On the Efficiency of Testing Procedures in the Linear Model for Multivariate Longitudinal Data

by

Catherine Njue

A thesis
presented to the University of Manitoba
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in
Statistics

Winnipeg, Manitoba, Canada, 2001

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

ON THE EFFICIENCY OF TESTING PROCEDURES IN THE LINEAR MODEL FOR MULTIVARIATE LONGITUDINAL DATA

BY

CATHERINE NJUE

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

of

DOCTOR OF PHILOSOPHY

Abstract

Multivariate data collected over time on the same experimental unit, referred to as multivariate longitudinal data, are typical of many agricultural, biological, clinical and medical studies. One way to account for the correlations that exist both within and across time is to express the variance-covariance matrix as the Kronecker product of two matrices. These matrices, denoted by Δ and Ω, reflect the characteristic and time dimensions underlying multivariate longitudinal data. The purpose of this thesis is to investigate the asymptotic relative efficiency (ARE) of hypothesis tests in the linear model for multivariate longitudinal data, evaluated through the trace asymptotic relative efficiency (TARE) and curvature asymptotic relative efficiency (CARE).

The gain in efficiency from exploiting a Kronecker product covariance structure when it is appropriate is investigated. To estimate the TARE and CARE, a Monte-carlo simulation study is conducted. The loss of efficiency from imposing a Kronecker product model when it is not appropriate is also considered. Using a class of non-Kronecker product covariance matrices and an index, which quantifies how far a given matrix departs from Kronecker product structure, a Monte-carlo simulation study is conducted. Ordinary least squares and generalised least squares procedures were also compared under a Kronecker product model.

For the designs and covariance matrices considered, the gain in efficiency from exploiting the Kronecker product covariance structure is most pronounced when there is high correlation across time. For the class of non-Kronecker product covariance matrices defined, a noticeable loss of efficiency occurs when the covariance matrix is far from Kronecker product structure, in particular when there is a moderate departure from the null hypothesis under consideration. The use of ordinary least squares, which ignores cross-sectional and longitudinal correlations, is shown to be inefficient, especially when these correlations are high in absolute value.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis supervisor Dr. John F. Brewster for the role he has played in my academic life for the past few years. I thank him for helping me define the topic of this dissertation and for many thoughtful discussions at any time during the research. I also thank him for his unfailing patience and guidance throughout the preparation of this dissertation, and for being an excellent teacher and mentor.

I also want to thank the members of my examining committee, whose encouragement and comments greatly improved this dissertation: Dr. Kenneth Mount of the Department of Statistics; Dr. Annette Gupton of the Faculty of Nursing and Dr. William D. Johnson of the Biostatistics Division, Department of Preventive Medicine, University of Mississippi Medical Center. A special thank you to Dr. Mount for suggesting the measures of asymptotic relative efficiency used in the dissertation and for many useful discussions during the course of this program.

I also wish to thank three other individuals who contributed to this work: Anthony Yeung of Statistics Canada, for always taking the time to answer a lot of questions regarding the UNIX operating system; Dr. Dennis Murphy of the Statistical Advisory Service, who patiently helped me resolve many issues relating to LaTeX and S-plus; and Nick Donaldson of Academic Computing and Networking, with whom I had many discussions regarding the SAS programs used in the dissertation.

I also wish to acknowledge the following sources of financial support: Statistical Advisory Service, for giving me the opportunity to work as a Statistical consultant throughout the entire course of this program; Department of Statistics, for various teaching positions assigned to me over the years. Thank you Dr. Brewster, Dr. Mount and Dr. Cheng.

A special thank you to Dr. A. Thavaneswaran of the Department of Statistics for funding me during the summer of 2000. Dr. Jure Manfreda at the Respiratory Hospital, Health Sciences Center, provided the data used and to him, I am very grateful. Special

thanks are also due to Dr. Erich Kliewer of the Department of Preventive Oncology and Epidemiology, CancerCare Manitoba, for allowing me the flexibility I needed with my job in the last year to complete this work.

The Department of Statistics, University of Manitoba, has been my "home" away from home. I thank the faculty, graduate students and staff for creating an environment that has allowed me to complete this work.

A special thank you to Justine Kiwanuka and her family of Winnipeg, the Holdbrook family of Winnipeg and their grandmother Mrs Regina Djan, the Mayenga family now in Toronto and the Nyachoti family of Winnipeg. Your unfailing friendship and words of encouragement have been very important to me during the course of this program. Thanks also for always being there to take care of Andrew when there was need.

