ESTIMATION AND TESTING OF PARAMETERS UNDER CONSTRAINTS FOR CORRELATED DATA

Laura Farnan

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, Gillings School of Global Public Health.

Chapel Hill
2011

Approved by:

Shyamal D. Peddada, PhD

Anastasia Ivanova, PhD

Pranab K. Sen, PhD

Bahjat F. Qaqish, PhD

Amanda H. Corbett, Pharm. D.

ABSTRACT

LAURA FARNAN: Estimation and Testing of Parameters under Constraints for Correlated Data
(Under direction of Dr. Shyamal D. Peddada and Dr. Anastasia Ivanova)

This dissertation work is motivated by problems encountered in the analysis of some toxicological and clinical trials data, where repeated measurements are made on each subject, and the investigator expects trends in mean response among dose groups and/or time points. There are two components to this research. The first component focuses on estimation of parameters subject to inequality constraints when the covariance matrix of the unrestricted estimator is non-diagonal. In particular, statistical properties of several available constrained estimators are investigated theoretically and via simulations under different covariance structures. The second component is developing a simple, yet statistically appropriate methodology for testing hypotheses in a linear mixed effects model with an inequality constraint in the alternative. Since in many applications one cannot be certain about the normality of the data, a bootstrap based methodology using MINQUE-Williams' type test is implemented for testing the above hypotheses. The resulting methodology is illustrated by re-analyzing the blood mercury level data provided in Cao et al. (2011).

ACKNOWLEDGEMENTS

Now, that my dissertation is drawing to a close, I would like to acknowledge all those people who have made it such a success. Firstly, I would like to thank my advisors Dr. Shyamal D. Peddada and Dr. Anastasia Ivanova. Not only did they provide guidance in research, but they also offered a lot of emotional support and encouragement along the way.

I would also like to sincerely thank Dr. Pranab K. Sen, Dr. Bahjat F. Qaqish, and Dr. Amanda H. Corbett for being on my advisory committee and for providing insightful comments that have definitely made my dissertation better. I certainly appreciate NIEHS researcher Dr. Walter Rogan for providing the data for the analysis.

I am extremely thankful to Dr. Marijus Radavičius, a professor at Vilnius University. He encouraged me to start research in a field that I was truly passionate about (applications of statistics in medicine) and to pursue my doctoral studies abroad.

I would also like to thank Dr. Jeannette Bensen, a Co-Director of the North Carolina-Louisiana Prostate Cancer Study (PCaP), for being such an awesome employer. Beyond the mere financial support her employment provided, I truly appreciate her understanding, patience and support at times when I needed it the most.

My parents have always been there to provide me with help and support, and even more so while I was working on the dissertation. My mother spent two summers with us to help us out, and this surely helped me to make substantial progress in my research.

Finally, I would like to thank my husband and daughter, as I certainly would not have been able to complete this journey without them - my husband Kirby for all his encouragement and support during the difficult times, and my daughter Katie for her loving smile and for keeping me focused on what life is truly all about.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION – LITERATURE REVIEW

1.1. Motivation

This dissertation work is motivated by problems encountered in the analysis of some toxicological data and clinical trials data, where repeated measurements are made on each subject, and the investigator expects trends in mean response among dose groups and/or time points. For example, Cao et al., 2011 were interested whether succimer, a mercaptan compound known to reduce blood lead concentration in children, also reduces blood mercury concentration. They used samples from a randomized placebo-controlled, double-blind trial clinical trial of succimer for lead poisoning in 780 children aged 12-33 months, called the Treatment of Lead-exposed Children trial, or TLC (Rogan, 1998). In TLC, 384 children were assigned to the placebo group and 396 to the succimer group. Up to three 26-day courses of succimer or placebo therapy were administered, depending on response to treatment in those, who were given succimer. For children in each group, blood lead concentrations were obtained twice before randomization and then on days 7, 28, and 42 after the beginning of each course of treatment. After treatment was stopped, blood lead levels were measured every three to four months until 36 months after the initiation of treatment. Cao et al. (2011) measured mercury in pre-treatment samples from 393 children given succimer and 374 given placebo. They also measured mercury

in 1-week post-treatment blood samples (N = 768) and in a 20% random sample of the 338 children who received the maximum 3 courses of treatment.

In addition to the presence of variance components, the data can be potentially heteroscedastic since the variability across time may not necessarily be constant. Very little literature exists on constrained inference in linear mixed models even under homoscedasticity, let alone under heteroscedasticity. Silvapulle (1997) proposed a methodology for testing linear constraints regarding fixed effects parameters under some conditions on the design matrices. The resulting test procedure does not depend upon the unknown variance components, thus ignores correlations within the subject over time. Thus the methodology developed in Silvapulle (1997) is restrictive and is not applicable to the present context. As observed in Hoferkamp and Peddada (2002), the biggest challenge in linear mixed models with or without heteroscedasticity is the derivation of restricted maximum likelihood estimators (RMLE) for various parameters of the model. Consequently, the derivation of the likelihood ratio test is non-trivial and has not been derived in the literature so far.

Examples such as the above one are rather common in applications, and often researchers tend to use the classical mixed effects analysis of variance followed by "post-hoc" analyses to make pair-wise comparisons rather than testing for the desired order restriction. There is clearly a demand for well developed theory and methodology for such problems.

Another example that motivated this research is a recently published proof-of-concept clinical trial (Ivanova, Liu, Snyder and Snavely, 2009). The clinical trial was a three period crossover trial, where each subject received placebo, active control and a

dose of an investigational drug. The objectives of this trial were estimation of the mean response (measured in minutes) under the assumption that mean responses are constrained by an umbrella order and comparing the best dose with placebo and control. Since each subject received two different doses (one of which was dose 0 mg, placebo), unrestricted estimates of mean response were correlated. Instead of maximizing the likelihood under restrictions while taking into account correlation structure, the investigators obtained unrestricted estimates first, while taking into account correlation structure and then obtained parameter estimates using a simpler method that is based on a well known pool adjacent violators algorithm (PAVA) (Silvapulle and Sen, 2005). The PAVA is used when non-decreasing order is assumed and proceeds as follows. If a pair of adjacent unrestricted estimates violates the hypothesized order, then, according to the algorithm, each such pair of estimates is replaced by their average. The process is repeated until all estimates satisfy the hypothesized order. Although PAVA is a very convenient methodology to implement, as described in the following sections, very little is known about theoretical properties of PAVA.

Motivated by the above applications, in this dissertation research we will focus on two aspects of constrained inference; (a) estimation of parameters for correlated data subject to inequality constraints, and (b) testing hypothesis for correlated data under inequality constraints. In section 1.2, we will review the literature on the estimation of parameters under constraints, when the underlying data are correlated, and in section 1.3 we will review the literature on the testing problem.

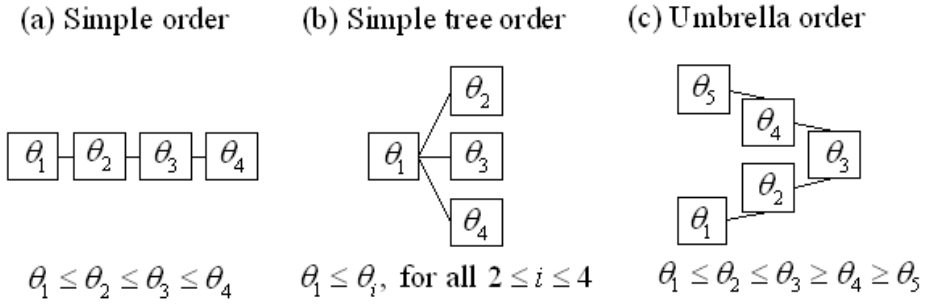
1.2. Methods of estimation

Let $\boldsymbol{\theta} = \left(\theta_1, \theta_2, .., \theta_p\right)'$ denote an unknown parameter vector whose components satisfy inequality constraints. Problem of estimating parameters under constraints arises for a variety of reasons. In some applications, such as in dose-finding clinical trials, constrained estimation plays an important role for determining dose at which the next patient needs to be treated (Stylianou and Flournoy, 2002; Ivanova et al., 2009; Conaway, Dunbar and Peddada, 2004). In other situations, such as in toxicology, researchers are often interested in testing for patterns of response. Again, in all such situations one needs to perform constrained testing of parameters. For example, toxicologists are usually interested in testing the hypothesis that the tumor incidence rate increases with the dose of a toxin, i.e. testing $H_0 : \theta_1 = \theta_2 = \ldots = \theta_p$ against $H_A : \theta_1 \le \theta_2 \le \ldots \le \theta_p$, known as *simple order restriction* (Peddada, Dinse and Kissling, 2007). Similarly, when comparing multiple toxins with the control, toxicologists often test the hypothesis that the tumor incidence due to a toxin is larger than the tumor incidence due to control, i.e., test $H_0 : \theta_1 = \theta_2 = \ldots = \theta_p$ against $H_A : \theta_1 \le \theta_i$, $i \ge 2$, known as *simple tree order*. In all such situations, the test statistic requires the estimation of parameters under the inequalities specified by the alternative hypothesis.

A variety of inequality constraints have been discussed in the literature, such as the simple order $\theta_1 \le \theta_2 \le \ldots \le \theta_p$, the umbrella order $\theta_1 \le \theta_2 \le \ldots \le \theta_i \ge \theta_{i+1} \ge \ldots \ge \theta_p$, the single loop order, etc. (Stylianou and Flournoy, 2002; Ivanova et al., 2009; Conaway et al., 2004).

Often order restrictions can be expressed using graphs as shown in Figure 1. In Chapter 3 we will focus on simple order restriction (a).

Figure 1. Examples of some order restrictions.



There exists over 50 years of literature on the estimation and testing of hypothesis under inequality constraints on the parameters $\theta_1, \theta_2, .., \theta_p$ in a variety of settings. For a comprehensive review on estimation and testing of parameters under constraints, one may refer to Silvapulle and Sen (2005) and van Eeden (2006). Much of the literature is based on the likelihood principle. However, as reviewed in van Eeden (2006) and Silvapulle and Sen (2005), several alternative estimation and testing procedures have been proposed in the literature. They are often computationally simpler to implement and are designed for the specific parametric model and specific order restrictions. Among these methods, PAVA is one of the most popular methods for estimating parameters under simple order restriction.

Suppose $\hat{\boldsymbol{\theta}}^{UMLE} = \left(\hat{\theta}_1^{UMLE}, \hat{\theta}_2^{UMLE}, .., \hat{\theta}_p^{UMLE}\right)'$ is an unrestricted estimator of $\boldsymbol{\theta}$, where the components of $\hat{\boldsymbol{\theta}}^{UMLE} = \left(\hat{\theta}_1^{UMLE}, \hat{\theta}_2^{UMLE}, .., \hat{\theta}_p^{UMLE}\right)'$ are independently distributed, then the constrained estimator of $\boldsymbol{\theta}$ is usually obtained by solving the minimization problem

$$\min_{\boldsymbol{\theta}\in C} \sum_{i=1}^{p} w_i \left(\hat{\theta}_i^{UMLE} - \theta_i\right)^2, \tag{1}$$

where C is the set of known inequalities satisfied by the components of $\boldsymbol{\theta}$, and w_i is some known weight, usually taken to be the reciprocal of the variance of $\hat{\theta}_i^{UMLE}$.

If C is a subset of the parameter space satisfying simple order constraints, then the above minimization problem (1) is often solved by using the well-known pool adjacent violator algorithm (PAVA) (cf. Silvapulle and Sen, 2005). Analytically, the PAVA estimator for $\theta_i, i = 1,2,..., p$ under $\theta_1 \le \theta_2 \le ... \le \theta_p$ is given by the following equivalent formulae (cf van Eeden, 2006):

$$\hat{\theta}_i^{PAVA(p)} = \min_{i\le t\le p} \max_{1\le s\le i} \frac{\sum_{j=s}^{t} w_j \hat{\theta}_j^{UMLE}}{\sum_{j=s}^{t} w_j} \tag{2}$$

$$= \max_{1\le s\le i} \min_{i\le t\le p} \frac{\sum_{j=s}^{t} w_j \hat{\theta}_j^{UMLE}}{\sum_{j=s}^{t} w_j},$$

where $w_j = \dfrac{1}{Var\left(\hat{\theta}_j^{UMLE}\right)}$. The superscript (p) in $\hat{\theta}_i^{PAVA(p)}$ denotes the PAVA estimate of θ_i based on p groups.

If the components of $\hat{\boldsymbol{\theta}}^{UMLE}$ are independently and normally distributed with known variances, then PAVA results in the restricted maximum likelihood estimator (RMLE) of $\boldsymbol{\theta}$ under the inequality constraints. PAVA provides a valid methodology for estimating parameters under constraints even when the data are not normally distributed. Thus it is "robust" to non-normality. For correlated normally distributed data, the RMLE

is derived by solving the following constrained minimization problem, where $\boldsymbol{\Sigma}$ is the (known) covariance matrix of $\hat{\boldsymbol{\theta}}^{UMLE}$:

$$\min_{\boldsymbol{\theta}\in \mathrm{C}} \left(\hat{\boldsymbol{\theta}}^{UMLE} - \boldsymbol{\theta}\right)' \boldsymbol{\Sigma}^{-1} \left(\hat{\boldsymbol{\theta}}^{UMLE} - \boldsymbol{\theta}\right). \tag{3}$$

Diaz and González (1988) identified some sufficient conditions on $\boldsymbol{\Sigma}$ for which PAVA and RMLE are the same. For example, the sufficient conditions are satisfied when $\boldsymbol{\Sigma}$ is an intra-class correlation matrix. In general, however, they are not the same. If the unconstrained estimator is multivariate normally distributed, then the above minimization problem results in RMLE. Again, we emphasize on the fact that PAVA, as well as the solution to (3), provides robust estimators to $\boldsymbol{\theta}$ by not relying on the knowledge of the underlying likelihood function which may not always be known.

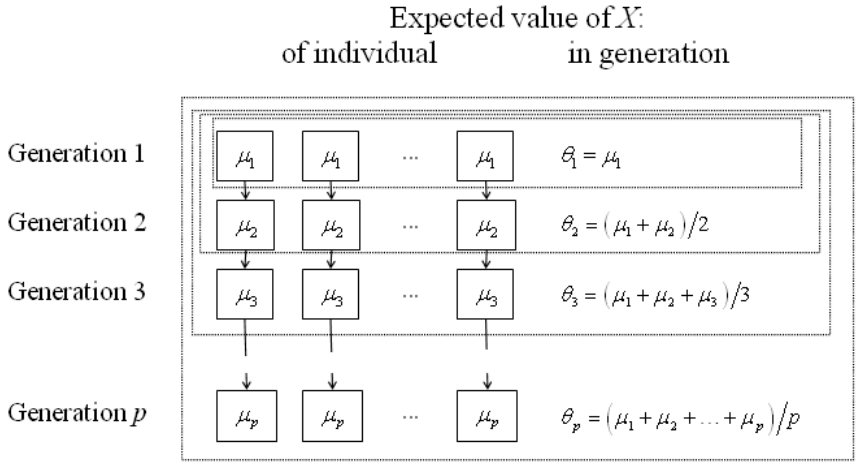
Furthermore, the computation of the RMLE may not always be straightforward, especially when $\boldsymbol{\Sigma}$ is unknown (Shi, Zheng and Guo, 2005; Hoferkamp and Peddada, 2002). Also, as observed by several authors (cf. Lee, 1988; Fernandez, Rueda and Salvador, 1999), the RMLE may not always perform well in terms of the mean squared error even when $\boldsymbol{\Sigma}$ is known. Hwang and Peddada (1994) argued that the RMLE is not only universally dominated by the UMLE under certain conditions, but any fixed width confidence interval centered at the RMLE may actually have a zero coverage probability as p increases. They surprisingly note that the RMLE may fail even in the case of simple order when the underlying covariance matrix is non-diagonal.

As an example of the RMLE failing in the case of some specific covariance matrix, let us discuss an example of star-shaped ordering presented by Shaked (1979). Consider a species consisting of k individuals each of which has a quantitative characteristic of interest X. Denote the expected value of X for each offspring of the ith

generation by $\mu_{i+1} \geq 0$. The expected value of *X* of the population in generation *i* is denoted by θ_i. Assume that *k* new individuals are produced in each generation, are added to the population, and that $\mu_{i+1} \geq \theta_i$ (i.e. *X* is improving on the average). The expected value of *X* in generation *i* is

$$\theta_i = (\mu_1 + \mu_2 + ... + \mu_i)/i \text{ and } \theta_{i+1} \geq \theta_i \text{ for all } i = 1, \ldots, p-1. \quad (4)$$

Figure 2. Star-shaped ordering



$$\mu_1 \leq (\mu_1 + \mu_2)/2 \leq ... \leq (\mu_1 + \mu_2 + ... + \mu_p)/p$$
$$\theta_1 \leq \theta_2 \leq ... \leq \theta_p$$

In this example $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$ satisfies star-shaped order restriction $\mu_1 \leq (\mu_1 + \mu_2)/2 \leq ... \leq (\mu_1 + \mu_2 + ... + \mu_p)/p$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ satisfies the simple order restriction $\theta_1 \leq \theta_2 \leq ... \leq \theta_p$. Thus, if $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{I})$ and $\hat{\theta}_i = (X_1 + ... + X_i)/i$, we have $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$ is non-diagonal and is given by the equation (6) on page 14. We performed a simulation study with $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \mathbf{I})$, where the components of $\boldsymbol{\mu}$ follow star-shaped ordering. Simulation

results indicate that RMLE is dominated by PAVA and, as the dimension p increases, it is also dominated by UMLE. These results are consistent with findings of Hwang and Peddada (1994). This example is revisited on page 13.

PAVA is widely used even in situations where the unrestricted estimators are not independently distributed, due to its computational simplicity. For instance, in clinical trials involving repeated measurements on the same subject, Ivanova et al. (2009) describe a proof-of-concept trial with crossover allocation, where PAVA was used to estimate the target dose at the end of the trial. Other examples where PAVA was used for correlated data include multidimensional scaling (Robertson, Wright and Dykstra, 1988), non-parametric semi-variogram estimation (Kim and Boos, 2004), linear models with covariates (Bretz, 2006), general constrained smoothing (Mammen, Marron, Turlach and Wand, 2001), estimation of the baseline survivor function in a proportional hazard model (Young, Jewell and Samuels, 2008; Li and Tseng, 2008), analysis of functional magnetic resonance imaging (FMRI) data (Woolrich, Ripley, Brady and Smith, 2001), and ranking and selection (Huang, 1984).

Shin et al. (1996) demonstrated that the solution to (3) is asymptotically equal to the solution of $\min_{\boldsymbol{\theta}\in C} \left(\hat{\boldsymbol{\theta}}^{UMLE} - \boldsymbol{\theta}\right)' \boldsymbol{\Omega}_1 \left(\hat{\boldsymbol{\theta}}^{UMLE} - \boldsymbol{\theta}\right)$, where $\boldsymbol{\Omega}_1$ is a suitable diagonal matrix. Thus by choosing suitable weights, one may solve the simpler isotonic regression problem (1) using standard PAVA, which only requires a simple hand held calculator rather than solving the optimization problem (3).

Although PAVA is widely used even when the components of $\hat{\boldsymbol{\theta}}^{UMLE}$ are not independent, there do not seem to exist any results in the literature on the performance of PAVA in such situations. Also, not much is known regarding the relationship between

PAVA and solution to (3), with the exception of Diaz and González (1988) who identify sufficient conditions under which PAVA provides the solution to (3).

In Chapter 2, PAVA will be evaluated in terms of mean squared error and universal domination criterion (Hwang, 1985) under the assumption, that the unrestricted estimator is multivariate normally distributed. For a pair of univariate estimators $\hat{\eta}_1$ and $\hat{\eta}_2$ of a parameter η, $\hat{\eta}_1$ is said to universally dominate (also known as stochastically dominate) $\hat{\eta}_2$ if for all η and all $c > 0$, $P\left(|\hat{\eta}_1 - \eta| < c\right) \geq P\left(|\hat{\eta}_2 - \eta| < c\right)$ with a strict inequality for some η. Equivalently, $\hat{\eta}_1$ is said to universally dominate $\hat{\eta}_2$ if $E\left[\phi\left(|\hat{\eta}_1 - \eta|\right)\right] \leq E\left[\phi\left(|\hat{\eta}_2 - \eta|\right)\right]$ for all non-decreasing functions ϕ with a strict inequality for some η. We demonstrate, that under certain conditions the RMLE dominates the PAVA estimator in terms of the mean squared error when $p = 2$, and that under certain conditions the PAVA estimator dominates the UMLE. In the case of $p > 2$ we will consider a variety of covariance structures, that are commonly encountered in the theory of experimental designs, clinical trials, econometrics, etc. Under certain conditions on the elements of the covariance matrix we demonstrate that the PAVA estimator universally dominates the UMLE. Since for certain patterns of covariance matrices, the PAVA estimator and the RMLE are the same (Diaz and González, 1988), thus, in such situations we actually derive universal domination results for the RMLE. All proofs are provided in the Appendix A of this document.

1.2.1. Some special covariance structures

Since it may not be possible to investigate the properties of the PAVA estimator for arbitrary covariance matrices, in this dissertation the universal domination of the PAVA estimator over the UMLE will be explored under some special covariance structures. The covariance structures considered in this dissertation are described below.

1.2.1.1. Supplemented balance designs (SBD)

There exists an exhaustive amount of literature on experimental designs for comparing treatment groups against a control group. For an efficient design it is well known that the number of replicates for the control group should be larger than the total number of treatment groups (Pearce, 1960). The basic idea of SBD is to "supplement" (also referred to as "augment" or "reinforce") a block design consisting of $p-1$ treatment groups by *m* replicates of the control group in each block. Typically these designs are such, that every pair of treatments occurs with equal frequency in all blocks, and the frequency of co-occurrence of a treatment and the control is constant in all blocks. Pearce (1960) termed these designs *Supplemented Balance Designs (SBD)*. Properties of such designs have been well studied in the literature (Stufken, 1987; Hedayat, Jacroux and Majumdar, 1988; Gupta, 1989). Consider a SBD where observations are taken in blocks of size $m+(p-1)n$, with *m* observations in a block taken on the control treatment, $i=1$, and *n* observations taken on each of the treatments $i=2,\ldots,p$. Observations are normally distributed with variance σ^2. While observations from different blocks are independently distributed, observations within

each block are assumed to be correlated with a correlation coefficient $\tilde{\rho}$. Denote $\sigma_1^2 = \sigma^2\left[1+(m-1)\tilde{\rho}\right]/m$, $\sigma_2^2 = \sigma^2\left[1+(n-1)\tilde{\rho}\right]/n$ and

$$\rho = \frac{\tilde{\rho}}{\sqrt{\left[1+(m-1)\tilde{\rho}\right]/m}\sqrt{\left[1+(n-1)\tilde{\rho}\right]/n}}.$$

Let $\hat{\boldsymbol{\theta}}^{UMLE}$ denote the UMLE of $\boldsymbol{\theta}$. The first component of $\boldsymbol{\theta}$ is the mean of the control group, and the remaining $p-1$ components are the means of the treatment groups. Then the variance-covariance matrix of the vector $\hat{\boldsymbol{\theta}}^{UMLE}$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \boldsymbol{1}' \\ \rho\sigma_1\sigma_2 \boldsymbol{1} & \sigma_2^2 \mathbf{K} \end{pmatrix}, \tag{5}$$

where $\boldsymbol{1}$ is a vector of 1s, $\rho \geq -1/(p-2)$ and $\mathbf{K} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$. As usual, $\mathbf{I}$ denotes the $(p-1)\times(p-1)$ identity matrix, and $\mathbf{J}$ is a $(p-1)\times(p-1)$ matrix of 1s (see Nigam et al., 1988, for more details).

Note that the above covariance structure also arises naturally in other contexts as well, such as graphical models (Whittaker, 1990; Lauritzen, 1996), and lattice model (Andersson and Perlman, 1993; Dempster, 1972). For a review on $\boldsymbol{\Sigma}$ one may refer to Sun and Sun (2005), where the authors consider a more general form of this matrix.

In this dissertation we focus on estimation of the control mean θ_1, the largest mean θ_p and elementary contrasts of treatment means with the control mean $\theta_i - \theta_1$, $i = 2, 3, \ldots, p$ under the constraint that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$. In section 2.2 (page 23) we present Theorem 2.2 demonstrating that in case of SBD covariance structure (5) and simple order restriction on the mean components, PAVA performs better than UMLE for

the control and the highest dose groups. Supporting simulation results are presented in Figure 5 and Figure 6.

We also argue that, since Theorem 2.2 also holds in the case of simple tree order restriction as well, PAVA may perform better than UMLE for the control group in this case. Results of supporting simulation studies are presented in Figure 7.

1.2.1.2. Designs where 1 is an eigenvector of the covariance matrix

There are many designs used in clinical trials and in other applications where every principal sub-matrix of $\boldsymbol{\Sigma}$ has $\boldsymbol{1} = (1,1,...,1)'$ as an eigenvector. Some common examples include: (a) the intra-class covariance matrix of the form $\boldsymbol{\Sigma} = \alpha\mathbf{I} + \beta\boldsymbol{J}$, where $\mathbf{J}$ is a matrix of 1's, (b) cross-over designs where patients in Group A receive treatments 1, 3, 5, etc, patients in Group B receive treatments 2, 4, 6, etc; all observations have the same variance, and the correlation coefficient within subject is same in both groups, and (c) covariance matrix of elementary contrasts

$\hat{\boldsymbol{\delta}} = \left(\hat{\theta}_2^{UMLE} - \hat{\theta}_1^{UMLE}, \hat{\theta}_3^{UMLE} - \hat{\theta}_1^{UMLE}, ..., \hat{\theta}_p^{UMLE} - \hat{\theta}_1^{UMLE}\right)'$ in a SBD.

1.2.1.3. Covariance matrix in a star-shaped ordering

Suppose $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \mathbf{I})$, then the components of $\boldsymbol{\mu}$ are said to satisfy a star-shaped order if $\mu_1 \le (\mu_1 + \mu_2)/2 \le ... \le (\mu_1 + \mu_2 + ... + \mu_p)/p$. As discussed in Shaked (1979) and Dykstra and Robertson (1982), star-shaped order restriction arises naturally in many applications. Performing a liner transformation $\hat{\theta}_i = (X_1 + ... + X_i)/i$, we have $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ with $\theta_i = (\mu_1 + ... + \mu_i)/i$ satisfying the simple order restriction $\theta_1 \le \theta_2 \le ... \le \theta_p$. Note that the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/2 & 1/3 & \dots & 1/p \\ 1/2 & 1/2 & 1/3 & \dots & 1/p \\ 1/3 & 1/3 & 1/3 & \dots & 1/p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/p & 1/p & 1/p & \dots & 1/p \end{pmatrix}. \quad (6)$$

In section 2.2 (page 29), we present Theorem 2.3 demonstrating that in case of star-shaped order restriction on the mean components, PAVA may perform better than UMLE for the control group. Supporting simulation results are presented in Figure 3.

1.3. Testing of hypotheses

Testing hypotheses under inequality constraints is a well researched area. For a comprehensive account, one may refer to the recent book by Silvapulle and Sen (2005). In addition to the standard likelihood ratio test, a variety of alternative tests are available for comparing means of two or more independent normal populations when no covariates are present. Suppose $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{I})$ and suppose one is interested in testing the following hypotheses regarding the components of $\boldsymbol{\theta}$, where the alternative hypothesis is the simple tree order

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_p \text{ versus } H_A : \theta_1 \le \theta_i,\ 2 \le i \le p$$

(with at least one strict inequality). In addition to the classical likelihood ratio test, a popular alternative test for testing the above hypotheses is the Dunnett's test (Dunnett, 1955), which is defined as follows. For an elementary contrast $\theta_i - \theta_1$, the UMLE is $\hat{\theta}_i^{UMLE} - \hat{\theta}_1^{UMLE}$, $i \ge 2$. An estimator of the variance of this estimator is given by $\hat{V}ar\left(\hat{\theta}_i^{UMLE} - \hat{\theta}_1^{UMLE}\right) = 2\hat{\sigma}^2$, where $\hat{\sigma}^2$ is the usual mean residual sum of squares in the

linear model. The Dunnett's test statistic is given by $\max_i \left|\hat{\theta}_i^{UMLE} - \hat{\theta}_1^{UMLE}\right| \Big/ \hat{\sigma}\sqrt{2}$. As noted by Marcus and Talpaz (1992), a potential weakness of this statistic is that its numerator does not use the inequality constraint specified by the alternative hypothesis. Accordingly, they modified the numerator of the statistic by replacing the UMLE by the RMLE of $\boldsymbol{\theta}$ under the simple tree order constraint. Although the resulting test improves the power of Dunnett's test for certain choices of $\boldsymbol{\theta}$, it is surprising that it does not improve the power uniformly for all $\boldsymbol{\theta}$. More recently Tang and Lin (1997) introduced an Approximate Likelihood Ratio (ALR) test that can be used for testing the above hypothesis. A distinct advantage of ALR is that it is computationally simple to implement for any *p* and it performs very well in terms of power in comparison to both the classical likelihood ratio test as well as Dunnett's test for certain choices of $\boldsymbol{\theta}$. The procedure of Marcus and Talpaz (1992) was inspired by the earlier papers of Williams (1971, 1972, 1977). In his 1971 and 1972 papers, Williams discussed the problem of comparing means of the treatment groups with the control group in a dose response study where the population means are assumed to be non-decreasing with dose (i.e. $\theta_1 \le \theta_2 \le ... \le \theta_p$). In Williams (1971) the test statistic was $\dfrac{\left(\hat{\theta}_p^{RMLE} - \hat{\theta}_1^{UMLE}\right)}{\hat{\sigma}\sqrt{2}}$, where as in Williams (1972) he used $\dfrac{\left(\hat{\theta}_p^{RMLE} - \hat{\theta}_1^{RMLE}\right)}{\hat{\sigma}\sqrt{2}}$, where $\hat{\theta}^{RMLE}$ is the RMLE under the simple order constraint. In his 1977 paper, Williams considered the problem of testing $H_0 : \theta_1 = \theta_2 = ... = \theta_p$ versus $H_a : \theta_1 \le \theta_2 \le ... \le \theta_p$ (with at least one strict inequality) using the statistic $\dfrac{\left(\hat{\theta}_p^{RMLE} - \hat{\theta}_1^{RMLE}\right)}{\hat{\sigma}\sqrt{2}}$.

Nonparametric versions of Dunnett's test and Williams' test were also developed in the literature using rank based methods by Dunn (1964) and Shirley (1977), respectively. These methods are widely used in practice for their practical simplicity, and simulation studies reported in the literature suggest, that these methods compete very well against the likelihood ratio tests.

It is reasonable to anticipate or assume a monotonic mean response in dose response studies conducted by toxicologists. In such situations, the Williams' test (1972, 1977) tends to have a higher power than the Dunnett's test when comparing the mean response at the highest dose with that of the control group. However, there are instances where, perhaps due to toxicity at high doses, the mean response at the higher doses may change direction resulting in a down-turn (or up-turn) in the mean response. In such situations, the Williams' test (1972, 1977) loses power in comparison to the Dunnett's test when comparing the mean of the highest dose with the control. Typically, in the analysis of their 90 day pre-chronic rodent cancer studies, the National Toxicology Program (NTP) uses either Dunnett's test or Williams' test depending upon the data – which is unsatisfactory. Since in practice it is not feasible to determine a priori whether the departure from monotonic response will take place or not, it is important to develop a method that would be robust to both possibilities. In Peddada et al. (2006) such a robust procedure was developed using the point estimators developed in Hwang and Peddada (1994). The resulting methodology seems to perform as well as the Williams' test when the mean responses are monotonic in dose and performs as well as Dunnett's test when the mean responses depart from monotonicity at the higher doses.

In many applications such as in epidemiology, researchers are often interested in testing for the equality of mean responses of various groups against the alternative hypothesis that the means are constrained by some inequality constraints, after adjusting for various covariates. Often such problems can be formulated using fixed effects linear models and applying the likelihood ratio methodology as detailed in Silvapulle and Sen (2005). Several variations to the likelihood ratio principle have been proposed in the literature. As previously stated, the likelihood ratio principle provides a rich framework to conduct the analyses of such data. In the presence of covariates (whether continuous or categorical), the unrestricted estimators of treatment means are not necessarily independently distributed. In such a case, as noted in the previous sections and in Hwang and Peddada (1994) and others, the RMLE may not perform well as an estimator of the mean vector. Consequently, one cannot assume that the likelihood ratio based methods would perform well in terms of power since they use RMLE. For this reason, Betcher and Peddada (2009) developed a Dunnett-type test statistic that uses a modified RMLE as the point estimator of the mean vector. Based on the simulation studies reported in Betcher and Peddada (2009), in the case of simple order, their new method provides better confidence intervals than those based on RMLE.

Constrained inference in linear mixed effects models arises naturally in many applications, such as the ones described in this Chapter1. Specifically, they arise naturally in the context of repeated measurement designs. Silvapulle (1997) proposed a simple methodology for testing linear constraints regarding fixed effects parameters under some conditions on the design matrices. The resulting test procedure does not seem to depend upon the unknown variance components.

In the case of general mixed effects models, a natural strategy for testing for inequalities among treatment effects after adjusting for covariates would be to develop a likelihood ratio test. Such a strategy necessarily requires the derivation of RMLE under inequality constraints. As noted in the literature, this is a very challenging problem. Hoferkamp and Peddada (2002) proposed an EM based algorithm for estimating regression parameters under constraints, when the error variances are heteroscedastic and potentially subject to inequality constraints. Under some conditions on design matrices, they discussed the convergence of their algorithm. Recently, Shi et al. (2005) addressed the problem in a slightly different context. They considered the usual fixed effects model but allowed the error variance to be multivariate normally distributed with an unknown non-diagonal covariance matrix $\boldsymbol{\Sigma}$. Thus, unlike the linear mixed effects model where $\boldsymbol{\Sigma}$ has a special structure, in Shi et al. (2005) it was not constrained by a particular structure.

They developed EM algorithm to estimate regression parameters subject to inequality constraints in such a linear model and identified conditions under which the EM algorithm converges. None of these papers discusses the problem of testing regression parameters under constraints in a linear mixed effects model. They are all limited to the constrained estimation problem and none of these papers address the testing problem.

Apart from the earlier attempts in some special cases, tests for linear mixed models under inequality constraints on the fixed effects parameters has not been well developed in the literature. For example, Mukerjee (1988) noted that the usual tests for order restrictions on the means of independent normal populations can be extended to the

case when normal populations are correlated as in a repeated measurements design. In this paper, the author does not include any covariates. Silvapulle (1997) generalized Mukerjee (1988) to some unbalanced designs with incomplete data. He noted that within-subject correlations make it difficult to generalize some tests into repeated models. Earlier, Singh and Wright (1990) considered order restricted inference on fixed effects in a two-factor mixed model. They presented an analogue to the usual *F*-test for homogeneity and obtained several closed-form results.

There did not exist a systematic general methodology for the analysis of linear mixed effects models, when the regression parameters are subject to inequality constraints, until Davidov and Rosen (2011), who derived the likelihood ratio test for testing the hypotheses of the type $H_0 : \boldsymbol{\eta} = 0$ vs $H_A : \boldsymbol{\eta} \geq 0$ when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$. In Section 3.2, the likelihood ratio test of Davidov and Rosen (2011) is reviewed, and in Section 3.3, the likelihood ratio test under heteroscedastic error structure $\boldsymbol{\Sigma} = diag\left[\sigma_1^2 \mathbf{I}_{n_1} : \sigma_2^2 \mathbf{I}_{n_2} : \ldots : \sigma_k^2 \mathbf{I}_{n_k} \right]$ is derived. Motivated by various limitations of these likelihood ratio tests, in Section 3.4 an EBLUP bootstrap methodology is described under homoscedastic as well as heteroscedastic error structures. An alternative method analogous to Williams (1971) and based on Rao's MINQUE theory (1970, 1971, 1972) is explored in Section 3.5. In Chapter 4, extensive simulation studies are performed to evaluate the performance of various tests in terms of the Type I error and power. Since very limited literature is available for this very important practical problem, this dissertation work extends the existing knowledge in this field substantially. The proposed methodologies are illustrated in Chapter 5 using the recently published succimer data (Cao et al., 2011).

CHAPTER 2

CONSTRAINED ESTIMATION AND THE PERFORMANCE OF PAVA

In this section, we will describe some of the theoretical and simulation results obtained so far in this dissertation with regards to the constrained estimation problem. Proofs of theorems are presented in Appendix A.

