in the data. For simplicity in the analysis, random effects are often assumed to follow a normal distribution. However, the estimated model parameters obtained by assuming normality for the random effects distribution can give misleading conclusions when the actual distribution of the random effects does not follow the normal distribution.

It has been suggested that misspecification of the random effects distribution in GLMMs may lead to

(i) bias in the estimated fixed effects parameters and variance components (Heagerty and Kurland (2001), Agresti et al. (2004), Litière et al. (2008)),

(ii) incorrect type-I and type-II error rates (Litière et al. (2007, 2008)),

(iii) reduce the power of statistical test (e.g., Wald test) of the model parameters of interest (Litière et al., 2007, 2008).

The main motivation of the thesis is to develop a statistical method for estimating the model parameters in GLMMs without assuming any specific parametric form of the random effects distribution. The thesis focuses on the randomization justification for deriving certain GLMMs associated with underlying designs, in particular the CRD and the RCBD. In this framework, no parametric

distribution is assumed for the random effects arising from the randomization for estimating the model parameters. As a result the randomization-based approach for estimating the model parameters may be an useful alternative when the normality assumption for the random effects distribution is not valid.

A GLMM based on the randomization approach for the CRD, RB-GLMM (2.26), is derived in Chapter 2. This model is also known as the individual level random effects GLMM, as can be seen in Rabe-Hesketh and Skrondal (2012, p.706-707) where the number of observations is equal to the number of random errors. Rabe-Hesketh and Skrondal (2012, p.706-707) considered this model for analyzing the data with overdispersion using the assumption of normality for the random effects distribution. Moreover, the results of the derived moments for the random effects in Section 2.1.5 show that there is an exchangeable (compound symmetric) correlation structure among the random effects. The same correlation structure can be found in Lee et al. (2006, p.256) for modelling the correlation arising due to common family-environment and genetic effects in family data. Moreover, we derive the likelihood function (2.24) for the RB-GLMM for the CRD in such a way that no parametric distribution for the random effects is required. Rather, the discrete uniform distribution on the symmetric group S_n is used for the random effects.

As the summation over all possible permutations in the symmetric group S_n is involved in the likelihood function, the direct maximization of the likelihood function is complicated as the sum does not commute with the natural logarithm. We develop an algorithm for estimating the model parameters of the RB-GLMM for the CRD in Chapter 3, where we maximize a minorization function rather than the log-likelihood function. The randomization-based algorithm combines the IWLS algorithm for estimating the fixed treatment effects parameters and BLP for predicting the random effects. We conduct simulation studies to assess the estimates of the model parameters for small samples. The results show that when there is a misspecification of the random effects distribution, the randomization-based algorithm gives better estimates of the model parameters in most cases than the standard GLMMs where normality is assumed for the random effects distribution. However, the randomization-based algorithm can only be applied for small samples at this stage because of computational limitations, as it requires all $n!$ permutations in the symmetric group S_n.

In Chapter 4, we extend the randomization approach of Brien and Bailey (2006) for deriving linear models for the RCBD with random block effects to GLMMs, and the resulting model (4.19) is found to be a GLMM with two random effects, one for random block effects and the other for the random unit errors (Section 1.4.2). The variance-covariance matrices are also derived from the randomization and we show that the random block effects and errors are correlated due to randomization.

In Chapter 5, we conduct simulation studies in order to investigate the impact of misspecification of the random effects distribution and of the correlation structure among the random effects. The data are simulated from the GLMM for the RCBD with random block effects, derived in Chapter 4, considering different random effects distributions. However, in the analysis different parametric models are fitted to these data with standard assumptions for the random effects, such as normality and no correlation. The simulation results show that the estimates of the fixed treatment effect parameters are not affected in terms of their biases and standard errors due to misspecification of both the correlation structure and of the random effects distribution. Similar results were found by McCulloch and Neuhaus (2011) and Neuhaus et al. (2013). However, the estimates of the variance components are severely biased due to misspecification of the correlation structure. Moreover, when the random effects were generated from t_3 distribution while, in the analysis, these are assumed to be normally distributed, the estimates of variance components are also affected in terms of the biases and standard errors for the misspecified random effects distribution. Neuhaus et al. (2011, 2013) also reported similar results.

6.2 Future Work

We implemented the randomization-based algorithm for the RB-GLMM for the CRD for small samples in Chapter 3 using all $n!$ permutations in the symmetric group S_n. However, because of computational limitations, as described in Section 3.9, the algorithm becomes infeasible for large samples as the number of permutations in S_n, $n!$, rapidly increases as n, the sample size, increases. As a result, one can implement the randomization-based algorithm for large samples using a random subset of permutations instead of all possible permu-

tations in S_n.

In the thesis, we have considered two types of randomization for the RCBD following Brien and Bailey (2006) for random block effects using the wreath product $S_t \wr S_b$ and Kempthorne (1955) for fixed block effects using the direct product S_t^b, described in Section 1.4.2. Extending these randomization approaches for deriving linear models to GLMMs, we derived the RB-GLMMs for the RCBD with both random and fixed block effects, moments of the random effects and the likelihood functions.

It is possible to extend the randomization-based estimation algorithm of the RB-GLMM for the CRD, to RB-GLMMs for the RCBD with both random and fixed block effects. More precisely, one could derive the minorization functions for the likelihood functions C.3 and C.4, similar to Lemma 8 in Section 3.2, as before. One could then maximize these minorization functions for estimating the RB-GLMMs parameters combining IWLS algorithm and BLP of the random effects, as done similarly for the RB-GLMM for the CRD in Section 3.6. However, at the moment, it does not seem to be possible to implement this approach in a practically useful way because of computational constraints.

Moreover, in principle, one could consider more complex experimental designs for which other groups will be needed. One could then derive the RB-GLMMs based on the randomization approach for these designs. One could also then derive the likelihood functions and minorization functions for the RB-GLMMs. Based on that one could extend the current randomization-based algorithm to develop algorithms for estimating the model parameters of the RB-GLMMs for the more complex designs.

One could also extend the randomization-based algorithm to a regression problem where multiple factors, including continuous and categorical, can be analyzed using an extended form of the design matrix. Finally, our plan is to develop an R package based on the randomization approach to have all relevant functions in the thesis, so that one can run the randomization-based algorithm for the data analysis.

Appendix A

Background Material

Background concepts and results related to my PhD project are described here. More specifically, the relevant concepts of groups and measure theory are described in Sections A.1 and A.2 respectively. The permanent of a square matrix is also described in Section A.4.

A.1 Groups

To describe the randomization approach for the models of interest, we consider three specific groups, namely, the *symmetric group* S_n of all permutations of n objects, the *direct product* S_t^b of b instances of S_t, and the *wreath product* $S_t \wr S_b$ of the *symmetric groups* S_t and S_b, which are particularly relevant to the PhD project. The summary description of these groups is given in the following sections.

We begin with a group. A group G is defined as a set with a binary operation $\circ$ which satisfies the following axioms for example, (Cameron, 1998):

(i) Closure: If $a, b \in G$ then $a \circ b \in G$ for all a and b.

(ii) Associativity: For all $a, b, c \in G$, $(a \circ b) \circ c = a \circ (b \circ c)$.

(iii) Identity: There exists an identity element $e \in G$ such that $e \circ a = a$ for every $a \in G$.

(iv) Inverse: For every $a \in G$ there exists an inverse element $a^{-1} \in G$ such that $a^{-1} \circ a = e$. The identity element e is unique, as is the inverse a^{-1}

for every $a \in G$.

A.1.1 Symmetric Group

Let $\Omega = \{1, 2, \ldots, n\}$ be the set of the first n natural numbers. The *symmetric group* S_n (see e.g., Cameron, 1998) is the set of all permutations of Ω with the operation $\circ$ being the composition of functions. That is $(\sigma \circ \pi)(i) = \sigma(\pi(i))$ for every $i \in \Omega$ and $\sigma, \pi \in S_n$. The number of elements of the *symmetric group* S_n is $n!$.

A.1.2 Direct Product of Symmetric Groups

Let S_t be the *symmetric group* of all permutations of the set $\{1, 2, \ldots, t\}$. The *direct product* S_t^b is the Cartesian product $S_t^b = S_t \times S_t \times \ldots \times S_t$ of b instances of S_t. The group elements are b-tuples $(\pi_1, \ldots, \pi_b)$ where $\pi_k \in S_t$ for $k = 1, 2, \ldots, b$. Suppose $\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)} \in S_t^b$ then the group operation on S_t^b is defined component wise as

$$\begin{aligned} \boldsymbol{\pi}^{(2)} \circ \boldsymbol{\pi}^{(1)} &= \left(\pi_1^{(2)}, \ldots, \pi_b^{(2)}\right) \circ \left(\pi_1^{(1)}, \ldots, \pi_b^{(1)}\right) \\ &= \left(\pi_1^{(2)}\pi_1^{(1)}, \ldots, \pi_b^{(2)}\pi_b^{(1)}\right), \end{aligned}$$

where for every $k = 1, \ldots, b$ and for every $i \in \{1, \ldots, t\}$ we have

$$\pi_k^{(2)}\pi_k^{(1)}(i) = \pi_k^{(2)}(\pi_k^{(1)}(i)).$$

Moreover, for every element $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_b) \in S_t^b$ the inverse is

$$\boldsymbol{\pi}^{-1} = (\pi_1, \ldots, \pi_b)^{-1} = \left(\pi_1^{-1}, \ldots, \pi_b^{-1}\right).$$

The number of elements of the *direct product* S_t^b is $|S_t^b| = t!^b$.

A.1.3 Wreath Product

Let S_t and S_b be the two symmetric groups of permutations of the sets $\{1, 2, \ldots, t\}$ and $\{1, 2, \ldots, b\}$ respectively. The elements of the wreath product $S_t \wr S_b$ (Cameron, 1999) are the elements of $S_t^b \times S_b$. Therefore, the number of elements of $S_t \wr S_b$ is $|S_t \wr S_b| = t!^b b!$. The wreath product $S_t \wr S_b$ consists of elements

$(\boldsymbol{\pi}, \delta)$ where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_b) \in S_t^b$ and $\delta \in S_b$. Let $\left(\boldsymbol{\pi}^{(1)}, \delta_1\right), \left(\boldsymbol{\pi}^{(2)}, \delta_2\right) \in S_t \wr S_b$ then the group operation on $S_t \wr S_b$ is defined as

$$\begin{aligned}\left(\boldsymbol{\pi}^{(2)}, \delta_2\right) \circ \left(\boldsymbol{\pi}^{(1)}, \delta_1\right) &= \left(\pi_1^{(2)}, \ldots, \pi_b^{(2)}, \delta_2\right) \circ \left(\pi_1^{(1)}, \ldots, \pi_b^{(1)}, \delta_1\right) \\ &= \left(\pi_1^{(2)} \pi_{\delta_2(1)}^{(1)}, \ldots, \pi_b^{(2)} \pi_{\delta_2(b)}^{(1)}, \delta_2 \delta_1\right).\end{aligned}$$

Moreover, the inverse of every element $(\boldsymbol{\pi}, \delta)$ is given by

$$(\boldsymbol{\pi}, \delta)^{-1} = (\pi_1, \ldots, \pi_b, \delta)^{-1} = \left(\pi_{\delta^{-1}(1)}^{-1}, \ldots, \pi_{\delta^{-1}(b)}^{-1}, \delta^{-1}\right).$$

A.2 Measure Theory

Some measure theoretic concepts and results are particularly relevant to my PhD project and these are briefly summarized below.