My family in Kenya, in spite of being so far away, has always been a constant source of love and support. In particular, I want to thank my parents, Valerian and Sylvester Njue, for always being there for me and helping me to be the best that I can be. Thanks are due to my only brother Major Dennis Kamuri, who helped me discover the fun in numbers at a very early age.

I would like to thank my husband, Ernest Aryee, for his steadfast love and unfailing support. He has been my greatest source of strength. Thanks for always being so patient and giving me the freedom to follow my dreams. Finally, I want to thank my son, Andrew Aryee, for giving me a reason to stay focused and reminding me that indeed, there is life beyond the numbers.

This thesis is dedicated to my parents, Valerian and Sylvester Njue of Embu, Kenya.

Executive summary

The multivariate longitudinal design, in which multiple characteristics are measured over time on the same experimental unit, is typical of many agricultural, biological, clinical and medical studies. For example, in a medical study, measurements on systolic blood pressure and diastolic blood pressure may be taken on each subject at a number of points in time. In such studies, it is important to account for both cross-sectional and longitudinal correlations. In some problems, it may be reasonable to express the within-subject variance-covariance matrix as the Kronecker product of two matrices, that is, $\Sigma_o = \Delta \otimes \Omega$. The matrices Δ and Ω reflect the characteristic and time dimensions underlying multivariate longitudinal data. Implicit in such a representation is that the covariance matrix for the different characteristics measured at each time point is constant with respect to time, and the correlation matrix for the longitudinal measurements on a given characteristic is the same for all characteristics.

The purpose of this thesis is to investigate the gain from exploiting the Kronecker product structure when it is appropriate. The converse of this situation is also considered, that is, the loss from imposing the Kronecker product structure when it is not appropriate. This will be accomplished by investigating the asymptotic relative efficiency (ARE) of hypothesis tests for the mean vector in the linear model for multivariate longitudinal data. For the purpose of this thesis, efficiency will be evaluated through the trace asymptotic relative efficiency (TARE) and curvature asymptotic relative efficiency(CARE), two measures of asymptotic relative efficiency. They can be applied to compare competing test statistics with limiting non-central chi-square distributions through a suitable Pitman alternative.

Chapter 2 reviews the existing literature on areas that are relevant to this dissertation. This includes the linear model for correlated data and estimation thereof, models for multivariate longitudinal data and the comparison of tests both in the one parameter and multi-parameter testing problem. Chapter 3 presents a detailed review of existing results that are useful in this dissertation. These include estimation in the linear model for correlated data assuming normally distributed errors using maximum

likelihood and restricted maximum likelihood estimation. Measuring test efficiency in the one parameter case and the concept of Pitman efficiency are also presented, including an example of measuring test efficiency in the one parameter case. Test efficiency in the multi-parameter case is reviewed and an example illustrated using two parameters.

Chapter 5 focuses on the potential gain in efficiency that would result from exploiting a Kronecker structured within-subject variance-covariance matrix when it is appropriate. This is done by evaluating the efficiency of a test based on a completely unstructured covariance matrix relative to one based on a Kronecker structured covariance matrix. Using the TARE and CARE to estimate efficiency, a Monte-carlo simulation study is conducted. A second goal of Chapter 5 is to describe a preliminary likelihood ratio test of the hypothesis $H_o : \Sigma_o = \Delta \otimes \Omega$ versus $H_a : \Sigma_o = \Sigma_a$, where Σ_a is an arbitrary covariance matrix. From the simulation study, efficiency is demonstrated to be a function of the covariance parameters defining Δ and Ω. For the design and covariance matrices considered, a gain in efficiency occurs from exploiting the Kronecker product structure. The parameter defining Ω was found to have the greatest impact on efficiency. For testing the hypotheses $H_o : \Sigma_o = \Delta \otimes \Omega$ versus $H_a : \Sigma_o = \Sigma_a$, a likelihood ratio test is incorporated and applied to data on two measures of lung function capacity recorded on subjects in two groups over five years.

Chapter 6 investigates the converse of the situation considered in Chapter 5; specifically, the loss of efficiency from imposing a Kronecker structured covariance matrix in hypothesis testing when it is not appropriate is investigated. To accomplish this, a class of matrices with some degree of departure from the Kronecker product model is introduced. A measure, called the Kronecker product deviation index, is defined. It is used to quantify how far a given variance-covariance matrix departs from Kronecker product structure. A Monte-carlo simulation study using this class of covariance matrices is performed to compare the impact of the Kronecker product deviation index on a test based on imposing a Kronecker product structure, relative to one based on a unstructured covariance matrix. For the design and class of non-Kronecker product covariance matrices considered, a loss of efficiency occurs from imposing the Kronecker product structure.