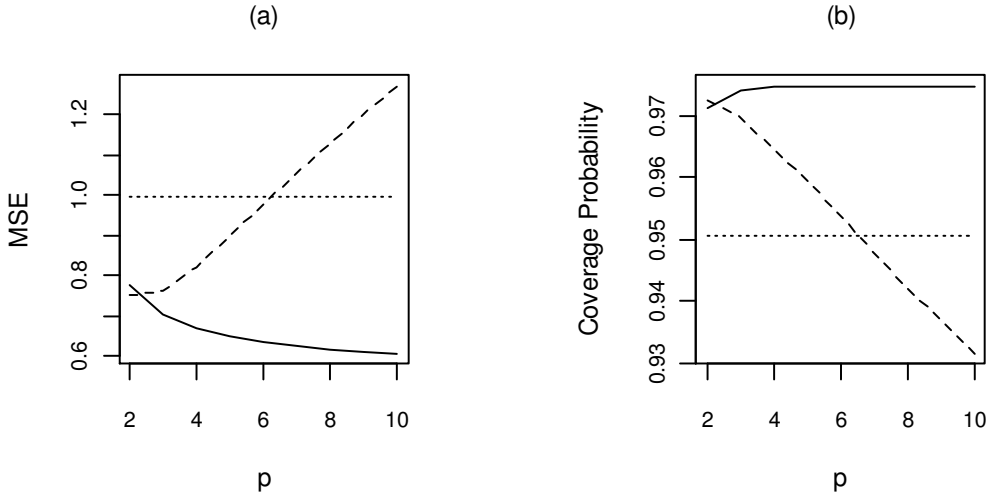
We assume that UMLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^{UMLE}$, is distributed according to a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$. As often done, without loss of generality, unless stated otherwise, we assume the sample size of 1, because it can be absorbed in $\boldsymbol{\Sigma}$. In general, the order restricted estimators (whether RMLE or other constrained estimators) do not always perform well in all settings. Their performance depends upon the type of inequality constraints as well as the covariance structure and the dimension p (cf. Lee, 1988; Hwang and Peddada, 1994; Fernandez et al., 1999).

To illustrate this point, we provide results of a small simulation study in Figure 3. In this study, we simulated data from a p-variate normal distribution with mean vector $\boldsymbol{\theta} = 0$ and covariance matrix $\boldsymbol{\Sigma}$ given by (6).

Under the constraint $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$, we estimated the MSE of the UMLE, RMLE and PAVA estimator of θ_1 (Figure 3 (a)). We also simulated the coverage probabilities of fixed width confidence intervals centered at the UMLE, RMLE and

PAVA estimator, i.e. $P\left(\left|\hat{\theta}_1^{UMLE} - \theta_1\right| < 1.96\right)$, $P\left(\left|\hat{\theta}_1^{RMLE} - \theta_1\right| < 1.96\right)$ and $P\left(\left|\hat{\theta}_1^{PAVA} - \theta_1\right| < 1.96\right)$ (Figure 3 (b)). We chose the value of 1.96, because for this value of the critical constant, the confidence interval centered at UMLE has a coverage probability of 0.95. All results are based on 100,000 simulation runs.

Figure 3. MSE and Coverage Probability of θ_1: UMLE (dotted line), RMLE (dashed line), PAVA estimator (solid line). The data are simulated with $\boldsymbol{\theta} = 0$ and $\boldsymbol{\Sigma}$ given by (6); parameters are estimated under the constraint $\theta_1 \le \theta_2 \le \ldots \le \theta_p$.



From the Figure 3, it is clear that for the covariance matrix considered in this example, the RMLE performs poorly both in terms of MSE as well as the coverage probability, as p increases, while the PAVA estimator performs the best. In view of the above illustration and the fact that PAVA is widely used in practice even for correlated data, we investigate its performance relative to UMLE in terms of universal domination criterion.

2.1. Performance of PAVA in the case of $p = 2$

We begin by comparing the MSE of the UMLE, RMLE and PAVA estimator for $p = 2$ to demonstrate that, even in this simple setting, the performance of PAVA can depend upon the underlying correlation structure. We assume that $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, $\theta_1 \le \theta_2$ and the elements of $\boldsymbol{\Sigma}$ have no special structure with $Var\left(\hat{\theta}_1^{UMLE}\right) = \sigma_1^2$, $Var\left(\hat{\theta}_2^{UMLE}\right) = \sigma_2^2$, and $Cov\left(\hat{\theta}_1^{UMLE}, \hat{\theta}_2^{UMLE}\right) = \rho\sigma_1\sigma_2$.

Theorem 2.1:

(a) If $\rho \le 0$ and either $\theta_2 \ge \theta_1 \ge 0$ or $0 \ge \theta_2 \ge \theta_1$, then

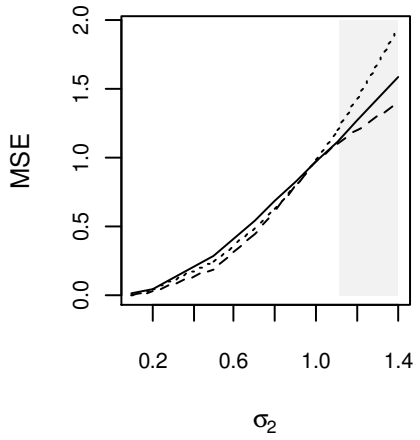
$$E(\hat{\theta}_2^{PAVA} - \theta_2)^2 \le E(\hat{\theta}_2 - \theta_2)^2;$$

(b) if $\theta_1 \le \theta_2$ and $\rho(\sigma_2 - \sigma_1)(\rho\sigma_2 - \sigma_1) \ge 0$, then

$$E(\hat{\theta}_2^{RMLE} - \theta_2)^2 \le E(\hat{\theta}_2^{PAVA} - \theta_2)^2.$$

To understand the performance of PAVA in the case, when the sufficient conditions of Theorem 2.1 are not true, we simulated the data from bivariate normal distributions with mean vector $(0,0)'$, $\sigma_1 = 1$, $\rho = 0.9$ and σ_2 ranging from 0.1 to 1.4. Under the constraint $\theta_1 \le \theta_2$, the estimated MSE of UMLE, RMLE and PAVA estimator are provided in Figure 4.

Figure 4. MSE of θ_2 as a function of σ_2: UMLE (dotted line), RMLE (dashed line), PAVA estimator (solid line). The data are simulated with $\theta_1 = \theta_2 = 0$, $\rho = 0.9$, $\sigma_1 = 1$; parameters are estimated under the constraint $\theta_1 \le \theta_2$. Shaded area shows the values of σ_2, where the conditions of Theorem 2.1 are satisfied, i.e., $\sigma_2 \ge \sigma_1/\rho$.



Theorem 2.1, together with Figure 4, suggests that the performance of PAVA depends upon the underlying correlation structure even in the case of $p = 2$. We note from Figure 4, that RMLE performs better than PAVA for the choice of parameters considered in this simulation study. In view of the above findings, we deduce that domination results may not exist for arbitrary covariance structures when $p > 2$. Therefore in section 2.2 we consider some covariance structures that arise naturally in many applications and investigate the performance of PAVA relative to UMLE for those structures.

2.2. Performance of PAVA for p > 2 under various covariance structures

2.2.1. SBD covariance structure

It is well-known that in many situations the total MSE of the RMLE is smaller than that of the UMLE (cf. Fernandez et al., 1999). Surprisingly, based on a small simulation study reported in Figure 5, we discover that the total MSE of the PAVA

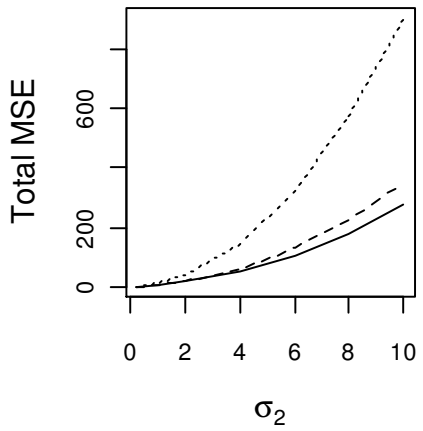
estimator is not only smaller than the total MSE of the UMLE but it is almost as small as the total MSE of the RMLE, if not smaller. We find this to be an interesting and a surprising result. In this simulation experiment, $\hat{\boldsymbol{\theta}}^{UMLE}$, the UMLE of $\boldsymbol{\theta}$, was generated according to a multivariate normal distribution with mean $\boldsymbol{\theta}$ and the covariance matrix given by (5), $p = 10$, $\sigma_1 = 1$ and $\rho = 0.4$. However, since it is well-known that reduction in the total MSE does not necessarily imply a reduction in the MSE of individual coordinates (Lee, 1988; Fernandez et al., 1999), in Figure 6 we investigated the performance of the PAVA estimator of the control mean θ_1 and the largest mean θ_p in terms of universal domination criterion. We identify sufficient conditions, under which PAVA performs better than UMLE. Recall from Hwang (1985) that universal domination is equivalent to domination in terms of all monotonic functions of quadratic loss and hence implies domination in terms of MSE. Analytical comparisons between PAVA and RMLE appear to be intractable and hence are not discussed here.

Theorem 2.2: Suppose $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is of the form (5) and suppose $\theta_1 \le \theta_2 \le \ldots \le \theta_p$.

(a) If either $-\dfrac{(p-1+\sigma_1^2/\sigma_2^2)}{(p-1)(p-2+2\sigma_1/\sigma_2)} < \rho < 0,\ \sigma_1 < \sigma_2$ or $\rho > 0,\ \sigma_1 > \sigma_2$, then for all $c > 0$, $P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right) \ge P\left(\left|\hat{\theta}_1^{UMLE} - \theta_1\right| < c\right)$.

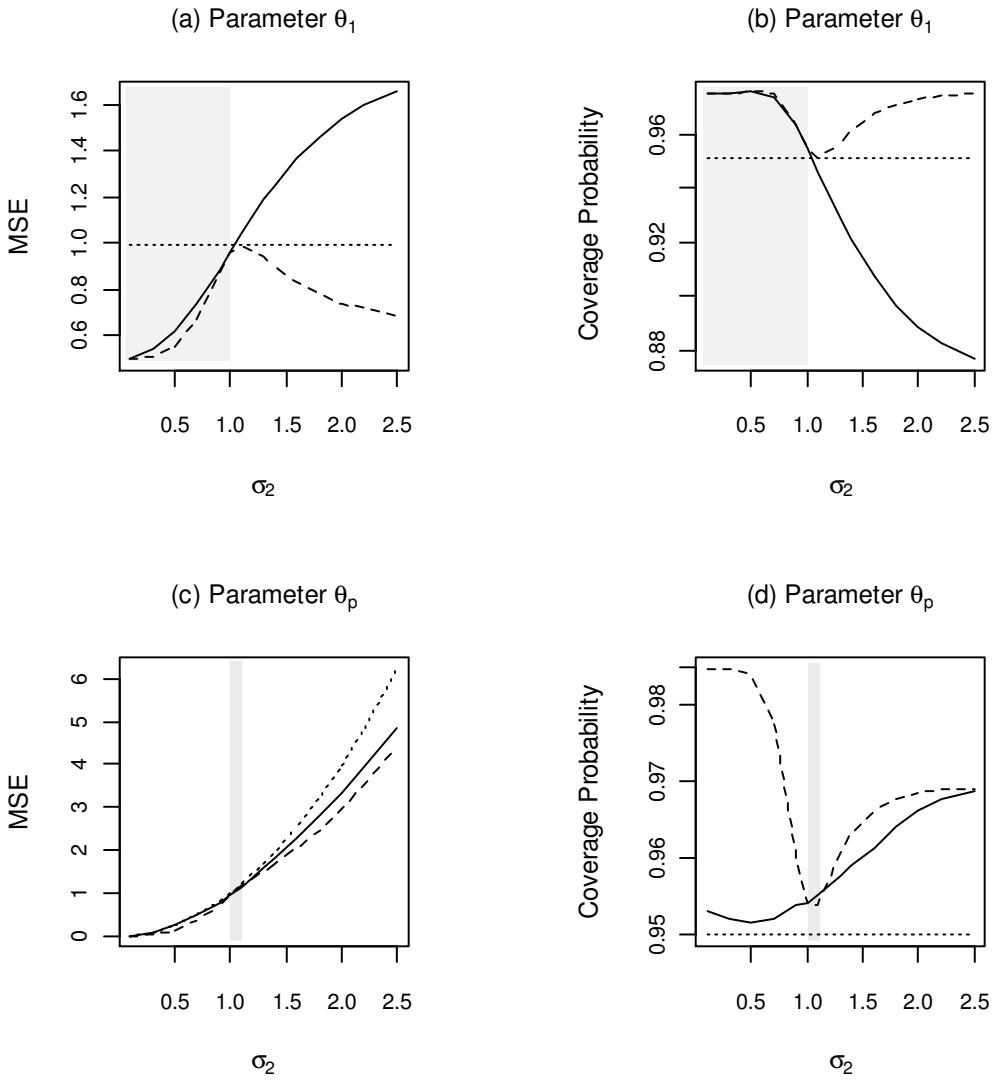
(b) If either $-\dfrac{(p-1+\sigma_1^2/\sigma_2^2)}{(p-1)(p-2+2\sigma_1/\sigma_2)} < \rho < 0,\ \sigma_1 > \sigma_2$ or $0 < \rho < \dfrac{\sigma_1}{\sigma_2} < 1$, then for all $c > 0$, $P\left(\left|\hat{\theta}_p^{PAVA(p)} - \theta_p\right| < c\right) \ge P\left(\left|\hat{\theta}_p^{UMLE} - \theta_p\right| < c\right)$.

Figure 5. Total MSE as a function of σ_2: UMLE (dotted line), RMLE (dashed line), PAVA estimator (solid line). The data are simulated with $p = 10$, $\theta_i = 0$, $i = 1, \ldots, p$, $\mathbf{\Sigma}$ given by (5), $\rho = 0.4$, $\sigma_1 = 1$; parameters are estimated under the constraint $\theta_1 \le \theta_2 \le \ldots \le \theta_p$.



We performed extensive simulation studies to compare UMLE, RMLE and PAVA including the situations where the sufficient conditions of the above theorem are not satisfied. A small sample of the results is provided in Figure 6. As expected, PAVA performs well in terms of MSE as well as the coverage probability, when the sufficient conditions of Theorem 2.2 are satisfied. However, its performance can be rather poor when the sufficient conditions are not satisfied. It is important to recognize that the sufficient conditions provided in parts (a) and (b) of Theorem 2.2 are disjoint. Together with the fact that simulation results suggest these conditions may even be necessary, we conclude that universal domination results for PAVA of θ_i, $1 < i < p$ over the corresponding UMLE may not exist.

Figure 6. MSE and Coverage Probability of θ_1 and θ_p as a function of σ_2: UMLE (dotted line), RMLE (dashed line), PAVA estimator (solid line). The data are simulated with $p=5$, $\theta_i=0$, $i=1,\ldots,p$, $\mathbf{\Sigma}$ given by (5), $\rho=0.9$, $\sigma_1=1$; parameters are estimated under the constraint $\theta_1 \le \theta_2 \le \ldots \le \theta_p$. Shaded area shows the values of σ_2, where the conditions of Theorem 2.2 are satisfied, i.e. $\sigma_2 < \sigma_1$ for (a), (b) and $\sigma_1 < \sigma_2 < \sigma_1/\rho$ for (c), (d).



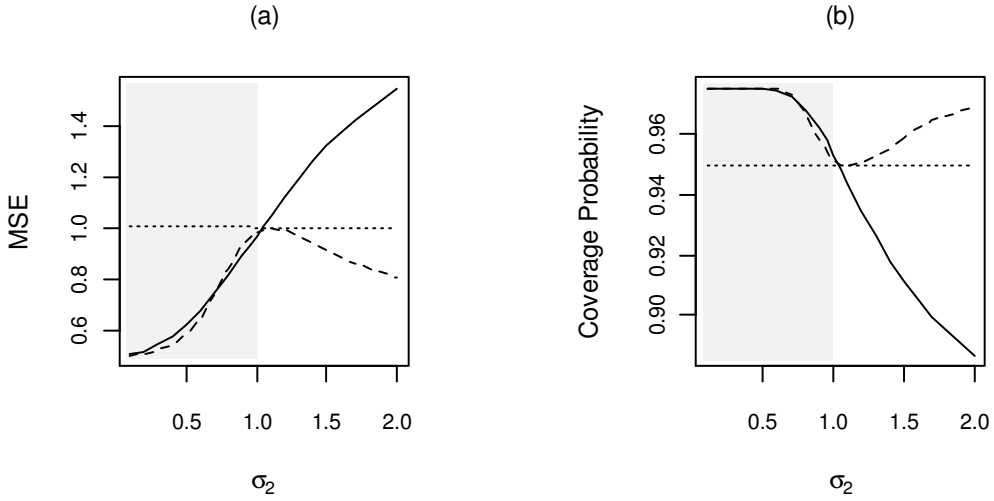
Note that the proof of Theorem 2.2 does not use any information regarding the inequalities among $\theta_2, \theta_3, ..., \theta_p$ but only uses the information that $\theta_1 \le \theta_i$, $i=2,3,\ldots,p$. Therefore, under a simple tree order constraint, we obtain the following corollary for the following PAVA based estimator derived from (2):

$$\hat{\theta}_1^{PAVA(p)} = \min_{1 \le t \le p} \frac{\sum_{j=1}^{t} w_j \hat{\theta}_j^{UMLE}}{\sum_{j=1}^{t} w_j}.$$

Corollary 1: Suppose $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is of the form (5) and suppose $\theta_1 \le \theta_i$, $i \ge 2$. If either $-\frac{(p-1+\sigma_1^2/\sigma_2^2)}{(p-1)(p-2+2\sigma_1/\sigma_2)} < \rho < 0$, $\sigma_1 < \sigma_2$ or $\rho > 0$, $\sigma_1 > \sigma_2$, then for all $c > 0$, $P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right) \ge P\left(\left|\hat{\theta}_1^{UMLE} - \theta_1\right| < c\right)$.

Again, as above, the simulation results provided in Figure 7 suggest that PAVA performs very well relative to both UMLE and RMLE when the sufficient conditions of Corollary 1 are satisfied (shaded area). Otherwise, its performance can be very poor (unshaded area).

Figure 7. MSE and Coverage Probability of θ_1 as a function of σ_2: UMLE (dotted line), RMLE (dashed line), PAVA estimator (solid line). The data are simulated with $p = 5$, $\theta_i = 0$, $i = 1, \ldots, p$, $\boldsymbol{\Sigma}$ given by (5), $\rho = 0.9$, $\sigma_1 = 1$; parameters are estimated under the constraint $\theta_1 \le \theta_i$, $i \ge 2$. Shaded area shows the values of σ_2 where the conditions of Corollary 1 are satisfied, i.e. $\sigma_2 < \sigma_1$.



2.2.2. Covariance matrices where 1 is an eigenvector

Recall the design described in section 1.2 on page 13, where every principal sub-matrix of Σ has $\mathbf{1} = (1,1,...,1)'$ as an eigenvector. Following arguments similar to those in the proof of Theorem 2.2 or by appealing to Hwang and Peddada (1994), we deduce the following important corollary. Note that, different from Theorem 2.2, the following result applies to all coordinates of the mean vector $\boldsymbol{\theta}$.

Corollary 2: Suppose $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. If every principal sub-matrix of $\boldsymbol{\Sigma}$ has $1 = (1,1,...,1)'$ as an eigenvector, and suppose that $\theta_1 \le \theta_2 \le ... \le \theta_p$, then for all $i = 1,2,...,p$ and $c > 0$, $P\left(\left|\hat{\theta}_i^{PAVA(p)} - \theta_i\right| < c\right) \ge P\left(\left|\hat{\theta}_i^{UMLE} - \theta_i\right| < c\right)$.

In the case of intra-class covariance structure, from Theorem 2.2 of Diaz and González (1988), we deduce that RMLE and PAVA are identical. Hence in that case the above corollary applies to RMLE as well.

2.2.3. Star-shaped order covariance structure

Recall the star-shaped order restriction (defined in section 1.2.1 on page 13) with covariance matrix $\boldsymbol{\Sigma}$ given by (6). Appealing to Theorem 2.2 in Diaz and González (1988) we note that PAVA and RMLE of μ are the same, but PAVA and RMLE of θ are not the same. As observed in the simulation study reported in Figure 3, RMLE of θ_1 can perform very poorly as p increases, but PAVA performs very well for all p. In the following theorem we demonstrate this fact analytically.

Theorem 2.3: Suppose $\hat{\boldsymbol{\theta}}^{UMLE} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$, $\theta_1 \le \theta_2 \le ... \le \theta_p$, and $\boldsymbol{\Sigma}$ is given by (6). For all $c > 0$, $P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right) \ge P\left(\left|\hat{\theta}_1^{UMLE} - \theta_1\right| < c\right)$.

2.3. Conclusions and recommendations

Often in clinical trials repeated measurements are made on each subject, the investigator expects trends in the mean response among dose groups and/or time points, and the problem of interest is to estimate and test parameters under such constraints on the mean responses. For example, Ivanova et al. (2009) describe such a Phase II trial, where each patient received control and two doses of the drug. The Pool Adjacent Violators Algorithm (PAVA) was designed for estimating parameters under the simple order restriction (i.e. increasing or decreasing order among the mean responses), when unrestricted estimators of parameters are independent. However, PAVA is also often used even when unrestricted estimators are correlated. Based on the results obtained in this research, it appears that simple PAVA based algorithms may be reasonable even if the unrestricted estimators are correlated.

For example, in a Supplemented Balance Design (SBD), where a researcher is interested in estimating elementary contrasts of each dose group with the control group (under the constraint that the mean responses are monotonic in dose), we found that the confidence interval centered at PAVA estimate of an elementary contrast between the dose group and the control group will have larger coverage probability than the confidence interval centered at the UMLE of the contrast. Thus, PAVA is recommended over UMLE for estimating all elementary contrasts of dose groups with the control group. If variances of the control group and treatment group under a SBD are equal, then the

covariance matrix of the sample mean vector has an intra-class covariance structure and satisfies the conditions of Corollary 2; thus, PAVA is recommended over UMLE for estimating all treatment means under the simple order constraint, and also, all elementary contrasts when they are subject to the simple order constraint. We also note that for estimating the control group mean (under either simple order or simple tree order restriction on treatment means), the PAVA performs better than the UMLE if: 1) the variance in the control group is smaller than the variance in the treatment group and the correlation between groups is negative or 2) the variance in the control group is larger than the variance in the treatment group, and the correlation between groups is positive.

There are situations in clinical trials, when a large number of treatments need to be compared, and not all treatments can be present in each block. In such case, a balanced incomplete block design (BIBD) may be considered. For illustration, consider a dose-response study consisting of control, low-dose and high-dose groups of a drug, and litters of mice are taken to be blocks. A BIBD can be constructed as follows. Suppose each block (litter) consists of two pups. The pups in the first block are randomly assigned to either control or low-dose group; pups in the second block are randomly assigned to either control or high-dose group; and the pups in the third block are randomly assigned to the low or high-dose group. The resulting design is a BIBD.

A feature of a BIBD is that all blocks have the same number of treatments, and all treatments, as well as all pairs of treatments, are observed the same number of times in the experiment. Furthermore, the correlation coefficient between sample means is constant for any pair of treatments within a block and is zero across blocks. As a consequence, the covariance matrix of the sample mean vector of all treatment means in a

BIBD has an intra-class covariance structure. Thus, the covariance matrix satisfies the conditions of Corollary 2. Note that the randomized complete block design (RCBD) can be thought as a special case of BIBD. Thus, Corollary 2 also applies to an RCBD. Thus, in these cases PAVA is recommended over UMLE for estimating all treatment means under the simple order constraint, and more importantly, all elementary contrasts when they are subject to simple order constraint. Again, a confidence interval centered at PAVA of any such contrast will have larger coverage probability than the confidence interval centered at the UMLE of the contrast.

Analytical comparisons between PAVA and RMLE appeared to be intractable and hence were not discussed in Chapter 2. However, performed simulations indicate that the RMLE might perform better or worse than PAVA, depending on the covariance matrix of the UMLE. It is known that in many situations the total mean squared error (MSE) of the RMLE is smaller than that of the UMLE. Surprisingly, based on a small simulation study under SBD, when the variance in the control group is smaller than the variance in the treatment group, and the correlation between groups is positive, we discover that the total MSE of the PAVA estimator is not only smaller than the total MSE of the UMLE, but it is also smaller than the total MSE of the RMLE. Note, that if variances of the control group and treatment group under a SBD are equal, RMLE and PAVA estimates of treatment means under the simple order constraint are the same. Star-shaped order is a known example where the RMLE does not perform well. We have shown that PAVA is superior to UMLE as well as RMLE for estimating the control mean in the case, where treatment means are under a star-shaped order constraint.

In general, it may not be possible to recommend an estimation procedure for an arbitrary experimental design.

CHAPTER 3

CONSTRAINED TESTING IN A LINEAR MIXED EFFECTS MODEL

Motivated by the data of Cao et al. (2011) discussed in Chapter 1, the focus of this chapter is to develop statistical methodology for performing constrained inference on the location parameters of a linear mixed effects model, where covariance structure is of the form $Cov(\boldsymbol{Y}) = \mathbf{UTU}' + \boldsymbol{\Sigma}$, where $\mathbf{T}$ and $\boldsymbol{\Sigma}$ are diagonal matrices. Although such a structure is reasonable in the motivating example and is often used when analyzing repeated measures data (cf. Khattree and Naik, 1999), in general, however, depending upon the application, the covariance structures may be more complicated or unspecified. For example, in a random slopes model for repeated measurement designs, it is common to have the structure of $\mathbf{T}$ to be of the form $\mathbf{T} = \mathbf{I} \otimes \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a non-diagonal matrix. A common choice for $\boldsymbol{\Omega}$ is the auto-correlation structure. There are also instances where the structure of the covariance matrix $Cov(\boldsymbol{Y})$ may not be pre-specified.

In Sections 3.2 and 3.3 we describe the likelihood ratio tests (LRT), developed in Davidov and Rosen (2011) for constrained inference in linear models with covariance structure of the form $Cov(\boldsymbol{Y}) = \mathbf{UTU}' + \boldsymbol{\Sigma}$, where $\mathbf{T}$ and $\boldsymbol{\Sigma}$ are diagonal matrices. However, as will be seen, the LRT depends upon unknown parameters $\mathbf{T}$ and $\boldsymbol{\Sigma}$ and hence cannot be used directly in most biostatistical and other applications. Consequently,

in Sections 3.4 and 3.5 of this dissertation, a bootstrap methodology is introduced, which can be used for most problems commonly encountered in biostatistics and public health. Some concluding remarks are made in Section 3.6, where we summarize the limitations of the proposed methodology and suggest simple modifications for some alternate covariance structures. As noted in Chapter 1, very limited literature exists on the analysis of linear mixed effects models, when the regression parameters are subject to inequality constraints. Thus this dissertation research makes an important contribution to the literature for some special covariance structures. The proofs of theorems are provided in Appendix B.

3.1. The model and notations

Let

$$\boldsymbol{Y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 + \mathbf{U}\boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{7}$$

denote a linear mixed effects models where $\boldsymbol{\theta}_1$ is the vector of treatment effects of the order $p_1 \times 1$, $\mathbf{X}_1$ is a design matrix of the order $N \times p_1$ consisting of 0s and 1s, $\mathbf{X}_2$ is a known matrix of covariates of the order $N \times p_2$ with corresponding (unknown) regression parameter vector $\boldsymbol{\theta}_2$ of the order $p_2 \times 1$, and $\mathbf{U}$ is a $N \times c$ matrix of known design constants. For convenience, we denote $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ and $\mathbf{U} = (\mathbf{U}_1 : \mathbf{U}_2 : ... : \mathbf{U}_q)$, where $\mathbf{U}_i$ is of order $N \times c_i$, with $\sum_{i=1}^{q} c_i = c$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ of order $p \times 1$, where $p = p_1 + p_2$. The observation vector $\boldsymbol{Y}$ is of the order $N \times 1$ and the unobservable random vectors $\boldsymbol{\xi} = (\boldsymbol{\xi}_1' : \boldsymbol{\xi}_2' : ... : \boldsymbol{\xi}_q')'$ and $\boldsymbol{\varepsilon}$ are independently and normally distributed

with mean $\boldsymbol{0}$ and covariance matrices $\mathbf{T}$ and $\boldsymbol{\Sigma}$, respectively, with $\mathbf{T} = Cov\left(\boldsymbol{\xi}_1' : \boldsymbol{\xi}_2' : \ldots : \boldsymbol{\xi}_q'\right) = diag\left(\tau_1^2\mathbf{I}_{c_1} : \ldots : \tau_q^2\mathbf{I}_{c_q}\right)$. Each $\boldsymbol{\xi}_i$, $i = 1,2,...,c_q$, is a random vector of order $c_i \times 1$. Motivated by applications, two different structures of $\boldsymbol{\Sigma}$ are considered in Chapter 3, namely, homoscedastic error structure, where $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_N$, and heteroscedastic error structure, where $\boldsymbol{\Sigma} = diag\left[\sigma_1^2\mathbf{I}_{n_1} : \sigma_2^2\mathbf{I}_{n_2} : \ldots : \sigma_k^2\mathbf{I}_{n_k}\right]$ and $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$ are unknown variances with $\sum_{i=1}^{k} n_i = N$.

Let $\mathbf{A}$ denote a $r \times p$ matrix of known constants, such that $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\theta}$ is an $r \times 1$ estimable linear function (i.e. $C(\mathbf{A}) \subseteq C(\mathbf{X})$, where C denotes the column space of a matrix). The problem of the interest is to test hypotheses of the form:

$$H_0 : \boldsymbol{\eta} = 0 \text{ versus } H_A : \boldsymbol{\eta} \geq 0, \tag{8}$$

where the inequalities are component-wise, with at least one strict inequality. For example, if one is interested in testing a simple order among the components of $\boldsymbol{\theta}_1$ then $\mathbf{A} = \left[\mathbf{A}_1 : \mathbf{0}\right]$ where

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix}$$

and $\mathbf{0}$ is the null matrix of suitable order.

3.2. The likelihood ratio test under homoscedastic errors

We begin this section by deriving the RMLE of $\boldsymbol{\theta}$, $\mathbf{T}$ and $\boldsymbol{\Sigma}$ under the constraint $\boldsymbol{\eta} \geq 0$. Let $L(\boldsymbol{\theta}, \mathbf{T}, \boldsymbol{\Sigma})$ denote the log-likelihood, then derivation of RMLE of $\boldsymbol{\theta}$, $\mathbf{T}$ and $\boldsymbol{\Sigma}$ entails the following maximization problem:

$$\max_{\boldsymbol{\eta} \geq 0} L(\boldsymbol{\theta}, \mathbf{T}, \boldsymbol{\Sigma}). \tag{9}$$

Davidov and Rosen (2011) addressed this problem by providing three asymptotically equivalent algorithms. Among the three algorithms, their Algorithm 3.3 is an E-M type algorithm. This algorithm is similar to a previously published algorithm of Hoferkamp and Peddada (2002), who discussed the problem of estimating $\boldsymbol{\theta}$, $\mathbf{T}$ and $\boldsymbol{\Sigma}$ under the heteroscedastic variance structure with variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ subject to inequality constraints. The methodology in Hoferkamp and Peddada (2002) was motivated by situations, where same experiments are repeated in multiple labs with unequal precision in observed data. Hoferkamp and Peddada (2002) did not impose constraints on the regression vector $\boldsymbol{\theta}$.

In the optimization problem (9), the basic idea underlying the E-M type algorithm is to perform the following constrained optimization

$$\min_{\mathbf{A}\boldsymbol{\theta} \geq 0} \left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\theta}\right)' \hat{\mathbf{V}}_i^{-1} \left(\hat{\boldsymbol{\theta}}^i - \boldsymbol{\theta}\right), \tag{10}$$

where $\hat{\boldsymbol{\theta}}^i$ is the estimate of $\boldsymbol{\theta}$ at the ith iterate of the algorithm and $\hat{\mathbf{V}}_i$ is the corresponding covariance matrix of $\hat{\boldsymbol{\theta}}$, at the end of the M-step during the ith iterate of the algorithm. Observe that, asymptotically, the constrained optimization problem (9) is equivalent to the constrained optimization problem in Algorithm 3.3 in Davidov and

Rosen (2011). The main difference between the two is in the objective function (10). Davidov and Rosen (2011) minimize

$$\min_{\mathbf{A}\boldsymbol{\theta}\geq 0}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i)'\hat{\boldsymbol{\Psi}}^{-1(i)}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i), \tag{11}$$

where $\hat{\boldsymbol{\Psi}}^{(i)}$ is the estimated covariance matrix of $\boldsymbol{Y}$ at the *i*th iterate of the algorithm. We note a minor typographical error in Algorithm 3.3, either the quadratic form should have a negative sign in front of the summation or the authors should state it as a minimization problem rather than maximization. As proved in Davidov and Rosen (2011), the constrained estimators derived from this algorithm are consistent.

Below we describe the algorithm for solving problem (9).

Algorithm A:

Let $\hat{\boldsymbol{\theta}}_1^{(m)}$, $\hat{\boldsymbol{\theta}}_2^{(m)}$, $\hat{\boldsymbol{\tau}}^{2(m)}$, $\hat{\sigma}^{2(m)}$ denote the *r*th iterate estimates of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\tau}^2$, and σ^2 respectively.

Step 0. Let $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. Compute $\boldsymbol{\theta}^{(0)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y}$, the ordinary least squares estimator for $\boldsymbol{\theta}$. Compute $\sigma^{2(0)} = \frac{1}{n}\left\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\theta}^{(0)}\right\|^2$. For $\hat{\boldsymbol{\tau}}^{2(0)}$ we use MINQUE (Rao, 1972).

Step 1. Set $m = m + 1$. Fix $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\tau}^2$ at $\hat{\boldsymbol{\theta}}_1^{(m-1)}$, $\hat{\boldsymbol{\theta}}_2^{(m-1)}$, $\hat{\boldsymbol{\tau}}^{2(m-1)}$ respectively, and iteratively estimate σ^2:

$$\begin{aligned}\hat{\sigma}^{2(r)} &= \sigma^{2(r-1)} \\ &+ \frac{1}{n}\hat{\sigma}^{4(r-1)} tr\left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right. \\ &\left. - \left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right],\end{aligned}$$

where $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \hat{\boldsymbol{\Sigma}}$, $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2}\mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2}\mathbf{I}_{c_q}\right)$, $\hat{\sigma}^{2(0)} = \sigma^{2(m-1)}$, and $\mathbf{A}_{ii}$ indicates the (i, i)th block of $\mathbf{A}$.

Step 2. Fix σ^2 at $\sigma^{2(m)}$ and iteratively estimate $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\tau}^2$ using the following estimation equations:

$$\hat{\boldsymbol{\theta}}_1^{(r)} = \hat{\boldsymbol{\theta}}_1^{(r-1)} + \left(\mathbf{X}_1'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right),$$

$$\hat{\boldsymbol{\theta}}_2^{(r)} = \hat{\boldsymbol{\theta}}_2^{(r-1)} + \left(\mathbf{X}_2'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right),$$

$$\hat{\tau}_i^{2(r)} = \hat{\tau}_i^2 + \frac{\hat{\tau}_i^4}{c_i}tr\left\{\mathbf{U}'_i\left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)\times\right.\right.$$
$$\left.\left.\times\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1} - \left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right]\mathbf{U}_i\right\}, \quad i = 1,\ldots,q,$$

where $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}^{(m-1)}$, $\hat{\boldsymbol{\tau}}^{2(0)} = \boldsymbol{\tau}^{2(m-1)}$, $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \hat{\boldsymbol{\Sigma}}$, and $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2}\mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2}\mathbf{I}_{c_q}\right)$. Then $\min_{\mathbf{A}\boldsymbol{\theta}\geq 0}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i)'\hat{\boldsymbol{\Psi}}^{-1(i)}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i)$ is used to obtain $\hat{\boldsymbol{\theta}}_1^{(r)}$.

Steps 1 and 2 are iterated until convergence. Note that Algorithm A is equivalent to Algorithm B on page 39 with $k = 1$.