A.2.1 Measure Space

A fundamental notion in measure theory is the measure space. It is a triple $(\Omega, \mathscr{F}, M)$ where Ω is a non-empty set, $\mathscr{F}$ is a σ-field on Ω and M is a measure on $\mathscr{F}$. A σ-field $\mathscr{F}$ (Billingsley, 1985, p.17) is defined as a subset of the power set $\mathscr{P}(\Omega)$ such that

(a) $\Omega \in \mathscr{F}$;

(b) If $A \in \mathscr{F}$ then the complement of A is in $\mathscr{F}$, i.e. $A^{\complement} \in \mathscr{F}$;

(c) If a sequence $A_1, A_2, \ldots \in \mathscr{F}$ then the union $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

It follows that the intersection of a sequence $A_1, A_2, \ldots$ of elements of a σ-field $\mathscr{F}$ is also an element of $\mathscr{F}$, that is

$$\bigcap_{i=1}^{\infty} A_i \in \mathscr{F}. \tag{A.1}$$

The pair $(\Omega, \mathscr{F})$ is called a measurable space.

Usually a special σ-field is used for the real numbers which is known as the Borel σ-field (Billingsley, 1985, p.155). The Borel σ-field $\mathscr{B}(\mathbb{R})$ can be defined

as the smallest σ-field on $\mathbb{R}$ which contains all intervals (a, b), where $a, b \in \mathbb{R}$. The elements of the Borel σ-field are called Borel sets.

Now let $(\Omega, \mathscr{F})$ be a measurable space. Then a function $M : \mathscr{F} \to \mathbb{R}$ is a measure (Billingsley, 1985, p.157) if

(a) $M(A) \geq 0$ for every $A \in \mathscr{F}$;

(b) $M(\emptyset) = 0$, where $\emptyset$ is the empty set;

(c) If $A_1, A_2, \ldots \in \mathscr{F}$ is a disjoint sequence with $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$M\left(\bigcup_{i=1}^{\infty} A_i)\right) = \sum_{i=1}^{\infty} M(A_i).$$

A measure P is said to be a probability measure if $P(\Omega) = 1$ and the triple $(\Omega, \mathscr{F}, P)$ is then called a probability space.

A.2.2 Measurable Function

Let $(\Omega, \mathscr{F})$ and $(\Gamma, \mathscr{G})$ be two measurable spaces. A function $f : \Omega \to \Gamma$ is a said to be $\mathscr{F}$-$\mathscr{G}$-measurable (Billingsley, 1985, p.182) if for every $G \in \mathscr{G}$

$$f^{-1}(G) \in \mathscr{F}, \tag{A.2}$$

where $f^{-1}(G) = \{\omega \in \Omega : f(\omega) \in G\}$.

Let $(\Omega, \mathscr{F}, P)$ be a probability space. A real random variable (Billingsley, 1985, p.183) is an $\mathscr{F}$-$\mathscr{B}(\mathbb{R})$ measurable function $X : \Omega \to \mathbb{R}$.

A.2.3 σ-field Generated by a Set and the Smallest σ-field

Let Ω be a set and S be a non-empty subset of the power set $\mathscr{P}(\Omega)$ i.e. $S \subseteq \mathscr{P}(\Omega)$. The σ-field $\mathscr{F}(S)$ generated by S is the intersection of all the σ-fields containing S (Billingsley, 1985, p.19) i.e.

$$\mathscr{F}(S) = \bigcap_{\substack{\mathscr{F}:\mathscr{F} \text{ is a } \sigma-\text{field} \\ \text{on } \Omega \text{ with } S \subseteq \mathscr{F}}} \mathscr{F}. \tag{A.3}$$

Moreover, $\mathscr{F}(S)$ is the smallest σ-field containing S. Note that $\mathscr{F}(S) \subseteq \mathscr{F}$ for all $\mathscr{F}$ with $S \subseteq \mathscr{F}$.

A.2.4 σ-field Generated by Random Variables

Suppose $(\Omega, \mathscr{F}, P)$ is a probability space. Let $X : \Omega \to \mathbb{R}$ be a real random variable. The σ-field generated by X is denoted by $\mathscr{F}(X)$ and defined as (Billingsley, 1985, p.64)

$$\mathscr{F}(X) = \{X^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}. \tag{A.4}$$

Since by definition we have $X^{-1}(B) \in \mathscr{F}$ for every $B \in \mathscr{B}(\mathbb{R})$ it follows that $\mathscr{F}(X) \subseteq \mathscr{F}$.

Now let $X_i : \Omega \to \mathbb{R}$ be a real random variable for $i = 1, \ldots, n$. Then the σ-field generated by the vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is denoted by $\mathscr{F}(\boldsymbol{X})$ (Billingsley, 1985, p.260) and defined as the σ-field $\mathscr{F}(S)$ generated by $S = \bigcup_{i=1}^{n} \mathscr{F}(X_i)$, where $\mathscr{F}(X_i)$ is the σ-field generated by X_i. It is the smallest σ-field for which every X_i is measurable.

A.2.5 General Definition of Conditional Expectation

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ be a real random variable then the expectation of X (Billingsley, 1985, p.280) is defined as

$$E(X) = \int_\Omega X(\omega) P(d\omega).$$

Suppose $\mathscr{F}_0 \subseteq \mathscr{F}$ is a sub-σ-field of $\mathscr{F}$, then the conditional expectation of X given $\mathscr{F}_0$, denoted by $E(X|\mathscr{F}_0)$ (Billingsley, 1985, p.466), is any $\mathscr{F}_0$-$\mathscr{B}(\mathbb{R})$-measurable random variable $X_0 : \Omega \to \mathbb{R}$ such that

$$\int_H X_0(\omega) P(d\omega) = \int_H X(\omega) P(d\omega) \text{ for every } H \in \mathscr{F}_0. \tag{A.5}$$

It is important to note that $E(X|\mathscr{F}_0)$ is a random variable. Moreover, any two random variables, say X_0 and X_0', which satisfy (A.5) are equal with

probability one. We are particularly interested in situations where the σ-field $\mathscr{F}_0$ is generated by a vector $X = (X_1, \ldots, X_n)^\top$ of random variables.

A.2.6 Conditional Expectation When $\mathscr{F}_0$ is Generated by a Partition of Ω

Finding the conditional expectation $E(X|\mathscr{F}_0)$ of a random variable X on a probability space $(\Omega, \mathscr{F}, P)$ can be facilitated if the sub-σ-field $\mathscr{F}_0$ of $\mathscr{F}$ is generated by a partition of Ω into mutually disjoint sets. The conditional expectation of X given $\mathscr{F}_0$ can then be computed as follows (Billingsley, 1985, p.467).

Theorem 2. *Let X be a real random variable on a probability space $(\Omega, \mathscr{F}, P)$ and $S = \{Z_1, \ldots, Z_m\}$ be a partition of Ω with $P(Z_l) \geq 0$ for every $l = 1, \ldots, m$. Furthermore, let $\mathscr{F}_0 = \mathscr{F}(S)$ be the σ-field generated by S. Then the conditional expectation of X given $\mathscr{F}_0$ is*

$$E(X|\mathscr{F}_o) = \sum_{l=1}^{m} E_{Z_l}(X) 1_{Z_l}, \tag{A.6}$$

where $E_{Z_l}(X) = \dfrac{1}{P(Z_l)} \int_{Z_l} X(\omega) P(d\omega)$ and 1_{Z_l} is the indicator function.

A.2.7 Factorization Lemma

The factorization lemma (Lehmann, 1959, p.37) which is particularly relevant for our model derivations and applied in Section 2.1 is given below.

Lemma 18. *Suppose that $f : \Omega \to \mathbb{R}$ and $g : \Omega \to G$ are functions and that $(G, \mathscr{G})$ is a measurable space. Let $\mathscr{F}(g)$ be the σ-field generated by g. Then f is $\mathscr{F}(g)$-$\mathscr{B}(\mathbb{R})$-measurable if and only if there exists a $\mathscr{G}$-$\mathscr{B}(\mathbb{R})$-measurable $h : G \to \mathbb{R}$ such that*

$$f = h \circ g, \tag{A.7}$$

where $\mathscr{B}(\mathbb{R})$ is the Borel σ-field.

A.2.8 Measure with Density

Let $(\Omega, \mathscr{F}, N)$ be a measure space. Let f be a non-negative function. By using f another measure M on the measurable space $(\Omega, \mathscr{F})$ can be defined as

(Billingsley, 1985, p.216)

$$M(A) = \int_A f(w)N(dw) \quad \text{for every } A \in \mathscr{F}. \tag{A.8}$$

In general, a measure M for which equation (A.8) holds is said to have a density f with respect to the measure N. The measure M is a probability measure if $M(\Omega) = 1$. The question whether a measure M on $(\Omega, \mathscr{F})$ has a density with respect to a measure N on $(\Omega, \mathscr{F})$ is answered by the Radon-Nikodym Theorem (Billingsley, 1985, p.443). However, in the thesis we are only interested in situations where M is *defined* by (A.8). If M has a density then integrals of the function g with respect to M can be calculated as integrals of gf with respect to N, where f is the density in equation (A.8). More precisely, we have the following result (Billingsley, 1985, p.217).

Theorem 3. *If M has a density f with respect to N, then*

$$\int_\Omega g(w)M(dw) = \int_\Omega g(w)f(w)N(dw) \tag{A.9}$$

holds for every non-negative function g.

An important application of Theorem 3 occurs when we want to compute the expectation of a random variable X on a probability space $(\Omega, \mathscr{F}, D)$, where D has a density f with respect to another probability measure P on the measurable space $(\Omega, \mathscr{F})$. The expectation $E_D(X)$ of X with respect to D can then be computed as

$$E_D(X) = \int_\Omega X(\omega)f(\omega)P(d\omega), \tag{A.10}$$

where, as explained in Section 3.2, the subscript D indicates that the expectation is taken with regard to D.

The above definition of a measure with density (A.8) and the result (A.9) are used to construct the minorization function for the log-likelihood function in Section 3.2 .

A.3 Useful Lemma

In the derivations of RB-GLMMs, the relevant σ-field is generated by a vector of random variables. Lemma 19 is formulated in a way that unifies similar

arguments that are used in the derivation of all three specific RB-GLMMs for the CRD and RCBD with both random and fixed block effects.