The power of the test under an assumed Kronecker product model was consistently lower than that of the test based on a unstructured covariance matrix. Also, the difference in power between the two tests was found to increase as the Kronecker product deviation index increased.

Chapter 4 compares the efficiency of ordinary least squares which ignores both cross-sectional and longitudinal correlations to generalised least squares which utilises the within-subject variance-covariance matrix assumed to be of the Kronecker product form. To this end, the efficiency of a test procedure that ignores correlation relative to one that models correlation as the Kronecker product of two matrices is evaluated using the TARE and CARE. Results are presented for two designs (growth curve and repeated measures analysis of variance) and two covariance structures for Ω (compound symmetry and first-order autoregressive). For the designs and covariance matrices considered, a loss of efficiency occurs from ignoring the two sources of correlation. As expected, the loss is greatest when the correlations between the characteristics and between longitudinal measurements on a given characteristic are high in absolute value.

The primary advantage of using the Kronecker product approach to model correlation in multivariate longitudinal data is that it takes into account and separates cross-sectional and longitudinal correlations. It allows one to study the differences in the way characteristics change over time for subjects classified into different groups while simultaneously incorporating correlations that arise both within and across time. Results obtained in this dissertation emphasize the importance of appropriately modelling the variance-covariance matrix. For example, if the underlying Kronecker product covariance structure is exploited, a gain in efficiency will occur in hypothesis testing. Conversely, imposing the Kronecker product covariance structure will result in a loss of efficiency. The loss is most noticeable when the covariance matrix is far from Kronecker product structure, in particular when there is a moderate departure from the null hypothesis. Failing to model correlations that exist both within and across time is shown to be statistically less efficient than if one appropriately accounts for these correlations, especially when the cross-sectional and longitudinal correlations have high absolute values.

Chapter 1

Introduction

1.1 Introduction

Suppose we have data collected on C characteristics over T occasions for I individuals who may be divided into G groups. The analysis of these kind of data, known as multivariate longitudinal data, assuming a Kronecker structured covariance matrix, is considered. In general, multivariate longitudinal data models are concerned with data recorded on several occasions, on individuals receiving different treatments or divided into different classes, such that each record consists of measurements made on a number of response variables or characteristics. The term "multivariate longitudinal" points to the fact that the data are multivariate in the direction of distinct responses, as well as longitudinal. Longitudinal data is defined broadly as data arising from designs in which the response of each unit is observed on two or more occasions. In this context, repeated measures designs, cross-over designs and growth-curve designs are considered variations of the basic longitudinal design.

The longitudinal design is very useful because a wide variety of scientific questions can only be addressed by utilising longitudinal data, including questions concerning the processes of development and aging. For this reason, it is widely used in medical and social science research. This research was motivated by an interest in finding solutions to

commonly occuring problems in the analysis of quality of life data, specifically data arising from cancer clinical trials. Useful references include Olshewski and Schumacher [47]; Zwinderman [82]; Tandon [61]; Schumacher and Olschewski [56]; Cox et al [9] and Hopwood, Stephens and Machin [23].

If the outcome variable is univariate and approximately normally distributed, and the data are balanced and complete, a large class of linear models are available and data analysis is relatively straightforward. When the data are unbalanced and incomplete, as is often the case when dealing with human subjects, most analysis techniques involve an individual formulation of multivariate linear models which can explore tracking of individual characteristics such as the random effects models proposed by Laird and Ware [34] or the structured covariance matrix proposed by Jennrich and Schluchter [25]. Recently, the seemingly unrelated regression model has been applied to longitudinal data, as proposed by Park and Woolson [49].

In many longitudinal studies, it is often of interest to collect a number of different characteristics on each of several occasions. That is, for each individual, multiple measurements are recorded at each time point instead of one. Krzanowski and Marriott [30] note that when individuals are followed up over a period of time, the cost of data collection is almost unaffected by the number of measurements taken at each time. An example is given in Sy, Taylor and Cumberland [60], who describe the relationship between two important immunologic measurements in HIV/AIDS research, namely, CD4 and beta-2-microglobulin. Both variables are measured longitudinally using data from the Los Angeles section of the Multicenter AIDS Cohort study. The resulting data are unique in that correlation arises in two ways: (i) the different characteristics recorded at each time point, and (ii) the same characteristics measured on different occasions.