Let $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2'\right)'$, $\hat{\mathbf{T}}$ and $\hat{\boldsymbol{\Sigma}}$ denote the constrained estimators of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\mathbf{T}$ and $\boldsymbol{\Sigma}$ under the alternative hypothesis, and let $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)'$, $\tilde{\mathbf{T}}^0$, and $\tilde{\boldsymbol{\Sigma}}^0$ denote the corresponding estimators under the null hypothesis (Searle, Casella and McCulloch, 1992). Then for an estimable linear function $\mathbf{A}\boldsymbol{\theta}$ the likelihood ratio test for testing $H_0 : \mathbf{A}\boldsymbol{\theta} = 0$ versus $H_A : \mathbf{A}\boldsymbol{\theta} \geq 0$ is $S_{lrt} = 2\left(L\left(\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}, \hat{\boldsymbol{\Sigma}}\right) - L\left(\tilde{\boldsymbol{\theta}}^0, \tilde{\mathbf{T}}^0, \tilde{\boldsymbol{\Sigma}}^0\right)\right)$ (Davidov and

Rosen, 2011). Davidov and Rosen (2011) deduced that asymptotically under the null hypothesis,

$$\lim_{n\to\infty} P\left(S_{lrt} > c\right) = \sum_{i=0}^{r} w_i \chi_i^2 .$$

Unfortunately, in the above expression, the weights w_i depend upon the unknown variance components $\mathbf{T}$ and σ^2. Furthermore, even if the variance components are assumed to be known, the weights can be computed exactly only for the case $r \le 3$, otherwise they are computed approximately. Lastly, the simulation studies conducted in Davidov and Rosen (2011) suggest that unless the sample sizes are extremely large, the above likelihood ratio test can potentially be liberal, that is the true Type I error rates exceed the nominal levels even when a conservative upper bound for $\sum_{i=0}^{r} w_i \chi_i^2$ was used when rejecting the null hypothesis.

3.3. The likelihood ratio test under heteroscedastic errors

To handle heteroscedasticity, the following algorithm is derived along the lines of Hoferkamp and Peddada (2002).

Algorithm B: Let $\hat{\boldsymbol{\theta}}_1^{(m)}$, $\hat{\boldsymbol{\theta}}_2^{(m)}$, $\hat{\boldsymbol{\tau}}^{2(m)}$, $\hat{\boldsymbol{\sigma}}^{2(m)}$ denote the *r*th iterate estimates of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\tau}^2$, and $\boldsymbol{\sigma}^2$ respectively.

Step 0. Let $\mathbf{X} = \left[\mathbf{X}_1 : \mathbf{X}_2\right]$, $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$. Compute $\boldsymbol{\theta}^{(0)} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{Y}$, the ordinary least squares estimator for $\boldsymbol{\theta}$. Compute $\sigma_i^{2(0)} = \frac{1}{n_i}\left\|\boldsymbol{Y}_i - \mathbf{X}_i\boldsymbol{\theta}^{(0)}\right\|^2$ for $i = 1, \ldots, k$.

For $\hat{\tau}_i^{2(0)}$ we use MINQUE (Rao, 1972).

Step 1. Set $m = m + 1$. Fix $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\tau}^2$ at $\hat{\boldsymbol{\theta}}_1^{(m-1)}$, $\hat{\boldsymbol{\theta}}_2^{(m-1)}$, $\hat{\boldsymbol{\tau}}^{2(m-1)}$ respectively, and iteratively estimate $\boldsymbol{\sigma}^2$:

$$\begin{aligned}\hat{\sigma}_i^{2(r)} &= \sigma_i^{2(r-1)} \\ &+\frac{1}{n_i}\hat{\sigma}_i^{4(r-1)} tr\left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y}-\mathbf{X}_1\boldsymbol{\theta}_1-\mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y}-\mathbf{X}_1\boldsymbol{\theta}_1-\mathbf{X}_2\boldsymbol{\theta}_2\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right. \\ &\left.-\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right]_{ii},\ i=1,\ldots,k,\end{aligned} \tag{12}$$

where $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \hat{\boldsymbol{\Sigma}}$, $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2}\mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2}\mathbf{I}_{c_q}\right)$, $\hat{\boldsymbol{\sigma}}^{2(0)} = \boldsymbol{\sigma}^{2(m-1)}$, and $\mathbf{A}_{ii}$ indicates the (i, i)th block of $\mathbf{A}$.

Step 2. Fix $\boldsymbol{\sigma}^2$ at $\boldsymbol{\sigma}^{2(m)}$ and iteratively estimate $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\tau}^2$ using the following estimation equations:

$$\hat{\boldsymbol{\theta}}_1^{(r)} = \hat{\boldsymbol{\theta}}_1^{(r-1)} + \left(\mathbf{X}_1'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y}-\mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)}-\mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right), \tag{13}$$

$$\hat{\boldsymbol{\theta}}_2^{(r)} = \hat{\boldsymbol{\theta}}_2^{(r-1)} + \left(\mathbf{X}_2'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y}-\mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)}-\mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right), \tag{14}$$

$$\begin{aligned}\hat{\tau}_i^{2(r)} &= \hat{\tau}_i^2 + \frac{\hat{\tau}_i^4}{c_i} tr\left\{\mathbf{U}'_i\left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y}-\mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)}-\mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)\times\right.\right. \\ &\left.\left.\times\left(\boldsymbol{Y}-\mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)}-\mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}-\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right]\mathbf{U}_i\right\},\quad i=1,\ldots,q,\end{aligned} \tag{15}$$

where $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}^{(m-1)}$, $\hat{\boldsymbol{\tau}}^{2(0)} = \boldsymbol{\tau}^{2(m-1)}$, $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \hat{\boldsymbol{\Sigma}}$, and $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2}\mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2}\mathbf{I}_{c_q}\right)$. Then $\min_{\mathbf{A}\boldsymbol{\theta}\geq 0}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i)'\hat{\boldsymbol{\Psi}}^{-1(i)}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1^i - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^i)$ is used to obtain $\hat{\boldsymbol{\theta}}_1^{(r)}$.

Steps 1 and 2 are iterated until convergence.

The following theorems derive the estimation equations for the EM algorithm for the heteroscedastic case.

Theorem 3.1. The EM estimates, at the r*th* iteration, for $\boldsymbol{\sigma}^2$, when $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and $\boldsymbol{\tau}^2$ are known, are given by (12).

Theorem 3.2. The EM estimates, at the r*th* iteration, for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, and $\boldsymbol{\tau}^2$, when $\boldsymbol{\sigma}^2$ is known, are given by (13), (14), and (15) respectively.

Using the general theory established in Nettleton (1999), we note that the above constrained estimator $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2'\right)'$ is consistent. As before, let the estimators under the null hypothesis be denoted by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)'$, $\tilde{\mathbf{T}}^0$, and $\tilde{\boldsymbol{\Sigma}}^0$, then the likelihood ratio test for testing the hypothesis (8) is given by $S_{lrt} = 2\left(L\left(\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}, \hat{\boldsymbol{\Sigma}}\right) - L\left(\tilde{\boldsymbol{\theta}}^0, \tilde{\mathbf{T}}^0, \tilde{\boldsymbol{\Sigma}}^0\right)\right)$. Following the arguments in Davidov and Rosen (2011), asymptotically under the null hypothesis,

$$\lim_{n\to\infty} P\left(S_{lrt} > c\right) = \sum_{i=0}^{r} w_i \chi_i^2$$

As in the case of homoscedastic errors, the weights w_i in the above limiting distribution involve unknown variance components $\tau_1^2, \tau_2^2, ..., \tau_q^2$ and $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$. Hence, as stated in the previous section, the likelihood ratio test cannot be used in practice. This motivates us to develop a bootstrap based methodology described in sections 3.4 and 3.5.

3.4. Parametric EBLUP bootstrap in linear mixed models under inequality constraints

To create residuals that honor the data structure, we will use the parametric EBLUP (empirical best linear unbiased prediction) bootstrap, described below.

3.4.1. Homoscedastic errors

We begin this section with homoscedastic data, where $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$.

Step Hom1: Obtain the point estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$ under the null hypothesis. Denote it by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)$.

Step Hom2: Obtain the point estimators of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\mathbf{T}$, and $\boldsymbol{\Sigma}$ under no constraints on the parameters. Denote them by $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2'\right)$, $\tilde{\mathbf{T}}$, and $\tilde{\boldsymbol{\Sigma}}$.

Step Hom3: Generate a random vector $\boldsymbol{\eta}_i^*$ of size $c_i \times 1$, $i = 1, 2, \ldots, q$ from a multivariate normal distribution with mean 0 and covariance matrix I. Similarly, generate an independent random vector $\boldsymbol{\upsilon}^*$ of size $N \times 1$ from a multivariate normal distribution with mean 0 and covariance matrix I. Finally, let $\tilde{\boldsymbol{\xi}}_i^* = \tilde{\boldsymbol{\tau}}_i \boldsymbol{\eta}_i^*$ and $\tilde{\boldsymbol{\varepsilon}}^* = \tilde{\boldsymbol{\sigma}} \boldsymbol{\upsilon}^*$, then the EBLUP bootstrap sample is given by

$$\boldsymbol{Y}^* = \mathbf{X}\tilde{\boldsymbol{\theta}}^0 + \mathbf{U}\tilde{\boldsymbol{\xi}}^* + \tilde{\boldsymbol{\varepsilon}}^*. \tag{16}$$

Recently, the asymptotic properties of parametric EBLUP bootstrap, when the random errors are normally distributed, have been discussed in Chatterjee et al. (2008).

Thus, the above model honors the null hypothesis regarding the parameter $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, as well as the underlying variance components structure. Thus, one may derive the bootstrap null distribution of any test statistic, including the likelihood ratio test, by repeatedly generating the null data in (16) (say 1000 times) and computing the desired test statistic for each null data.

3.4.2. Heteroscedastic errors

In the case of heteroscedastic errors, the construction of bootstrap sample $\boldsymbol{Y}^*$ requires a minor modification from the homoscedastic case as follows.

Step Het1: Obtain the point estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$ under the null hypothesis. Denote it by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)$.

Step Het2: Obtain the point estimators of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\mathbf{T}$, and $\boldsymbol{\Sigma}$ under no constraints on the parameters. Denote them by $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2'\right)$, $\tilde{\mathbf{T}}$, and $\tilde{\boldsymbol{\Sigma}}$.

Step Het3: Generate a random vector $\boldsymbol{\eta}_i^*$ of size $c_i \times 1$, $i = 1, 2, \ldots, q$ from a multivariate normal distribution with mean 0 and covariance matrix I. Similarly, generate an independent random vector $\boldsymbol{\upsilon}_i^*$, $i = 1, 2, \ldots, k$ of size $n_i \times 1$ from a multivariate normal distribution with mean 0 and covariance matrix $\mathbf{I}$. Finally, let $\tilde{\boldsymbol{\xi}}_i^* = \tilde{\tau}_i \boldsymbol{\eta}_i^*$, $i = 1, 2, \ldots, q$ and $\tilde{\boldsymbol{\varepsilon}}_i^* = \tilde{\sigma}_i \boldsymbol{\upsilon}_i^*$, $i = 1, 2, \ldots, k$, then the EBLUP bootstrap sample is given by

$$\boldsymbol{Y}^* = \mathbf{X}\tilde{\boldsymbol{\theta}}^0 + \mathbf{U}\tilde{\boldsymbol{\xi}}^* + \tilde{\boldsymbol{\varepsilon}}^*. \tag{17}$$

3.5. MINQUE-Williams based methodology

Methodology described in previous sections assumed that $\boldsymbol{Y}$ is multivariate normally distributed. In many applications, this may not necessarily be true. Therefore, in this section, we develop a distribution free methodology for performing constrained inference in linear mixed effects models. For the rest of this dissertation we limit to $\mathbf{A}$ such that $\mathbf{A}\boldsymbol{\theta}$ is estimable and $\mathbf{A}\boldsymbol{\theta} \geq \boldsymbol{0}$ is a simple order cone.

We begin with the estimation of variance components in the linear model (1). There exists considerable literature on the estimation of variance components $\tau_1^2, \tau_2^2, ..., \tau_q^2$ and $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$. As an alternative to MLE, which was developed for normally distributed data, several distribution free methods have been proposed in the literature, such as the ANOVA based methods of Henderson (1953), the MINQUE (minimum norm quadratic unbiased estimation) theory of Rao (1970, 1971, 1972). Henderson's ANOVA based estimators are essentially method of moments type estimators, whereas Rao's MINQUE theory is based on some basic principles an estimator of variance components should possess, namely: (i) quadratic form of the data, since the parameter is quadratic, (ii) translation invariant, hence does not depend upon the location (or regression parameter), (iii) unbiasedness, and (iv) minimum norm. Several variations to MINQUE have been proposed in the literature, such as, I-MINQUE, MINQE, CMINQUE, MINQUE (SD), etc. A well established theory for MINQUE and related methods has been developed. For a comprehensive account on this subject, one may refer to Rao and Kleffe (1988).

We now describe MINQUE for estimating the variance components under heteroscedasticity case since the homoscedasticity is a special case and can be easily deduced. Thus in this section we describe the MINQUE methodology for estimating $\phi = \left(\tau_1^2, \tau_2^2, ..., \tau_q^2, \sigma_1^2, \sigma_2^2, ..., \sigma_k^2\right)'$. We rewrite the linear model (1) as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\zeta},$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ with $E(\boldsymbol{\zeta}) = 0$, $Var(\boldsymbol{\zeta}) = \mathbf{G}_\phi = \sum_{i=1}^{q+k} \phi_i \mathbf{F}_i$,

$\mathbf{F}_i = \mathbf{U}_i \mathbf{U}_i'$, $i = 1, 2, ..., q$, and $\mathbf{F}_i = Diag[0 : 0 : ... : \mathbf{I} : 0...0]$, $i = q+1, 2, ..., q+k$ with the identity matrix of order $n_{i-q} \times n_{i-q}$ located at the *i*th location. Each $\mathbf{F}_i$ is $N \times N$. Let

$$\mathbf{F} = \left[\mathbf{F}_1 : \mathbf{F}_2 : ... : \mathbf{F}_{q+k} \right],$$

$$\mathbf{W}_\phi = \mathbf{G}_\phi + \mathbf{X}\mathbf{X}',$$

$$\mathbf{R}_\phi = \mathbf{W}_\phi^- - \mathbf{W}_\phi^- \mathbf{X}(\mathbf{X}'\mathbf{W}_\phi^- \mathbf{X})^- \mathbf{X}'\mathbf{W}_\phi^-,$$

$z_\phi = (\boldsymbol{Y}'\mathbf{R}_\phi \mathbf{F}_1 \mathbf{R}_\phi' \boldsymbol{Y}, \boldsymbol{Y}'\mathbf{R}_\phi \mathbf{F}_2 \mathbf{R}_\phi' \boldsymbol{Y}, ..., \boldsymbol{Y}'\mathbf{R}_\phi \mathbf{F}_{q+k} \mathbf{R}_\phi' \boldsymbol{Y})'$, and

$$\mathbf{S}_\phi = \left[Tr(\mathbf{R}_\phi \mathbf{F}_i \mathbf{R}_\phi' \mathbf{F}_j),\ i, j = 1, 2, ..., q+k \right].$$

In the above expression, $\mathbf{A}^-$ denotes a *generalized inverse* (or *g-inverse*) of $\mathbf{A}$. The above expressions are invariant to the choice of g-inverse. Hence without loss of generality one may use the Moore-Penrose inverse $\mathbf{A}^+$. The MINQUE of $\boldsymbol{\phi}$ is then obtained by solving the following system of linear equations (Rao, 1972):

$$\mathbf{S}_{\boldsymbol{\phi}^{(0)}} \boldsymbol{\phi} = z_{\boldsymbol{\phi}^{(0)}}, \tag{18}$$

where $\boldsymbol{\phi}^{(0)}$ denotes an initial estimate of $\boldsymbol{\phi}$. Since the MINQUE depends upon the initial estimate, Rao and Kleffe (1988) recommend iterating (18) until convergence. The resulting estimator is known as the iterated MINQUE (or I-MINQUE). Denote the I-MINQUE of $\boldsymbol{\phi}$ by $\hat{\boldsymbol{\phi}}$. As discussed in Rao and Subrahmaniam (1971), estimated parameters can be negative. As is commonly done, in such cases we replace them with 0.01.

Let $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\theta}$ be an estimable linear function of $\boldsymbol{\theta}$, then its weighted least squares estimator is given by

$$\hat{\boldsymbol{\eta}} = \mathbf{A}(\mathbf{X}'\mathbf{W}_{\hat{\phi}}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{W}_{\hat{\phi}}^{+}\mathbf{Y}\,. \tag{19}$$

Under mild regularity conditions stated below, in Theorem 3.3 we note that the weighted least squares estimator is asymptotically normally distributed.

R1: $0 < \min\left(\phi_1, \phi_2, \ldots, \phi_{q+k}\right) \le \min\left(\phi_1, \phi_2, \ldots, \phi_{q+k}\right) < \infty$.

R2: $E\left\|\boldsymbol{\zeta}\right\|^4 < \infty$.

R3: $\dfrac{Tr\left(\mathbf{W}_{\phi}^{+}\mathbf{F}_i\mathbf{W}_{\phi}^{+}\mathbf{F}_j\right)}{Tr\left(\mathbf{W}_{\phi}^{+}\mathbf{F}_i\mathbf{W}_{\phi}^{+}\mathbf{F}_i\right)} \to d_{ij}$, where the matrix $\mathbf{D} = \left(d_{ij},\ i, j = 1, 2, \ldots, q+k\right)$ is non-singular.

R4: $\dfrac{\lambda_{\max}\left(\mathbf{W}_{\phi}^{+}\mathbf{F}_i\right)\lambda_{\max}\left(\mathbf{W}_{\phi}^{+}\mathbf{F}_j\right)}{Tr\left(\mathbf{W}_{\phi}^{+}\mathbf{F}_i\mathbf{W}_{\phi}^{+}\mathbf{F}_i\right)} \to 0$, where $\lambda_{\max}\left(\mathbf{H}\right)$ denotes the largest eigenvalue of a matrix $\mathbf{H}$.

R5: $\mathbf{X}\left(\mathbf{X}'\mathbf{W}_{\phi}^{+}\mathbf{X}\right)^{+}\mathbf{X}' \to \mathbf{M}(\boldsymbol{\phi})$, where $M(\boldsymbol{\phi})$ is a positive definite matrix.

R6: $\lambda_{\max}(\mathbf{X}\mathbf{X}'\mathbf{W}_{\phi}^{+}\mathbf{G}_{\phi}\mathbf{W}_{\phi}^{+}\mathbf{X}\mathbf{X}') \to 0$.

Theorem 3.3. For any estimable linear function $\mathbf{A}\boldsymbol{\theta}$ in a linear mixed model (1) satisfying the regularity conditions R1 to R6,

$$\mathbf{A}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \overset{asymptotically}{\sim} N\left(0,\ \mathbf{A}\mathbf{M}(\boldsymbol{\phi})^{-1}\mathbf{A}'\right).$$

As a consequence of asymptotic normality of $\mathbf{A}\hat{\boldsymbol{\theta}}$, one may appeal to Davidov and Rosen (2011) and derive a constrained estimator for $\mathbf{A}\boldsymbol{\theta}$ under the constraint $\mathbf{A}\boldsymbol{\theta} \ge \boldsymbol{\theta}$ and construct the likelihood ratio test. However, as noted in Section 3.3, the

asymptotic null distribution of the likelihood ratio test involves the unknown variance components. Furthermore, it is not easy to compute the null distribution when the dimension is large.

To keep the methodology distribution free and computationally simple we estimate $\boldsymbol{\theta}_1$ under the simple order constraint by applying PAVA on $\hat{\boldsymbol{\theta}}_1$. We denote the resulting constrained estimator by $\hat{\boldsymbol{\theta}}_1^{PAVA}$.

Owing to the simplicity of expression and superior power, the Williams' test (Williams, 1972) is used widely in applications. For example, the National Toxicology Program uses the Williams' test to test for trends in its pre-chronic 90-day rodent bioassay. Motivated by its popular use, in this thesis, as an alternative to the likelihood ratio test, we propose the following test statistic.

$$W = \frac{\hat{\theta}_{1p}^{PAVA} - \hat{\theta}_{11}^{PAVA}}{\sqrt{Var\left(\hat{\theta}_{1p} - \hat{\theta}_{11}\right)}}, \tag{20}$$

where $Var\left(\hat{\theta}_{1p} - \hat{\theta}_{11}\right)$ is the estimated variance of the contrast $\left(\hat{\theta}_{1p} - \hat{\theta}_{11}\right)$. We now describe the nonparametric EBLUP bootstrap methodology for deriving the null distribution of the test statistic (20).

3.5.1. Homoscedastic errors

As in section 4.3, we begin this section with homoscedastic data, where $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$. Let $\mathbf{C} = Cov\left(\boldsymbol{Y}, \boldsymbol{\xi}\right) = \mathbf{UT}$ and let $\boldsymbol{\Psi} = Cov(\boldsymbol{Y}) = \mathbf{UTU}' + \boldsymbol{\Sigma}$, then the best linear unbiased predictor (BLUP) of $\boldsymbol{\xi}$ is given by $\mathbf{C}'\boldsymbol{\Psi}^{-1}\left(\mathbf{I} - \mathbf{P}\right)\boldsymbol{Y}$, where $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Psi}^{-1}$. However, since $\mathbf{C}$ and $\boldsymbol{\Psi}$ are unknown, the empirical version of $\mathbf{C}'\boldsymbol{\Psi}^{-1}\left(\mathbf{I} - \mathbf{P}\right)\boldsymbol{Y}$, known

as the empirical best linear unbiased predictor (EBLUP) of $\boldsymbol{\xi}$, when no constraints are imposed on the parameters, is given by $\tilde{\boldsymbol{\xi}} = \left(\tilde{\boldsymbol{\xi}}_1' : \tilde{\boldsymbol{\xi}}_2' : \ldots : \tilde{\boldsymbol{\xi}}_q'\right)' = \tilde{\mathbf{C}}'\tilde{\boldsymbol{\Psi}}^{-1}\left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$ and the estimated residuals are given by $\tilde{\boldsymbol{\varepsilon}} = \left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$.

Step Hom1: Obtain the point estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$ under the null hypothesis. Denote it by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)$.

Step Hom2: Obtain the point estimators of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\mathbf{T}$, and $\boldsymbol{\Sigma}$ under no constraints on the parameters. Denote them by $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2'\right)$, $\tilde{\mathbf{T}}$, and $\tilde{\boldsymbol{\Sigma}}$ then compute $\tilde{\boldsymbol{\xi}} = \left(\tilde{\boldsymbol{\xi}}_1' : \tilde{\boldsymbol{\xi}}_2' : \ldots : \tilde{\boldsymbol{\xi}}_q'\right)' = \tilde{\mathbf{C}}'\tilde{\boldsymbol{\Psi}}^{-1}\left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$ and $\tilde{\boldsymbol{\varepsilon}} = \left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$.

Step Hom3: Let $\boldsymbol{\eta}_i = \dfrac{\tilde{\boldsymbol{\xi}}_i}{sd\left(\tilde{\boldsymbol{\xi}}_i\right)}$, $i = 1, 2, \ldots, q$ and let $\boldsymbol{\upsilon} = \dfrac{\tilde{\boldsymbol{\varepsilon}}}{sd(\tilde{\boldsymbol{\varepsilon}})}$, where $sd(\boldsymbol{\zeta})$ represents the usual sample standard deviation of the elements in the vector $\boldsymbol{\zeta}$.

Step Hom4: Let $\boldsymbol{\eta}_i^*$, $i = 1, 2, \ldots, q$ denote a random vector obtained by taking simple random sample (with replacement) of size $c_i \times 1$, $i = 1, 2, \ldots, q$ from the elements of $\boldsymbol{\eta}_i$. Similarly, let $\boldsymbol{\upsilon}^*$ denote a random vector obtained by taking a simple random sample (with replacement) of size $N \times 1$ from the elements of $\boldsymbol{\upsilon}$. Finally, let $\tilde{\boldsymbol{\xi}}_i^* = \tilde{\tau}_i \boldsymbol{\eta}_i^*$, $i = 1, 2, \ldots, q$ and $\tilde{\boldsymbol{\varepsilon}}^* = \tilde{\sigma}\boldsymbol{\upsilon}^*$, then the EBLUP bootstrap sample is given by

$$\boldsymbol{Y}^* = \mathbf{X}\tilde{\boldsymbol{\theta}}^0 + \mathbf{U}\tilde{\boldsymbol{\xi}}^* + \tilde{\boldsymbol{\varepsilon}}^*. \tag{21}$$

Thus, the above model honors the null hypothesis regarding the parameter $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, as well as the underlying variance components structure. Repeatedly generating the null data in (21) and computing (20) for each null data yields the null distribution of (20).

While preparing this dissertation, in a personal communication with Dr. Peddada, Dr. Chatterjee informed that he is currently investigating the asymptotic properties of the nonparametric EBLUP, and the manuscript is being completed.

3.5.2. Heteroscedastic errors

As in the case of parametric EBLUP bootstrap, the construction of bootstrap sample $\boldsymbol{Y}^*$ requires a minor modification from the homoscedastic case as follows.

Step Het1: Obtain the point estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$ under the null hypothesis. Denote it by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)$.

Step Het2: Obtain the point estimators of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\mathbf{T}$, and $\boldsymbol{\Sigma}$ under no constraints on the parameters. Denote them by $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2'\right)$, $\tilde{\mathbf{T}}$, and $\tilde{\boldsymbol{\Sigma}}$, then compute $\tilde{\boldsymbol{\xi}} = \left(\tilde{\boldsymbol{\xi}}_1' : \tilde{\boldsymbol{\xi}}_2' : \ldots : \tilde{\boldsymbol{\xi}}_q'\right)' = \tilde{\mathbf{C}}'\tilde{\boldsymbol{\Psi}}^{-1}(\mathbf{I} - \tilde{\mathbf{P}})\boldsymbol{Y}$ and $\tilde{\boldsymbol{\varepsilon}} = (\mathbf{I} - \tilde{\mathbf{P}})\boldsymbol{Y}$.

Step Het3: Let $\boldsymbol{\eta}_i = \dfrac{\tilde{\boldsymbol{\xi}}_i}{sd\left(\tilde{\boldsymbol{\xi}}_i\right)}$, $i = 1, 2, \ldots, q$ and let $\boldsymbol{\upsilon}_i = \dfrac{\tilde{\boldsymbol{\varepsilon}}_i}{sd(\tilde{\boldsymbol{\varepsilon}}_i)}$, $i = 1, 2, \ldots, k$ where $sd(\boldsymbol{\zeta})$ represents the usual sample standard deviation of the elements in the vector $\boldsymbol{\zeta}$. Note that unlike in the homoscedastic case, here $Var(\boldsymbol{\varepsilon}_i) = \sigma_i^2 \mathbf{I}_{n_i}$, $i = 1, 2, ..., k$.

Step Het4: Let $\boldsymbol{\eta}_i^*$, $i = 1,2,\ldots,q$ denote a random vector obtained by taking simple random sample (with replacement) of size $c_i \times 1$, $i = 1,2,\ldots,q$ from the elements of $\boldsymbol{\eta}_i$. Similarly, let $\boldsymbol{\upsilon}_i^*$, $i = 1,2,\ldots,k$ denote a random vector obtained by taking a simple random sample (with replacement) of size $\mathrm{n_i} \times 1$ from the elements of $\boldsymbol{\upsilon}_i$. Finally, let $\tilde{\boldsymbol{\xi}}_i^* = \tilde{\tau}_i \boldsymbol{\eta}_i^*$, $i = 1,2,\ldots,q$ and $\tilde{\boldsymbol{\varepsilon}}_i^* = \tilde{\sigma}_i \boldsymbol{\upsilon}_i^*$, $i = 1,2,\ldots,k$, then the EBLUP bootstrap sample is given by

$$\boldsymbol{Y}^* = \mathbf{X}\tilde{\boldsymbol{\theta}}^0 + \mathbf{U}\tilde{\boldsymbol{\xi}}^* + \tilde{\boldsymbol{\varepsilon}}^*. \tag{22}$$

The above model honors the null hypothesis regarding the parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ as well as honors the underlying variance components structure. Repeatedly generating the null data in (22) and computing (20) for each null data yields the null distribution of (20).

3.6. Some concluding remarks

In this dissertation we developed a bootstrap based methodology for performing constrained inference in linear mixed effects models with covariance structure of the form $Cov(\boldsymbol{Y}) = \mathbf{UTU}' + \boldsymbol{\Sigma}$, where $\mathbf{T}$ and $\boldsymbol{\Sigma}$ are diagonal matrices. As noted earlier, although this covariance structure is encountered in many applications, in general, however, depending upon the application, the covariance structures may be more complicated or unspecified. In such situations, the bootstrap methodology proposed in Sections 3.4 and 3.5 may not be robust in achieving the desired nominal type I error rate or may also potentially lose power. This is not surprising and is not unique to the present situation. For example, even in the classical Behrens-Fisher problem of comparing

means of two independent univariate normal populations with unequal and unknown variances (Lehmann, 1975), the pooled two-sample t-test could result in an inflated type I error rate. On the other hand, Welch's t-test (Welch, 1938) would lose power to the pooled t-test when the two population variances are equal. More recently, a similar issue was discussed in great detail in Lim, Sen and Peddada (2010, 2011) in the context of non-linear regression models under homoscedastic and heteroscedastic errors. Thus, performance of a method depends highly on the underlying assumptions regarding the model and the covariance structure. In Chapter 4, we evaluate the robustness of the proposed bootstrap based methodology, when the structure of the underlying covariance matrix is either unspecified or it has an auto-correlation structure.

Although the proposed methodology depends upon the underlying covariance structure, it can be easily be adapted to a given covariance structure. We illustrate this by considering two common covariance structures.

Unspecified covariance structure

Recently researchers at NIEHS conducted a large study to understand the factors associated with growth of fibroids (benign smooth muscle tumors of the uterus) in premenopausal women (Peddada et al., 2008). Since the researchers collected data on multiple tumors within each woman, it is reasonable to assume that the growth rates of tumors within the same woman are correlated. However, one cannot be sure about the correlation structure a priori. A question of biological interest is whether the rate of growth of a tumor depends upon the tumor size. Often very large tumors tend to experience necrosis since blood supply to the tumor may be cut off or reduced. Consequently, one may be interested in testing the hypothesis that the rate of growth of a

tumor decreases or stays same with tumor size. One may want to test such hypothesis after adjusting for a variety of covariates such as the location of the tumor (e.g. located in the fundus, corpus or in the lower segment of the uterus), tumor type (submucosal, intramural or subserosal), age of the subject, race of the subject, etc. Often for the convenience of interpretation biologists categorize tumors into p_1 size categories, such as small, medium, large, etc., and are interested in comparing these categories in terms of tumor growth rates. Thus, if $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12}, ..., \theta_{1p_1})'$ is a vector of parameters describing the growth rates of tumors in the p_1 size categories, then one may be interested in testing the following hypothesis:

$$H_0 : \theta_{11} = \theta_{12} = ... = \theta_{1p_1} \text{ versus } H_A : \theta_{11} \geq \theta_{12} \geq ... \geq \theta_{1p_1}. \tag{23}$$

Data and the problems such as the above ones can be described using classical fixed effects linear regression model with unknown and unstructured covariance matrix. More precisely, for the *i*th subject, $i = 1, 2, ..., n,$ let $\boldsymbol{Y}_i$ denote the growth rate of p_1 tumors in a given interval of interest, and let $\mathbf{X}_i$ denote the $p_1 \times p_2$ model matrix of covariates with the corresponding $p_2 \times 1$ unknown regression parameter $\boldsymbol{\theta}_2$. Then the linear model corresponding to the *i*th subject is given by:

$$\boldsymbol{Y}_i = \boldsymbol{\theta}_1 + \mathbf{X}_i \boldsymbol{\theta}_2 + \boldsymbol{\varepsilon}_i. \tag{24}$$

We assume that the random errors $\boldsymbol{\varepsilon}_i \sim N(0, \boldsymbol{\Sigma})$, where the structure of the covariance matrix $\boldsymbol{\Sigma}$ is unknown. The general bootstrap methodology described in the previous sections can be easily modified by re-sampling suitable residuals as follows. Let $\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\Sigma}}\right)$ denote the UMLE of $\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\Sigma}\right)$ under no restrictions on the parameters

and let $\left(\hat{\boldsymbol{\theta}}_1^0, \hat{\boldsymbol{\theta}}_2^0\right)$ denote the maximum likelihood estimator of $\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)$ under the null hypothesis. Let $\hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{Y}_i - \hat{\boldsymbol{\theta}}_1 - \mathbf{X}_i\hat{\boldsymbol{\theta}}_2$ denote the residual vector corresponding to the *i*th subject, $i = 1, 2, ..., n$. Draw a simple random sample (with replacement) of n subjects from the sample of n subjects in the study. Denote the residuals corresponding to these re-sampled subjects by $\left\{\boldsymbol{\eta}_1^*, \boldsymbol{\eta}_2^*, ..., \boldsymbol{\eta}_n^*\right\}$. Then the bootstrap data corresponding to the *i*th subject selected in the bootstrap sample is given by:

$$\boldsymbol{Y}_i^* = \hat{\boldsymbol{\theta}}_1^0 + \mathbf{X}_i\hat{\boldsymbol{\theta}}_2^0 + \boldsymbol{\eta}_i^*. \tag{25}$$

As in Section 3.3, let $\hat{\boldsymbol{\theta}}_1^{PAVA}$ denote the PAVA estimator of $\boldsymbol{\theta}_1$ under the inequality constraints specified by the alternative hypothesis. Then, as before, the hypotheses in (23) can be tested using the following test statistic using the null distribution derived according to the above bootstrap methodology:

$$W = \frac{\hat{\theta}_{11}^{PAVA} - \hat{\theta}_{1p}^{PAVA}}{\sqrt{Var\left(\hat{\theta}_{11}^{PAVA} - \hat{\theta}_{1p}^{PAVA}\right)}}.$$

Random slopes model

In many applications researchers model repeated measurement data using random slopes model as follows. For the *i*th subject, $i = 1, 2, 3, ..., n,$ the model (24) is modified as follows:

$$\boldsymbol{Y}_i = \boldsymbol{\theta}_1 + \mathbf{X}_i\boldsymbol{\theta}_2 + \mathbf{U}_i\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i \tag{26}$$

where $\boldsymbol{Y}_i, \boldsymbol{\theta}_1, \mathbf{X}_i,$ and $\boldsymbol{\theta}_2$ are as defined above. The random vectors $\boldsymbol{\xi}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be independently distributed with unspecified non-diagonal covariance

structures, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$, respectively. Stacking the models for all subjects together, we have the following linear mixed effects model:

$$\boldsymbol{Y} = \boldsymbol{I} \otimes \boldsymbol{\theta}_1 + \mathbf{X}\boldsymbol{\theta}_2 + \mathbf{U}\boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{27}$$

where $\boldsymbol{Y} = \left[\boldsymbol{Y}_1' : \boldsymbol{Y}_2' : \ldots : \boldsymbol{Y}_n'\right]'$, $\mathbf{X} = \left[\mathbf{X}_1' : \mathbf{X}_2' : \ldots : \mathbf{X}_n'\right]'$, $\mathbf{U} = diag\left[\mathbf{U}_1 : \mathbf{U}_2 : \ldots : \mathbf{U}_n\right]$,

$\boldsymbol{\xi} = \left[\boldsymbol{\xi}_1' : \boldsymbol{\xi}_2' : \ldots : \boldsymbol{\xi}_n'\right]'$, $\boldsymbol{\varepsilon} = \left[\boldsymbol{\varepsilon}_1' : \boldsymbol{\varepsilon}_2' : \ldots : \boldsymbol{\varepsilon}_n'\right]'$, $\mathbf{T} = Cov\left(\boldsymbol{\xi}\right) = \mathbf{I} \otimes \boldsymbol{\Lambda}$, and $Cov(\boldsymbol{\varepsilon}) = \mathbf{I} \otimes \boldsymbol{\Sigma}$.

Thus the covariance matrix of $\boldsymbol{Y}$ is given by $\boldsymbol{\Psi} = Cov\left(\boldsymbol{Y}\right) = \mathbf{U}\left(\mathbf{I} \otimes \boldsymbol{\Lambda}\right)\mathbf{U}' + \mathbf{I} \otimes \boldsymbol{\Sigma}$. The bootstrap methodology described in Section 3.5.1 can be suitably modified for this covariance structure as follows. Notations used in Section 3.5.1 are slightly different from those used here, since the vectors are stacked here by subject rather than by time point, as was done in Section 3.5.1. The overall notations are still the same. That is, $\mathbf{C} = Cov\left(\boldsymbol{Y}, \boldsymbol{\xi}\right) = \mathbf{UT}$, $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Psi}^{-1}$.