Lemma 19. *Let S be a finite non-empty set and $\mathscr{F}$ be a σ-field on S. Also let $\tilde{V}_1, \ldots, \tilde{V}_m$ be real random variables on the measurable space $(S, \mathscr{F})$ and Ω be a non-empty set. Assume that $\{s\} \in \mathscr{F}$ for every $s \in S$. For $i = 1, \ldots, m$ define $V_i : S \times \Omega \to \mathbb{R}$ by $V_i(s, \omega) = \tilde{V}_i(s)$ for every $(s, \omega) \in S \times \Omega$. Let $\mathscr{F}(\boldsymbol{V})$ be the σ-field generated by the vector $\boldsymbol{V} = (V_1, \ldots, V_m)^\top$ and define $\tilde{\mathscr{F}} = \{A \times \Omega : A \in \mathscr{F}\}$. Then*

(i) $\tilde{\mathscr{F}}$ is a σ-field on $S \times \Omega$,

(ii) Every V_i is $\tilde{\mathscr{F}}$-$\mathscr{B}(\mathbb{R})$-measurable, where $\mathscr{B}(\mathbb{R})$ is the Borel σ-field,

(iii) $\tilde{\mathscr{F}}$ is the σ-field generated by the set $\{\{s\} \times \Omega : s \in S\}$ and

(iv) $\mathscr{F}(\boldsymbol{V}) \subseteq \tilde{\mathscr{F}}$.

Proof. (i) To show that $\tilde{\mathscr{F}}$ is a σ-field it is necessary to check all three properties of a σ-field (Section A.2.1): (a) Since $\mathscr{F}$ is a σ-field on S it contains S. Hence $S \times \Omega \in \tilde{\mathscr{F}}$; (b) Let $A \times \Omega \in \tilde{\mathscr{F}}$. Then $(A \times \Omega)^{\complement} = \{(s, \omega) \in S \times \Omega : (s, \omega) \notin A \times \Omega\} = \{s \in S : s \notin A\} \times \Omega = A^{\complement} \times \Omega$. Since $\mathscr{F}$ is a σ-field so $A \in \mathscr{F}$ implies that $A^{\complement} \in \mathscr{F}$ and hence $A^{\complement} \times \Omega \in \tilde{\mathscr{F}}$; (c) Let $A_1 \times \Omega, A_2 \times \Omega, \ldots$ be a sequence of elements of $\tilde{\mathscr{F}}$. Then $\bigcup_{i=1}^{\infty}(A_i \times \Omega) = (\bigcup_{i=1}^{\infty} A_i) \times \Omega$. Since each of $A_1, A_2, \ldots$ is an element of $\mathscr{F}$ and $\mathscr{F}$ is a σ-field, it follows that $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ and so $\bigcup_{i=1}^{\infty}(A_i \times \Omega) \in \tilde{\mathscr{F}}$. Hence $\tilde{\mathscr{F}}$ is a σ-field on $S \times \Omega$.

(ii) Let $B \in \mathscr{B}(\mathbb{R})$. We need to show that $V_i^{-1}(B) \in \tilde{\mathscr{F}}$. Now by definition we can write $V_i^{-1}(B) = \{(s, \omega) \in S \times \Omega : V_i(s, \omega) \in B\} = \{(s, \omega) \in S \times \Omega : \tilde{V}_i(s) \in B\} = \{s \in S : \tilde{V}_i(s) \in B\} \times \Omega = \tilde{V}_i^{-1}(B) \times \Omega$. Since $\tilde{V}_i$ is a real random variable on $(S, \mathscr{F})$, and hence $\mathscr{F}$-$\mathscr{B}(\mathbb{R})$- measurable, it follows that $\tilde{V}_i^{-1}(B) \in \mathscr{F}$. Therefore $V_i^{-1}(B) \in \tilde{\mathscr{F}}$ and hence every V_i is $\tilde{\mathscr{F}}$-$\mathscr{B}(\mathbb{R})$-measurable.

(iii) For every $s \in S$ we have $\{s\} \times \Omega \in \tilde{\mathscr{F}}$ since by assumption $\{s\} \in \mathscr{F}$. Hence the σ-field generated by the set $\{\{s\} \times \Omega : s \in S\}$ is a subset of the σ-field $\tilde{\mathscr{F}}$. Conversely, if $A \times \Omega \in \tilde{\mathscr{F}}$ then $A \times \Omega = \left(\bigcup_{a \in A} \{a\}\right) \times \Omega = \bigcup_{a \in A}(\{a\} \times \Omega)$ is an element of the σ-field generated by the set $\{\{s\} \times \Omega : s \in S\}$ since S and, hence, A is finite. So $\tilde{\mathscr{F}}$ is a subset of

the σ-field generated by the set $\{\{s\} \times \Omega : s \in S\}$. Thus $\tilde{\mathscr{F}}$ is equal to the σ-field generated by the set $\{\{s\} \times \Omega : s \in S\}$.

(*iv*) From parts (*i*) and (*ii*) it follows that the σ-field $\mathscr{F}(\boldsymbol{V})$ generated by the vector $\boldsymbol{V}$ is a subset of $\tilde{\mathscr{F}}$ i.e., $\mathscr{F}(\boldsymbol{V}) \subseteq \tilde{\mathscr{F}}$ as we know $\mathscr{F}(\boldsymbol{V})$ is the smallest σ-field for which every V_i is measurable (Section A.2.4).

■

A.4 Permanent of a Matrix

The permanent of a square matrix is related to the likelihood function of the RB-GLMM for the CRD as can be seen in Corollary 3 (Section 2.2.1). For an $n \times n$ matrix $\boldsymbol{A} = (a_{i,j})$ the permanent of $\boldsymbol{A}$ is defined as (Marcus and Minc, 1964, p.18)

$$\text{per}\boldsymbol{A} = \sum_{\pi \in S_n} a_{\pi(1),1}\, a_{\pi(2),2} \ldots a_{\pi(n),n} = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{\pi(i),i}, \tag{A.11}$$

where S_n is the symmetric group of all permutations π of the set $\{1, 2, \ldots, n\}$. It can be shown that $\text{per}\,\boldsymbol{A} = \text{per}\,\boldsymbol{A}^\top$ and so the permanent can equivalently be defined by

$$\text{per}\boldsymbol{A} = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{i,\pi(i)}$$

which is also common in many books. On the other hand, the determinant of $\boldsymbol{A}$ is defined by (Marcus and Minc, 1964, p.12)

$$\det\boldsymbol{A} = \sum_{\pi \in S_n} \text{sign}(\pi) \prod_{i=1}^{n} a_{\pi(i),i}, \tag{A.12}$$

where $\text{sign}(\pi) = \pm 1$. The sign of the permutation is computed by the number of interchanges $(even, odd)$ to obtain natural order as: $\text{sign}(\pi) = +1$ for *even* and $\text{sign}(\pi) = -1$ for *odd*.

Appendix B

Proof of Lemmas 11 and 12

Proof of Lemma 11 . Let $M_{il} = E(X_{i,j_l})$ and $V_{il} = Var(X_{i,j_l})$. Recall that

$$\begin{aligned}\boldsymbol{\Sigma} &= \frac{1}{(n-1)n}\Big((n-1)\,\mathrm{diag}(\boldsymbol{V}^\top \mathbf{1}_n) + n\,\mathrm{diag}(\boldsymbol{a_M}) - \mathrm{diag}^2(\boldsymbol{M}^\top \mathbf{1}_n) \\ &\quad -\boldsymbol{M}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{M}\Big). \end{aligned} \tag{B.1}$$

We have that

$$\mathrm{diag}(\boldsymbol{V}^\top \mathbf{1}_n) = \begin{pmatrix} \sum_{i^*=1}^{n} v_{i^*1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sum_{i^*=1}^{n} v_{i^*n} \end{pmatrix}, \tag{B.2}$$

$$\mathrm{diag}(\boldsymbol{a_M}) = \begin{pmatrix} \sum_{i^*=1}^{n} M_{i^*1}^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sum_{i^*=1}^{n} M_{i^*n}^2 \end{pmatrix}, \tag{B.3}$$

$$\mathrm{diag}^2(\boldsymbol{M}^\top \mathbf{1}_n) = \begin{pmatrix} \left(\sum_{i^*=1}^{n} M_{i^*1}\right)^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \left(\sum_{i^*=1}^{n} M_{i^*n}\right)^2 \end{pmatrix} \tag{B.4}$$

and

$$\boldsymbol{M}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{M} = \boldsymbol{G}\boldsymbol{M} \tag{B.5}$$

with

$$\boldsymbol{G} = \boldsymbol{M}^\top (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top).$$

Now the elements of $\boldsymbol{G}$ are

$$\begin{aligned}
\boldsymbol{G}_{il} &= \frac{1}{n}\left((n-1)M_{li} - \sum_{k\neq l}^{n} M_{ki}\right) \\
&= \frac{1}{n}\left((n-1)M_{li} + M_{li} - M_{li} - \sum_{k\neq l}^{n} M_{ki}\right) \\
&= \frac{1}{n}\left(nM_{li} - \sum_{i^*=1}^{n} M_{i^*i}\right) = M_{li} - \frac{1}{n}\sum_{i^*=1}^{n} M_{i^*i}. \qquad \text{(B.6)}
\end{aligned}$$

Therefore, the elements of $\boldsymbol{GM}$ are

$$\begin{aligned}
(\boldsymbol{GM})_{il} &= \sum_{i^*=1}^{n} G_{ii^*}M_{i^*l} \\
&= \sum_{i^*=1}^{n}\left(M_{i^*i} - \frac{1}{n}\sum_{i^*=1}^{n} M_{i^*i}\right)M_{i^*l} \\
&= \sum_{i^*=1}^{n} M_{i^*i}M_{i^*l} - \frac{1}{n}\left(\sum_{i^*=1}^{n} M_{i^*i}\right)\left(\sum_{i^*=1}^{n} M_{i^*l}\right). \qquad \text{(B.7)}
\end{aligned}$$

For $i = l$ it follows that

$$(\boldsymbol{GM})_{ii} = \sum_{i^*=1}^{n} M_{i^*i}^2 - \frac{1}{n}\left(\sum_{i^*=1}^{n} M_{i^*i}\right)^2. \qquad \text{(B.8)}$$