The bootstrap methodology can be described as follows:

Step 1: Obtain the UMLE of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$ under the null hypothesis. Denote it by $\tilde{\boldsymbol{\theta}}^0 = \left(\tilde{\boldsymbol{\theta}}_1^{0\prime}, \tilde{\boldsymbol{\theta}}_2^{0\prime}\right)$.

Step 2: Obtain the UMLE of $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ under no constraints on the parameters. Denote them by $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2'\right)'$, $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Sigma}}$ respectively. Let $\tilde{\mathbf{C}} = \mathbf{U}\left(\mathbf{I} \otimes \tilde{\boldsymbol{\Lambda}}\right)$, $\tilde{\boldsymbol{\Psi}} = \mathbf{U}\left(\mathbf{I} \otimes \tilde{\boldsymbol{\Lambda}}\right)\mathbf{U}' + \mathbf{I} \otimes \tilde{\boldsymbol{\Sigma}}$ and $\tilde{\mathbf{P}} = \mathbf{X}\left(\mathbf{X}'\tilde{\boldsymbol{\Psi}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\tilde{\boldsymbol{\Psi}}^{-1}$. As in Section 3.5.1, compute the residuals $\tilde{\boldsymbol{\xi}} = \left[\tilde{\boldsymbol{\xi}}_1' : \tilde{\boldsymbol{\xi}}_2' : \ldots : \tilde{\boldsymbol{\xi}}_n'\right]' = \tilde{\mathbf{C}}'\tilde{\boldsymbol{\Psi}}^{-1}\left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$ and $\tilde{\boldsymbol{\varepsilon}} = \left[\tilde{\boldsymbol{\varepsilon}}_1' : \tilde{\boldsymbol{\varepsilon}}_2' : \ldots : \tilde{\boldsymbol{\varepsilon}}_n'\right]' = \left(\mathbf{I} - \tilde{\mathbf{P}}\right)\boldsymbol{Y}$.

Step 3: Unlike in Section 3.4.1, randomly select subjects and use the residuals of the selected subject. Thus, corresponding to the subject *i*, we randomly select (with replacement) the subject i^*, whose residuals are denoted by $\tilde{\boldsymbol{\xi}}_{i^*}$ and $\tilde{\boldsymbol{\varepsilon}}_{i^*}$ respectively. Then the bootstrap sample may be constructed as follows:

$$\boldsymbol{Y}_i^* = \tilde{\boldsymbol{\theta}}_1^0 + \mathbf{X}_i\boldsymbol{\theta}_2 + \mathbf{U}_i\tilde{\boldsymbol{\xi}}_{i^*} + \tilde{\boldsymbol{\varepsilon}}_{i^*},\ i = 1, 2, ..., n.$$

Thus, the above model honors the null hypothesis regarding the parameter $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, as well as the underlying variance components structure.

Once the bootstrap samples are obtained, a test statistic similar to the one described in Section 3.5.1 may be constructed, and its bootstrap null distribution may be derived. Thus, the general framework described in this chapter can be modified for other commonly observed covariance structures.

CHAPTER 4

SIMULATION STUDIES FOR CONSTRAINED TESTING IN LINEAR MIXED EFFECTS MODELS

4.1. Normally distributed data

4.1.1. Study design

Extensive simulations studies were performed to evaluate the performance of various tests in terms of the Type I error and power. The data were simulated using the following model:

$$\boldsymbol{Y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 + \mathbf{U}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \tag{28}$$

where $\boldsymbol{\theta}_1$ denotes the $p\times 1$ vector of treatment effects, $\mathbf{X}_1$ is a $N\times p$ design matrix consisting of 0's and 1's, $\mathbf{X}_2$ is a known $N\times 1$ matrix of covariates, $\boldsymbol{\theta}_2 = 2$ is a corresponding regression parameter, $\mathbf{U}$ is a known matrix of design constants, where $\mathbf{U}$ is of order $N\times c$, $\boldsymbol{\xi}$ is a $c\times 1$ vector of independent subject random effects.

The random vectors $\boldsymbol{\xi}, \boldsymbol{\varepsilon}$ were independently and normally distributed with means 0 and covariance matrices $\mathbf{T}$ and $\boldsymbol{\Sigma}$, where $\mathbf{T} = Cov(\boldsymbol{\xi}) = \tau^2\mathbf{I}_c$. Two different structures of $\boldsymbol{\Sigma}$ were considered, namely, homoscedastic error structure with $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_N$ and heteroscedastic error structure with $\boldsymbol{\Sigma} = diag\left[\sigma_1^2\mathbf{I}_{n_1} : \sigma_2^2\mathbf{I}_{n_2} : \ldots : \sigma_k^2\mathbf{I}_{n_p}\right]$, where

$\sigma_1^2, \sigma_2^2, ..., \sigma_p^2$ are unknown variances with $\sum_{i=1}^{p} n_i = N$. Simulations were performed for p = 3, 5 and 10 treatment groups, c = 10, 30, 50 subjects per treatment and four different patterns of treatment means $\boldsymbol{\theta}_1$:

(1) 0, ... , 0, a,

(2) 0, a, ... , a,

(3) $a + d$, $a + 2d$, ... , $a + (p\text{-}1)d$,

(4) 0, a, ... , a, b.

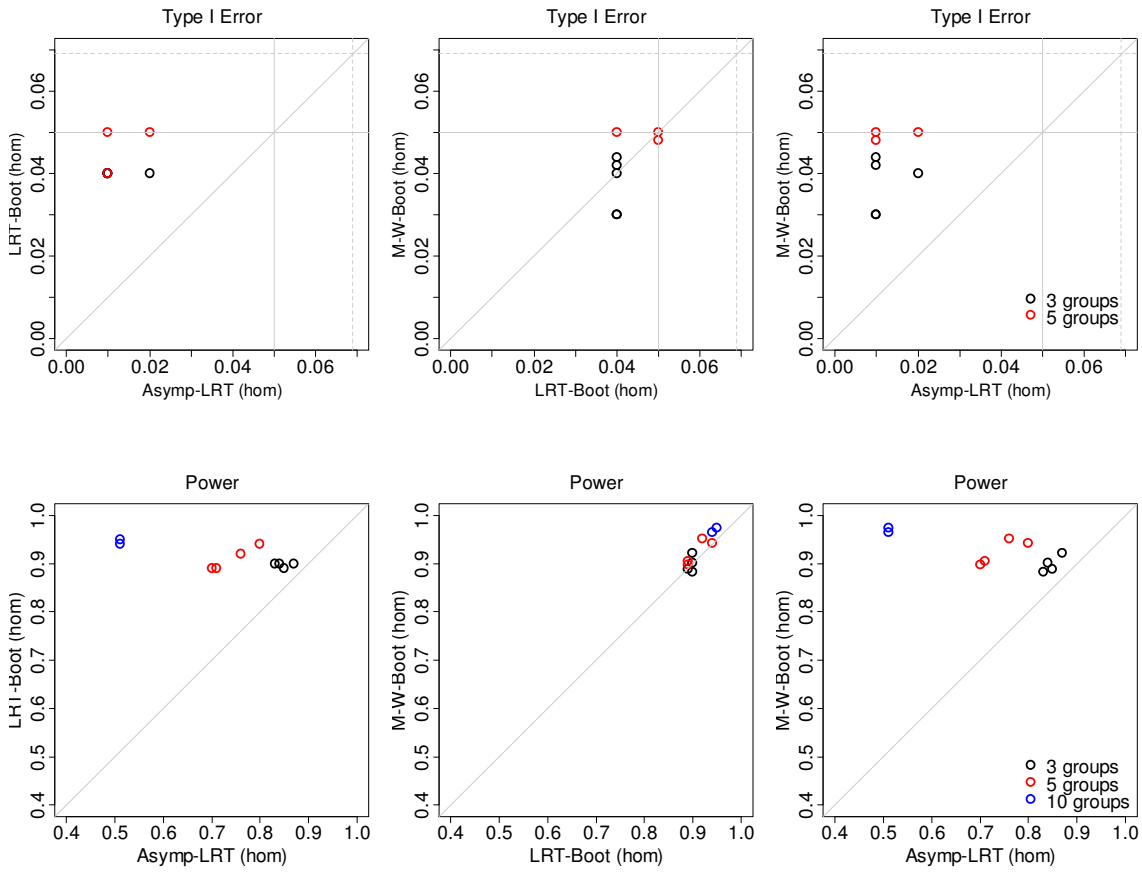
The components of $\boldsymbol{\theta}_1$ were restricted to satisfy a simple order constraint $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$. We compared the type I error and power of two proposed tests (MINQUE-based bootstrap test and likelihood ratio bootstrap test) with other methods, namely the asymptotic likelihood ratio test (Davidov and Rosen, 2011), unconstrained F test, unconstrained F test assuming linear regression on treatment parameters, as well as parametric and non-parametric bootstrap implementations of these tests. Complete simulation results are presented in Appendix C.

Simulation results are based on 500 data simulation runs and 500 bootstrap runs (where applicable). If the true type I error rate is 0.05, the estimated type I error rate is within (0.031, 0.069) 95% of the time. On type I error plots, we show the nominal alpha level of 0.05 as a solid line and the upper bound of the 95% confidence interval, 0.069, as a dashed line. For conciseness, only results for 10 and 50 subjects per treatment are presented.

4.1.2. Results for homoscedastic case

Simulations were performed with $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$ ($\sigma = 1$), $\tau^2 = 0.2,\ 1,\ 2$, and treatment mean patterns (1)-(3). Type I error and power of the proposed MINQUE-based Williams bootstrap test, the likelihood ratio bootstrap test, and the asymptotic likelihood ratio test (Davidov and Rosen, 2011) are compared in Figure 8. Type I errors of all tests attain the nominal level of 0.05. Both bootstrap tests have similar power (MINQUE-based one is slightly higher) and gain in power over the asymptotic likelihood ratio test. Note that power of asymptotic likelihood ratio test is descending with increasing number of treatment groups.

Figure 8. Type I Error and Power of homoscedastic tests on the normally distributed homoscedastic data



We also explored performance of heteroscedastic tests on the data simulated assuming homoscedastic errors. Results are presented in Figure 9. Note, that while type I errors still attain the nominal level of 0.05, both asymptotic likelihood ratio test and likelihood ratio bootstrap test lose power comparing to homoscedastic tests. Pairwise power comparisons of three tests indicate that MINQUE-based bootstrap test performs the best, while asymptotic likelihood ratio test performs the worst. Simulation results for the tests considered in section 4.1.3 are provided in Table 7.2 and Table 7.3 of Appendix C.

Figure 9. Comparison of Type I Error and power of heteroscedastic tests on the normally distributed homoscedastic data

A. Type I error comparison

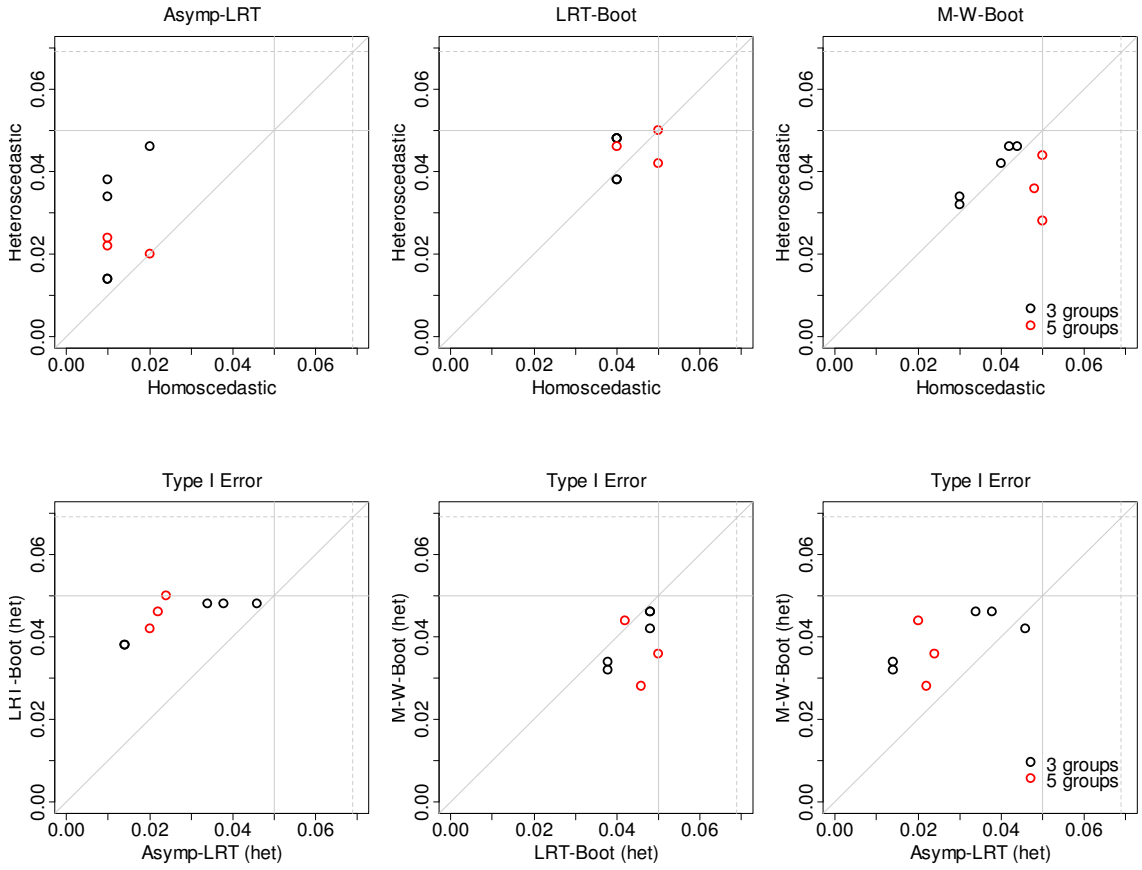
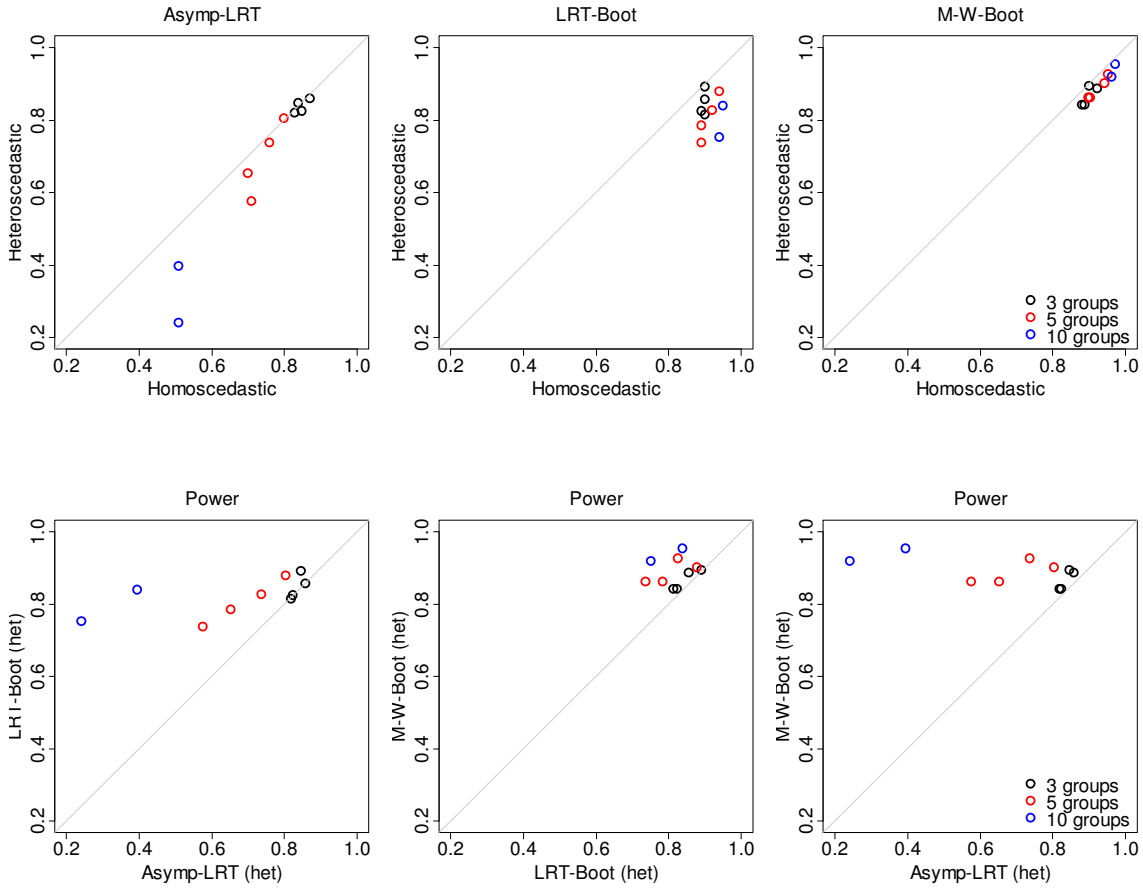


Figure 9.

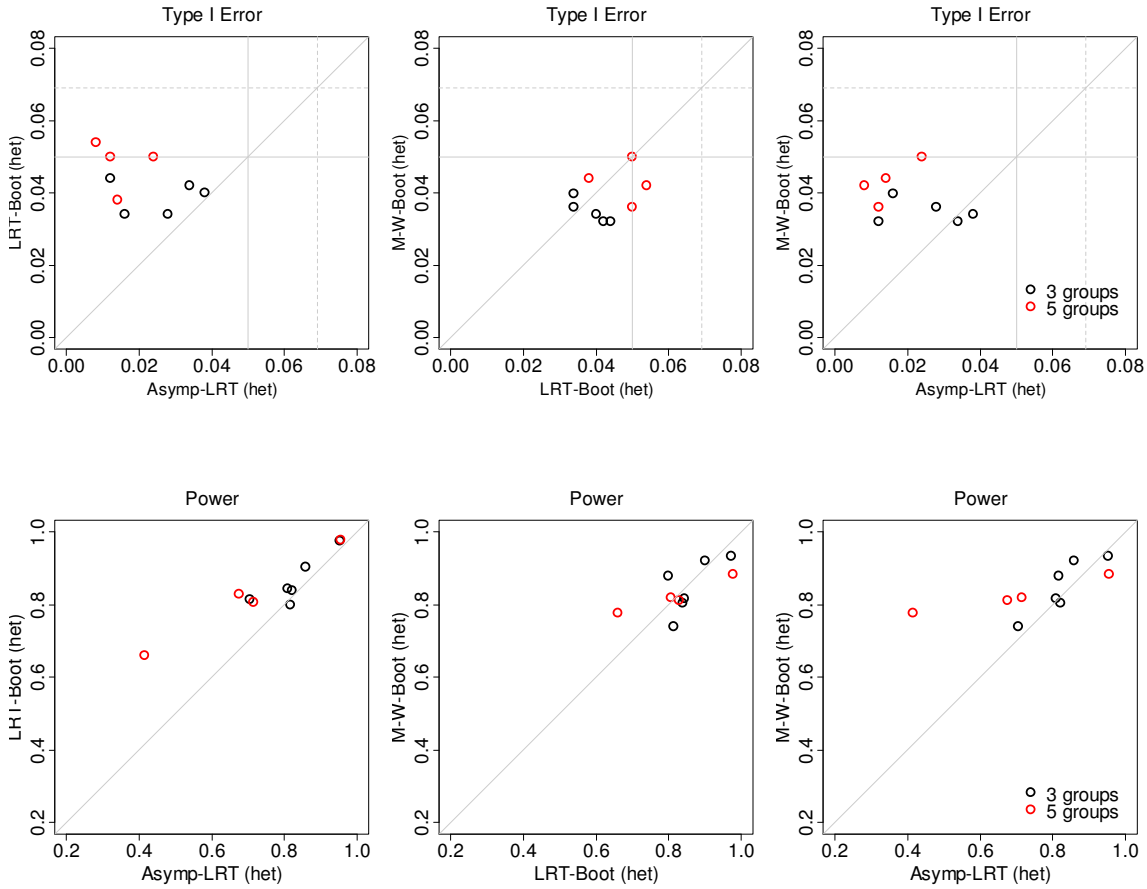
B. Power comparison



4.1.3. Results for heteroscedastic case

Simulations were performed with $\tau^2 = 1$ and $\boldsymbol{\Sigma} = diag\left[\sigma_1^2\mathbf{I}_{n_1} : \sigma_2^2\mathbf{I}_{n_2} : \ldots : \sigma_k^2\mathbf{I}_{n_p}\right]$, $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2\right)$, $N = \sum_{i=1}^{p} n_i$, where $\sigma_i^2 = \theta_i^2$ (in the case of $\theta_i = 0$, $\sigma_i^2 = 0.1$). Type I error and power of the proposed MINQUE-based bootstrap test, likelihood ratio bootstrap test, and asymptotic likelihood ratio test (Davidov and Rosen, 2011) are compared in Figure 10. Again, type I errors of all tests attain nominal level of 0.05. Both bootstrap tests have similar power and gain in power over the likelihood ratio test.

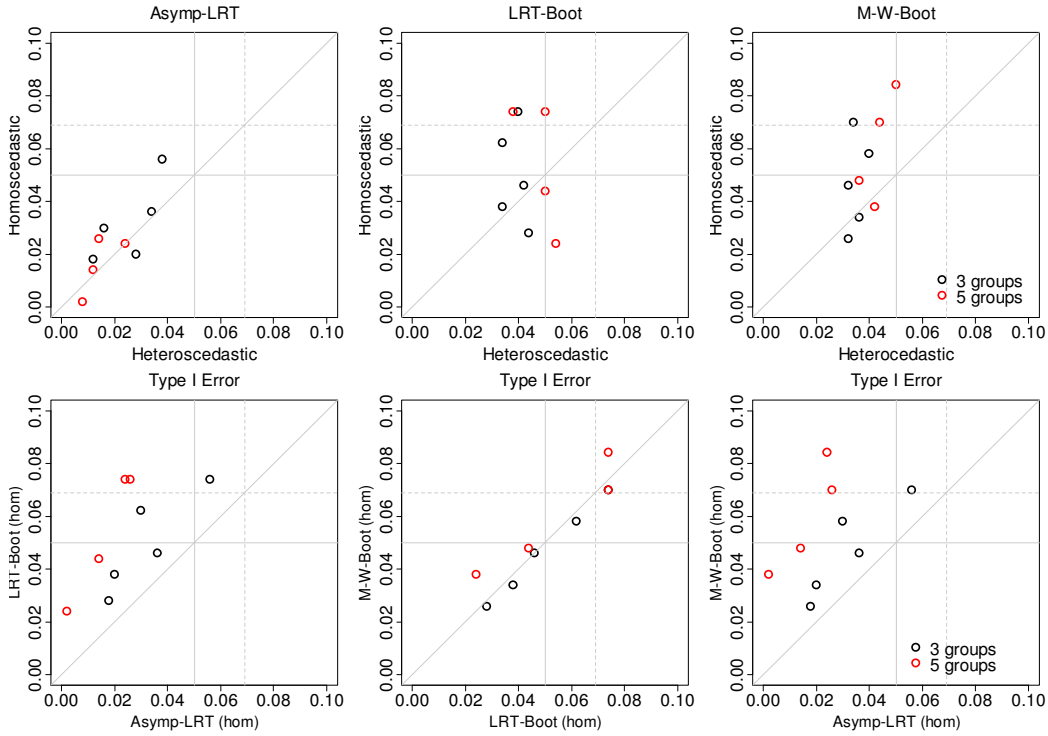
Figure 10. Type I Error and Power of heteroscedastic tests on the normally distributed heteroscedastic data



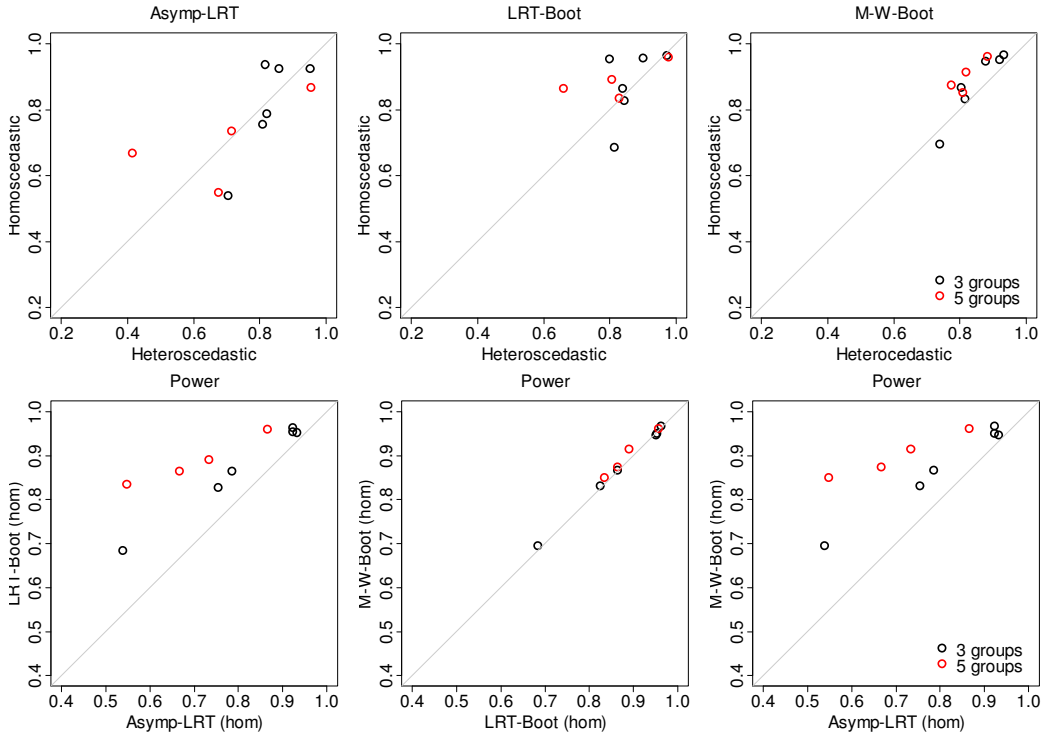
Even though the data were simulated assuming heteroscedastic errors, we also explored performance of homoscedastic tests and presented them in Figure 11. We see elevated type I errors in both bootstrap tests, thus their use is not recommended. Simulation results for the tests considered in section 4.1.3 are provided in Table 7.4 and Table 7.5 of Appendix C.

Figure 11. Comparison of Type I Error and Power of homoscedastic tests on the normally distributed heteroscedastic data

A. Type I error comparison



B. Power comparison



4.1.4. Robustness under the misspecified covariance structure

As noted in Section 3.6, there are many cases in which the underlying covariance structure will be different from the one considered in the linear mixed effects model described in (7). In this section, we evaluate the robustness of the proposed bootstrap based methodology when for a given subject *i*, the structure of the underlying covariance matrix is either unspecified or it has an auto-correlation structure. More precisely, using the same notations as before, for the *i*th subject, $i = 1, 2, 3, ..., c$, suppose the response vector $\boldsymbol{Y}_i$ is modeled as $\boldsymbol{Y}_i = \boldsymbol{\theta}_1 + \mathbf{X}_i\boldsymbol{\theta}_2 + \boldsymbol{\varepsilon}_i$. We consider two structures for $Cov(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Omega}$, namely, (a) $\boldsymbol{\Omega}$ has no pre-specified structure and (b) $\boldsymbol{\Omega}$ is an auto-correlation matrix of the form:

$$\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^p \\ \rho & 1 & \rho & \dots & \rho^{p-1} \\ & \ddots & \ddots & \ddots & \\ \rho^{p-1} & \dots & \rho & 1 & \rho \\ \rho^p & \dots & \rho^2 & \rho & 1 \end{pmatrix}.$$

In contrast, the covariance structure considered earlier in the simulation studies induced an intra-class covariance structure for each subject.

We considered two different values for the dimension *p*, namely, *p* = 3, 5. In the case of (a) we generated $\boldsymbol{\Omega}$ using a Wishart distribution with degrees of freedom *df* and a pattern of the scale matrix constructed from the data provided in Cao et al. (2011). We used two scale matrices in our simulation study as follows:

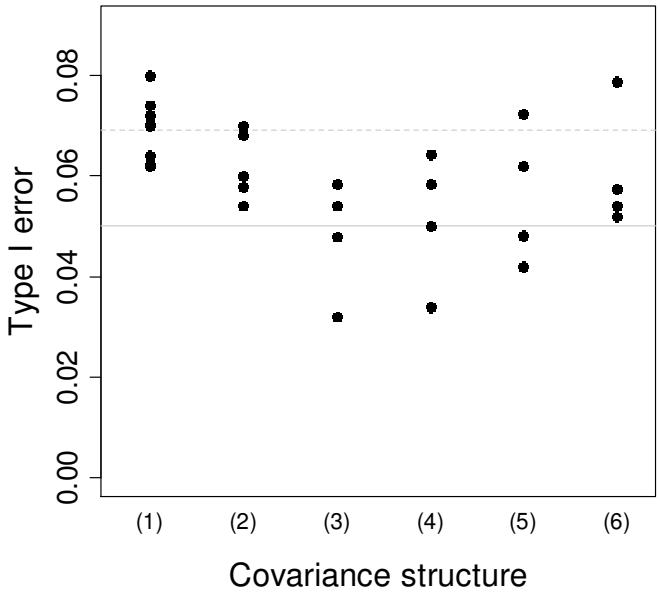
(1) $\begin{pmatrix} 0.31 & 0.21 & 0.11 \\ 0.21 & 0.29 & 0.13 \\ 0.11 & 0.13 & 0.27 \end{pmatrix}$ (placebo group) or

$$(2)\ \begin{pmatrix} 0.34 & 0.26 & 0.13 \\ 0.26 & 0.33 & 0.10 \\ 0.13 & 0.10 & 0.27 \end{pmatrix} \text{ (succimer group).}$$

To generate the data for the case $p = 5$, we augmented the above matrices with two additional rows and columns such that the diagonal elements were 0.25 and 0.23, while off-diagonal elements were 0.2.

In the case of (b), we again considered two different values for the dimension p, namely, $p = 3, 5$, and we also considered four different patterns of correlation coefficients, namely, (3) $\rho = 0.2$, (4) $\rho = 0.4$, (5) $\rho = 0.6$, (6) $\rho = 0.9$. Thus we considered a total of six different patterns of covariance structures, which are labeled as (1), (2), (3), (4), (5) and (6). For each pattern, we considered two patterns of dimensions p, i.e. $p = 3$ or 5. Since the focus of this study is to determine whether the proposed methodology is robust to departures from the assumed covariance structure in terms of type I error, in Figure 12 we summarized the type I errors of all patterns considered in this simulation study. The X-axis denotes the six patterns under consideration, and the Y-axis denotes the type I error. The nominal level was taken to 0.05. As before, the results are based on 500 simulation runs using 500 bootstrap samples. As expected, when the covariance matrix is very different from the presumed covariance matrix, the type I error exceeds the nominal level. This happens in the case of the unspecified covariance matrix and the auto-correlation structure with patterns (5) and (6). The type I errors corresponding to the auto-correlation structure seem to be below the nominal level in patterns (3) and (4). Thus, in general, as one would expect, the proposed methodology may not be robust to departures from the presumed covariance structure. Complete set of results are provided in Tables 7.6, 7.7, 7.8 and 7.9 in the Appendix C.

Figure 12. Type I errors under the misspecified covariance matrix



4.2. Non-normally distributed data

In this dissertation research, we also explored the robustness of the proposed methodology for departures from normality. We simulated data according to the following mixed effects model, where the elements of the random error term $\boldsymbol{r} = (r_1, r_2, \ldots, r_N)$ followed one of the following commonly studied distributions, namely, the log-normal, gamma and mixture of two normally distributed variables:

$$\boldsymbol{Y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 + \mathbf{U}\boldsymbol{\xi} + \boldsymbol{r}, \tag{29}$$

where the terms $\mathbf{X}_1, \boldsymbol{\theta}_1, \mathbf{X}_2, \boldsymbol{\theta}_2, \mathbf{U}$ and $\boldsymbol{\xi}$ are defined in section 4.1.1.

4.2.1. Log-normally distributed data

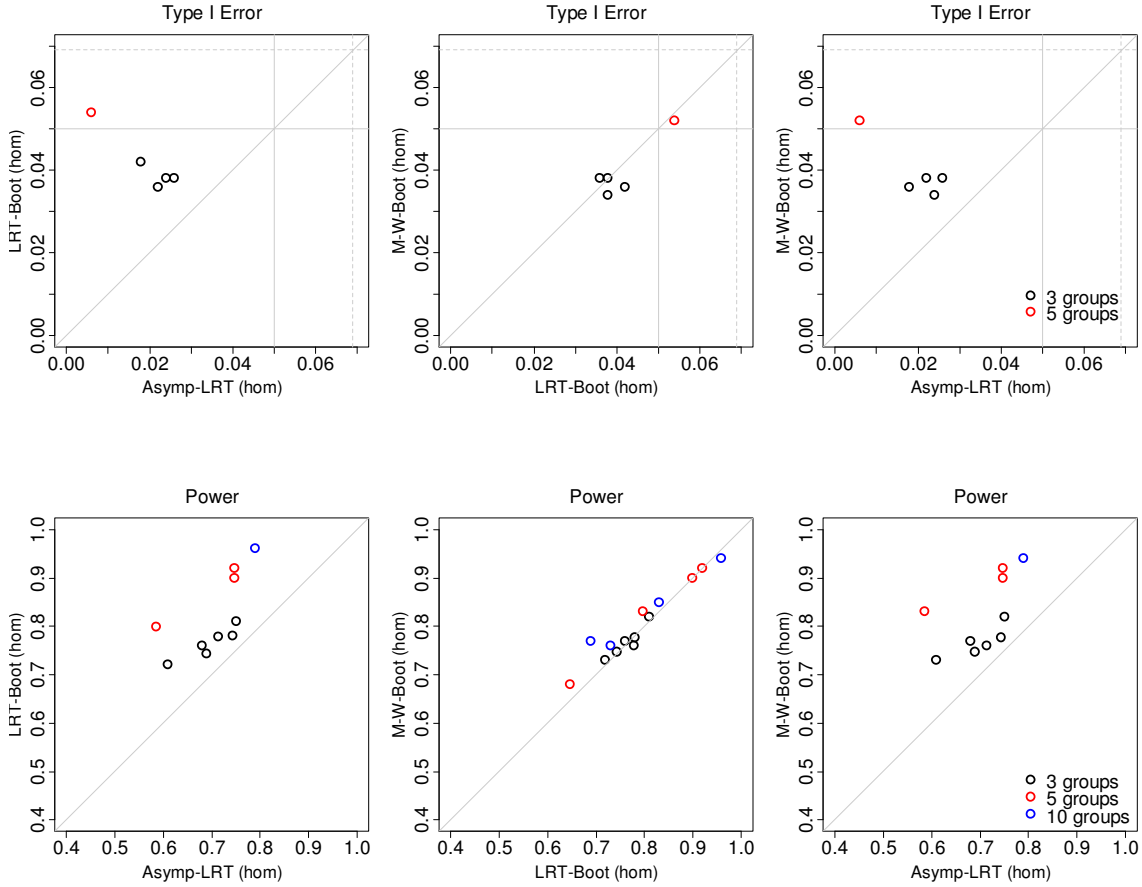
4.2.1.1. Study design

Suppose $\boldsymbol{r} = \exp(\boldsymbol{\varepsilon})$, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, with other definitions presented in section 4.1.1 on page 56. We compared type I error and power of the proposed MINQUE-based bootstrap test, likelihood ratio bootstrap test and asymptotic likelihood ratio test (Davidov and Rosen, 2011).

4.2.1.2. Results for homoscedastic case

Simulations were performed with $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$ ($\sigma = 1$), $\tau^2 = 0.2,\ 1,\ 2$, and treatment mean patterns (1)-(3). Results are presented in Figure 13. As in the case of normal data, type I errors of all three homoscedastic tests attain the nominal level of 0.05. Both bootstrap tests have similar power (though MINQUE-based one is higher in most cases) and gain in power over the asymptotic likelihood ratio test.

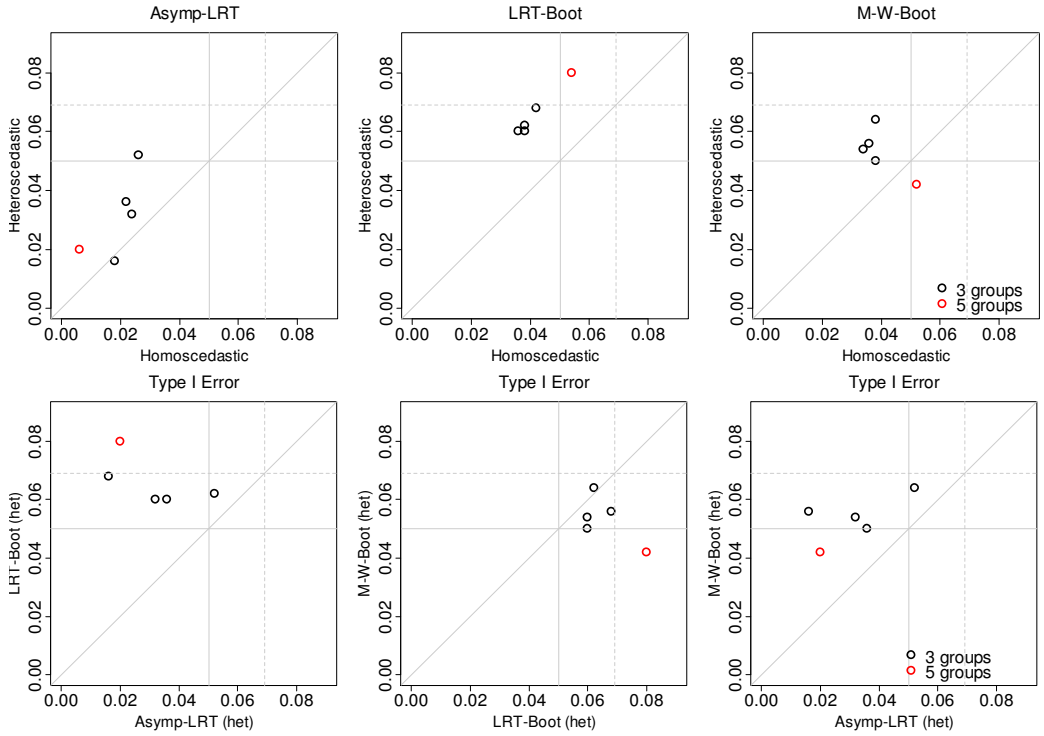
Figure 13. Type I Error and Power of homoscedastic tests on the log-normally distributed homoscedastic data



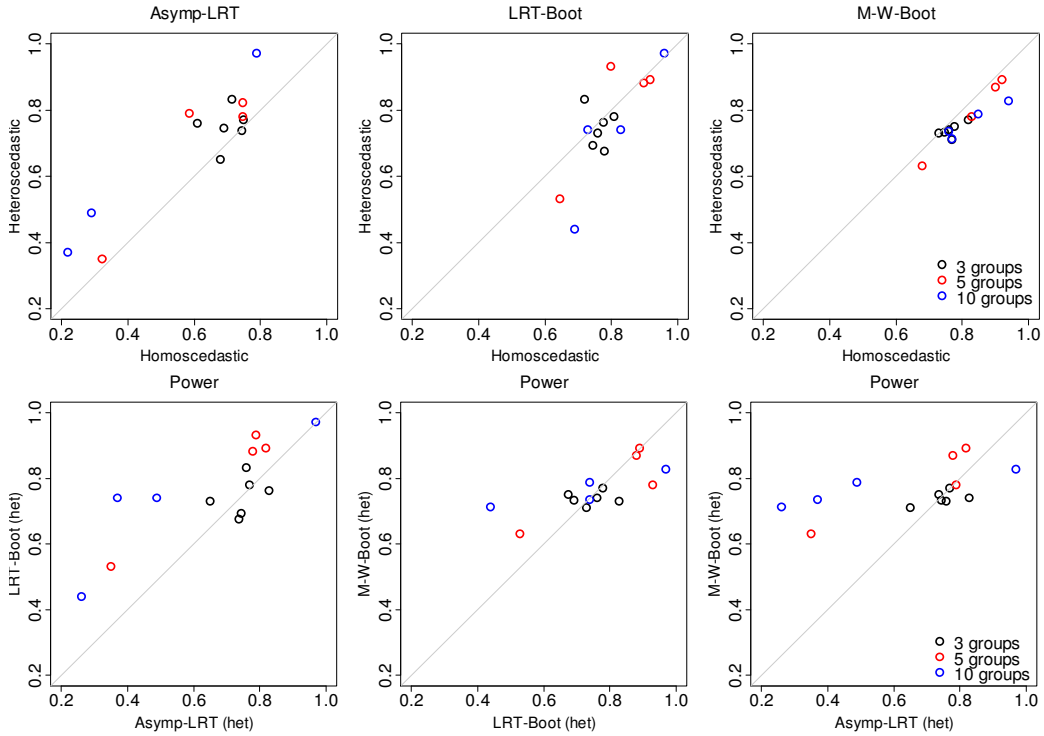
We also explored how heteroscedastic tests perform on the homoscedastic non-normal data. Results are presented in Figure 14. Note that type I errors are elevated for likelihood ratio bootstrap test while attaining the nominal level of 0.05 for the asymptotic likelihood ratio and the MINQUE-based bootstrap tests. In the most cases, MINQUE-based bootstrap test gains in power over the asymptotic likelihood ratio test. Simulation results for the tests considered in section 4.2.1.2 are provided in Table 7.10 and Table 7.11 of Appendix C.

Figure 14. Comparison of Type I Error and power of heteroscedastic tests on the log-normally distributed homoscedastic data

A. Type I error comparison



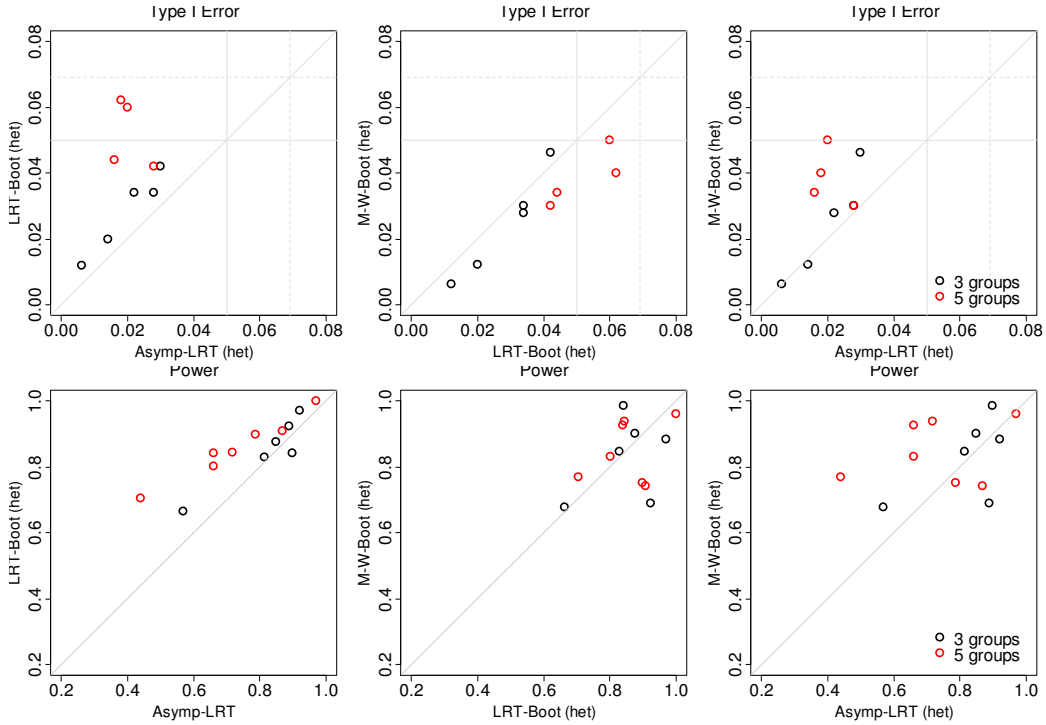
B. Power comparison



4.2.1.3. Results for heteroscedastic case

Simulations were performed with $\tau^2 = 1$, $\boldsymbol{\Sigma} = diag\left[\sigma_1^2\mathbf{I}_{n_1} : \sigma_2^2\mathbf{I}_{n_2} : \ldots : \sigma_k^2\mathbf{I}_{n_p}\right]$, $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2\right)$, $N = \sum_{i=1}^{p} n_i$, where $\sigma_i^2 = \theta_i^2$ (in the case of $\theta_i = 0$, $\sigma_i^2 = 0.1$). Type I error and power of the proposed MINQUE-based bootstrap test, likelihood ratio bootstrap test, and likelihood ratio test (Davidov and Rosen, 2011) are compared in Figure 15. Again, type I errors of all tests attain the nominal level of 0.05 and are comparable in power.

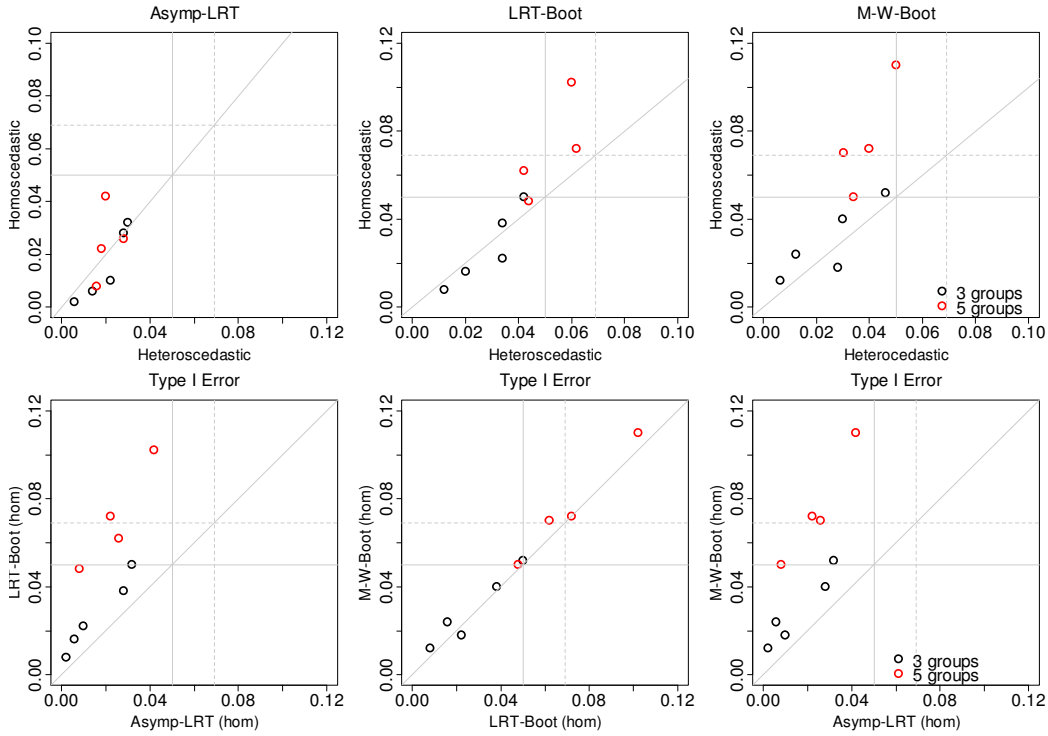
Figure 15. Type I Error and Power of heteroscedastic tests on the log-normally distributed heteroscedastic data



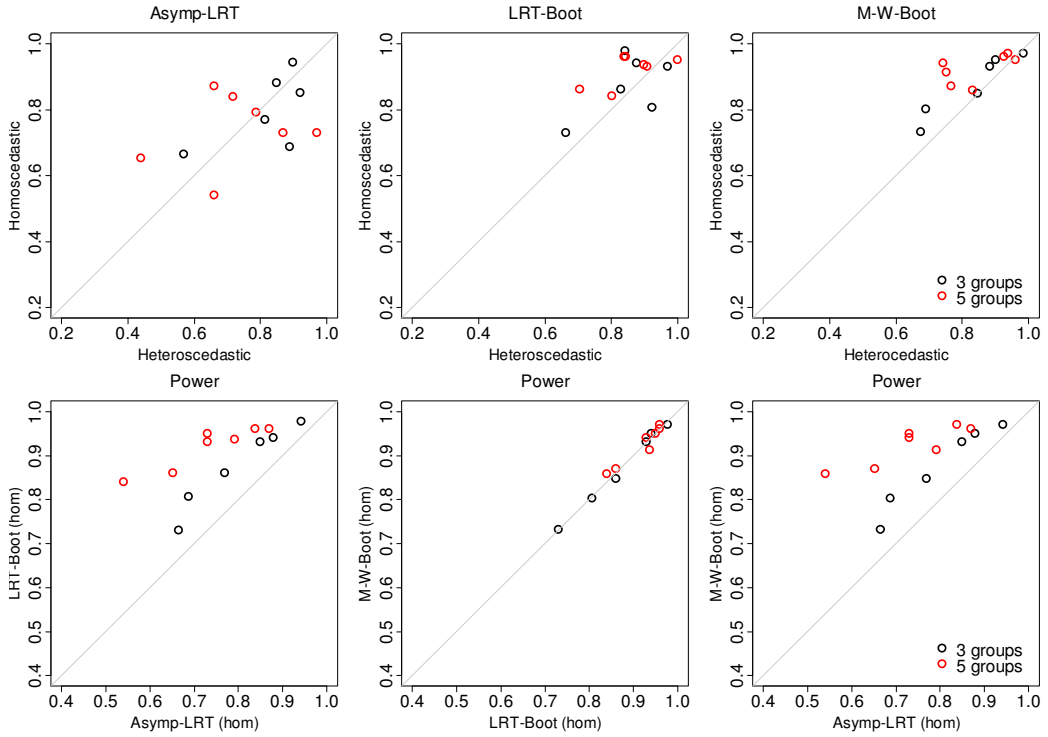
We also explored performance of homoscedastic tests when the data were simulated assuming heteroscedastic errors (Figure 16). We note that MINQUE-based bootstrap test can have elevated type I errors. Simulation results for the tests considered in section 4.2.1.3 are provided in Table 7.12 and Table 7.13 of Appendix C.

Figure 16. Comparison of Type I Error and Power of homoscedastic tests on the log-normally distributed heteroscedastic data

A. Type I error comparison



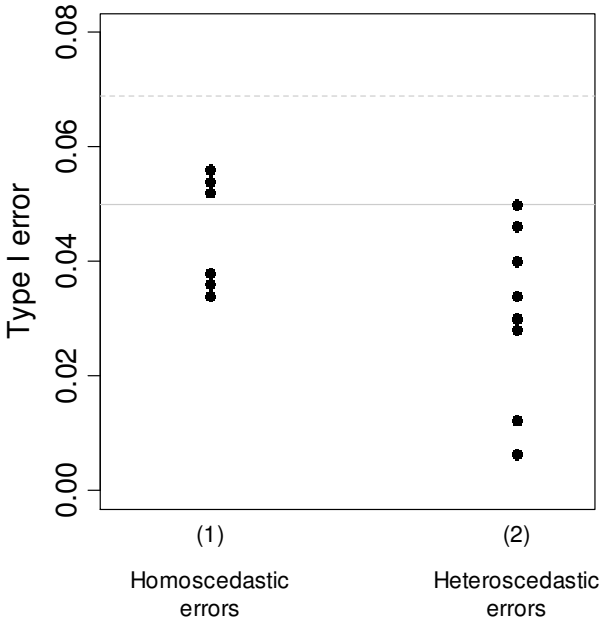
B. Power comparison



4.2.1.4. Type I errors

Type I errors for of the proposed test are plotted in Figure 17. In conclusion, the proposed methodology seems to maintain the nominal level for type I errors for log-normally distributed data assuming both homoscedastic and heteroscedastic random errors.

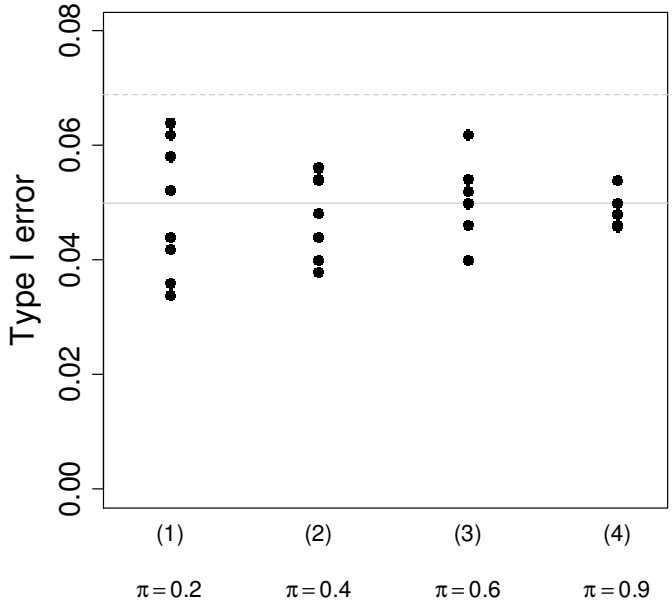
Figure 17. Type I errors of a proposed test for the log-normally distributed data



4.2.2. A mixture of two normally distributed random variables

The data were simulated according to (29) with $r_i = \pi e_1 + (1-\pi) e_2$, $i = 1,2,\ldots,N$, where $e_1 \sim N(0,0.5)$, $e_2 \sim N(0,s)$, $s = 1,5$ and (1) $\pi = 0.2$, (2) $\pi = 0.4$, (3) $\pi = 0.6$, (4) $\pi = 0.9$. Type I errors of the proposed test presented in Figure 18 indicate that the nominal level is maintained. Complete simulation results are presented in Table 7.14 and Table 7.15 of Appendix C.

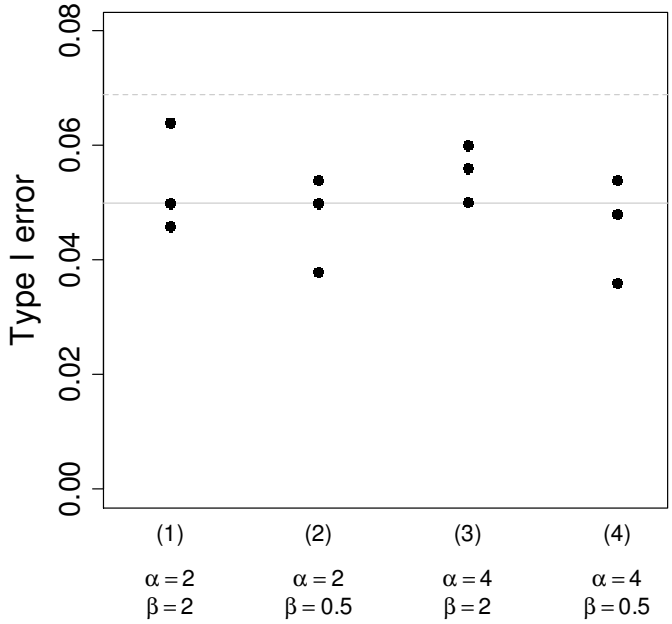
Figure 18. Type I errors of a proposed test for a mixture of two normally distributed random variables



4.2.3. Gamma-distributed random errors

The data were simulated according to (29) with r_i $(i = 1,\ldots,N)$ following a gamma distribution with a shape parameter α and a scale parameter β with the density function $f(x) = x^{\alpha-1}\frac{e^{-x/\beta}}{\beta^{\alpha}\Gamma(k)}$, $x \geq 0,\ \alpha,\beta > 0$ under the following patterns: (1) $\alpha = 2,\ \beta = 2$, (2) $\alpha = 2,\ \beta = 0.5$, (3) $\alpha = 4,\ \beta = 2$, (4) $\alpha = 4,\ \beta = 0.5$. Type I errors of the proposed test presented in Figure 19 indicate that the nominal level is maintained. Complete simulation results are presented in Table 7.16 and Table 7.17 of Appendix C.

Figure 19. Type I errors of a proposed test when random errors follow the gamma distribution



In conclusion, simulations performed in section 4.2 indicate that when the data are not normally distributed, the proposed bootstrap methodology maintains the nominal level for type I errors.

4.3. Concluding remarks and recommendations

Analyzing simulations for normally distributed data, we noted that the heteroscedastic MINQUE-based bootstrap test performs the best. For both homoscedastic and heteroscedastic data, its type I errors attain the nominal level of 0.05. For heteroscedastic data, it has similar power as the heteroscedastic likelihood ratio bootstrap test and gains in power over the asymptotic likelihood ratio test. In the case of the homoscedastic data, it gains in power over both asymptotic likelihood ratio and likelihood ratio bootstrap tests. Also, for the homoscedastic data, heteroscedastic

MINQUE-based bootstrap test does not lose much in power in comparison to the homoscedastic MINQUE-based bootstrap test.

Analyzed cases of non-normally distributed data also support the usage of the heteroscedastic MINQUE-based bootstrap test. For heteroscedastic data, its type I errors attain the nominal level of 0.05 and power is similar to other tests. If the data are generated assuming homoscedasticity, MINQUE-based bootstrap test's type I errors achieve the nominal level of 0.05, while being elevated for the likelihood ratio bootstrap test; also in most cases MINQUE-based bootstrap test gains in power over the asymptotic likelihood ratio test.

In conclusion, we recommend the heteroscedastic MINQUE-based bootstrap test for performing constrained inference in linear mixed effects models. For heteroscedastic data, its type I errors attain the nominal level of 0.05 and its power is similar to the power of other tests. In the case of normally distributed data, it gains in power over the heteroscedastic likelihood ratio test. If the data are homoscedastic, it gains in power over the likelihood ratio bootstrap test while its type I errors still attain the nominal level of 0.05.

For the cases when the covariance structure is either unspecified or has some other special structure, we recommend modifying the methodology by bootstrapping suitable residuals as described in Section 3.6.

CHAPTER 5

ILLUSTRATION

To illustrate the methodology, we used the real data provided Dr. Walter Rogan, Epidemiology Branch, NIEHS. The data were collected during a randomized placebo-controlled, double-blind trial clinical trial of succimer for lead poisoning, called the Treatment of Lead-exposed Children trial, or TLC (Rogan, 1998). In TLC, 384 children aged 12-33 months were assigned to the placebo group and 396 to the succimer group. Up to three 26-day courses of succimer or placebo therapy were administered, depending on response to treatment in those, who were given succimer. Cao et al. (2011) were interested whether succimer, a mercaptan compound known to reduce blood lead concentration in children, also reduces blood mercury concentration. At the baseline, blood mercury levels were obtained in 767 samples (393 succimer group and 374 placebo group) and detected and quantified in 657 samples (86%; 338 succimer and 319 placebo). At 1-week post treatment, total mercury concentration was measured in 768 samples (389 succimer and 379 placebo) and detected and quantified in 623 samples (81%; 313 succimer and 310 placebo). After 5 months of treatment, blood mercury levels were obtained from a 20% random sample of 338 children completing 3 courses of treatment. Total mercury was detected and quantified in 61 samples: 30 (out of 393) succimer treated children and 31 (out of 374) placebo treated children. Cao et al. (2011) used an

ad-hoc bootstrap-based isotonic regression method to compare the trend over time in the difference between the adjusted mean mercury concentrations in the succimer group and the placebo group. Their analysis adjusted for child's age, sex, race and the study center the child belonged to. Cao et al. (2011) hypothesized a monotonic trend in the difference between the adjusted mean mercury concentrations in the succimer group and the placebo group. Authors used point-wise confidence intervals for the mean differences at each time point to describe the differences between the two groups. Although the bootstrap methodology used by the authors exploits the underlying dependence structure due to repeated measurements, the test statistic they used ignores it.

Cao et al. (2011) implicitly inferred that the trend they observed could potentially be due to the fact that, over time, there is no real difference between the succimer and the placebo groups. This motivated us to re-analyze their data for the succimer and placebo groups separately to understand the trend in mean mercury levels in the two treatment groups. As done in Cao et al. (2011), we adjusted for child's age, sex, race and the study center the child belonged to. The variable of interest was organic mercury level in log-scale. The normal quantile-quantile (Q-Q) plots of studentized residuals for the placebo and the succimer groups (Figure 20) suggest that the data are potentially non-normally distributed. They suggest heavy tails. To illustrate heavy tails, in Figure 21 we present histograms of studentized residuals.

Figure 20. Normal probability plots of organic mercury level in log-scale for placebo and succimer groups.

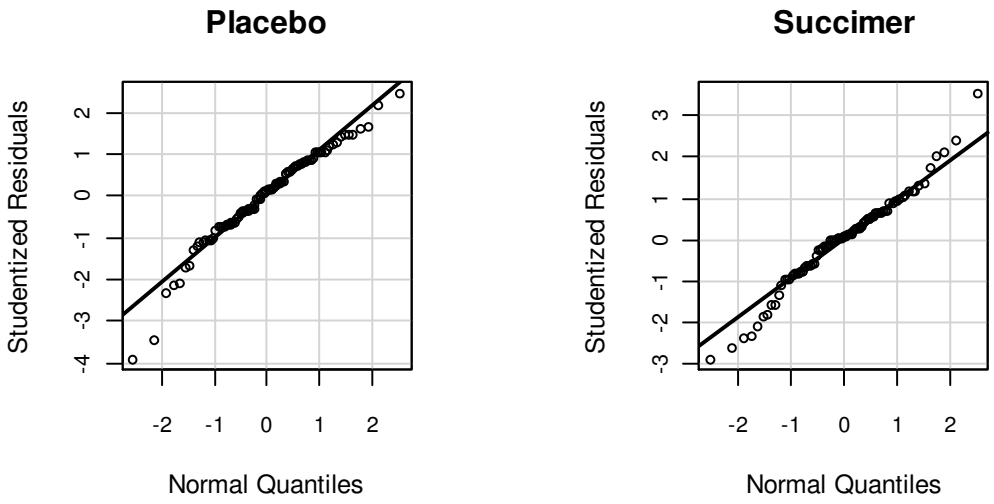
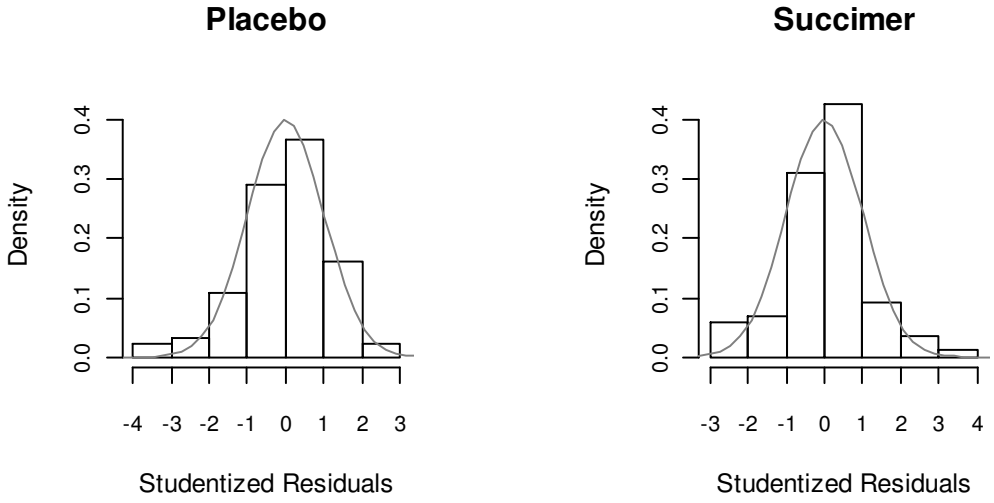
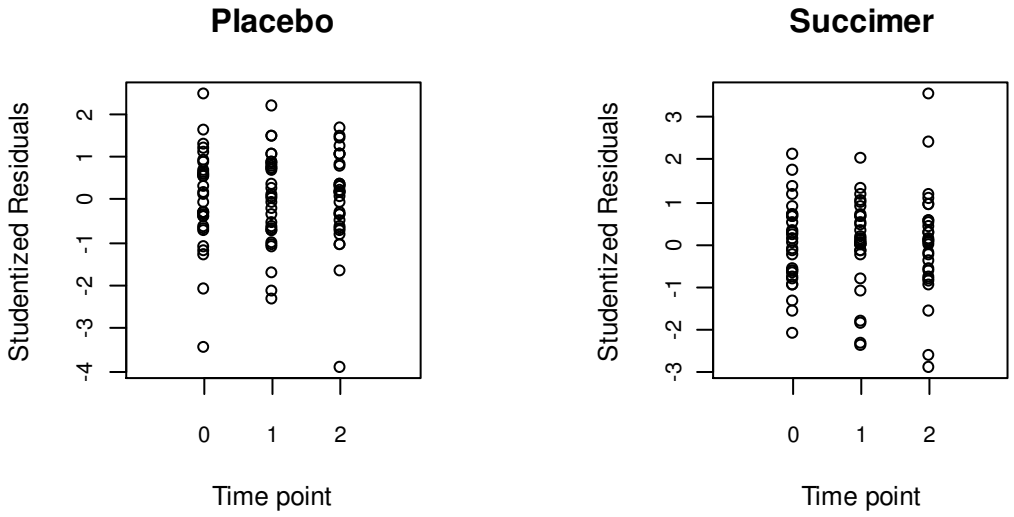


Figure 21. Histograms of organic mercury level in log-scale for placebo and succimer groups.



Studentized residuals were plotted against time in Figure 22. Visually the variability in the data seems to be constant across time, with a few outliers in the succimer treated group's 5-month data.

Figure 22. Studentized residuals by time point (0 – baseline, 1 – 1-week post-treatment, 2 – 5-month post-treatment).



Since the data appear to be somewhat non-normal and possibly heteroscedastic, we used both heteroscedastic and homoscedastic MINQUE-Williams based non-parametric bootstrap methods introduced in this dissertation. In Tables 5.1 and 5.2, we provide UMLE and MINQUE-based PAVA estimates of organic mercury level in log-scale. In Figure 23, we present the estimates and standard errors for heteroscedastic methods.

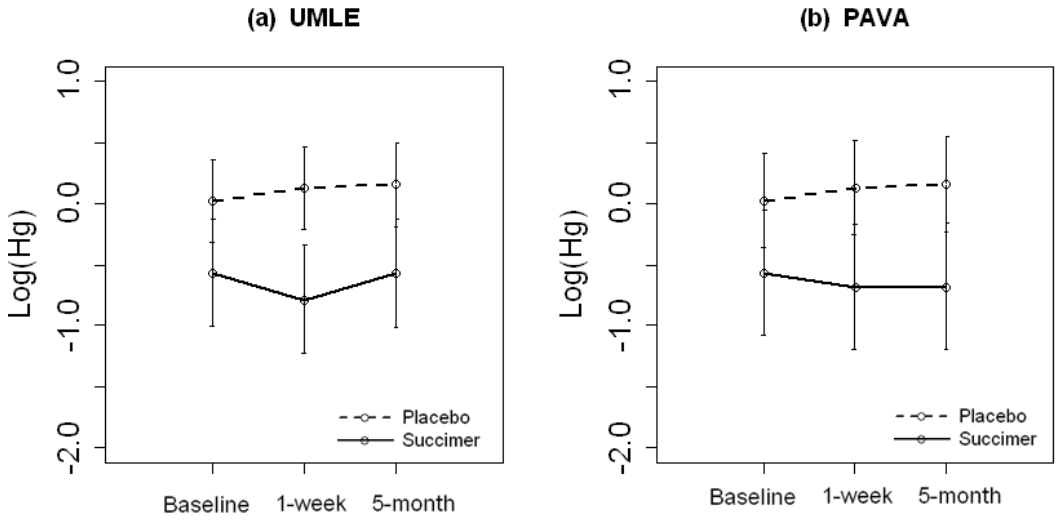
Table 5.1. Mean blood concentration of organic mercury in children given placebo (in log-scale)

Time	n	Heteroscedastic errors				Homoscedastic errors			
		UMLE		PAVA		UMLE		PAVA	
		Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
Baseline	31	0.03	0.34	0.03	0.38	0.10	0.33	0.10	0.37
1-week	31	0.13	0.34	0.13	0.38	0.21	0.33	0.21	0.37
5-month	31	0.16	0.34	0.16	0.39	0.23	0.33	0.23	0.37

Table 5.2. Mean blood concentration of organic mercury in children given succimer (in log-scale)

Time	n	Heteroscedastic errors				Homoscedastic errors			
		UMLE		PAVA		UMLE		PAVA	
		Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
Baseline	29	-0.57	0.44	-0.57	0.51	-0.51	0.43	-0.51	0.49
1-week	29	-0.78	0.45	-0.68	0.51	-0.73	0.43	-0.62	0.49
5-month	29	-0.57	0.45	-0.68	0.52	-0.51	0.43	-0.62	0.49

Figure 23. Estimated mean blood log concentration of organic mercury in children given succimer or placebo; (a) UMLE, (b) PAVA. Error bars demonstrate ± 1 standard error.



Denote the mean mercury levels at the tth time for the placebo and succimer groups as $\theta_t^{placebo}$ and $\theta_t^{succimer}$ respectively ($t = 1$ for baseline, $t = 2$ for 1-week and $t = 3$ for 5-months). In the placebo group we hypothesized the increase of blood mercury level over time, i.e. tested the hypothesis:

$$H_0 : \theta_1^{placebo} = \theta_2^{placebo} = \theta_3^{placebo} \quad \text{versus}$$
$$H_a : \theta_1^{placebo} \le \theta_2^{placebo} \le \theta_3^{placebo},$$

with at least one strict inequality among parameters. To test this hypothesis, we performed a heteroscedastic MINQUE-Williams test. The data were not strong enough

to reject the null hypothesis at 0.05 level (p-value was 0.152). For comparison purposes we also considered other tests. A heteroscedastic likelihood ratio test had a p-value of 0.163. Homoscedastic MINQUE-Williams test had a p-value of 0.122 and homoscedastic likelihood ratio test had a p-value of 0.139.

In the succimer group we hypothesized the decrease of blood mercury level over time. The hypothesis was the following:

$$H_0 : \theta_1^{succimer} = \theta_2^{succimer} = \theta_3^{succimer} \quad \text{versus}$$
$$H_a : \theta_1^{succimer} \geq \theta_2^{succimer} \geq \theta_3^{succimer},$$

with at least one strict inequality among parameters. To test this hypothesis, we performed a heteroscedastic MINQUE-Williams test. The data were not strong enough to reject the null hypothesis at 0.05 level (p-value was 0.192). For comparison purposes we also considered other tests. A heteroscedastic likelihood ratio test rejected the null hypothesis (p-value of 0.048). Homoscedastic MINQUE-Williams test had a p-value of 0.202 and homoscedastic likelihood ratio test had a p-value of 0.205.

In conclusion, the data were not strong enough to support the hypotheses of the trend in placebo and succimer groups. This is consistent with conclusions of Cao et al. (2011), that succimer chelation for low level organic mercury exposure in children has limited efficacy.

CHAPTER 6

SUMMARY AND CONCLUDING REMARKS

Inequality constraints arise naturally in many applications, such as toxicology, where researchers are interested in studying dose-response of a chemical, gene expression studies in oncology, where a researcher may be interested in understanding the changes in gene expression according to cancer stage, etc. There exists an extensive literature on statistical inference under inequality constraints, including four excellent books on the subject. For a detailed review of the estimation of parameters subject to inequality constraints, one may refer to van Eeden (2006), while a comprehensive account on testing problems is provided in Silvapulle and Sen (2005). This dissertation research has two components, estimation and testing under inequality constraints, with focus on simple order constraint where inequalities among all unknown parameters are known a priori.

As summarized in van Eeden (2006), there are numerous methods available in the literature to estimate parameters under inequality constraints, the popular ones being the restricted maximum likelihood estimator (RMLE) and the pool adjacent violators algorithm (PAVA) type estimators. The performance of RMLE is well understood for both independent and correlated data. However, even though PAVA is widely used even when the underlying data are correlated, there does not exist any literature on its

performance in such cases. This motivated the present dissertation work. In this dissertation, the performance of PAVA estimator was evaluated using the universal domination (also known as stochastic domination) criterion. It was demonstrated that performance of PAVA depends upon the underlying covariance matrix. Under suitable sufficient conditions derived in this dissertation, it is shown that PAVA estimator universally dominates the unrestricted maximum likelihood estimator (UMLE). Interestingly, extensive simulation studies conducted in this dissertation work suggest that these sufficient conditions are also potentially necessary conditions. Observe that for $p > 2$ under the simple order cone, the sufficient conditions obtained in Chapter 2 for the largest and the smallest parameters are disjoint.

Consequently, it may not be possible to obtain similar domination theorems for the intermediate population means.