Therefore, the diagonal elements of $\mathbf{\Sigma}$ (B.1) are

$$\begin{aligned}
\mathbf{\Sigma}_{ii} &= \frac{1}{(n-1)n}\Bigg[(n-1)\sum_{i^*=1}^{n} V_{i^*i} + n\sum_{i^*=1}^{n} M_{i^*i}^2 - \left(\sum_{i^*=1}^{n} M_{i^*i}\right)^2 \\
&\quad - \sum_{i^*=1}^{n} M_{i^*i}^2 + \frac{1}{n}\left(\sum_{i^*=1}^{n} M_{i^*i}\right)^2\Bigg] \\
&= \frac{1}{n}\sum_{i^*=1}^{n}\left(V_{i^*i} + M_{i^*i}^2\right) - \frac{1}{n^2}\left(\sum_{i^*=1}^{n} M_{i^*i}\right)^2 \\
&= \frac{1}{n}\sum_{i^*=1}^{n}\left(Var(X_{i^*,j_i}) + (E(X_{i^*,j_i}))^2\right) - \left(\frac{1}{n}\sum_{i^*=1}^{n} E(X_{i^*,j_i})\right)^2 \\
&= \frac{1}{n}\sum_{i^*=1}^{n}\left(Var(X_{i^*,j_i}) + (E(X_{i^*,j_i}))^2\right) - m_{j_i}^2 \\
&= Var(Y_{i,j_i}) \qquad \text{(B.9)}
\end{aligned}$$

which are exactly the same as in Lemma 9(ii). Similarly, the off-diagonal elements of $\mathbf{\Sigma}$ for $k \neq l$ are

$$\begin{aligned}
\mathbf{\Sigma}_{kl} &= \frac{1}{(n-1)n}\left[-\left(\sum_{k^*=1}^{n} M_{k^*k}M_{k^*l} - \frac{1}{n}\left(\sum_{k^*=1}^{n} M_{k^*k}\right)\left(\sum_{k^*=1}^{n} M_{k^*l}\right)\right)\right] \\
&= \frac{1}{(n-1)n^2}\left(\sum_{k^*=1}^{n} M_{k^*k}\right)\left(\sum_{k^*=1}^{n} M_{k^*l}\right) - \frac{1}{(n-1)n}\sum_{k^*=1}^{n} M_{k^*k}M_{k^*l} \\
&= \frac{1}{(n-1)}\left(\frac{1}{n}\sum_{k^*=1}^{n} E(X_{k^*,j_k})\right)\left(\frac{1}{n}\sum_{k^*=1}^{n} E(X_{k^*,j_l})\right) \\
&\quad - \frac{1}{(n-1)n}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l}) \\
&= \frac{1}{n-1}m_{j_k}m_{j_l} - \frac{1}{n(n-1)}\sum_{k^*=1}^{n} E(X_{k^*,j_k})E(X_{k^*,j_l}) \\
&= Cov(Y_{k,j_k}, Y_{l,j_l})
\end{aligned}$$

which are also the same as in Lemma 9(iii).

■

Proof of Lemma 12 . Recall that $M_{il} = E(X_{i,j_l})$ and

$$\boldsymbol{C} = \frac{1}{(n-1)n}(n\,\mathrm{diag}(\boldsymbol{e}^\top \boldsymbol{M}) - \mathbf{1}_n \boldsymbol{e}^\top \boldsymbol{M}).$$

We have that

$$\mathrm{diag}(\boldsymbol{e}^\top \boldsymbol{M}) = \begin{pmatrix} \sum_{i^*=1}^{n} e_{i^*} M_{i^*1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sum_{i^*=1}^{n} e_{i^*} M_{i^*n} \end{pmatrix} \tag{B.10}$$

and

$$\mathbf{1}_{\boldsymbol{n}} \boldsymbol{e}^\top \boldsymbol{M} = \begin{pmatrix} \sum_{i^*=1}^{n} e_{i^*} M_{i^*1} & \cdots & \sum_{i^*=1}^{n} e_{i^*} M_{i^*n} \\ \vdots & & \vdots \\ \sum_{i^*=1}^{n} e_{i^*} M_{i^*1} & \cdots & \sum_{i^*=1}^{n} e_{i^*} M_{i^*n} \end{pmatrix}. \tag{B.11}$$

Therefore, the diagonal elements of $\boldsymbol{C}$ are

$$\begin{aligned} C_{ii} &= \frac{1}{(n-1)n}\left(n \sum_{i^*=1}^{n} e_{i^*} M_{i^*i} - \sum_{i^*=1}^{n} e_{i^*} M_{i^*i} \right) = \frac{1}{n} \sum_{i^*=1}^{n} e_{i^*} M_{i^*i} \\ &= \frac{1}{n} \sum_{i^*=1}^{n} e_{i^*} E(X_{i^*,j_i}). \end{aligned} \tag{B.12}$$

These are exactly the same as in Lemma 10(i). Similarly, the off-diagonal elements for $i \neq l$ are

$$C_{il} = -\frac{1}{(n-1)n} \sum_{i^*=1}^{n} e_{i^*} M_{i^*l} = -\frac{1}{(n-1)n} \sum_{i^*=1}^{n} e_{i^*} E(X_{i^*,j_l}), \tag{B.13}$$

which are also the same as in Lemma 10(ii). ■

Appendix C

RB-GLMM for the RCBD with Fixed Block Effects and Likelihood Functions

C.1 Summary of RB-GLMM for the RCBD with Fixed Block Effects

The GLMM for the RCBD with fixed block effects is

$$g\left(E(Y_{i,j_i,k}|\boldsymbol{\epsilon})\right) = \mu + \alpha_{j_i} + \beta_k + \epsilon_{i,k} \tag{C.1}$$

for every $i = 1\ldots, t$, $j = 1, \ldots, t$ and $k = 1, \ldots, b$. Here g is the link function as before, $Y_{i,j_i,k}$ is the response for the unit with label i and treatment j_i in block k and $E(Y_{i,j_i,k}|\boldsymbol{\epsilon})$ is the conditional expectation of $Y_{i,j_i,k}$ given the vector of random errors $\boldsymbol{\epsilon} = (\epsilon_{1,1}, \ldots, \epsilon_{t,1}, \ldots, \epsilon_{1,b}, \ldots, \epsilon_{t,b})^\top$. Also μ is the grand mean, α_{j_i} is the j-th fixed treatment effect on i-th experimental unit, β_k is the k-th fixed block effect and $\epsilon_{i,k}$ is the random unit error for i-th unit of k-th block. The model form (C.1) is our generalization of the derived linear model of Hinkelmann and Kempthorne (2008, p.281) to a GLMM. One can see that the only random variable on the right hand side of (C.1) is $\epsilon_{i,k}$ and the treatment j is fixed for every block k by the specific design $d = (j_1, \ldots, j_t)$.

In matrix notation, the RB-GLMM with fixed block effects (C.1) is

$$\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{C.2}$$

where $\boldsymbol{g}\left(\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{\epsilon})\right)$ is a vector of conditional expectations with components $g\left(E(Y_{i,j_i,k}|\boldsymbol{\epsilon})\right)$, $\boldsymbol{Y}$ is the $bt\times 1$ vector of responses, $\boldsymbol{X}$ is the $bt\times(b+t+1)$ design matrix for fixed effects, $\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_t, \beta_1, \ldots \beta_b)^\top$ is a $(b+t+1)\times 1$ vector of fixed treatment and block effect parameters, and $\boldsymbol{\epsilon}$ is the $bt\times 1$ vector of random errors. The conditional distribution of $Y_{i,j_i,k}$ given $\boldsymbol{\epsilon}$ is a member of the exponential family. Moreover, we also derived $\boldsymbol{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and the variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\epsilon}) = \text{diag}(\tilde{\boldsymbol{V}}_1, \ldots, \tilde{\boldsymbol{V}}_b)$ with $\tilde{\boldsymbol{V}}_k = \dfrac{\sigma^2_{ku}}{t-1}\left(t\boldsymbol{I}_t - \boldsymbol{1}_t\boldsymbol{1}_t^\top\right) = \sigma^2_{ku}\boldsymbol{\mathcal{P}_t}$ for $k = 1, \ldots, b$ where $\boldsymbol{\mathcal{P}_t} = \dfrac{1}{t-1}\left(t\boldsymbol{I}_t - \boldsymbol{1}_t\boldsymbol{1}_t^\top\right)$ as before and $Var(\epsilon_{i,k}) = \sigma^2_{ku}$.

The likelihood function for the RB-GLMM (C.2) for the RCBD with fixed block effects is equal to

$$L(\boldsymbol{\theta}) = \frac{1}{t!^b}\sum_{\boldsymbol{\pi}\in S_t^b}\prod_{k=1}^{b}\prod_{i=1}^{t} f_{\pi_k(i),j_i,k}(y_{i,k};\boldsymbol{\theta}), \tag{C.3}$$

where $f_{\pi_k(i),j_i,k}(y_{i,k};\boldsymbol{\theta})$ is the probability density function of the random variable $Y_{i,j_i,k}$ and S_t^b is the direct product of b instances of the symmetric group S_t.

C.2 Likelihood Function of RB-GLMM for the RCBD with Random Block Effects

The likelihood function for the RB-GLM for the RCBD with random block effects is

$$L(\boldsymbol{\theta}) = \frac{1}{t!^b b!}\sum_{(\boldsymbol{\pi},\delta)\in S_t \wr S_b}\prod_{k=1}^{b}\prod_{i=1}^{t} f_{\pi_{\delta(k)}(i),j_i,\delta(k)}(y_{i,k};\boldsymbol{\theta}), \tag{C.4}$$

where $f_{\pi_{\delta(k)}(i),j_i,\delta(k)}(y_{i,k};\boldsymbol{\theta})$ is the probability density function of the random variable $Y_{i,j_i,k}$ and $S_t \wr S_b$ is the wreath product of b-copies of the symmetric group S_t and S_b.

Appendix D

Simulation Results under Scenario-II

D.1 Misspecified Correlation Structure

D.1.1 Bias and SEs of Fixed Effects Estimates

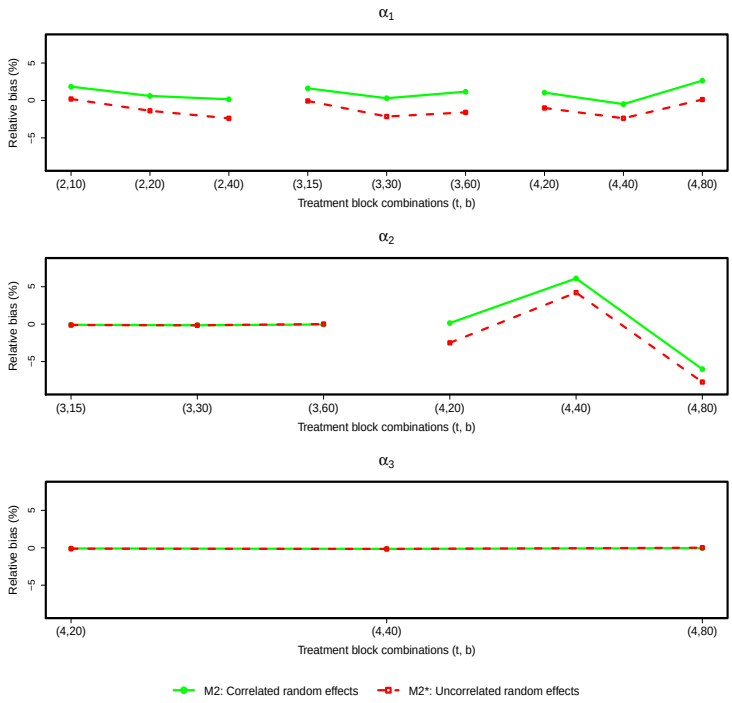

Figure D.1: Comparison of relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-II between $M2$ (true model) and $M2^*$(misspecified model).

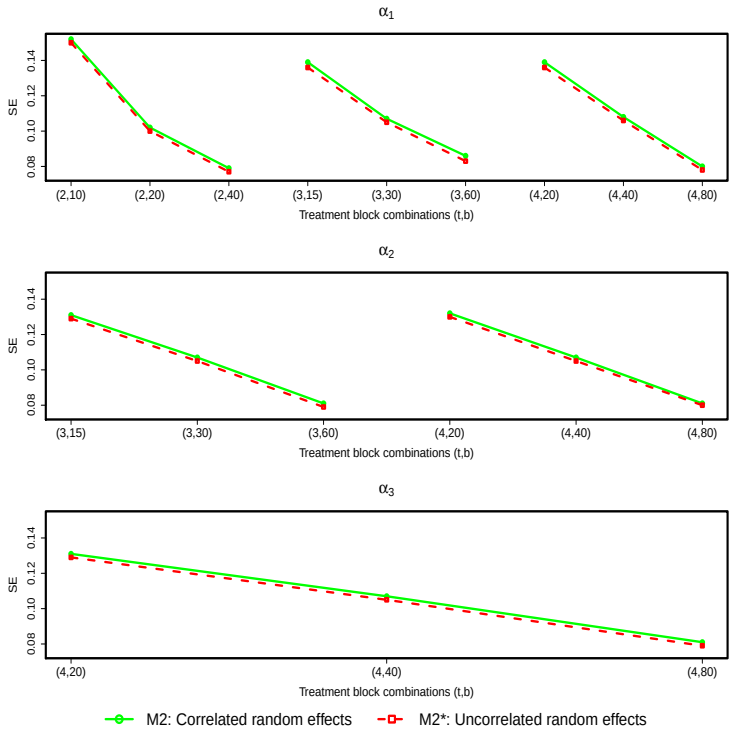

Figure D.2: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-II between $M2$ (true model) and $M2^*$(misspecified model)

D.1.2 Bias and SEs of Variance Components Estimates

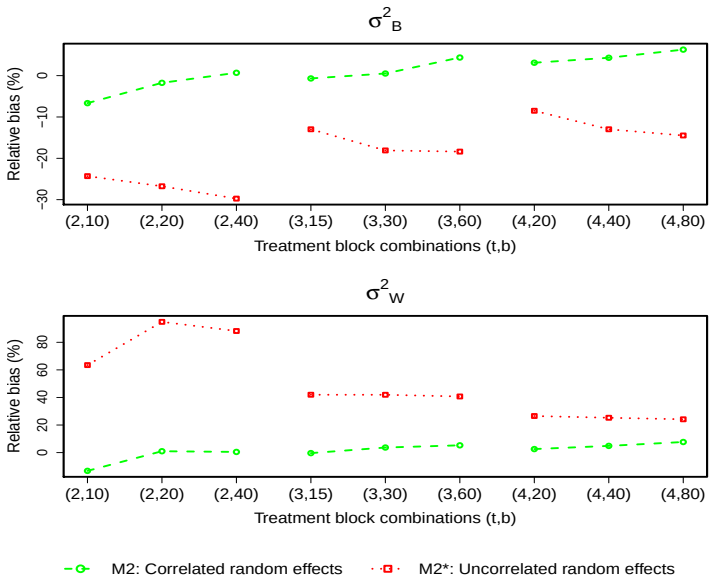

Figure D.3: Comparison of relative biases of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ under Scenario-II between $M2$ (true model) and $M2^*$(misspecified model).

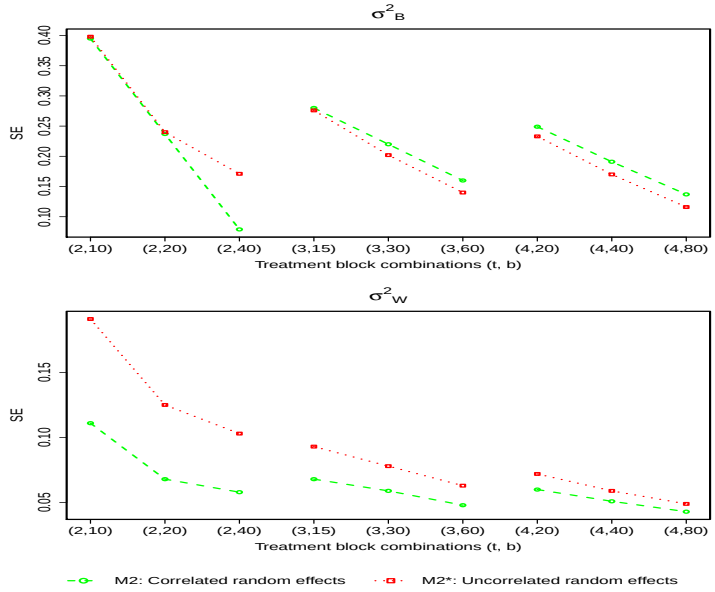

Figure D.4: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-II between $M2$ (true model) and $M2^*$(misspecified model).

D.2 Misspecified Random Effects Distributions

D.2.1 Bias and SEs of Fixed Effects Estimates

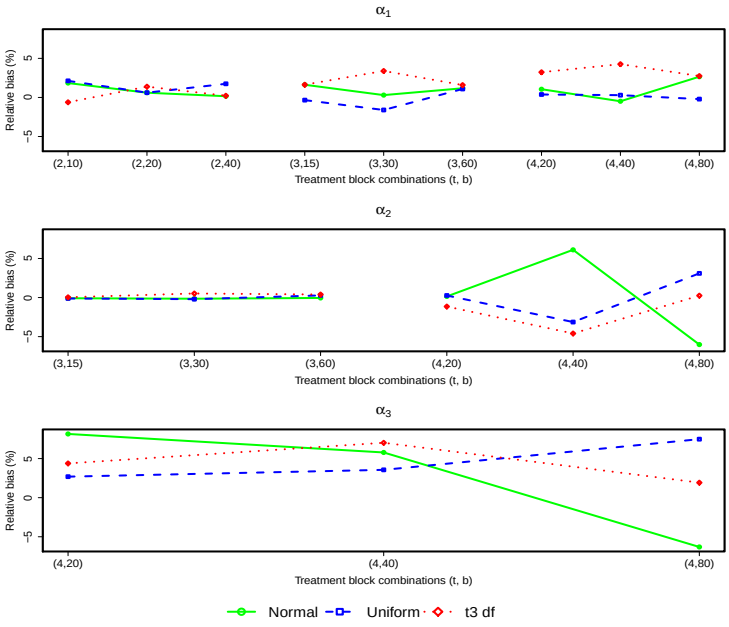

Figure D.5: Comparison of percent relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-II among normal, uniform and t_3 random effects distributions.

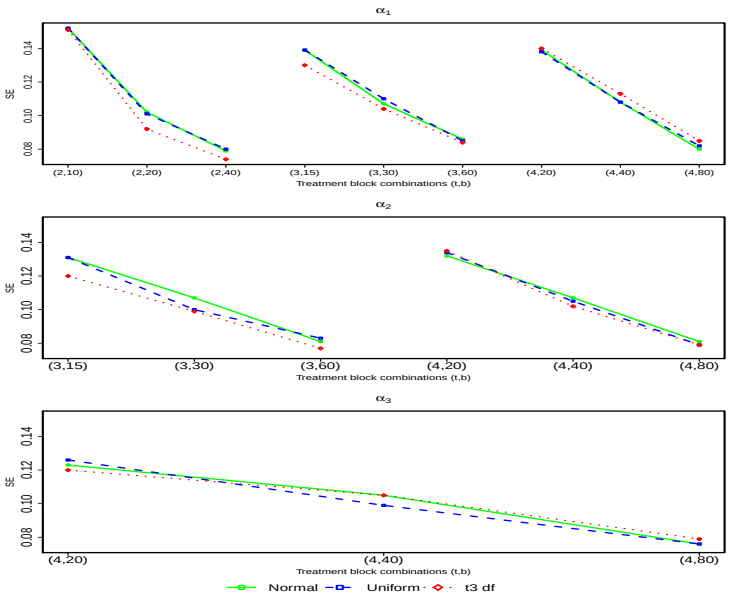

Figure D.6: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-II among normal, uniform and t_3 random effects distributions.

D.2.2 Bias and SEs of Variance Components Estimates

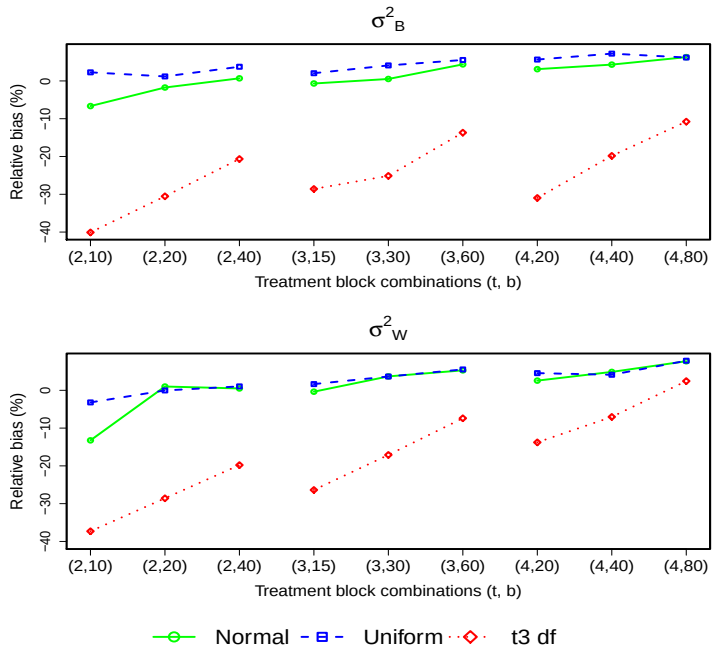

Figure D.7: Comparison of percent relative biases of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-II among normal, uniform and t_3 random effects distributions.