In view of the existing literature on RMLE (Hwang and Peddada, 1994; Peddada, Dunson and Tan, 2005; Betcher and Peddada, 2009) and the results obtained in this dissertation research, we conclude that none of the existing constrained estimators is expected to perform better than the UMLE for all covariance matrices, under all inequality constrains and for all parameters. Furthermore, even for a given covariance matrix, under simple order constraint, it is not possible to analytically determine which of the existing constrained estimators, namely, RMLE, PAVA, the covariance weighted PAVA (Hwang and Peddada, 1994), the modified covariance-weighted PAVA (Peddada et al., 2005) or the modified RMLE (Betcher and Peddada, 2009) is the best choice. Perhaps, for a given application, given covariance matrix and the parameter of interest

the investigator should perform extensive simulation studies under a variety of plausible scenarios and choose the best estimator for that application.

The second component of this dissertation work was statistical testing under inequality constraints when the underlying data are correlated. Again, this work is largely motivated by applications in toxicology and clinical trials. Although there is a well-developed asymptotic likelihood ratio based theory for general problems (Silvapulle and Sen, 2005), surprisingly, in the literature there is very little known about testing for specific covariance structures, such as those encountered in repeated measures type data. Very recently Davidov and Rosen (2011) were the first to provide a general framework for testing under inequality constraints in a linear mixed effects model. Furthermore, it is important to note that Davidov and Rosen's work was not known when this dissertation work was being prepared. Davidov and Rosen (2011) provide an asymptotic test, whereas in this dissertation a nonparametric bootstrap based method was developed. In the simple order restriction, extensive simulation studies conducted in this dissertation work suggest that the proposed methodology provides a better control of type I error than the asymptotic likelihood ratio test of Davidov and Rosen (2011) when the data are non-normally distributed. Since the proposed test uses Rao's MINQUE theory (1970, 1971, 1972) for estimating variance components and PAVA for estimating the means, it does not necessarily require normality.

Although the focus of Chapters 3 and 4 of this dissertation work was on the simple order cone, the proposed methodology can be easily extended to other order restrictions. The framework developed in this dissertation is very general and can be

easily adapted to other order restrictions by suitably choosing the elements of matrix A in the inequality (8).

In toxicology, researchers are often interested in dose×time response surfaces, which results in a two-way classification that can be expressed as a constrained inference problem with constraints on rows and columns of a matrix. The proposed testing procedure can be extended to such cases by generalizing the methodology developed in Teoh et al. (2008) along the lines of the non-parametric bootstrap developed in this dissertation work.

APPENDIX A

Proofs and additional lemmas of Chapter 2

We begin with the following lemma which can be derived using straightforward algebra. Corresponding to a 2x2 real matrix $\mathbf{C}$, let $\boldsymbol{a} = \dfrac{\mathbf{C}'1}{1'\mathbf{C}1}$, where $\mathbf{1} = (1,1)'$, and let $b = (1,-1)'$.

Lemma A1: Suppose $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is a 2x2 positive definite matrix and $\mathbf{C}$ is a real 2x2 matrix. Let $\boldsymbol{b} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ then $\boldsymbol{a}'\boldsymbol{\Sigma a} - \dfrac{(\boldsymbol{a}'\boldsymbol{\Sigma b})^2}{\boldsymbol{b}'\boldsymbol{\Sigma b}} = \dfrac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$, invariant of $\mathbf{C}$.

Lemma A2: Based on a random sample $\boldsymbol{U} = (U_1, U_2, ..., U_n) \in R^n$ suppose, for $i = 1,2$, $T_i(\boldsymbol{U}) = T_0(\boldsymbol{U})I(\boldsymbol{U} \in S) + T_i^*(\boldsymbol{U})I(\boldsymbol{U} \in S^c)$ are two estimators of a parameter $\boldsymbol{\mu}$, where $T_0(U)$ and $T_i^*(U)$ are arbitrary functions of U, $S \cup S^c = R^n$, and $I(\boldsymbol{U} \in R^n)$ is an indicator function taking value of 1 whenever $\boldsymbol{U} \in R^n$ and is zero otherwise. If

$$(a)\ E_{T_1^*|S^c}\left\{\left[T_1^*(\boldsymbol{U}) - E_{T_1^*|S^c}\left(T_1^*(\boldsymbol{U})\right)\right]^2 I\left(\boldsymbol{U} \in S^c\right)\right\}$$

$$= E_{T_2^*|S^c}\left\{\left[T_2^*(\boldsymbol{U}) - E_{T_2^*|S^c}\left(T_2^*(\boldsymbol{U})\right)\right]^2 I\left(\boldsymbol{U} \in S^c\right)\right\},$$

$$(b)\ E_{S^c}\left\{\left[E_{T_1^*|S^c}T_1^*(\boldsymbol{U}) - \mu\right]^2 I\left(\boldsymbol{U} \in S^c\right)\right\} \le E_{S^c}\left\{\left[E_{T_2^*|S^c}T_2^*(\boldsymbol{U}) - \boldsymbol{\mu}\right]^2 I\left(\boldsymbol{U} \in S^c\right)\right\},$$

then $E\left(T_1(\boldsymbol{U}) - \mu\right)^2 \le E\left(T_2(\boldsymbol{U}) - \mu\right)^2$.

Proof: For simplicity of notation we drop $\boldsymbol{U}$ from $T_i(\boldsymbol{U})$. Note that

$E(T_i-\mu)^2 = E_S\{E(T_0-\mu)^2 I(\boldsymbol{U}\in S)\}+E_{S^c}\{E(T_i^*-\mu)^2 I(\boldsymbol{U}\in S^c)\}$ therefore

$E(T_1-\boldsymbol{\mu})^2$ - $E(T_2-\mu)^2 = E_{S^c}\{E_{T_1^*|S^c}[(T_1^*-\mu)^2 I(\boldsymbol{U}\in S^c)]-E_{T_2^*|S^c}[(T_2^*-\mu)^2 I(\boldsymbol{U}\in S^c)]\}$.

Note that

$$E\left(T_i-\boldsymbol{\mu}\right)^2 = E_S\left\{E\left(T_0-\boldsymbol{\mu}\right)^2 I\left(\boldsymbol{U}\in S\right)\right\}+E_{S^c}\left\{E\left(T_i^*-\boldsymbol{\mu}\right)^2 I\left(\boldsymbol{U}\in S^c\right)\right\} \text{ and}$$

$$\begin{aligned}
&E_{T_i^*|S^c}\left[\left(T_i^*-\boldsymbol{\mu}\right)^2 I\left(\boldsymbol{U}\in S^c\right)\right]\\
&= E_{T_i^*|S^c}\left[\left\{T_i^*-E_{T_i^*|S^c}\left(T_i^* I\left(\boldsymbol{U}\in S^c\right)\right)+E_{T_i^*|S^c}\left(T_i^*\left(\boldsymbol{U}\in S^c\right)\right)-\boldsymbol{\mu}\right\}^2 I\left(\boldsymbol{U}\in S^c\right)\right]=\\
&= E_{T_i^*|S^c}\left[\left\{T_i^*-E_{T_i^*|S^c}\left(T_i^* I\left(\boldsymbol{U}\in S^c\right)\right)\right\}^2 I\left(\boldsymbol{U}\in S^c\right)\right]\\
&+E_{T_i^*|S^c}\left[\left\{E_{T_i^*|S^c}\left(T_i^*\left(\boldsymbol{U}\in S^c\right)\right)-\boldsymbol{\mu}\right\}^2 I\left(\boldsymbol{U}\in S^c\right)\right]\\
&+2E_{T_i^*|S^c}\left[\left\{T_i^*-E_{T_i^*|S^c}\left(T_i^* I\left(\boldsymbol{U}\in S^c\right)\right)\right\}\left\{E_{T_i^*|S^c}\left(T_i^*\left(\boldsymbol{U}\in S^c\right)\right)-\boldsymbol{\mu}\right\} I\left(\boldsymbol{U}\in S^c\right)\right]\\
&= E_{T_i^*|S^c}\left[\left\{T_i^*-E_{T_i^*|S^c}\left(T_i^* I\left(\boldsymbol{U}\in S^c\right)\right)\right\}^2 I\left(\boldsymbol{U}\in S^c\right)\right]\\
&+E_{T_i^*|S^c}\left[\left\{E_{T_i^*|S^c}\left(T_i^*\left(\boldsymbol{U}\in S^c\right)\right)-\boldsymbol{\mu}\right\}^2 I\left(\boldsymbol{U}\in S^c\right)\right].
\end{aligned}$$

Hence from (a) we have

$$E\left(T_1-\boldsymbol{\mu}\right)^2-E\left(T_2-\boldsymbol{\mu}\right)^2 = E_{S^c}\left\{E_{T_1^*|S^c}\left[\left(T_1^*-\boldsymbol{\mu}\right)^2 I\left(\boldsymbol{U}\in S^c\right)\right]-E_{T_2^*|S^c}\left[\left(T_2^*-\boldsymbol{\mu}\right)^2 I\left(\boldsymbol{U}\in S^c\right)\right]\right\}.$$

The result follows from (b).

□

For notational simplicity in the rest of this Appendix we will drop the subscripts from the expectation E. Let the $\boldsymbol{\Sigma}$ denote the covariance matrix of $\hat{\boldsymbol{\theta}}^{UMLE}=\left(\hat{\theta}_1^{UMLE},\hat{\theta}_2^{UMLE}\right)'$ and $\boldsymbol{\Sigma}_1$ be a diagonal matrix with the same diagonal elements as

$\boldsymbol{\Sigma}$. Recall that the UMLE, the RMLE, and the PAVA estimator of θ_1 under the constraint $\theta_1 \le \theta_2$ are of the form $\hat{\theta}_1^{UMLE} I(\hat{\theta}_1 < \hat{\theta}_2) + a_m' \hat{\theta}^{UMLE} I(\hat{\theta}_1 > \hat{\theta}_2)$, where $\boldsymbol{a}_{UMLE} = \dfrac{\boldsymbol{1}'\mathbf{V}}{\boldsymbol{1}'\mathbf{V}\boldsymbol{1}}$, $\boldsymbol{a}_{RMLE} = \dfrac{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}}{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}$, $\boldsymbol{a}_{PAVA} = \dfrac{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}}{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{1}}$, and $\mathbf{V} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Following notations of Lemma A2, $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2\right)$ corresponds to $\boldsymbol{U}$, $T_0\left(\hat{\boldsymbol{\theta}}\right) = \hat{\theta}_1$, $T_i^*(\boldsymbol{U}) = \boldsymbol{a}_m' \hat{\boldsymbol{\theta}}$, $S = \left\{\hat{\theta}_1 < \hat{\theta}_2\right\}$.

For a method *m*, let $\boldsymbol{Y} = (Y_{1m}, Y_2)' = \left(\boldsymbol{a}_m' \hat{\theta}^{UMLE}, \boldsymbol{b}' \hat{\theta}^{UMLE}\right)'$, then from Lemma A1 and $Var(Y_{1m} \mid Y_2) = \boldsymbol{a}_m' \boldsymbol{\Sigma} \boldsymbol{a}_m - \dfrac{(\boldsymbol{a}_m' \boldsymbol{\Sigma} \boldsymbol{b})^2}{\boldsymbol{b}' \boldsymbol{\Sigma} \boldsymbol{b}}$ we deduce the following lemma.

Lemma A3: $Var(Y_{1RMLE} \mid Y_2) = Var(Y_{1PAVA} \mid Y_2) = Var(Y_{1UMLE} \mid Y_2)$.

Proof of Theorem 2.1 (a):

There is no loss of generality in assuming $\theta_2 > 0$. If $\theta_2 < 0$ then one may perform a simple linear transformation $\hat{\theta}_i \to -\hat{\theta}_i$ and exploit the symmetry of a normal distribution and prove the domination theorem for estimating $-\theta_2$ under the constraint $-\theta_2 \le -\theta_1$. Thus in the following we assume $\theta_2 > 0$.

Note that

$$E\left(\hat{\theta}_2^{RMLE} \mid \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) = \frac{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}}{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} \text{ and } E\left(\hat{\theta}_2^{PAVA} \mid \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) = \frac{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\theta}}{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{1}} + \frac{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}\boldsymbol{b}}{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{1}\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}\boldsymbol{b}}\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} - \boldsymbol{b}'\boldsymbol{\theta}\right).$$

Therefore

$$\Delta = \left[E\left(\hat{\theta}_2^{PAVA} \mid \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) - \theta_2\right]^2 - \left[E\left(\hat{\theta}_2^{RMLE} \mid \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) - \theta_2\right]^2$$

$$= \left[\frac{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\theta}}{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{1}} + \frac{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}\boldsymbol{b}}{\boldsymbol{1}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{1}\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}\boldsymbol{b}}\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} - \boldsymbol{b}'\boldsymbol{\theta}\right) - \theta_2\right]^2 - \left[\frac{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\theta}{\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} - \theta_2\right]^2$$

The above expression can be simplified as

$$\Delta = \left[\frac{\sigma_1^2 \sigma_2}{\left(\sigma_1^2 + \sigma_1^2\right)^2 \left(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)^2} \right]$$
$$\times \left[\rho^2 \sigma_2 \left(\sigma_2^2 - \sigma_1^2\right)^2 \left(b'\hat{\theta} + \theta_2\right)^2 + 2\rho\left(\sigma_2^4 - \sigma_1^4\right)(\rho\sigma_2 - \sigma_1)(\theta_2 - \theta_1)\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} + \theta_2\right) \right]$$

Hence

$$E\left(\hat{\theta}_2^{PAVA} - \theta_2\right)^2 - E\left(\hat{\theta}_2^{RMLE} - \theta_2\right)^2 = E\left[\Delta \times I\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} > 0\right)\right]$$

$$= \left[\frac{\sigma_1^2 \sigma_2}{\left(\sigma_1^2 + \sigma_1^2\right)^2 \left(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2\right)^2} \right]$$
$$\times \left[\begin{array}{l} \rho^2 \sigma_2 \left(\sigma_2^2 - \sigma_1^2\right)^2 E\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} + \theta_2\right)^2 I\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} > 0\right) \\ +2\rho\left(\sigma_2^4 - \sigma_1^4\right)\left(\rho\sigma_2 - \sigma_1\right)\left(\theta_2 - \theta_1\right) E\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} + \theta_2\right) I\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} > 0\right) \end{array} \right]$$

Since $E\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} + \theta_2\right)^2 I\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} > 0\right) > 0$ and $E\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} + \theta_2\right) I\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}} > 0\right) > 0$, therefore the result follows from the sufficient conditions of the theorem.

□

Proof of Theorem 2.1 (b):

Similarly to the proof of Theorem 2.1 (a), it is sufficient to prove the theorem when $\theta_2 \geq \theta_1 \geq 0$; the proof in the case where $0 \geq \theta_2 \geq \theta_1$ follows by performing a simple linear transformation $\hat{\theta}_i \to -\hat{\theta}_i$ and by exploiting the symmetry of a normal distribution. As in the above proof, straightforward algebra results in

$$\Delta = \left[E\left(\hat{\theta}_2^{PAVA} \mid \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) - \theta_2 \right]^2 - \left[E\left(\hat{\theta}_2 \middle| \boldsymbol{b}'\hat{\boldsymbol{\theta}}\right) - \theta_2 \right]^2$$

$$= \frac{\sigma_1^2\left(\boldsymbol{b}'\hat{\boldsymbol{\theta}}\right)}{\left(\sigma_1^2+\sigma_2^2\right)^2\left(\sigma_1^2-2\rho\sigma_1\sigma_2\right)} \times$$

$$\left[\left(2\rho\sigma_1\sigma_2^3-\sigma_1^2\sigma_2^2-\sigma_2^4\right)\left(b'\hat{\theta}\right)+2\left(\theta_1+\theta_2\right)\sigma_1\sigma_2\left(\left(\sigma_1^2+\sigma_2^2\right)\rho-2\sigma_1\sigma_2\right)-2\sigma_1^4\theta_2-2\sigma_2^4\theta_1\right].$$

Since $\theta_2 \geq \theta_1 \geq 0$ and $\rho \leq 0$, then the above expression is negative. The result then follows by appealing to Lemma A2 and Lemma A3.

□

In the following $\boldsymbol{I} = (1,1,1...,1)'$, $\mathbf{J} = \mathbf{1}\mathbf{1}'$ and let $\mathbf{I}$ denote the identity matrix. The orders of the vectors and matrices would be apparent from the context.

Lemma A4: Suppose $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix defined as follows:

$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2\boldsymbol{I}' \\ \rho\sigma_1\sigma_2 1 & \sigma_2^2\mathbf{K}_1 \end{pmatrix}$, with $\mathbf{K}_1 = (1-\rho)\mathbf{I} + \rho\mathbf{J}$, $\rho \geq -1/(p-2)$. Then

$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \psi_1^2 & \rho_2\psi_1\psi_2 1' \\ \rho_2\psi_1\psi_2 1 & \psi_2^2 K_2 \end{pmatrix}$, where $\mathbf{K}_2 = (1-\rho_2)\mathbf{I} + \rho_2\mathbf{J}$, $\rho_2 = -\frac{\rho}{(p-2)\rho+1}$,

$k = \left(\frac{(p-2)\rho+1}{1-(p-2)\rho-(p-1)\rho^2}\right)^{1/2}$ and $\psi_i = \frac{k}{\sigma_i}$, $i = 1,2$.

Proof: Proof follows by verifying that $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \mathbf{I}$.

□

Lemma A5: Let $u = \psi_1^2 + (p-1)\rho_2\psi_1\psi_2$, $z = \psi_2^2 + (p-1)\rho_2\psi_1\psi_2 + (p-2)\rho_2\psi_2^2$ and $w_i = \sigma_i^{-2}, i = 1,2.$ Then

$u \geq 0$ if either $\rho < 0$ or $\rho > 0$, $\dfrac{\sigma_1}{\sigma_2} \leq \dfrac{(p-2)\rho + 1}{(p-1)\rho}$.

$u \leq 0$ if $\rho > 0, \dfrac{\sigma_1}{\sigma_2} \geq \dfrac{(p-2)\rho + 1}{(p-1)\rho}$

$z > 0$ if either $0 < \rho \leq \dfrac{\sigma_1}{\sigma_2}$ or $\rho < 0$

$z \leq 0$ if $\rho \geq \dfrac{\sigma_1}{\sigma_2}$

$u - \dfrac{w_1}{w_2} z \geq 0$ if either $\rho < 0, \dfrac{\sigma_1}{\sigma_2} \geq 1$ or $\rho > 0, \dfrac{\sigma_1}{\sigma_2} \leq 1$

$u - \dfrac{w_1}{w_2} z \leq 0$ if either $\rho < 0, \dfrac{\sigma_1}{\sigma_2} < 1$ or $\rho > 0, \dfrac{\sigma_1}{\sigma_2} > 1$

Proof:

Note that $\rho_2 \geq 0$ if and only if $\rho \leq 0$.

$u = \psi_1^2 + (p-1)\rho_2 \psi_1 \psi_2 = \dfrac{k^2}{\sigma_1^2} + (p-1)\dfrac{k^2 \rho_2}{\sigma_1 \sigma_2} = \dfrac{k^2}{\sigma_1^2}\left(1 + (p-1)\dfrac{\rho_2 \sigma_1}{\sigma_2}\right)$ is non-negative

whenever $\rho \leq 0$. Since $\rho_2 = -\dfrac{\rho}{(p-2)\rho + 1}$, then $u = \dfrac{k^2}{\sigma_1^2}\left(1 - (p-1)\dfrac{\rho \sigma_1}{(1 + (p-2)\rho)\sigma_2}\right)$,

which is non-negative if $\rho > 0$, $\dfrac{\sigma_1}{\sigma_2} \leq \dfrac{(p-2)\rho + 1}{(p-1)\rho}$ and is non-positive if $\rho > 0$ and

$\dfrac{\sigma_1}{\sigma_2} \geq \dfrac{(p-2)\rho + 1}{(p-1)\rho}$. Hence (a) and (b) are true. Note that

$$z = \psi_2^2 + \rho_2\psi_1\psi_2 + (p-2)\rho_2\psi_2^2 = \frac{k^2}{\sigma_2^2} + \frac{k^2\rho_2}{\sigma_1\sigma_2} + (p-2)\frac{k^2\rho_2}{\sigma_2^2} =$$

$$= \frac{k^2}{\sigma_2^2\sigma_1}\left(\sigma_1 + \rho_2\sigma_2 + (p-2)\rho_2\sigma_1\right) = \frac{k^2}{\sigma_2^2\sigma_1}\left(\sigma_1 + \rho_2\left(\sigma_2 + (p-2)\sigma_1\right)\right) =$$

$$= \frac{k^2}{\sigma_2^2\sigma_1}\left(\sigma_1 - \frac{\rho}{(p-2)\rho+1}\left(\sigma_2 + (p-2)\sigma_1\right)\right) =$$

$$= \frac{k^2}{\sigma_2^2\sigma_1}\left(\frac{\sigma_1\left((p-2)\rho+1\right) - \rho\left(\sigma_2 + (p-2)\sigma_1\right)}{(p-2)\rho+1}\right) =$$

$$= \frac{k^2}{\sigma_2^2\sigma_1}\left(\frac{\sigma_1(p-2)\rho + \sigma_1 - (p-2)\sigma_1\rho - \rho\sigma_2}{(p-2)\rho_1+1}\right) =$$

$$= \frac{k^2}{\sigma_2^2\sigma_1}\left(\frac{\sigma_1 - \rho\sigma_2}{(p-2)\rho+1}\right)$$

Hence $z \geq 0$ if either $\rho \leq 0$ or $0 \leq \rho \leq \frac{\sigma_1}{\sigma_2}$, and $z \leq 0$ if $\rho \geq \frac{\sigma_1}{\sigma_2}$, which proves (c) and (d). Similarly, it is straightforward to prove (e) and (f).

□

Lemma A6: Let $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p\right)' \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is not necessarily diagonal with variance $Var(\hat{\theta}_i) = \sigma_i^2 = w_i^{-1}$, $i = 1,2,..., p$. Then for any $c > 0$,

$$\frac{\partial}{\partial\theta_p} P\left(\left|\hat{\theta}_i^{PAVA(p)} - \theta_i\right| < c\right) = k_1 \int_{A_Y} \exp\left\{-\frac{1}{2}[R(-c) - R(c)]\right\} dY_{(1)},$$

where $Y = \left(Y_{(1)}', Y_p\right)' \sim N\left(-c\boldsymbol{1}, \boldsymbol{\Sigma}\right)$, $k_1 = (2\pi)^{-(p/2)}|\boldsymbol{\Sigma}|^{-1/2}$,

$$A_Y = \left\{\min_{i \leq t \leq p-1} \max_{1 \leq s \leq i} \sum_{j=s}^{t} w_j(Y_j + \theta_j) \geq 0\right\},\ R(c) = \left[\left(Y_{(1)}' : b\right) + c\boldsymbol{1}'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} + c\boldsymbol{1}\right] \text{ and}$$

$$b = -\max_{1 \leq s \leq i} \frac{\sum_{j=s}^{p-1} w_j\left(Y_j + \theta_j\right)}{w_p} - \theta_p .$$

Proof:

Note that

$$P\left(\left|\hat{\theta}_i^{PAVA(p)}-\theta_i\right|<c\right)=P\left(\hat{\theta}_i^{PAVA(p)}-\theta_i>-c\right)-P\left(\hat{\theta}_i^{PAVA(p)}-\theta_i>c\right)=Q(-c)-Q(c).$$

Hence in the following we first compute $\frac{\partial}{\partial\theta_p}Q(c)$ so that $\frac{\partial}{\partial\theta_p}P\left(\left|\hat{\theta}_i^{PAVA(p)}-\theta_i\right|<c\right)$

can be computed as $\frac{\partial}{\partial\theta_p}Q(-c)-\frac{\partial}{\partial\theta_p}Q(c).$

Denote $Ave_{\hat{\theta}}(s,t)=\frac{\sum_{j=s}^{t}w_j\hat{\theta}_j}{\sum_{j=s}^{t}w_j}$, then

$$\hat{\theta}_i^{PAVA(p)}=\min_{i\le t\le p}\ \max_{1\le s\le i}\frac{\sum_{j=s}^{t}w_j\hat{\theta}_j}{\sum_{j=s}^{t}w_j}=\min\left(\min_{i\le t\le p-1}\ \max_{1\le s\le i}Ave_{\hat{\theta}}(s,t),\ \max_{1\le s\le i}Ave_{\hat{\theta}}(s,p)\right)..$$

Performing a linear transformation $X^0=\left(\hat{\theta}_1-c,\ \hat{\theta}_2-c,...,\hat{\theta}_p-c\right)'$ we have

$X^0\sim N(\boldsymbol{\theta}-c\boldsymbol{1},\boldsymbol{\Sigma})$ and

$$\hat{\theta}_i^{PAVA(p)}-c=\min_{i\le t\le p}\ \max_{1\le s\le i}\frac{\sum_{j=s}^{t}w_j\hat{\theta}_j}{\sum_{j=s}^{t}w_j}-c$$

$$=\min\left(\min_{i\le t\le p-1}\ \max_{1\le s\le i}Ave_{X^0}(s,t),\ \max_{1\le s\le i}Ave_{X^0}(s,p)\right).$$

Let $S^{s,t}=\sum_{j=s}^{t}w_jX_j^0$ then $A_{\hat{\theta}}(s,p)\ge 0$ if and only if $X_p^0\ge-\max_{1\le s\le i}\frac{S^{s,p-1}}{w_p}$. Then we

have

$$Q(c) = P\left(\hat{\theta}_i^{PAVA(p)} - \theta_i > c\right)$$

$$= P\left\{\min\left(\min_{i\le t\le p-1} \max_{1\le s\le i} Ave_{X^0}(s,t),\ \max_{1\le s\le i} Ave_{X^0}(s,p)\right) \ge 0\right\}$$

$$= P\left\{\min_{i\le t\le p-1} \max_{1\le s\le i} Ave_{X^0}(s,t) \ge 0, \max_{1\le s\le i} Ave_{X^0}(s,p) \ge 0\right\}$$

$$= P\left\{\min_{i\le t\le p-1} \max_{1\le s\le i} Ave_{X^0}(s,t) \ge 0, X_p^0 \ge -\max_{1\le s\le i} \frac{S^{s,p-1}}{w_p}\right\}.$$

Performing another linear transformation $Y_j = X_j^0 - \theta_j,\ j = 1,2,...,\ p,$ we have

$Y \sim N(-c\mathbf{1}, \boldsymbol{\Sigma})$. Furthermore, without any loss we may assume that $\theta_i = 0$. Hence,

$$A_{X^0} = \left\{\min_{i\le t\le p-1} \max_{1\le s\le i} Ave_{X^0}(s,t) \ge 0\right\} = \left\{\min_{i\le t\le p-1} \max_{1\le s\le i} \sum_{j=s}^{t} w_j(Y_j + \theta_j) \ge 0\right\} = A_Y \text{ and}$$

$X_p^0 \ge -\max_{1\le s\le i} \frac{S^{s,p-1}}{w_p}$ is equivalent to $Y_p \ge -\max_{1\le s\le i} \frac{\sum_{j=s}^{p-1} w_j\left(Y_j + \theta_j\right)}{w_p} - \theta_p$.

Therefore we may express $Q(c)$ in terms of Y as follows:

$$Q(c) = P\left\{\min_{i\le t\le p-1} \max_{1\le s\le i} Ave_{X^0}(s,t) \ge 0, X_p^0 \ge -\max_{1\le s\le i} \frac{S^{s,p-1}}{w_p}\right\}$$

$$= P\left\{A_Y, Y_p \ge -\max_{1\le s\le i} \frac{\sum_{j=s}^{p-1} w_j(Y_j + \theta_j)}{w_p} - \theta_p\right\}$$

Since $Y \sim N(-c\mathbf{1}, \boldsymbol{\Sigma})$, then

$Q(c) = k_1 \int_{A_Y} \int_b \exp\left\{-\frac{1}{2}\left(Y_{(1)}' + c1' : Y_p + c\right)'\Sigma^{-1}\begin{pmatrix} Y_{(1)} + c1 \\ Y_p + c \end{pmatrix}\right\} dY_{(1)}$. Note that in the above

expression only b is a function of θ_p and hence using a fundamental theorem of calculus

we obtain the following

$$\frac{\partial}{\partial \theta_p} Q(c) = k_1 \int_{A_Y} \exp\left\{ -\frac{1}{2} \left(Y'_{(1)} + c1' : Y_p + c \right) \Sigma^{-1} \begin{pmatrix} Y_{(1)} + c1 \\ Y_p + c \end{pmatrix} \right\} dY_{(1)} = k_1 \int_{A_Y} \exp\left\{ -\frac{1}{2} R(c) \right\} dY_{(1)}.$$

Hence

$$\frac{\partial}{\partial \theta_p} P\left(\left| \hat{\theta}_i^{PAVA(p)} - \theta_i \right| < c \right) = k_1 \int_{A_Y} \left\{ \exp[-\frac{1}{2} R(-c)] - \exp[-\frac{1}{2} R(c)] \right\} dY_{(1)}.$$

□

Lemma A7: Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)' \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where Σ is not necessarily diagonal with variance $Var(\hat{\theta}_i) = \sigma_i^2 = w_i^{-1}$, $i = 1,2,..., p$. Then for any $c > 0$,

$$\frac{\partial}{\partial \theta_1} P\left(\left| \hat{\theta}_i^{PAVA(p)} - \theta_i \right| < c \right) = k_1 \int_{A_Y} \left\{ \exp[-\frac{1}{2} R(c)] - \exp[-\frac{1}{2} R(-c)] \right\} dY_{(1)},$$

where $Y = \left(Y_1, Y'_{(1)} \right)' \sim N\left(-c\boldsymbol{1}, \boldsymbol{\Sigma} \right)$, $k_1 = (2\pi)^{-(p/2)} |\boldsymbol{\Sigma}|^{-1/2}$,

$$A_Y = \left\{ \max_{2 \le s \le i} \min_{i \le t \le p} \sum_{j=s}^{t} w_j (Y_j + \theta_j) \ge 0 \right\}, \; R(c) = \left[\left(b, Y'_{(1)} \right) + c1' \right] \boldsymbol{\Sigma}^{-1} \left[\begin{pmatrix} b \\ Y_{(1)} \end{pmatrix} + c\boldsymbol{1} \right], \text{ and}$$

$$b = -\min_{i \le t \le p} \frac{\sum_{j=2}^{t} w_j \left(Y_j + \theta_j \right)}{w_1} - \theta_1.$$

Proof:

Proof follows along the same lines as the proof of Lemma A5 but it uses the fact that $\hat{\theta}_i^{PAVA(p)} = \max_{1 \le s \le i} \min_{i \le t \le p} \frac{\sum_{j=s}^{t} w_j \hat{\theta}_j}{\sum_{j=s}^{t} w_j}$ and that

$$P\left(\left| \hat{\theta}_i^{PAVA(p)} - \theta_i \right| < c \right) = P\left(\hat{\theta}_i^{PAVA(p)} - \theta_i < c \right) - P\left(\hat{\theta}_i^{PAVA(p)} - \theta_i < -c \right) = Q(c) - Q(-c).$$

Details of the proof are omitted as it follows exactly along the same lines as the proof of Lemma A5.

□

Proof of Theorem 2.2 (a):

The proof is deduced as follows. We demonstrate that $P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right)$ is a non-increasing function of θ_p. Since $\lim_{\theta_p \to \infty} \hat{\theta}_1^{PAVA(p)} = \hat{\theta}_1^{PAVA(p-1)}$, we therefore conclude that

$$P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right) \geq P\left(\left|\hat{\theta}_1^{PAVA(p-1)} - \theta_1\right| < c\right).$$

Following same set of arguments inductively, we then have

$$P\left(\left|\hat{\theta}_1^{PAVA(p)} - \theta_1\right| < c\right) \geq P\left(\left|\hat{\theta}_1^{PAVA(p-1)} - \theta_1\right| < c\right) \geq P\left(\left|\hat{\theta}_1^{PAVA(p-2)} - \theta_1\right| < c\right)$$
$$\geq \ldots \geq P\left(\left|\hat{\theta}_1^{UMLE} - \theta_1\right| < c\right),$$

proving the theorem. Thus to prove the theorem it is sufficient to demonstrate that:

$$\frac{\partial}{\partial \theta_p} P\left(\mid \hat{\theta}_1^{PAVA(p)} - \theta_1 \mid < c\right) < 0.$$

From Lemma A6, taking $\theta_1 = 0$, without loss of generality, we note that for any $c > 0$,

$$\frac{\partial}{\partial \theta_p} P\left(\left|\hat{\theta}_i^{PAVA(p)} - \theta_i\right| < c\right) = k_1 \int_{A_Y} \left\{\exp[-\frac{1}{2}R(-c)] - \exp[-\frac{1}{2}R(c)]\right\} dY_{(1)},$$

where $Y = \left(Y'_{(1)}, Y_p\right)' \sim N\left(-c\boldsymbol{1}, \boldsymbol{\Sigma}\right)$, $k_1 = (2\pi)^{-(p/2)}|\boldsymbol{\Sigma}|^{-1/2}$, $A_Y = \left\{\min_{1\le t\le p-1} \sum_{j=1}^{t} w_j\left(Y_j + \theta_j\right) \ge 0\right\}$,

$R(c) = \left[\left(Y'_{(1)} : b\right) + c\boldsymbol{1}'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} + c1\right]$, and $b = -\dfrac{\sum_{j=1}^{p-1} w_j\left(Y_j + \theta_j\right)}{w_p} - \theta_p$.

Hence it is sufficient to demonstrate that $R(-c) > R(c)$, which is equivalent to demonstrating the following inequality

$$\left[\left(Y'_{(1)} : b\right) - c1'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} - c1\right] - \left[\left(Y'_{(1)} : b\right) + c1'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} + c\boldsymbol{1}\right]$$
$$= -4c\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} > 0.$$

Equivalently, we need to demonstrate that $\boldsymbol{1}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} < 0.$

From Lemma A4 note that

$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \psi_1^2 & \rho_2\psi_1\psi_2 1' \\ \rho_2\psi_1\psi_2 1 & \psi_2^2 K_2 \end{pmatrix}$, where $\mathbf{K}_2 = (1-\rho_2)\mathbf{I} + \rho_2\mathbf{J}$, $\rho_2 = -\dfrac{\rho}{(p-2)\rho + 1}$,

$k = \left(\dfrac{(p-2)\rho + 1}{1-(p-2)\rho - (p-1)\rho^2}\right)^{1/2}$ and $\psi_i = \dfrac{k}{\sigma_i}$, $i = 1, 2$.

Hence $\boldsymbol{\Sigma}^{-1}\boldsymbol{1} = (u, z, z, ..., z)'$, where $u = \psi_1^2 + (p-1)\rho_2\psi_1\psi_2$ and

$z = \psi_2^2 + \rho_2\psi_1\psi_2 + (p-2)\rho_2\psi_2^2$.

Consequently,

$$\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} = uY_1 + z\left(Y_2 + Y_3 + ... + Y_{p-1} + b\right) = uY_1 + z\left(Y_2 + Y_3 + ... + Y_{p-1} - \frac{\sum_{j=1}^{p-1} w_j\left(Y_j + \theta_j\right)}{w_p} - \theta_p\right).$$

Since $w_2 = w_3 = ... = w_p$, the above expression simplifies to

$\boldsymbol{I}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} = Y_1(u - w_1 z / w_2) - z\sum_2^p \theta_i$. Since the region of the above integral is

$A_Y = \left\{ \min_{1 \le t \le p-1} \sum_{j=1}^{t} w_j (Y_j + \theta_j) \ge 0 \right\}$ therefore $Y_1 \ge 0$. Also, since each $\theta_i \ge \theta_1 = 0, i \ge 2,$

therefore to prove $\boldsymbol{I}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} < 0$ it is sufficient to verify that $z > 0$ and $u - \frac{w_1}{w_2} z < 0$.

From Lemma A5(c, f) we note that under the sufficient conditions of the theorem, $z > 0$

and $u - \frac{w_1}{w_2} z < 0$. Hence the proof of part (a).

Proof of Theorem 2.2 (b): Proof of part (b) follows along the same lines as the proof of part (a) and is hence omitted.