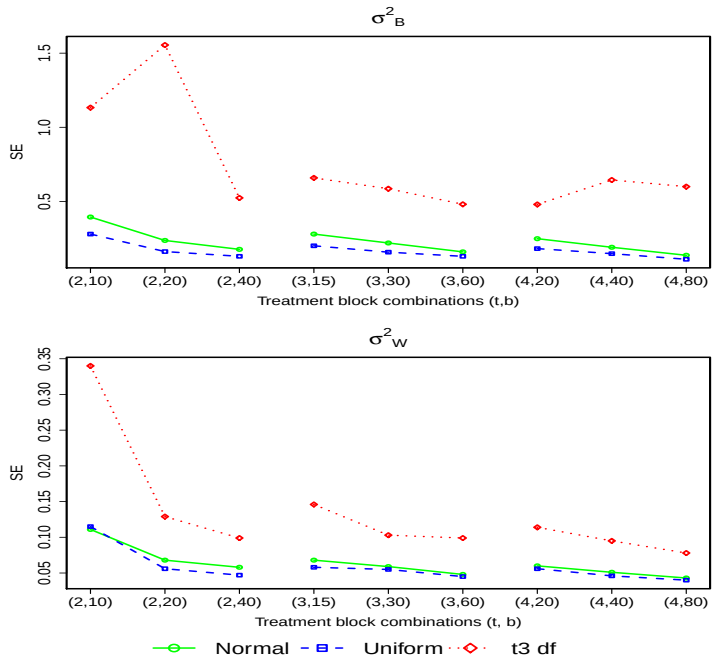

Figure D.8: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-II among normal, uniform and t_3 random effects distributions.

Appendix E

Simulation Results under Scenario-III

E.1 Misspecified Correlation Structure

E.1.1 Bias and SEs of Fixed Effects Estimates

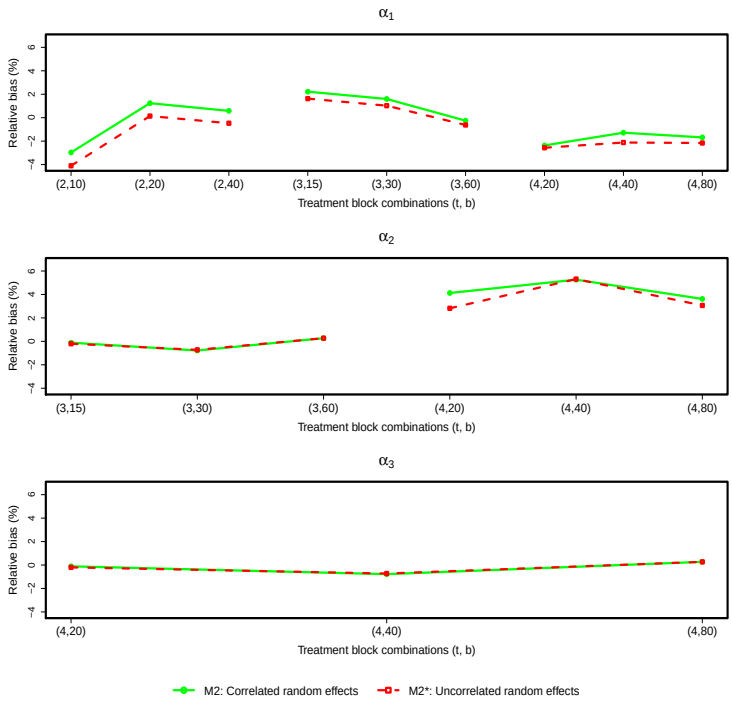

Figure E.1: Comparison of relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-III between $M2$ (true model) and $M2^*$(misspecified model).

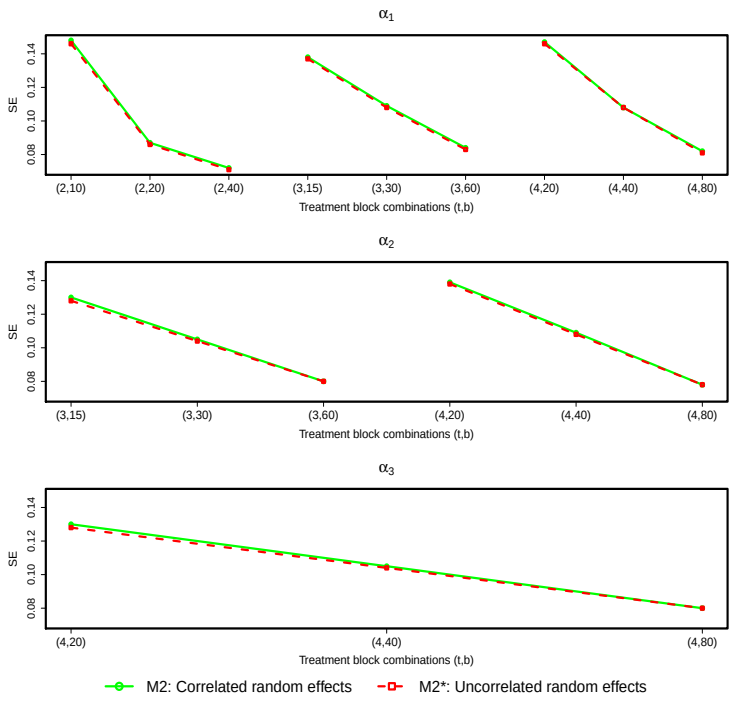

Figure E.2: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-III between $M2$ (true model) and $M2^*$(misspecified model)

E.1.2 Bias and SEs of Variance Components Estimates

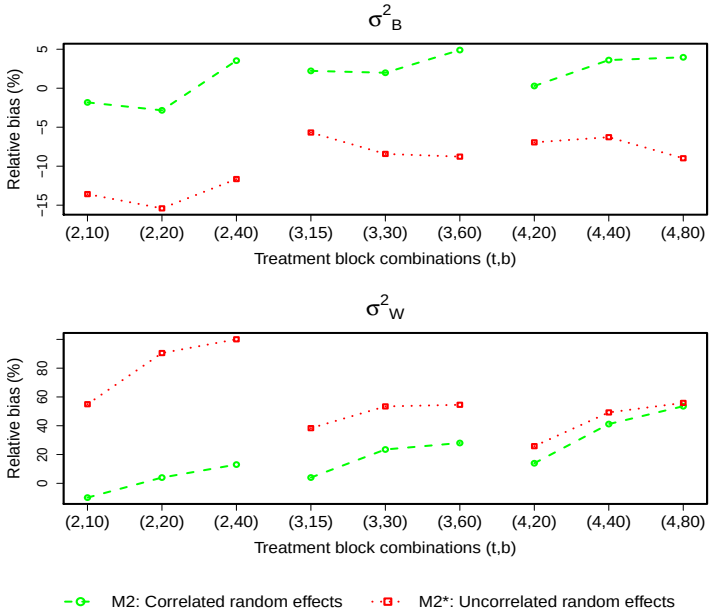

Figure E.3: Comparison of relative biases of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ under Scenario-III between $M2$ (true model) and $M2^*$(misspecified model).

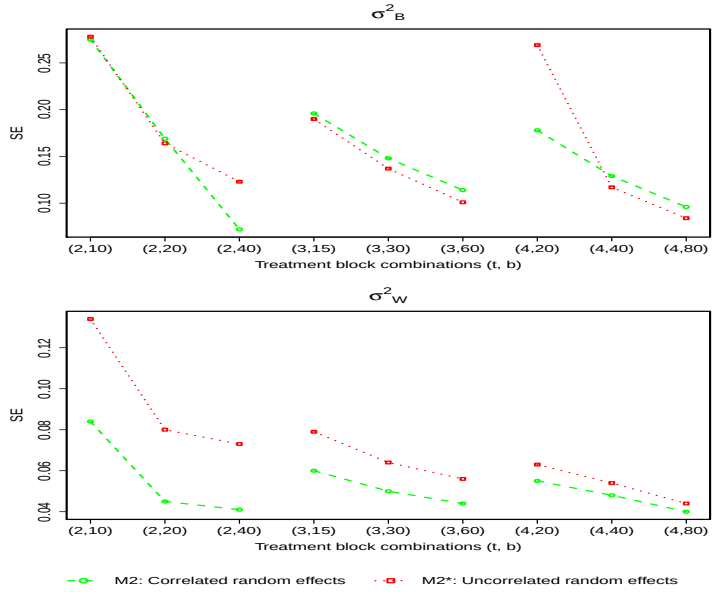

Figure E.4: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-III between $M2$ (true model) and $M2^*$(misspecified model).

E.2 Misspecified Random Effects Distribution

E.2.1 Bias and SEs of Fixed Effects Estimates

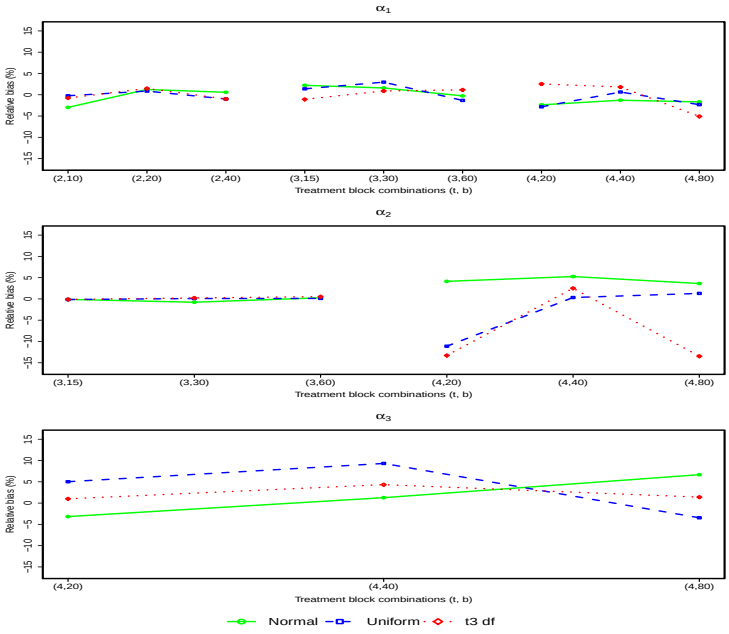

Figure E.5: Comparison of percent relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-III among normal, uniform and t_3 random effects distributions.

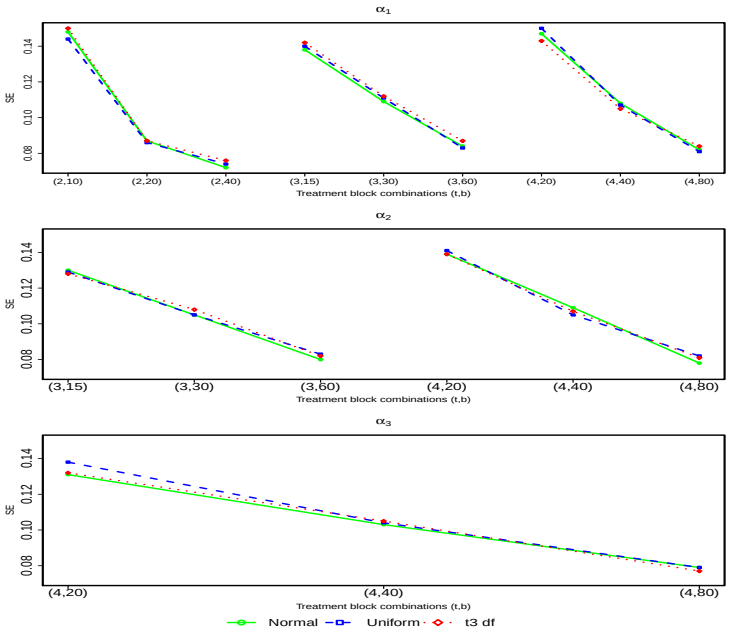

Figure E.6: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-III among normal, uniform and t_3 random effects distributions.