□

Proof of Theorem 2.3:

The proof follows exactly along the same lines as the proof of Theorem 2.2. Thus to prove the theorem it is sufficient to demonstrate that:

$$\frac{\partial}{\partial \theta_p} P\left(\mid \hat{\theta}_1^{PAVA(p)} - \theta_1 \mid < c \right) < 0.$$

From Lemma A6, taking $\theta_1 = 0$, without loss of generality, we note that for any

$c > 0$, $\frac{\partial}{\partial \theta_p} P\left(\left| \hat{\theta}_i^{PAVA(p)} - \theta_i \right| < c \right) = k_1 \int_{A_Y} \left\{ \exp\left[-\frac{1}{2} R(-c) \right] - \exp\left[-\frac{1}{2} R(c) \right] \right\} dY_{(1)}$,

where $Y = \left(Y'_{(1)}, Y_p\right)' \sim N\left(-c\mathbf{1}, \boldsymbol{\Sigma}\right)$, $k_1 = (2\pi)^{-(p/2)}|\boldsymbol{\Sigma}|^{-1/2}$, $A_Y = \left\{\min_{1\le t\le p-1} \sum_{j=1}^{t} w_j\left(Y_j + \theta_j\right) \ge 0\right\}$,

$R(c) = \left[(Y'_{(1)} : b) + c\mathbf{1}\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} + c\mathbf{1}\right]$ and $b = -\dfrac{\sum_{j=1}^{p-1} w_j\left(Y_j + \theta_j\right)}{w_p} - \theta_p$.

Hence it is sufficient to demonstrate that $R(-c) > R(c)$, which is equivalent to demonstrating the following inequality

$$\left[\left(Y'_{(1)} : b\right) - c\mathbf{1}'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} - c1\right] - \left[\left(Y'_{(1)} : b\right) + c\mathbf{1}'\right]\boldsymbol{\Sigma}^{-1}\left[\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} + c\mathbf{1}\right] = -4c\mathbf{1}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} > 0.$$

Since $\mathbf{1}'\boldsymbol{\Sigma}^{-1} = \left(0, 0, ..., 0, p\right)'$, therefore $\mathbf{1}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix} Y_{(1)} \\ b \end{pmatrix} = pb$. Hence it is sufficient to prove that $b < 0$. Note that $b = -\dfrac{\sum_{j=1}^{p-1} w_j(Y_j + \theta_j)}{w_p} - \theta_p$ and in the region

$A_Y = \left\{\min_{1\le t\le p-1} \sum_{j=1}^{t} w_j(Y_j + \theta_j) \ge 0\right\}$ it is negative since $\theta_p > 0$. Thus proving the theorem.

□

APPENDIX B

Proofs of Chapter 3

Proof of Theorem 3.1:

$\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\tau}^2$ are given, we need to estimate $\boldsymbol{\sigma}^2$.

The complete data are defined as $\mathcal{y} = [\boldsymbol{Y}' : \boldsymbol{\varepsilon}']'$. $\mathcal{y} \sim N(\boldsymbol{\mu}_{\mathcal{y}}, \boldsymbol{\Xi})$, where

$$\mu_{\mathcal{y}} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix}, \ \boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma} & \boldsymbol{\Sigma} \end{pmatrix} \text{ and } \boldsymbol{\Psi} = \mathbf{UTU}' + \boldsymbol{\Sigma}.$$

The likelihood function is

$$f\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{y}, \sigma^2\right) = (2\pi)^{-(N+N)/2} \left|\boldsymbol{\Xi}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mathcal{y} - \boldsymbol{\mu}_{\mathcal{y}}\right)' \boldsymbol{\Xi}^{-1} \left(\mathcal{y} - \boldsymbol{\mu}_{\mathcal{y}}\right)\right\}.$$

Since $\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{vmatrix} = |\mathbf{C}|\left|\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\right|$ (Searle et al., 1992),

$$|\boldsymbol{\Xi}| = |\boldsymbol{\Sigma}|\left|\boldsymbol{\Psi} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right| = |\boldsymbol{\Sigma}||\boldsymbol{\Psi} - \boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}||\mathbf{UTU}'| = \left(\prod_{i=1}^{k} \sigma_i^{2n_i}\right)|\mathbf{UTU}'|.$$

Since $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{C}^{-1}\mathbf{B}' \end{pmatrix}\left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\right)^{-1}\left(\mathbf{I}_N, -\mathbf{B}\mathbf{C}^{-1}\right)$ (Searle et al., 1992),

$$\boldsymbol{\Xi}^{-1} = \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma} & \boldsymbol{\Sigma} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{I}_N \end{pmatrix}\left(\boldsymbol{\Psi} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right)^{-1}\left(\mathbf{I}_N, -\mathbf{I}_N\right) =$$

$$= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{I}_N \end{pmatrix}\left(\boldsymbol{\Psi} - \boldsymbol{\Sigma}\right)^{-1}\left(\mathbf{I}_N, -\mathbf{I}_N\right) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{I}_N \end{pmatrix}\left(\mathbf{UTU}'\right)^{-1}\left(\mathbf{I}_N, -\mathbf{I}_N\right).$$

Since $\mathcal{Y} = [\boldsymbol{Y}' : \boldsymbol{\varepsilon}']'$ and $\mu_{\mathcal{Y}} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix}$, thus

$$\mathcal{Y} - \mu_{\mathcal{Y}} = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{\varepsilon} \end{pmatrix} - \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\varepsilon} \end{pmatrix}.$$

$$\begin{aligned}
&\left(\mathcal{Y} - \boldsymbol{\mu}_{\mathcal{Y}}\right)' \Xi^{-1} \left(\mathcal{Y} - \boldsymbol{\mu}_{\mathcal{Y}}\right) = \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\varepsilon}' \right) \Xi^{-1} \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\varepsilon} \end{pmatrix} \\
&= \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\varepsilon}' \right) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\varepsilon} \end{pmatrix} \\
&+ \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\varepsilon}' \right) \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{I}_N \end{pmatrix} \left(\mathbf{UTU}'\right)^{-1} \left(\mathbf{I}_N, -\mathbf{I}_N\right) \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\varepsilon} \end{pmatrix}) \\
&= \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right)' \left(\mathbf{UTU}'\right)^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right) \\
&= \sum_{i=1}^{k} \sigma_i^{-2} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right)' \left(\mathbf{UTU}'\right)^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right).
\end{aligned}$$

Simplified likelihood function is the following:

$$f\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{Y}, \sigma^2\right) = \left(2\pi\right)^{-N} \left[\left(\prod_{i=1}^{k} \sigma_i^{2n_i} \right) \left|\mathbf{UTU}'\right| \right]^{-1/2} \times$$

$$\times \exp\left\{ -\frac{1}{2} \left[\sum_{i=1}^{k} \sigma_i^{-2} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right)' \left(\mathbf{UTU}'\right)^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right) \right] \right\}.$$

The log likelihood is the following:

$$\begin{aligned}
l\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{Y}, \sigma^2\right) &= -N \ln\left(2\pi\right) - (1/2) \left(\sum_{i=1}^{k} n_i \ln \sigma_i^2 \right) - (1/2) \ln \left|\mathbf{UTU}'\right| \\
&- \frac{1}{2} \sum_{i=1}^{k} \sigma_i^{-2} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i - \frac{1}{2} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right)' \left(\mathbf{UTU}'\right)^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \boldsymbol{\varepsilon}\right).
\end{aligned}$$

The M (maximization) step of the EM algorithm finds maximum likelihood estimators of the unknown parameters as if the complete data were available.

$$\frac{\partial l}{\partial \sigma_i^2} = -(1/2) n_i \sigma_i^{-2} + \frac{1}{2} \sigma_i^{-4} \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i, \; i = 1, \ldots, k.$$

By setting the derivative to be equal to 0 we obtain:

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i,\ i = 1,\ldots,k.$$

The E (expectation) step of the EM algorithm replaces the unknown quantities $\boldsymbol{\varepsilon}_i$ with their expected values conditioned on the known values.

Since $\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Psi} & \sigma_i^2\boldsymbol{\Gamma}_i \\ \sigma_i^2\boldsymbol{\Gamma}_i' & \sigma_i^2\mathbf{I}_{n_i} \end{pmatrix}\right)$,

where $\boldsymbol{\Gamma}_i = \left[\mathbf{0}:\cdots:\mathbf{0}:\mathbf{I}_{n_i}:\mathbf{0}:\ldots:\mathbf{0}\right]'$, thus

$$E\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right) = \sigma_i^2\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right),$$

$$Cov\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right) = \sigma_i^2\mathbf{I}_{n_i} - \sigma_i^4\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i,\ i = 1,\ldots,k.$$

Since $E\left(\boldsymbol{y}'\mathbf{A}\boldsymbol{y}\right) = \left(E\left(\boldsymbol{y}\right)\right)'\mathbf{A}E\left(\boldsymbol{y}\right) + tr\left(\mathbf{A}Cov\left(\boldsymbol{y}\right)\right)$ and $\boldsymbol{y}'\boldsymbol{y} = tr\left(\boldsymbol{y}\boldsymbol{y}'\right)$,

$$\begin{aligned}
&E\left(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right) = E\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right)' E\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right) + tr\left(Cov\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{Y}\right)\right) \\
&= \sigma_i^4\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right) + tr\left(\sigma_i^2\mathbf{I}_{n_i} - \sigma_i^4\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\right) \\
&= \sigma_i^4\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right) + n_i\sigma_i^2 - \sigma_i^4 tr\left(\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\right) \\
&= \sigma_i^4 tr\left[\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\right] + n_i\sigma_i^2 - \sigma_i^4 tr\left(\boldsymbol{\Gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}_i\right) \\
&= n_i\sigma_i^2 + \sigma_i^4 tr\left[\boldsymbol{\Gamma}_i'\left\{\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\right\}\boldsymbol{\Gamma}_i\right] \\
&= n_i\sigma_i^2 + \sigma_i^4 tr\left[\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\right]_{ii},
\end{aligned}$$

where A_{ii} indicates the (i,i)th block of A.

Thus

$$\begin{aligned}
\hat{\sigma}_i^{2(r)} = \sigma_i^{2(r-1)} + \frac{1}{n_i}\hat{\sigma}_i^{4(r-1)} tr\Big[&\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1} \\
&- \left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\Big]_{ii},\ i = 1,\ldots,k,
\end{aligned}$$

where $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \boldsymbol{\Sigma}$ and $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2}\mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2}\mathbf{I}_{c_q}\right)$.

□

Proof of Theorem 3.2:

σ is given, we need to estimate $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, τ_i^2 with order restrictions on $\boldsymbol{\theta}_1$.

The complete data are defined as $\mathcal{y} = \left[\boldsymbol{Y}' : \boldsymbol{\xi}'\right]'$. $\mathcal{y} \sim N\left(\boldsymbol{\mu}_{\mathcal{y}}, \boldsymbol{\Xi}\right)$,

where

$$\mu_{\mathcal{y}} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix},\ \boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{UT} \\ \mathbf{TU}' & \mathbf{T} \end{pmatrix} \text{ and } \boldsymbol{\Psi} = \mathbf{UTU}' + \boldsymbol{\Sigma}.$$

The likelihood function is

$$f\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{y}, \sigma^2\right) = (2\pi)^{-(N+c)/2}\left|\boldsymbol{\Xi}\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathcal{y} - \boldsymbol{\mu}_{\mathcal{y}}\right)'\boldsymbol{\Xi}^{-1}\left(\mathcal{y} - \boldsymbol{\mu}_{\mathcal{y}}\right)\right\}. \tag{30}$$

Since $\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{vmatrix} = \left|\mathbf{C}\right|\left|\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}'\right|$ (Searle et al., 1992),

$$\left|\boldsymbol{\Xi}\right| = \left|\mathbf{T}\right|\left|\boldsymbol{\Psi} - \mathbf{UTT}^{-1}\mathbf{TU}'\right| = \left|\mathbf{T}\right|\left|\boldsymbol{\Psi} - \mathbf{UTU}'\right| = \left|\mathbf{T}\right|\left|\boldsymbol{\Sigma}\right| = \left(\prod_{i=1}^{q}\tau_i^{2c_i}\right)\left|\boldsymbol{\Sigma}\right|.$$

Since $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{C}^{-1}\mathbf{B}' \end{pmatrix}\left(\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}'\right)^{-1}\left(\mathbf{I}_N, -\mathbf{BC}^{-1}\right)$ (Searle et al., 1992),

$$\boldsymbol{\Xi}^{-1} = \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{UT} \\ \mathbf{TU}' & \mathbf{T} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{T}^{-1}\mathbf{TU}' \end{pmatrix}\left(\boldsymbol{\Psi} - \mathbf{UTT}^{-1}\mathbf{TU}'\right)^{-1}\left(\mathbf{I}_N, -\mathbf{UTT}^{-1}\right) =$$

$$= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{U}' \end{pmatrix}\left(\boldsymbol{\Psi} - \mathbf{UTU}'\right)^{-1}\left(\mathbf{I}_N - \mathbf{U}\right) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{U}' \end{pmatrix}\boldsymbol{\Sigma}^{-1}\left(\mathbf{I}_N, -\mathbf{U}\right).$$

Since $\mathcal{Y} = [\boldsymbol{Y}' : \boldsymbol{\xi}']'$ and $\mu_{\mathcal{Y}} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix}$, thus

$$\mathcal{Y} - \mu_{\mathcal{Y}} = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{\xi} \end{pmatrix} - \begin{pmatrix} \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\xi} \end{pmatrix}.$$

$$\begin{aligned}
&\left(\mathcal{Y} - \boldsymbol{\mu}_{\mathcal{Y}}\right)' \boldsymbol{\Xi}^{-1} \left(\mathcal{Y} - \boldsymbol{\mu}_{\mathcal{Y}}\right) = \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\xi}' \right) \boldsymbol{\Xi}^{-1} \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\xi} \end{pmatrix} \\
&= \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\xi}' \right) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\xi} \end{pmatrix} \\
&+ \left(\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right)' : \boldsymbol{\xi}' \right) \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{U}' \end{pmatrix} \boldsymbol{\Sigma}^{-1} \left(\mathbf{I}_N, -\mathbf{U}\right) \begin{pmatrix} \boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 \\ \boldsymbol{\xi} \end{pmatrix} \\
&= \boldsymbol{\xi}'\mathbf{T}^{-1}\boldsymbol{\xi} + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right) \\
&= \sum_{i=1}^{q} \tau_i^{-2} \boldsymbol{\xi}_i'\boldsymbol{\xi}_i + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right).
\end{aligned}$$

Simplified likelihood function is the following:

$$f\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{Y}, \sigma^2\right) = (2\pi)^{-(N+c)/2} \left[\left(\prod_{i=1}^{q} \tau_i^{2c_i} \right) |\boldsymbol{\Sigma}| \right]^{-1/2} \times$$

$$\times \exp\left\{ -\frac{1}{2} \left[\sum_{i=1}^{q} \tau_i^{-2} \boldsymbol{\xi}_i'\boldsymbol{\xi}_i + \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right) \right] \right\}.$$

The log likelihood is the following:

$$\begin{aligned}
l\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}^2; \mathcal{Y}, \sigma^2\right) &= \left[-\left(N + c\right)/2 \right] \ln\left(2\pi\right) - (1/2) \left(\sum_{i=1}^{q} c_i \ln \tau_i^2 \right) - (1/2) \ln |\boldsymbol{\Sigma}| \\
&- \frac{1}{2} \sum_{i=1}^{q} \tau_i^{-2} \boldsymbol{\xi}_i'\boldsymbol{\xi}_i - \frac{1}{2} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right).
\end{aligned}$$

The M (maximization) step of the EM algorithm finds maximum likelihood estimators of the unknown parameters as if the complete data were available.

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\theta}_1} &= -\frac{1}{2} \left[-2\mathbf{X}_1'\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right) + 2\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\boldsymbol{\theta}_1 \right] \\
&= \mathbf{X}_1'\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right) - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\boldsymbol{\theta}_1.
\end{aligned}$$

$$\frac{\partial l}{\partial \boldsymbol{\theta}_2} = \mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{U}\boldsymbol{\xi}\right) - \mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\boldsymbol{\theta}_2.$$

$$\frac{\partial l}{\partial \tau_i} = -\frac{1}{2}c_i\tau_i^{-2} + \frac{1}{2}\tau_i^{-4}\xi_i'\xi_i$$

Derivatives are set to be equal to 0:

$$\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\boldsymbol{\theta}_1 = \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\boldsymbol{\theta}_2 = \mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$c_i\tau_i^{-2} = \tau_i^{-4}\xi_i'\xi_i$$

Thus

$$\hat{\boldsymbol{\theta}}_1 = \left(\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$\hat{\boldsymbol{\theta}}_2 = \left(\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$\hat{\boldsymbol{\theta}}_1 = \left(\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\right)^{+}\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$\hat{\boldsymbol{\theta}}_2 = \left(\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\right)^{+}\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{U}\boldsymbol{\xi}\right)$$

$$\hat{\tau}_i^2 = \frac{1}{c_i}\xi_i'\xi_i,\ i = 1,\ldots,q.$$

The E (expectation) step of the EM algorithm replaces the unknown quantities ξ_i with their expected values conditioned on the known values.

Since $\begin{pmatrix}\boldsymbol{Y}\\ \xi_i\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2\\ 0\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Psi} & \tau_i^2\mathbf{U}_i\\ \tau_i^2\mathbf{U}_i & \tau_i^2\mathbf{I}_{c_i}\end{pmatrix}\right)$, thus

$$E\left(\xi_i \mid \boldsymbol{Y}\right) = \tau_i^2\mathbf{U}_i'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2\right),$$

$$Cov\left(\xi_i \mid \boldsymbol{Y}\right) = \tau_i^2\mathbf{I}_{c_i} - \tau_i^4\mathbf{U}_i'\boldsymbol{\Psi}^{-1}\mathbf{U}_i$$

$$E(\boldsymbol{\xi} \mid \boldsymbol{Y}) = \mathbf{T}\mathbf{U}'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)$$

Since $E(\boldsymbol{y}'\mathbf{A}\boldsymbol{y}) = (E(\boldsymbol{y}))' \mathbf{A}E(\boldsymbol{y}) + tr(\mathbf{A}Cov(\boldsymbol{y}))$ and $\boldsymbol{y}'\boldsymbol{y} = tr(\boldsymbol{y}\boldsymbol{y}')$,

$$E(\boldsymbol{\xi}_i'\boldsymbol{\xi}_i \mid \boldsymbol{Y}) = E(\boldsymbol{\xi}_i \mid \boldsymbol{Y})' E(\boldsymbol{\xi}_i \mid \boldsymbol{Y}) + tr(Cov(\boldsymbol{\xi}_i \mid \boldsymbol{Y}))$$
$$= \tau_i^4 (\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)' \boldsymbol{\Psi}^{-1}\mathbf{U}_i\mathbf{U}_i'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2) + tr(\tau_i^2\mathbf{I}_{c_i} - \tau_i^4\mathbf{U}_i'\boldsymbol{\Psi}^{-1}\mathbf{U}_i)$$
$$= \tau_i^4 tr\left\{\mathbf{U}_i'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)' \boldsymbol{\Psi}^{-1}\mathbf{U}_i\right\} + c_i\tau_i^2 - \tau_i^4 tr(\mathbf{U}_i'\boldsymbol{\Psi}^{-1}\mathbf{U}_i)$$
$$= c_i\tau_i^2 + \tau_i^4 tr\left\{\mathbf{U}_i'\left[\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)' \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\right]\mathbf{U}_i\right\}.$$

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\mathbf{T}\mathbf{U}'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2))$$
$$= (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{U}\mathbf{T}\mathbf{U}'(\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)\right)$$
$$= (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\left\{\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\boldsymbol{\theta}_2 - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left((\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma} - \boldsymbol{\Sigma})(\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)\right)\right\}$$
$$= (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\left\{\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\boldsymbol{\theta}_2 + \mathbf{X}_1'(\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)\right.$$

$$\left. - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\left((\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma})(\mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)\right)\right\}$$
$$= (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\left\{\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2\boldsymbol{\theta}_2 + \mathbf{X}_1'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2) - \mathbf{X}_1'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)\right\}$$
$$= (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\left\{\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1\boldsymbol{\theta}_1\right\} + (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2)$$
$$= \boldsymbol{\theta}_1 + (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \mathbf{X}_1\boldsymbol{\theta}_1 - \mathbf{X}_2\boldsymbol{\theta}_2).$$

$$\hat{\tau}_i^{2(r)} = \hat{\tau}_i^2 + \frac{\hat{\tau}_i^4}{c_i} tr\left\{\mathbf{U}'_i\left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right)'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right.\right.$$
$$\left.\left. - \left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\right]\mathbf{U}_i\right\}, \quad i = 1, \ldots, q,$$

Thus

$$\hat{\boldsymbol{\theta}}_1^{(r)} = \hat{\boldsymbol{\theta}}_1^{(r-1)} + (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right),$$

$$\hat{\boldsymbol{\theta}}_2^{(r)} = \hat{\boldsymbol{\theta}}_2^{(r-1)} + (\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\left(\hat{\boldsymbol{\Psi}}^{(r-1)}\right)^{-1}\left(\boldsymbol{Y} - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2^{(r-1)}\right),$$

$$\hat{\tau}_i^{2(r)} = \hat{\tau}_i^2 + \frac{\hat{\tau}_i^4}{c_i} tr\left\{ \mathbf{U}'_i \left[\left(\hat{\boldsymbol{\Psi}}^{(r-1)} \right)^{-1} \left(\boldsymbol{Y} - \mathbf{X}_1 \hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2 \hat{\boldsymbol{\theta}}_2^{(r-1)} \right) \left(\boldsymbol{Y} - \mathbf{X}_1 \hat{\boldsymbol{\theta}}_1^{(r-1)} - \mathbf{X}_2 \hat{\boldsymbol{\theta}}_2^{(r-1)} \right)' \left(\hat{\boldsymbol{\Psi}}^{(r-1)} \right)^{-1} \right.\right.$$
$$\left.\left. - \left(\hat{\boldsymbol{\Psi}}^{(r-1)} \right)^{-1} \right] \mathbf{U}_i \right\}, \quad i = 1, \ldots, q,$$

where $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \boldsymbol{\Sigma}$ and $\hat{\mathbf{T}}^{(r-1)} = diag\left(\tau_1^{(r-1)2} \mathbf{I}_{c_1} : \ldots : \tau_q^{(r-1)2} \mathbf{I}_{c_q} \right)$.

□

Proof of Theorem 3.3:

Under the regularity conditions R1 to R4, from Theorem 10.2.3 in Rao and Kleffe (1988) we deduce that the MINQUE $\hat{\boldsymbol{\phi}}$ is consistent for $\boldsymbol{\phi}$. Appealing to Noether's conditions (R5 and R6) we deduce the asymptotic normality of $\mathbf{A}\hat{\boldsymbol{\theta}}$ from the discussion in Chapter 10.7 in Rao and Kleffe (1988).

□

APPENDIX C

Complete simulation results

Table 7.1. Abbreviations for tests

Test abbreviation	Test
Hm1	Homoscedastic unrestricted F-test
Hm2	Homoscedastic unrestricted F-test (linear regression)
Hm3	Homoscedastic LR parametric bootstrap
Hm4	Homoscedastic parametric bootstrap F-test
Hm5	Homoscedastic parametric bootstrap F-test (linear regression)
Hm6	Homoscedastic LR non-parametric bootstrap
Hm7	Homoscedastic non-parametric bootstrap F-test
Hm8	Homoscedastic non-parametric bootstrap F-test (linear regression)
Hm9	Homoscedastic asymptotic LR test
Hm10	Homoscedastic MINQUE-based Williams non-parametric bootstrap
Ht1	Heteroscedastic unrestricted F-test
Ht2	Heteroscedastic unrestricted F-test (linear regression)
Ht3	Heteroscedastic LR parametric bootstrap
Ht4	Heteroscedastic parametric bootstrap F-test
Ht5	Heteroscedastic parametric bootstrap F-test (linear regression)
Ht6	Heteroscedastic LR non-parametric bootstrap
Ht7	Heteroscedastic non-parametric bootstrap F-test
Ht8	Heteroscedastic non-parametric bootstrap F-test (linear regression)
Ht9	Heteroscedastic asymptotic LR test
Ht10	Heteroscedastic MINQUE-based Williams non-parametric bootstrap

Table 7.2. Type I errors for homoscedastic normally distributed data (section 4.1.2).

p	subj	τ^2	Hm6	Hm9	Hm10	Ht6	Ht9	Ht10
3	10	1	0.04	0.01	0.04	0.05	0.03	0.05
3	50	1	0.04	0.01	0.03	0.04	0.01	0.03
3	10	0.2	0.04	0.02	0.04	0.05	0.05	0.04
3	10	2	0.04	0.01	0.04	0.05	0.04	0.05
3	50	2	0.04	0.01	0.03	0.04	0.01	0.03
5	10	1	0.04	0.01	0.05	0.05	0.02	0.03
5	10	0.2	0.05	0.02	0.05	0.04	0.02	0.04
5	10	2	0.05	0.01	0.05	0.05	0.02	0.04

Table 7.3. Power for homoscedastic normally distributed data (section 4.1.2).

p	subj	θ_1									τ^2	Hm6	Hm9	Hm10	Ht6	Ht9	Ht10
3	10	0	0.00	1.25							1	0.90	0.83	0.88	0.81	0.82	0.84
3	10	0	1.26	1.26							1	0.89	0.85	0.89	0.82	0.82	0.84
3	50	0	0.55	0.55							1	0.90	0.84	0.90	0.89	0.85	0.89
3	10	0	0.73	1.45							1	0.90	0.87	0.92	0.86	0.86	0.89
5	10	0	0.00	0.00	0.00	1.27					1	0.89	0.70	0.90	0.78	0.65	0.86
5	10	0	1.24	1.24	1.24	1.24					1	0.89	0.71	0.90	0.74	0.58	0.86
5	10	0	0.37	0.74	1.11	1.48					1	0.94	0.80	0.94	0.88	0.80	0.90
5	10	0	0.81	0.81	0.81	1.62					1	0.92	0.76	0.95	0.83	0.74	0.93
10	10	0	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1	0.94	0.51	0.96	0.75	0.24	0.92
10	10	0	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	1.72	1	0.95	0.51	0.97	0.84	0.40	0.95

Table 7.4. Type I errors for heteroscedastic normally distributed data (section 4.1.3).

p	subj	σ^2					τ^2	Ht6	Ht9	Ht10	Hm6	Hm9	Hm10
3	10	0.1	0.10	2.37			1	0.04	0.04	0.03	0.07	0.06	0.07
3	10	0.1	0.20	0.20			1	0.03	0.03	0.04	0.04	0.02	0.03
3	10	0.1	0.09	0.36			1	0.04	0.03	0.03	0.05	0.04	0.05
3	50	0.1	0.10	0.01			1	0.04	0.01	0.03	0.03	0.02	0.03
3	50	0.1	0.02	0.02			1	0.03	0.02	0.04	0.06	0.03	0.06
5	10	0.1	0.10	0.10	0.10	0.16	1	0.05	0.01	0.04	0.04	0.01	0.05
5	10	0.1	0.20	0.20	0.20	0.20	1	0.05	0.01	0.04	0.02	0.00	0.04
5	10	0.1	0.11	0.44	0.99	1.76	1	0.04	0.01	0.04	0.07	0.03	0.07
5	10	0.1	0.11	0.11	0.11	0.45	1	0.05	0.02	0.05	0.07	0.02	0.08

Table 7.5. Power for heteroscedastic normally distributed data (section 4.1.3).

p	subj	θ_1					τ^2	Ht6	Ht9	Ht10	Hm6	Hm9	Hm10
3	10	0	0.00	1.54			1	0.80	0.82	0.88	0.95	0.93	0.95
3	10	0	0.45	0.45			1	0.84	0.81	0.82	0.83	0.75	0.83
3	10	0	0.30	0.60			1	0.84	0.82	0.80	0.86	0.79	0.87
3	50	0	0.00	0.10			1	0.81	0.70	0.74	0.68	0.54	0.69
3	50	0	0.15	0.15			1	0.90	0.86	0.92	0.95	0.92	0.95
3	50	0	0.08	0.16			1	0.97	0.95	0.93	0.96	0.92	0.97
5	10	0	0.00	0.00	0.00	0.40	1	0.66	0.42	0.78	0.86	0.67	0.87
5	10	0	0.45	0.45	0.45	0.45	1	0.83	0.68	0.81	0.83	0.55	0.85
5	10	0	0.33	0.66	1.00	1.33	1	0.98	0.96	0.88	0.96	0.87	0.96
5	10	0	0.34	0.34	0.34	0.67	1	0.81	0.71	0.82	0.89	0.73	0.91

Table 7.6. Type I errors for normally distributed data with an unspecified covariance matrix (defined in section 4.1.4, scale matrices (1), (2) are taken from page 63, method = Ht10).

Scale matrix	p	c	*df*	Type I error
(1)	3	10	4	0.07
(1)	3	10	7	0.08
(1)	3	10	10	0.07
(1)	3	10	13	0.07
(1)	3	50	5	0.07
(1)	3	50	7	0.07
(1)	3	50	10	0.07
(1)	3	50	13	0.06
(1)	3	10	53	0.06
(1)	3	10	103	0.06
(1)	3	50	53	0.06
(1)	3	50	103	0.06
(2)	3	10	4	0.07
(2)	3	10	7	0.05
(2)	3	10	10	0.07
(2)	3	10	13	0.07
(2)	3	50	5	0.06
(2)	3	50	7	0.06
(2)	3	50	10	0.05
(2)	3	50	13	0.05

Table 7.7. Power for normally distributed data with an unspecified covariance matrix (defined in section 4.1.4, scale matrices (1), (2) are taken from page 63, method = Ht10).

Scale matrix	p	c	θ_1					*df*	Power
(1)	3	10	0.000	0.000	1.540			13	0.62
(1)	5	10	0.000	0.332	0.664	0.996	1.328	10	0.93
(2)	3	10	0.000	0.000	1.540			13	0.59
(2)	5	10	0.000	0.332	0.664	0.996	1.328	10	0.89

Table 7.8. Type I errors for normally distributed data with the auto-correlation covariance matrix (defined in section 4.1.4, method = Ht10).

ρ	p	c	Type I error
0.2	3	10	0.05
0.2	3	50	0.05
0.2	5	10	0.03
0.2	5	50	0.06
0.4	3	10	0.05
0.4	3	50	0.06
0.4	5	10	0.03
0.4	5	50	0.06
0.6	3	10	0.05
0.6	3	50	0.06
0.6	5	10	0.04
0.6	5	50	0.07
0.9	3	10	0.05
0.9	3	50	0.06
0.9	5	10	0.05
0.9	5	50	0.08

Table 7.9. Power for normally distributed data with the auto-correlation covariance matrix (defined in section 4.1.4, method = Ht10).

ρ	p	c	θ_1					Power
0.2	3	10	0.00	0.00	1.00			0.73
0.2	3	10	0.00	1.00	1.00			0.70
0.2	3	10	0.00	0.50	1.00			0.66
0.2	3	50	0.00	0.00	0.45			0.78
0.2	3	50	0.00	0.50	0.50			0.86
0.2	5	10	0.00	0.00	0.00	0.00	1.00	0.77
0.2	5	10	0.00	1.00	1.00	1.00	1.00	0.70
0.2	5	10	0.00	0.30	0.60	0.90	1.20	0.82
0.2	5	50	0.00	0.00	0.00	0.00	0.50	0.90
0.2	5	50	0.00	0.50	0.50	0.50	0.50	0.89
0.2	5	50	0.00	0.15	0.30	0.45	0.60	0.95
0.2	5	50	0.00	0.30	0.30	0.30	0.60	0.93
0.6	3	10	0.00	0.00	0.70			0.62
0.6	3	10	0.00	0.80	0.80			0.71
0.6	3	10	0.00	0.40	0.80			0.66
0.6	3	50	0.00	0.00	0.35			0.77
0.6	3	50	0.00	0.35	0.35			0.77
0.6	3	50	0.00	0.20	0.40			0.81
0.6	5	10	0.00	0.00	0.00	0.00	0.90	0.74
0.6	5	10	0.00	0.80	0.80	0.80	0.80	0.63
0.6	5	10	0.00	0.25	0.50	0.75	1.00	0.71
0.6	5	50	0.00	0.00	0.00	0.00	0.50	0.95
0.6	5	50	0.00	0.50	0.50	0.50	0.50	0.92
0.6	5	50	0.00	0.30	0.30	0.30	0.60	0.94

Table 7.10. Type I errors for homoscedastic log-normally distributed data (section 4.2.1.2).

p	subj	Hm6	Hm9	Hm10	Ht6	Ht9	Ht10
3	10	0.04	0.02	0.03	0.06	0.03	0.05
3	50	0.04	0.02	0.04	0.07	0.02	0.06
3	10	0.04	0.03	0.04	0.06	0.05	0.06
3	10	0.04	0.02	0.04	0.06	0.04	0.05
5	10	0.05	0.01	0.05	0.08	0.02	0.04

Table 7.11. Power for homoscedastic log-normally distributed data (section 4.2.1.2).

p	subj	θ_1									τ^2	Hm6	Hm9	Hm10	Ht6	Ht9	Ht10
3	10	0	0.00	1.90							1	0.78	0.71	0.76	0.76	0.83	0.74
3	50	0	0.00	0.80							1	0.72	0.61	0.73	0.83	0.76	0.73
3	10	0	2.00	2.00							1	0.78	0.74	0.78	0.67	0.74	0.75
3	50	0	0.80	0.80							1	0.76	0.68	0.77	0.73	0.65	0.71
3	10	0	1.00	2.00							1	0.74	0.69	0.75	0.69	0.74	0.73
3	50	0	0.50	1.00							1	0.81	0.75	0.82	0.78	0.77	0.77
5	50	0	0.70	0.70	0.70	0.70					1	0.65	0.32	0.68	0.53	0.35	0.63
5	50	0	0.70	0.70	0.70	1.40					1	0.92	0.75	0.92	0.89	0.82	0.89
10	10	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	1	0.73	0.22	0.76	0.74	0.37	0.73
10	10	0	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1	0.69	0.13	0.77	0.44	0.26	0.71
10	10	0	0.31	0.62	0.93	1.24	1.55	1.86	2.17	2.48	2.79	1	0.96	0.79	0.94	0.97	0.97	0.83
10	10	0	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	2.60	1	0.83	0.29	0.85	0.74	0.49	0.79

Table 7.12. Type I errors for heteroscedastic log-normally distributed data (section 4.2.1.3).

p	subj	σ^2					τ^2	Ht6	Ht9	Ht10	Hm6	Hm9	Hm10
3	10	0.10	0.10	0.04			1	0.03	0.02	0.03	0.02	0.01	0.02
3	10	0.10	0.20	0.20			1	0.04	0.03	0.05	0.05	0.03	0.05
3	10	0.10	0.01	0.04			1	0.03	0.03	0.03	0.04	0.03	0.04
3	50	0.10	0.02	0.02			1	0.01	0.01	0.01	0.01	0.00	0.00
3	50	0.10	0.01	0.03			1	0.02	0.01	0.01	0.02	0.01	0.01
5	10	0.10	0.10	0.10	0.10	0.16	1	0.06	0.02	0.04	0.07	0.02	0.07
5	10	0.10	0.04	0.04	0.04	0.04	1	0.04	0.02	0.03	0.05	0.01	0.05
5	10	0.10	0.01	0.04	0.09	0.16	1	0.06	0.02	0.05	0.10	0.04	0.11
5	10	0.10	0.01	0.01	0.01	0.04	1	0.04	0.03	0.03	0.06	0.03	0.07

Table 7.13. Power for heteroscedastic log-normally distributed data (section 4.2.1.3).

p	nsubj	θ_1				τ^2	Ht6	Ht9	Ht10	Hm6	Hm9	Hm10
3	10	0	0.00	1.54		1	0.84	0.90	0.98	0.98	0.94	0.97
3	10	0	0.45	0.45		1	0.83	0.81	0.85	0.86	0.77	0.85
3	10	0	0.30	0.60		1	0.88	0.85	0.90	0.94	0.88	0.95
3	50	0	0.00	0.20		1	0.97	0.92	0.88	0.93	0.85	0.93
3	50	0	0.15	0.15		1	0.66	0.57	0.68	0.73	0.67	0.73
3	50	0	0.08	0.16		1	0.92	0.89	0.69	0.81	0.69	0.80
5	10	0	0.00	0.00	0.00	0.40	1	0.70	0.44	0.77	0.86	0.65	0.87
5	10	0	0.45	0.45	0.45	0.45	1	0.80	0.66	0.83	0.84	0.54	0.86
5	10	0	0.10	0.20	0.30	0.40	1	0.90	0.79	0.75	0.94	0.79	0.91
5	10	0	0.34	0.34	0.34	0.67	1	0.84	0.72	0.94	0.96	0.84	0.97
5	50	0	0.00	0.00	0.00	0.20	1	1.00	0.97	0.96	0.95	0.73	0.95
5	50	0	0.20	0.20	0.20	0.20	1	0.84	0.66	0.92	0.96	0.87	0.96
5	50	0	0.08	0.08	0.08	0.16	1	0.91	0.87	0.74	0.93	0.73	0.94

Table 7.14. Type I errors for the mixture of two normally distributed random variables (defined in section 4.2.2, method = Ht10).