E.2.2 Bias and SEs of Variance Components Estimates

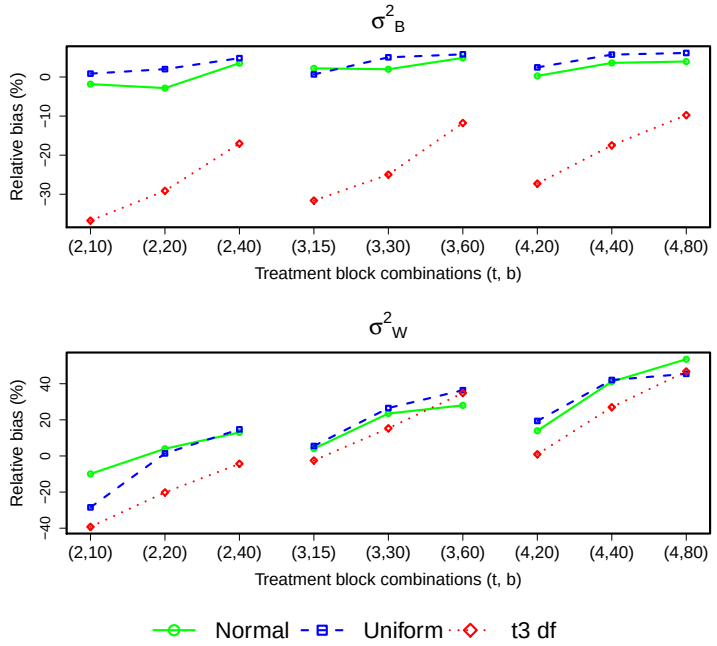

Figure E.7: Comparison of percent relative biases of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-III among normal, uniform and t_3 random effects distributions.

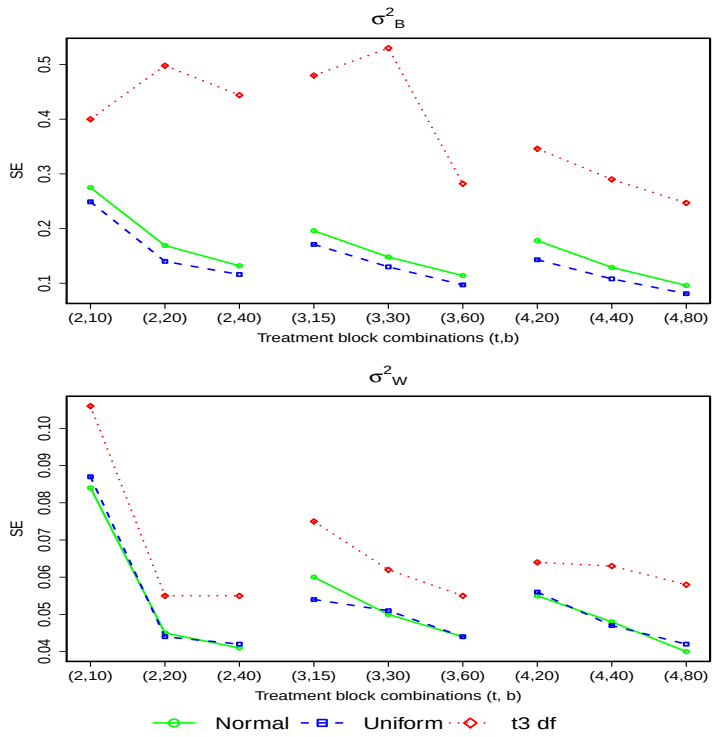

Figure E.8: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-III among normal, uniform and t_3 random effects distributions.

Appendix F

Simulation Results under Scenario-IV

F.1 Misspecified Correlation Structure

F.1.1 Bias and SEs of Fixed Effects Estimates

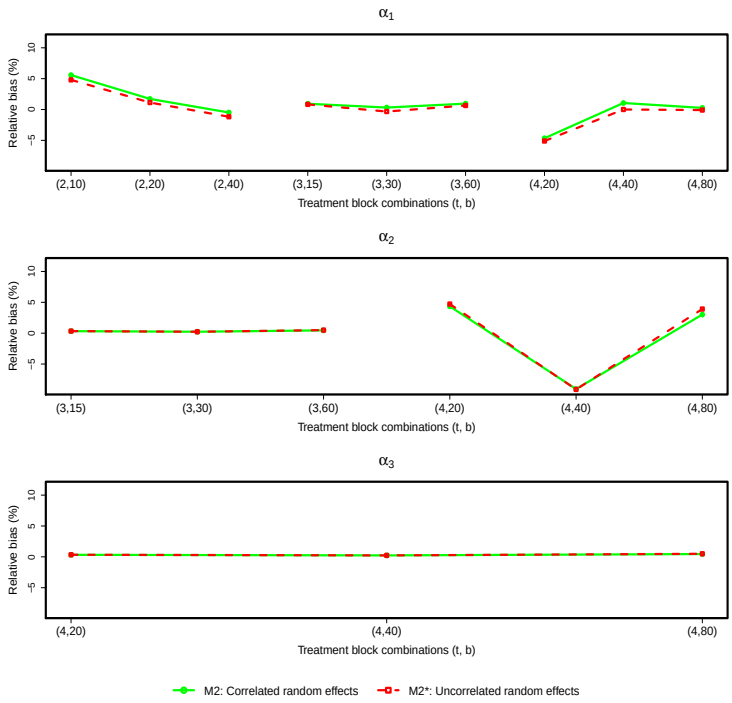

Figure F.1: Comparison of relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-IV between $M2$ (true model) and $M2^*$(misspecified model).

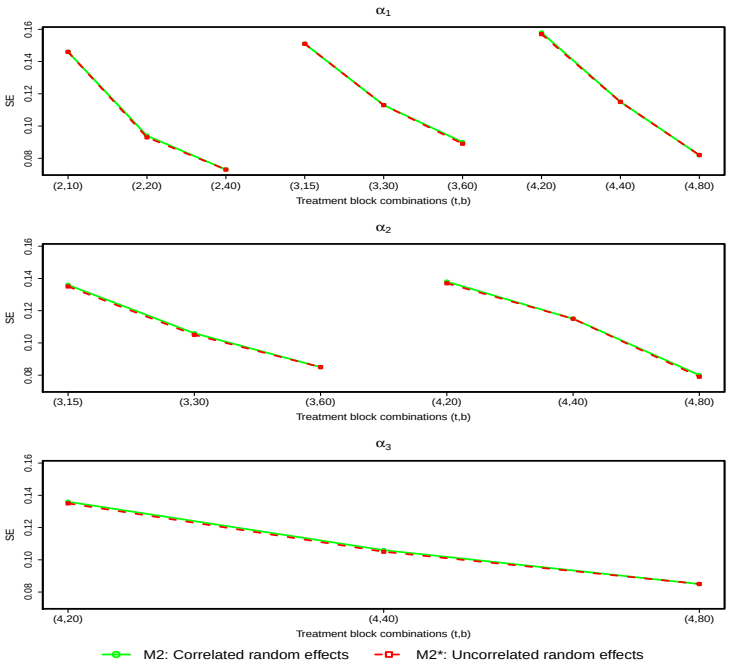

Figure F.2: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-IV between $M2$ (true model) and $M2^*$(misspecified model)

F.1.2 Bias and SEs of Variance Components Estimates

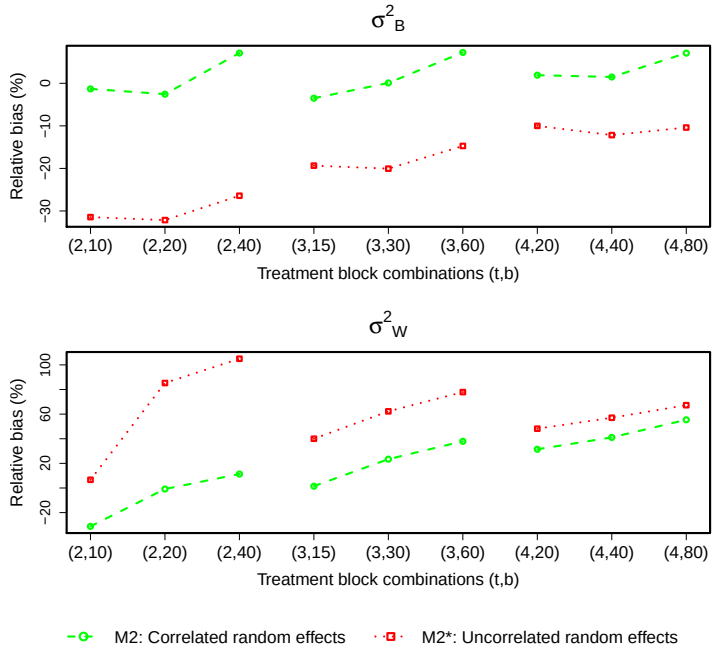

Figure F.3: Comparison of relative biases of $\hat{\sigma}^2_B$ and $\hat{\sigma}^2_W$ under Scenario-IV between $M2$ (true model) and $M2^*$(misspecified model).

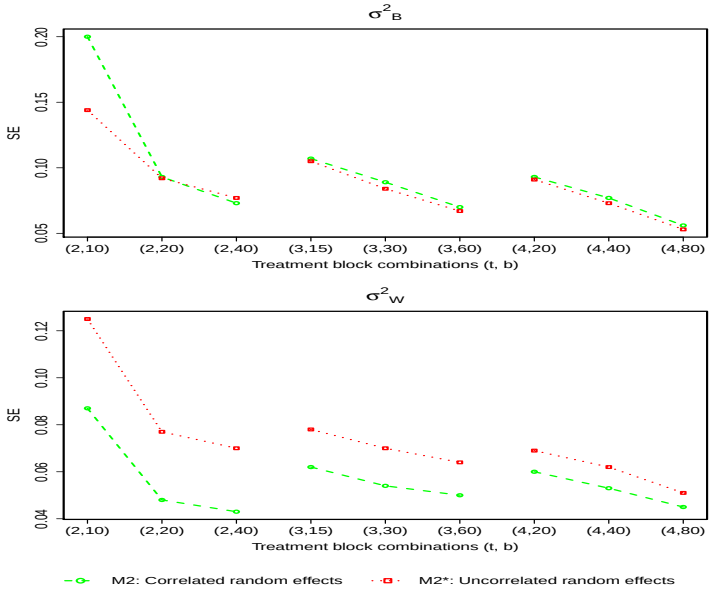

Figure F.4: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-IV between $M2$ (true model) and $M2^*$(misspecified model).

F.2 Misspecified Random Effects Distribution

F.2.1 Bias and SEs of Fixed Effects Estimates

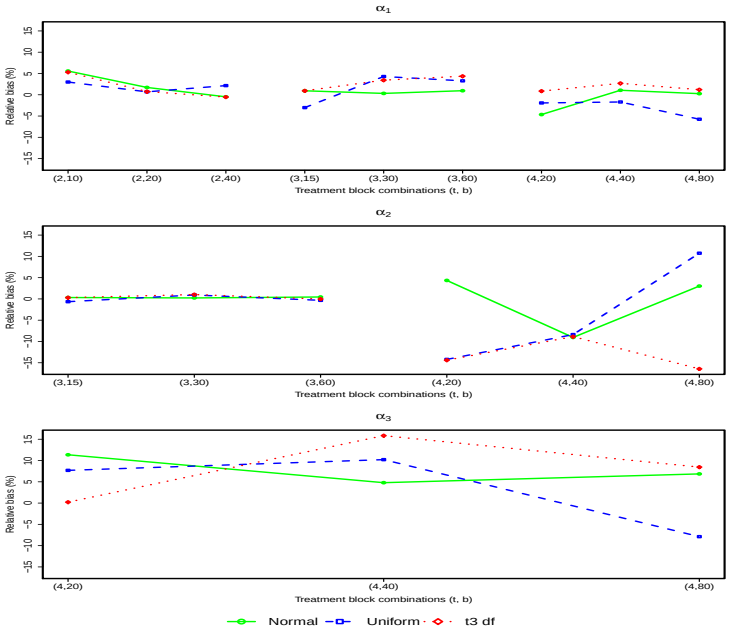

Figure F.5: Comparison of percent relative biases of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-IV among normal, uniform and t_3 random effects distributions.

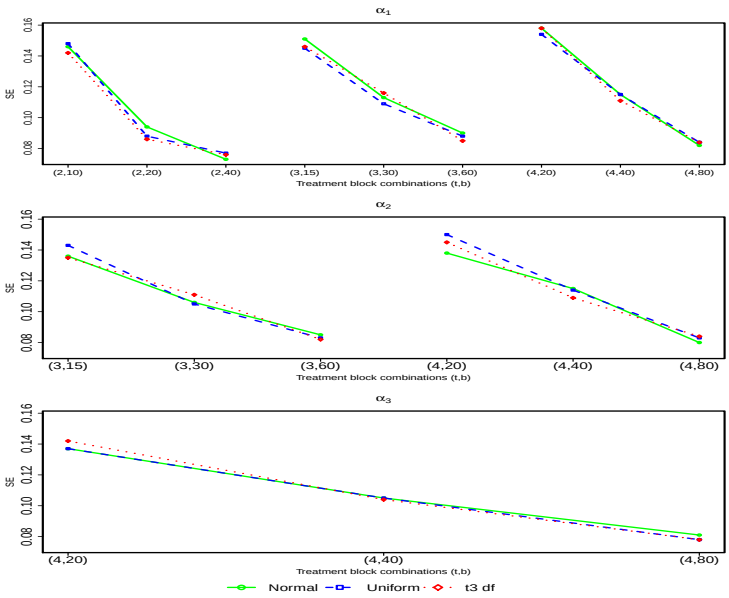

Figure F.6: Comparison of SEs of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ under Scenario-IV among normal, uniform and t_3 random effects distributions.

F.2.2 Bias and SEs of Variance Components Estimates

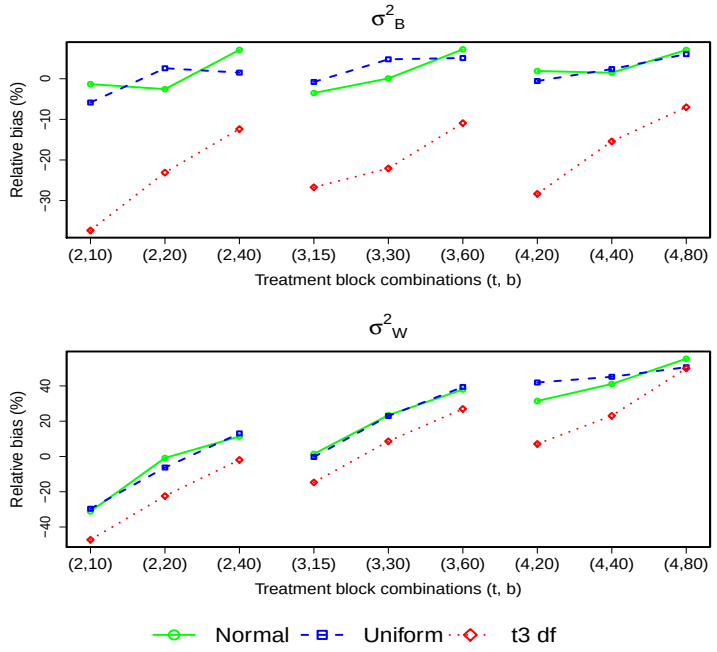

Figure F.7: Comparison of percent relative biases of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-IV among normal, uniform and t_3 random effects distributions.

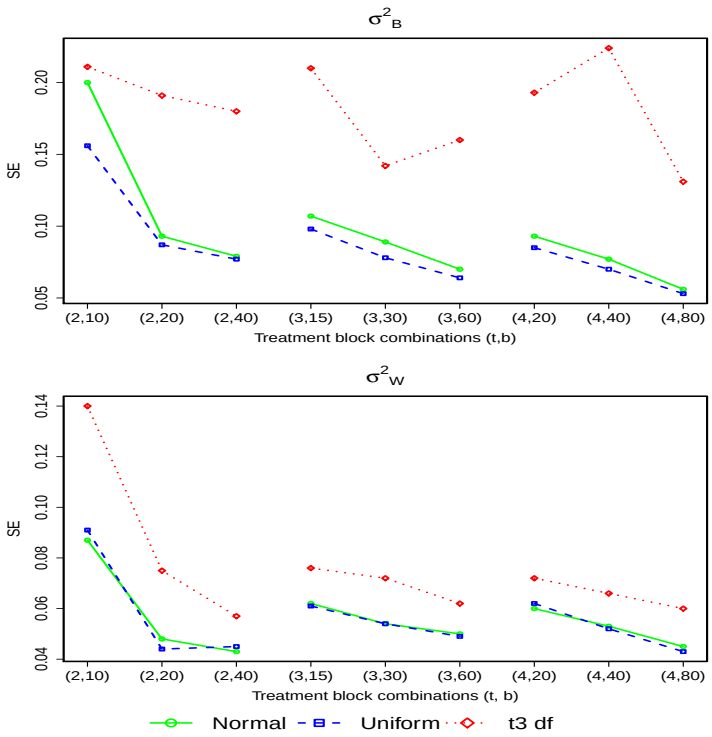

Figure F.8: Comparison of SEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_W^2$ under Scenario-IV among normal, uniform and t_3 random effects distributions.

Bibliography

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, 47:639–653.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128.

Arora, S. and Barak, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press.

Bezáková, I., Štefankovic, D., Vazirani, V. V., and Vigoda, E. (2008). Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM Journal on Computing*, 37(5):1429–1454.

Bickel, P. J. and Doksum, K. A. (2001). *Mathematical Statistics, volume I*. Prentice–Hall, New Jersey, 2nd edition.

Billingsley, P. (1985). *Probability and Measure*. Wiley & Sons, New York, 2nd edition.

Brien, C. J. and Bailey, R. (2006). Multiple randomizations. *Journal of the Royal Statistical Society: Series B*, 68:571–609.

Butler, S. M., Louis, T. A., et al. (1997). Consistency of maximum likelihood estimators in general random effects models for binary data. *The Annals of Statistics*, 25(1):351–377.

Cameron, P. J. (1998). *Introduction to Algebra*. Oxford University Press, Oxford.

Cameron, P. J. (1999). *Permutation Groups*. Cambridge University Press, Cambridge.

Christensen, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer-Verlag, New York.

Cox, D. (2009). Randomization in the design of experiments. *International Statistical Review*, 77(3):415–429.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88:973–985.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge University Press, 2nd edition.

Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments: Volume 1, Introduction to Experimental Design*. Wiley & Sons, Hoboken, NJ, 2nd edition.

Huber, M. and Law, J. (2008). Fast approximation of the permanent for very dense problems. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 681–689. Society for Industrial and Applied Mathematics.

Hunter, D. R. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37.

Jerrum, M. (2003). *Counting, Sampling and Integrating: Algorithms and Complexity*. Springer.

Jerrum, M., Sinclair, A., and Vigoda, E. (2004). A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697.

Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association*, 50(271):946–967.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811.

Lange, K. (2013). *Optimization*. Springer, 2nd edition.

Lange, K., Hunter, D. R., and Yang, I. (2000). Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, 9(1):1–20.

Lee, W. and Lee, Y. (2012). Modifications of REML algorithm for HGLMs. *Statistics and Computing*, 22(4):959–966.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B*, 58(4):619–678.

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

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley & Sons, New York.

Litière, S., Alonso, A., and Molenberghs, G. (2007). Type I and type II errors under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63:1038–1044.

Litière, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 27:3125–3144.

Magder, L. S. and Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of gaussians. *Journal of the American Statistical Association*, 91(435):1141–1151.

Marcus, M. and Minc, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn & Bacon Inc., Boston.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, 92(437):162–170.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, 26:388–402.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. Wiley & Sons, New York, 2nd edition.

Minc, H. (1984). *Permanents*, volume 6. Cambridge University Press.

Neuhaus, J. M., McCulloch, C. E., and Boylan, R. (2011). A note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics*, 67:654–660.

Neuhaus, J. M., McCulloch, C. E., and Boylan, R. (2013). Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. *Statistics in Medicine*, 32:2419–2429.

Noh, M., Yip, B., Lee, Y., and Pawitan, Y. (2006). Multicomponent variance estimation for binary traits in family-based studies. *Genetic Epidemiology*, 30(1):37–47.

Piepho, H.-P. and McCulloch, C. E. (2004). Transformations in mixed models: Application to risk analysis for a multienvironment trial. *Journal of agricultural, biological, and environmental statistics*, 9(2):123–137.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata Volume II: Catergorical Responses, Counts, and Survival*. STATA press, 3rd edition.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley & Sons, New York, 2nd edition.

Rönnegård, L., Shen, X., and Alam, M. (2010). hglm: A package for fitting hierarchical generalized linear models. *The R Journal*, 2(2):20–28.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.

Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC press.

Valiant, L. G. (1979). The complexity of computing the permanent. *Theoretical computer science*, 8(2):189–201.

Ver Hoef, J. M. (2012). Who invented the delta method? *The American Statistician*, 66(2):124–127.

Waite, T. W. and Woods, D. C. (2015). Designs for generalized linear models with random block effects via information matrix approximations. *Biometrika*, 102(3):677–693.

Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):185–198.

Woods, D. C. and Van de Ven, P. (2011). Blocked designs for experiments with correlated non-normal response. *Technometrics*, 53(2):173–182.

Wu, T. T., Lange, K., et al. (2010). The MM alternative to EM. *Statistical Science*, 25(4):492–505.