π	s	p	c	Type I error
0.2	1	3	10	0.06
0.2	1	3	50	0.04
0.2	1	5	10	0.03
0.2	1	5	50	0.05
0.2	5	3	10	0.06
0.2	5	3	50	0.04
0.2	5	5	10	0.04
0.2	5	5	50	0.06
0.4	1	3	10	0.06
0.4	1	3	50	0.05
0.4	1	5	10	0.04
0.4	5	3	10	0.05
0.4	5	3	50	0.04
0.4	5	5	10	0.04
0.4	5	5	50	0.05
0.6	1	3	10	0.05
0.6	1	3	50	0.05
0.6	1	5	10	0.04
0.6	5	3	10	0.06
0.6	5	5	10	0.04
0.9	1	3	10	0.05
0.9	1	3	50	0.05
0.9	1	5	10	0.05
0.9	1	5	50	0.05
0.9	5	3	10	0.05
0.9	5	3	50	0.05
0.9	5	5	10	0.05

Table 7.15. Power for the mixture of two normally distributed random variables (defined in section 4.2.2, method = Ht10).

π	s	p	c	θ_1					Power
0.2	1	3	10	0.00	0.00	1.54			0.99
0.2	1	5	10	0.00	0.33	0.66	1.00	1.33	0.95
0.6	1	3	10	0.00	0.00	1.54			1.00
0.6	1	3	10	0.00	0.45	0.45			0.52
0.6	1	3	10	0.00	0.30	0.60			0.69
0.6	1	5	10	0.00	0.45	0.45	0.45	0.45	0.51
0.6	1	5	10	0.00	0.33	0.66	1.00	1.33	1.00
0.6	1	5	10	0.00	0.34	0.34	0.34	0.67	0.73

Table 7.16. Type I errors when random errors follow gamma distribution (defined in section 4.2.3, method = Ht10).

α	β	p	c	Type I error
2	0.5	3	10	0.05
2	0.5	3	50	0.04
2	0.5	5	10	0.05
2	0.5	5	50	0.04
2	2	3	10	0.05
2	2	3	50	0.05
2	2	5	10	0.06
4	0.5	3	10	0.05
4	0.5	3	50	0.05
4	0.5	5	10	0.04
4	2	3	10	0.05
4	2	3	50	0.06
4	2	5	10	0.06

Table 7.17. Power when random errors follow gamma distribution(defined in section 4.2.3, method = Ht10).

α	β	p	c	θ_1					Power
2	2	3	10	0.000	0.000	1.540			0.338
4	0.5	3	10	0.000	0.000	1.540			0.936
4	0.5	3	10	0.000	0.300	0.600			0.338
4	0.5	5	10	0.000	0.332	0.664	0.996	1.328	0.840
4	0.5	5	10	0.000	0.335	0.335	0.335	0.670	0.382

Table 7.18. Type I errors for homoscedastic normally distributed data.

p	subj	τ^2	Hm1	Hm2	Hm3	Hm4	Hm5	Hm6	Hm7	Hm8	Hm9
3	10	0.2	0.05	0.05	0.04	0.03	0.05	0.04	0.04	0.04	0.02
3	10	1	0.05	0.05	0.03	0.03	0.05	0.04	0.03	0.05	0.01
3	10	2	0.05	0.05	0.04	0.03	0.05	0.04	0.04	0.05	0.01
3	30	0.2	0.06	0.06	0.07	0.06	0.06	0.07	0.06	0.07	0.04
3	30	1	0.06	0.06	0.07	0.06	0.06	0.07	0.06	0.06	0.04
3	30	2	0.06	0.06	0.07	0.06	0.06	0.07	0.06	0.06	0.04
3	50	0.2	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.02
3	50	1	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.01
3	50	2	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.01
5	10	0.2	0.06	0.07	0.05	0.04	0.07	0.05	0.05	0.06	0.02
5	10	1	0.06	0.07	0.05	0.04	0.07	0.04	0.04	0.06	0.01
5	10	2	0.06	0.07	0.05	0.04	0.07	0.05	0.04	0.06	0.01
5	30	0.2	0.07	0.05	0.06	0.05	0.05	0.06	0.07	0.05	0.02
5	30	1	0.07	0.05	0.06	0.05	0.05	0.06	0.07	0.05	0.02
5	30	2	0.07	0.05	0.06	0.05	0.05	0.07	0.06	0.05	0.02
5	50	0.2	0.04	0.05	0.04	0.04	0.05	0.05	0.04	0.05	0.01
5	50	1	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.05	0.01
5	50	2	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.05	0.01
10	10	0.2	0.06	0.05	0.05	0.04	0.05	0.06	0.04	0.04	0.00
10	10	1	0.06	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.00
10	10	2	0.06	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.00
10	30	0.2	0.04	0.06	0.06	0.03	0.06	0.06	0.04	0.05	0.00
10	30	1	0.04	0.06	0.06	0.03	0.06	0.06	0.04	0.05	0.00
10	30	2	0.04	0.06	0.06	0.03	0.06	0.06	0.04	0.06	0.00
10	50	1	0.04	0.04	0.06	0.03	0.04	0.06	0.04	0.04	0.00
10	50	0.21	0.04	0.04	0.06	0.03	0.04	0.06	0.04	0.04	0.00
10	50	2	0.04	0.04	0.06	0.03	0.04	0.06	0.04	0.04	0.00

Table 7.19. Power for homoscedastic normally distributed data.

(a) $p = 3$

p	subj	$\boldsymbol{\theta}_1$			τ^2	Hm1	Hm2	Hm3	Hm4	Hm5	Hm6	Hm7	Hm8	Hm9
3	10	0.00	0.00	1.25	1	0.80	0.74	0.90	0.90	0.74	0.72	0.83	0.74	0.72
3	10	0.00	0.00	1.24	0.2	0.80	0.75	0.90	0.89	0.74	0.72	0.86	0.74	0.72
3	10	0.00	0.00	1.25	2	0.80	0.74	0.90	0.90	0.74	0.72	0.83	0.74	0.72
3	10	0.00	1.26	1.26	1	0.80	0.74	0.89	0.89	0.76	0.72	0.85	0.76	0.70
3	10	0.00	1.22	1.22	0.2	0.80	0.73	0.89	0.89	0.75	0.69	0.84	0.75	0.69
3	10	0.00	1.26	1.26	2	0.80	0.73	0.89	0.89	0.76	0.72	0.85	0.77	0.70
3	10	0.00	0.73	1.45	1	0.80	0.90	0.90	0.90	0.74	0.87	0.87	0.76	0.88
3	10	0.00	0.71	1.42	0.2	0.80	0.90	0.90	0.90	0.74	0.89	0.88	0.74	0.89
3	10	0.00	0.73	1.45	2	0.80	0.90	0.90	0.90	0.75	0.88	0.87	0.76	0.88
3	30	0.00	0.00	0.69	1	0.80	0.76	0.89	0.89	0.78	0.77	0.83	0.78	0.76
3	30	0.00	0.00	0.69	0.2	0.80	0.77	0.90	0.89	0.78	0.76	0.83	0.78	0.76
3	30	0.00	0.00	0.69	2	0.80	0.76	0.90	0.89	0.79	0.77	0.83	0.79	0.77
3	30	0.00	0.69	0.69	1	0.80	0.78	0.92	0.92	0.79	0.77	0.84	0.79	0.77
3	30	0.00	0.69	0.69	0.2	0.80	0.77	0.92	0.91	0.79	0.77	0.84	0.79	0.76
3	30	0.00	0.69	0.69	2	0.80	0.78	0.92	0.91	0.80	0.78	0.84	0.79	0.77
3	30	0.00	0.40	0.80	1	0.80	0.89	0.91	0.91	0.80	0.88	0.85	0.80	0.88
3	30	0.00	0.40	0.79	0.2	0.80	0.88	0.91	0.91	0.79	0.88	0.86	0.79	0.88
3	30	0.00	0.40	0.80	2	0.80	0.89	0.91	0.91	0.80	0.88	0.85	0.80	0.88
3	50	0.00	0.00	0.53	1	0.80	0.75	0.91	0.90	0.79	0.75	0.84	0.80	0.76
3	50	0.00	0.00	0.53	0.2	0.80	0.75	0.92	0.90	0.79	0.75	0.84	0.81	0.76
3	50	0.00	0.00	0.53	2	0.80	0.75	0.91	0.91	0.79	0.76	0.84	0.80	0.76
3	50	0.00	0.55	0.55	1	0.80	0.78	0.90	0.90	0.79	0.79	0.84	0.80	0.79
3	50	0.00	0.55	0.55	0.2	0.80	0.79	0.90	0.91	0.79	0.79	0.86	0.81	0.79
3	50	0.00	0.55	0.55	2	0.80	0.78	0.90	0.90	0.80	0.79	0.84	0.80	0.79
3	50	0.00	0.31	0.62	1	0.80	0.89	0.92	0.93	0.79	0.89	0.85	0.80	0.89
3	50	0.00	0.31	0.62	0.2	0.80	0.89	0.93	0.93	0.80	0.89	0.87	0.80	0.89
3	50	0.00	0.31	0.62	2	0.80	0.89	0.92	0.93	0.79	0.88	0.85	0.80	0.89

(b) $p = 5$

p	subj	$\boldsymbol{\theta}_1$					τ^2	Hm1	Hm2	Hm3	Hm4	Hm5	Hm6	Hm7	Hm8	Hm9
5	10	0.00	0.00	0.00	0.00	1.27	1	0.80	0.63	0.89	0.75	0.62	0.89	0.76	0.64	0.70
5	10	0.00	0.00	0.00	0.00	1.26	0.2	0.80	0.64	0.90	0.76	0.61	0.89	0.75	0.62	0.73
5	10	0.00	0.00	0.00	0.00	1.27	2	0.80	0.63	0.89	0.75	0.62	0.90	0.76	0.63	0.70
5	10	0.00	1.24	1.24	1.24	1.24	1	0.80	0.61	0.89	0.74	0.60	0.89	0.75	0.60	0.71
5	10	0.00	1.24	1.24	1.24	1.24	0.2	0.80	0.61	0.89	0.74	0.61	0.91	0.74	0.60	0.73
5	10	0.00	1.25	1.25	1.25	1.25	2	0.80	0.61	0.90	0.75	0.61	0.90	0.76	0.60	0.71
5	10	0.00	0.37	0.74	1.11	1.48	1	0.80	0.95	0.95	0.76	0.95	0.94	0.77	0.95	0.80
5	10	0.00	0.37	0.73	1.10	1.46	0.2	0.80	0.94	0.94	0.75	0.94	0.94	0.76	0.94	0.80
5	10	0.00	0.37	0.74	1.11	1.48	2	0.80	0.95	0.95	0.76	0.95	0.94	0.76	0.95	0.80
5	10	0.00	0.81	0.81	0.81	1.62	1	0.80	0.85	0.92	0.75	0.84	0.92	0.75	0.83	0.76
5	10	0.00	0.80	0.80	0.80	1.60	0.2	0.80	0.85	0.92	0.74	0.83	0.92	0.75	0.81	0.77
5	10	0.00	0.81	0.81	0.81	1.62	2	0.80	0.85	0.92	0.75	0.84	0.92	0.76	0.84	0.76
5	30	0.00	0.00	0.00	0.00	0.70	1	0.80	0.67	0.93	0.79	0.67	0.92	0.79	0.67	0.74
5	30	0.00	0.00	0.00	0.00	0.70	0.2	0.80	0.67	0.93	0.79	0.67	0.92	0.80	0.67	0.74
5	30	0.00	0.00	0.00	0.00	0.70	2	0.80	0.67	0.93	0.79	0.67	0.93	0.79	0.67	0.74
5	30	0.00	0.70	0.70	0.70	0.70	1	0.80	0.68	0.91	0.78	0.68	0.92	0.78	0.67	0.73
5	30	0.00	0.70	0.70	0.70	0.70	0.2	0.80	0.68	0.91	0.78	0.68	0.91	0.78	0.67	0.74
5	30	0.00	0.70	0.70	0.70	0.70	2	0.80	0.68	0.91	0.78	0.68	0.91	0.78	0.66	0.73
5	30	0.00	0.20	0.40	0.60	0.80	1	0.80	0.93	0.95	0.77	0.93	0.95	0.78	0.93	0.81
5	30	0.00	0.20	0.40	0.60	0.80	0.2	0.80	0.93	0.95	0.77	0.93	0.95	0.79	0.93	0.82
5	30	0.00	0.20	0.40	0.60	0.80	2	0.80	0.93	0.95	0.77	0.93	0.95	0.78	0.93	0.81
5	30	0.00	0.44	0.44	0.44	0.89	1	0.80	0.86	0.94	0.79	0.86	0.93	0.79	0.85	0.78
5	30	0.00	0.44	0.44	0.44	0.89	0.2	0.80	0.86	0.94	0.79	0.86	0.93	0.79	0.86	0.79
5	30	0.00	0.44	0.44	0.44	0.89	2	0.80	0.86	0.94	0.79	0.86	0.93	0.78	0.86	0.78
5	50	0.00	0.00	0.00	0.00	0.54	1	0.80	0.67	0.91	0.78	0.67	0.92	0.78	0.67	0.73
5	50	0.00	0.00	0.00	0.00	0.54	0.2	0.80	0.67	0.91	0.79	0.68	0.92	0.79	0.66	0.74
5	50	0.00	0.00	0.00	0.00	0.54	2	0.80	0.67	0.91	0.78	0.67	0.91	0.78	0.67	0.73
5	50	0.00	0.55	0.55	0.55	0.55	1	0.80	0.69	0.92	0.78	0.70	0.91	0.79	0.68	0.76
5	50	0.00	0.55	0.55	0.55	0.55	0.2	0.80	0.69	0.92	0.78	0.70	0.91	0.79	0.68	0.76
5	50	0.00	0.55	0.55	0.55	0.55	2	0.80	0.69	0.92	0.78	0.70	0.91	0.79	0.69	0.76
5	50	0.00	0.15	0.30	0.46	0.61	1	0.80	0.93	0.93	0.79	0.93	0.94	0.79	0.92	0.79
5	50	0.00	0.15	0.30	0.46	0.61	0.2	0.80	0.93	0.94	0.79	0.93	0.95	0.78	0.92	0.80
5	50	0.00	0.15	0.30	0.46	0.61	2	0.80	0.93	0.93	0.79	0.93	0.95	0.80	0.93	0.79
5	50	0.00	0.34	0.34	0.34	0.68	1	0.80	0.85	0.92	0.79	0.85	0.92	0.79	0.86	0.79
5	50	0.00	0.34	0.34	0.34	0.68	0.2	0.80	0.85	0.92	0.79	0.85	0.92	0.79	0.85	0.80
5	50	0.00	0.34	0.34	0.34	0.68	2	0.80	0.85	0.92	0.79	0.85	0.92	0.79	0.86	0.79

(c) p = 10, part I

p	subj	Design	θ_1										τ^2
10	10	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.28	1
10	10	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.28	0.2
10	10	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.28	2
10	10	4	0.00	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1
10	10	5	0.00	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	0.2
10	10	6	0.00	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31	2
10	10	7	0.00	0.14	0.28	0.41	0.55	0.69	0.83	0.97	1.10	1.24	1
10	10	8	0.00	0.14	0.28	0.41	0.55	0.69	0.83	0.97	1.10	1.24	0.2
10	10	9	0.00	0.14	0.28	0.41	0.55	0.69	0.83	0.97	1.10	1.24	2
10	10	10	0.00	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	1.72	1
10	10	11	0.00	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	1.71	0.2
10	10	12	0.00	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	1.72	2
10	30	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	1
10	30	14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.2
10	30	15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	2
10	30	16	0.00	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	1
10	30	17	0.00	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.2
10	30	18	0.00	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	2
10	30	19	0.00	0.08	0.15	0.23	0.31	0.39	0.46	0.54	0.62	0.69	1
10	30	20	0.00	0.08	0.15	0.23	0.31	0.39	0.46	0.54	0.62	0.69	0.2
10	30	21	0.00	0.08	0.15	0.23	0.31	0.39	0.46	0.54	0.62	0.69	2
10	30	22	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.99	1
10	30	23	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.99	0.2
10	30	24	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.99	2
10	50	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59	1
10	50	26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59	0.2
10	50	27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59	2
10	50	28	0.00	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	1
10	50	29	0.00	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.2
10	50	30	0.00	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	2
10	50	31	0.00	0.06	0.12	0.19	0.25	0.31	0.37	0.44	0.50	0.56	1
10	50	32	0.00	0.06	0.12	0.19	0.25	0.31	0.37	0.44	0.50	0.56	0.2
10	50	33	0.00	0.06	0.12	0.19	0.25	0.31	0.37	0.44	0.50	0.56	2
10	50	34	0.00	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.79	1
10	50	35	0.00	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.79	0.2
10	50	36	0.00	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.79	2

(c) $p = 10$, part II

groups	subj	Design	Hm1	Hm2	Hm3	Hm4	Hm5	Hm6	Hm7	Hm8	Hm9
10	10	1	0.80	0.48	0.95	0.70	0.48	0.94	0.70	0.49	0.46
10	10	2	0.80	0.48	0.95	0.69	0.48	0.95	0.70	0.49	0.47
10	10	3	0.80	0.48	0.95	0.69	0.48	0.95	0.70	0.49	0.46
10	10	4	0.80	0.49	0.94	0.73	0.49	0.94	0.71	0.50	0.51
10	10	5	0.80	0.49	0.94	0.74	0.49	0.94	0.73	0.49	0.53
10	10	6	0.80	0.49	0.94	0.74	0.50	0.94	0.73	0.51	0.52
10	10	7	0.80	0.97	0.98	0.75	0.97	0.99	0.74	0.97	0.70
10	10	8	0.80	0.97	0.98	0.75	0.97	0.99	0.75	0.97	0.70
10	10	9	0.80	0.97	0.98	0.75	0.97	0.98	0.74	0.97	0.70
10	10	10	0.80	0.77	0.95	0.72	0.76	0.95	0.73	0.75	0.51
10	10	11	0.80	0.76	0.94	0.72	0.76	0.95	0.72	0.76	0.52
10	10	12	0.80	0.77	0.95	0.72	0.76	0.95	0.72	0.76	0.51
10	30	13	0.80	0.53	0.93	0.77	0.52	0.94	0.78	0.53	0.54
10	30	14	0.80	0.53	0.93	0.77	0.52	0.94	0.78	0.54	0.55
10	30	15	0.80	0.53	0.93	0.77	0.52	0.94	0.77	0.53	0.54
10	30	16	0.80	0.55	0.95	0.78	0.53	0.95	0.79	0.55	0.57
10	30	17	0.80	0.54	0.95	0.78	0.52	0.95	0.77	0.55	0.57
10	30	18	0.80	0.55	0.95	0.78	0.53	0.95	0.79	0.55	0.57
10	30	19	0.80	0.97	0.98	0.78	0.98	0.98	0.78	0.97	0.69
10	30	20	0.80	0.97	0.98	0.78	0.98	0.98	0.77	0.97	0.69
10	30	21	0.80	0.97	0.98	0.78	0.98	0.98	0.78	0.97	0.69
10	30	22	0.80	0.80	0.94	0.78	0.79	0.95	0.77	0.80	0.60
10	30	23	0.80	0.80	0.94	0.78	0.79	0.94	0.79	0.80	0.60
10	30	24	0.80	0.80	0.94	0.78	0.79	0.94	0.78	0.80	0.60
10	50	25	0.86	0.62	0.96	0.84	0.63	0.95	0.84	0.60	0.54
10	50	26	0.80	0.54	0.95	0.80	0.54	0.95	0.79	0.54	0.54
10	50	27	0.80	0.54	0.95	0.80	0.54	0.95	0.80	0.53	0.54
10	50	28	0.80	0.54	0.94	0.79	0.53	0.95	0.79	0.53	0.53
10	50	29	0.80	0.54	0.94	0.79	0.53	0.95	0.78	0.54	0.53
10	50	30	0.80	0.54	0.94	0.79	0.53	0.95	0.78	0.53	0.53
10	50	31	0.79	0.98	0.99	0.80	0.97	0.98	0.79	0.98	0.72
10	50	32	0.80	0.98	0.98	0.79	0.98	0.98	0.78	0.98	0.71
10	50	33	0.79	0.98	0.99	0.80	0.97	0.98	0.80	0.98	0.72
10	50	34	0.79	0.88	0.97	0.81	0.87	0.95	0.79	0.87	0.60
10	50	35	0.80	0.80	0.95	0.79	0.79	0.96	0.79	0.79	0.60
10	50	36	0.79	0.88	0.97	0.81	0.87	0.95	0.80	0.87	0.60

Table 7.20. Type I errors for heteroscedastic normally distributed data

p	nsubj	σ^2					Ht1	Ht2	Ht3	Ht4	Ht5	Ht6	Ht7	Ht8	Ht9
3	10	0.10	0.10	2.37			0.07	0.08	0.04	0.04	0.05	0.04	0.03	0.03	0.04
3	10	0.10	0.20	0.20			0.09	0.07	0.04	0.03	0.04	0.03	0.03	0.03	0.03
3	10	0.10	0.09	0.36			0.09	0.09	0.04	0.03	0.05	0.04	0.03	0.04	0.03
3	30	0.10	0.10	0.02			0.08	0.07	0.05	0.05	0.05	0.05	0.05	0.05	0.03
3	30	0.10	0.04	0.04			0.08	0.07	0.07	0.06	0.06	0.07	0.07	0.06	0.05
3	30	0.10	0.01	0.03			0.08	0.07	0.06	0.05	0.05	0.06	0.05	0.06	0.04
3	50	0.10	0.10	0.01			0.04	0.05	0.04	0.03	0.05	0.04	0.03	0.05	0.01
3	50	0.10	0.02	0.02			0.04	0.05	0.04	0.03	0.05	0.03	0.02	0.04	0.02
3	50	0.10	0.01	0.03			0.04	0.04	0.05	0.03	0.04	0.05	0.04	0.04	0.03
5	10	0.10	0.10	0.10	0.10	0.16	0.16	0.15	0.05	0.04	0.05	0.05	0.04	0.05	0.01
5	10	0.10	0.20	0.20	0.20	0.20	0.16	0.14	0.06	0.05	0.05	0.05	0.03	0.05	0.01
5	10	0.10	0.11	0.44	0.99	1.76	0.12	0.13	0.05	0.04	0.05	0.04	0.03	0.04	0.01
5	10	0.10	0.11	0.11	0.11	0.45	0.15	0.14	0.06	0.04	0.06	0.05	0.04	0.05	0.02
5	30	0.10	0.10	0.10	0.10	0.01	0.09	0.06	0.03	0.05	0.04	0.04	0.06	0.04	0.00
5	30	0.10	0.03	0.03	0.03	0.03	0.09	0.06	0.06	0.05	0.04	0.06	0.06	0.04	0.02
5	30	0.10	0.00	0.00	0.00	0.00	0.07	0.06	0.05	0.05	0.04	0.05	0.05	0.03	0.02
5	30	0.10	0.01	0.01	0.01	0.05	0.07	0.06	0.05	0.05	0.04	0.05	0.05	0.04	0.02
5	50	0.10	0.10	0.10	0.10	0.01	0.07	0.08	0.04	0.05	0.08	0.05	0.06	0.05	0.01

Table 7.21. Power for heteroscedastic normally distributed data.

p	nsubj	θ_1					Ht1	Ht2	Ht3	Ht4	Ht5	Ht6	Ht7	Ht8	Ht9
3	10	0	0.00	1.54			0.79	0.15	0.82	0.67	0.11	0.80	0.59	0.09	0.82
3	10	0	0.45	0.45			0.83	0.83	0.85	0.69	0.67	0.84	0.65	0.64	0.81
3	10	0	0.30	0.60			0.77	0.85	0.83	0.64	0.72	0.84	0.60	0.71	0.82
3	30	0	0.00	0.15			0.79	0.79	0.84	0.74	0.74	0.85	0.74	0.75	0.77
3	30	0	0.20	0.20			0.80	0.67	0.90	0.77	0.63	0.90	0.76	0.61	0.84
3	30	0	0.09	0.17			0.80	0.88	0.90	0.75	0.85	0.90	0.75	0.85	0.85
3	50	0	0.00	0.10			0.73	0.79	0.81	0.70	0.77	0.81	0.70	0.77	0.70
3	50	0	0.15	0.15			0.81	0.62	0.91	0.80	0.59	0.90	0.81	0.59	0.86
3	50	0	0.08	0.16			0.92	0.97	0.97	0.91	0.96	0.97	0.92	0.96	0.95
5	10	0	0.45	0.45	0.45	0.45	0.84	0.75	0.84	0.63	0.60	0.83	0.60	0.57	0.68
5	10	0	0.33	0.66	1.00	1.33	0.96	0.99	0.98	0.90	0.97	0.98	0.84	0.96	0.96
5	10	0	0.00	0.00	0.00	0.40	0.70	0.40	0.67	0.41	0.23	0.66	0.36	0.21	0.42
5	10	0	0.34	0.34	0.34	0.67	0.81	0.79	0.83	0.60	0.63	0.81	0.55	0.61	0.71
5	30	0	0.00	0.00	0.00	0.12	0.79	0.84	0.86	0.72	0.79	0.85	0.71	0.79	0.60
5	30	0	0.18	0.18	0.18	0.18	0.71	0.32	0.80	0.61	0.28	0.80	0.63	0.28	0.61
5	30	0	0.01	0.02	0.03	0.04	1.00	1.00	0.75	1.00	1.00	0.75	0.99	1.00	0.47
5	30	0	0.11	0.11	0.11	0.21	0.79	0.58	0.90	0.70	0.53	0.90	0.70	0.51	0.74
5	50	0	0.00	0.00	0.00	0.08	0.76	0.90	0.90	0.78	0.87	0.86	0.71	0.87	0.52

REFERENCES

Andersson, S. A. and Perlman, M. D. (1993). Lattice Models for Conditional Independence in a Multivariate Normal Distribution, *The Annals of Statistics* **21**, 1318-1358.

Betcher, J. and Peddada, S. D. (2009). Statistical inference under order restrictions in analysis of covariance using a modified restricted maximum likelihood estimator, *Sankhya: The Indian Journal of Statistics* **71**, 79-96.

Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models, *Computational Statistics & Data Analysis* **50**, 1735-1748.

Cao, Y., Chen, A., Jones, R. L., Radcliffe, J., Dietrich, K. N., Caldwell, K. L., Peddada, S. and Rogan, W. J. (2011). Efficacy of Succimer Chelation of Mercury at Background Exposures in Toddlers: A Randomized Trial, *The Journal of Pediatrics* **158**, 480-485.e481.

Chatterjee, S., Lahiri, P. and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models, *The annals of statistics.* **36**, 1221-1245.

Conaway, M. R., Dunbar, S. and Peddada, S. D. (2004). Designs for single- or multiple-agent phase I trials, *Biometrics* **60**, 661-669.

Davidov, O. and Rosen, S. (2011). Constrained inference in mixed-effects models for longitudinal data with application to hearing loss, *Biostatistics* **12**, 327-340.

Dempster, A. P. (1972). Covariance Selection, *Biometrics* **28**, 157-175.

Diaz, M. M. and González, B. S. (1988). The validity of the "Pool-Adjacent-Violator" algorithm, *Statistics & Probability Letters* **6**, 143-145.

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums, *Technometrics* **6**, 241-252.

Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association* **50**, 1096-1121.

Dykstra, R. L. and Robertson, T. (1982). An Algorithm for Isotonic Regression for Two or More Independent Variables, *The Annals of Statistics* **10**, 708-716.

Fernandez, M. A., Rueda, C. and Salvador, B. (1999). The Loss of Efficiency Estimating Linear Functions under Restrictions, *Scandinavian Journal of Statistics* **26**, 579-592.

Gupta, S. (1989). Efficient Designs for Comparing Test Treatments with a Control, *Biometrika* **76**, 783-787.

Hedayat, A. S., Jacroux, M. and Majumdar, D. (1988). Optimal Designs for Comparing Test Treatments with Controls, *Statistical Science* **3**, 462-476.

Henderson, C. R. (1953). Estimation of Variance and Covariance Components, *Biometrics* **9**, 226-252.

Hoferkamp, C. L. and Peddada, S. D. (2002). Parameter Estimation in Linear Models with Heteroscedastic Variances Subject to Order Restrictions, *Journal of Multivariate Analysis* **82**, 65-87.

Huang, W.-T. (1984), "Nonparametric isotonic selection rules under a prior ordering," in *Design of Experiments - Ranking and Selection* (Vol. 56), eds. T. J. Santner and A. C. Tamhane, New York: Marcel Decker Inc., pp. 95-112.

Hwang, J. T. (1985). Universal Domination and Stochastic Domination: Estimation Simultaneously Under a Broad Class of Loss Functions, *The Annals of Statistics* **13**, 295-314.

Hwang, J. T. G. and Peddada, S. D. (1994). Confidence Interval Estimation Subject to Order Restrictions, *The Annals of Statistics* **22**, 67-93.

Ivanova, A., Liu, K., Snyder, E. and Snavely, D. (2009). An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose-finding study, *Statistics in Medicine* **28**, 2941-2951.

Khattree, R. and Naik, D. N. (1999). *Applied multivariate statistics with SAS software*. Cary, NC; [New York]: SAS Institute ; J. Wiley & Sons.

Kim, H.-J. and Boos, D. D. (2004). Variance Estimation in Spatial Regression Using a Non-parametric Semivariogram Based on Residuals, *Scandinavian Journal of Statistics* **31**, 387-401.

Lauritzen, S. L. (1996). *Graphical models*. Oxford; New York: Clarendon Press ; Oxford University Press.

Lee, C.-I. C. (1988). Quadratic Loss of Order Restricted Estimators for Treatment Means with a Control, *The Annals of Statistics* **16**, 751-758.

Lehmann, E. L. (1975). *Nonparametrics statistical methods based on ranks*. San Francisco: Holden-Day.

Li, G. and Tseng, C.-H. (2008). Non-parametric Estimation of a Survival Function with Two-stage Design Studies, *Scandinavian Journal of Statistics* **35**, 193-211.

Lim, C., Sen, P. and Peddada, S. (2010). Statistical inference in nonlinear regression under heteroscedasticity, *Sankhya B - Applied and Interdisciplinary Statistics* **72**, 202-218.

Lim, C., Sen, P. K. and Peddada, S. (2011). Accounting for uncertainty in heteroscedasticity in nonlinear regression, *Journal of Statistical Planning and Inference*, In press.

Mammen, E., Marron, J. S., Turlach, B. A. and Wand, M. P. (2001). A General Projection Framework for Constrained Smoothing, *Statistical Science* **16**, 232-248.

Marcus, R. and Talpaz, H. (1992). Further results on testing homogeneity of normal means against simple tree alternatives, *Communications in Statistics - Theory and Methods* **21**, 2135 - 2149.

Mukerjee, H. (1988). Order Restricted Inference in a Repeated Measures Model, *Biometrika* **75**, 616-617.

Nettleton, D. (1999). Convergence Properties of the EM Algorithm in Constrained Parameter Spaces, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **27**, 639-648.

Nigam, A. K., Puri, P. D. and Gupta, V. K. (1988). *Characterizations and analysis of block designs*. New York: Wiley.

Pearce, S. C. (1960). Supplemented Balance, *Biometrika* **47**, 263-271.

Peddada, S. D., Dinse, G. E. and Kissling, G. E. (2007). Incorporating Historical Control Data When Comparing Tumor Incidence Rates, *Journal of the American Statistical Association* **102**, 1212-1220.

Peddada, S. D., Dunson, D. B. and Tan, X. (2005). Estimation of order-restricted means from correlated data, *Biometrika* **92**, 703-715.

Peddada, S. D., Haseman, J. K., Tan, X. and Travlos, G. (2006). Tests for a simple tree order restriction with application to dose-response studies, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**, 493-506.

Peddada, S. D., Laughlin, S. K., Miner, K., Guyon, J.-P., Haneke, K., Vahdat, H. L., Semelka, R. C., Kowalik, A., Armao, D., Davis, B. and Baird, D. D. (2008). Growth of uterine leiomyomata among premenopausal black and white women, *PNAS* **105**, 19887-19892.

Rao, C. R. (1970). Estimation of Heteroscedastic Variances in Linear Models, *Journal of the American Statistical Association* **65**, 161-172.

Rao, C. R. (1971). Estimation of variance and covariance components--MINQUE theory, *Journal of Multivariate Analysis* **1**, 257-275.

Rao, C. R. (1972). Estimation of Variance and Covariance Components in Linear Models, *Journal of the American Statistical Association* **67**, 112-115.

Rao, C. R. and Kleffe, J. (1988). *Estimation of variance components and applications*. Amsterdam; New York; New York, N.Y., U.S.A.: North-Holland.

Rao, J. N. K. and Subrahmaniam, K. (1971). Combining Independent Estimators and Estimation in Linear Regression with Unequal Variances, *Biometrics* **27**, 971-990.

Robertson, T., Wright, F. T. and Dykstra, R. (1988). *Order Restricted Statistical Inference*. Chichester; New York: Wiley.

Rogan, W. J. (1998). The Treatment of Lead-exposed Children (TLC) trial: design and recruitment for a study of the effect of oral chelation on growth and development in toddlers, *Paediatric and Perinatal Epidemiology* **12**, 313-333.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

Shaked, M. (1979). Estimation of Starshaped Sequences of Poisson and Normal Means, *The Annals of Statistics* **7**, 729-741.

Shi, N.-Z., Zheng, S.-R. and Guo, J. (2005). The restricted EM algorithm under inequality restrictions on the parameters, *Journal of Multivariate Analysis* **92**, 53-76.

Shin, D. W., Park, C. G. and Park, T. (1996). Testing for ordered group effects with repeated measurements, *Biometrika* **83**, 688-694.

Shirley, E. (1977). A Non-Parametric Equivalent of Williams' Test for Contrasting Increasing Dose Levels of a Treatment, *Biometrics* **33**, 386-389.

Silvapulle, M. J. (1997). On order restricted inference in some mixed linear models, *Statistics & Probability Letters* **36**, 23-27.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained statistical inference: inequality, order, and shape restrictions*. Hoboken, N.J.: Wiley-Interscience.

Singh, B. and Wright, F. T. (1990). Testing for and against an order restriction in mixed-effects models, *Statistics & Probability Letters* **9**, 195-200.

Stufken, J. (1987). A-Optimal Block Designs for Comparing Test Treatments with a Control, *The Annals of Statistics* **15**, 1629-1638.

Stylianou, M. and Flournoy, N. (2002). Dose Finding Using the Biased Coin Up-and-Down Design and Isotonic Regression, *Biometrics* **58**, 171-177.

Sun, X. and Sun, D. (2005). Estimation of the Cholesky decomposition of the covariance matrix for a conditional independent normal model, *Statistics & Probability Letters* **73**, 1-12.

Tang, D.-I. and Lin, S. P. (1997). An Approximate Likelihood Ratio Test for Comparing Several Treatments to a Control, *Journal of the American Statistical Association* **92**, 1155-1162.

Teoh, E., Nyska, A., Wormser, U. and Peddada, S. D. (2008). Statistical inference under order restrictions on both rows and columns of a matrix, with an application in toxicology, *IMS collections* **1**, 62-77.

van Eeden, C. (2006). *Restricted parameter space estimation problems: admissibility and minimaxity properties*. New York: Springer.

Welch, B. L. (1938). The Significance of the Difference Between Two Means when the Population Variances are Unequal, *Biometrika* **29**, 350-362.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester [England]; New York: Wiley.

Williams, D. A. (1971). A Test for Differences between Treatment Means When Several Dose Levels are Compared with a Zero Dose Control, *Biometrics* **27**, 103-117.

Williams, D. A. (1972). The Comparison of Several Dose Levels with a Zero Dose Control, *Biometrics* **28**, 519-531.

Williams, D. A. (1977). Some Inference Procedures for Monotonically Ordered Normal Means, *Biometrika* **64**, 9-14.

Woolrich, M. W., Ripley, B. D., Brady, M. and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data, *NeuroImage* **14**, 1370-1386.

Young, J. G., Jewell, N. P. and Samuels, S. J. (2008). Regression Analysis of a Disease Onset Distribution Using Diagnosis Data, *Biometrics* **64**, 20-28.