The techniques for analysing multivariate longitudinal data must in some way take into account these two sources of correlation. However, when faced with multivariate longitudinal data, most researchers tend to analyse each variable or characteristic that has been measured over time separately. Apart from the issues raised by multiple testing, this approach does not in any way take into account the correlation that may exist

between the different variables on each occasion. For researchers interested in analysing this type of longitudinal design from a multivariate perspective, using a model with a Kronecker structured covariance matrix may provide a possible alternative.

Continous data from multivariate longitudinal data designs are sometimes analysed using ordinary least squares. If different subjects are being measured at different times, this might be a reasonable approach. However, when we have the same subjects being measured over time, it is more realistic to assume that the observations within a subject are correlated. The Kronecker product approach represents one way of modelling this correlation. One major advantage of using this approach in the analysis of multivariate longitudinal data is that it takes into account both cross-sectional and longitudinal correlations. Hence, the model allows one to study the differences in the way C characteristics change over time for subjects classified into different groups while simultaneously incorporating correlations that arise both within and across time.

The Kronecker product model assumes that the within-subject variance-covariance matrix can be expressed as the Kronecker product of two matrices. For the Kronecker product covariance structure to be valid, we should be able to determine from the data that the within-subject variance-covariance matrix can be modelled as the Kronecker product of a $C \times C$ matrix and a $T \times T$ matrix. The $C \times C$ matrix represents the covariance matrix between the C characteristics at each time point and is assumed to remain constant over time. This covariance matrix captures the cross-sectional (and consequently the multivariate) component of the data. The $T \times T$ matrix represents the covariance matrix for each of the C characteristics measured on T occasions and is assumed to remain constant for all C characteristics. This covariance matrix captures the longitudinal component of the data. Additionally, homegeneity of the covariance matrices across the levels of the between subjects or grouping factor is assumed.

The model can be written as $y = (\Theta \otimes I_C \otimes X)\lambda + e$, where Θ is the $I \times G$ between subject design matrix for I subjects in G treatment groups; I_C is the $C \times C$ identity matrix; X is the $T \times p$ within subject design matrix where p represents the number of columns in the design matrix; λ is the $pCG \times 1$ vector of unknown parameters and e

is the error vector with covariance matrix $\Sigma = I_I \otimes \Sigma_o$, where I_I is the $I \times I$ identity matrix. Σ_o is the $CT \times CT$ within-subject variance-covariance matrix and takes the form $\Sigma_o = \Delta \otimes \Omega$; Δ is the $C \times C$ covariance matrix for C dependent variables and Ω is the $T \times T$ covariance matrix for T repeated measures on each dependent variable. Σ_o depends on γ_1 and γ_2, the parameter vectors defining Δ and Ω respectively. If one does not assume that the within-subject variance-covariance matrix has the Kronecker product structure, another approach would be to let this matrix be an unstructured $CT \times CT$ matrix. Using this approach, no restrictions are placed on the structure of the common within-subject variance-covariance matrix and it need not be Kronecker product.

Galecki [18] states that one of the advantages of the Kronecker product approach is that it simplifies computation in what might otherwise be a very difficult situation. Partial derivatives, inversion and Cholesky decomposition of the overall variance-covariance matrix are reduced to operations on factor specific matrices with smaller dimensions. Other advantages to using the Kronecker product approach, as outlined by Galecki [18] include:

1. clear and useful interpretation in terms of the contribution of the dimensions involved (characteristics and time) to the overall within-subject variance-covariance matrix,

2. reduction in number of covariance parameters that need to be estimated,

3. and an enrichment of the class of covariance structures available for modelling multivariate longitudinal data.

1.2 Statement of the problem

In the context of multivariate longitudinal data, as in many other settings, one is sometimes faced with the problem of comparing the relative performance of two (or more) tests for testing some multi-parameter hypothesis of interest. For example, in a growth curve setting involving two or more groups, one may wish to test for parallelism.

In a repeated measures analysis of variance setting, one may wish to test for equality of two or more treatment groups. If there are two or more tests available for testing the null hypothesis of interest and one has information available on the relative performance of the tests under consideration, then an informed decision can be made on which test to use for the purpose of statistical inference.

Two criteria of asymptotic relative efficiency have been proposed by Woolson and Sen [79] for the multi-parameter testing problem. The criteria are known as the CARE (Curvature Asymptotic Relative Efficiency) and the TARE (Trace Asymptotic Relative Efficiency). Both criteria may be applied for the comparison of competing test statistics, each with limiting noncentral chi-square distributions utilising a suitable Pitman alternative. The criteria are products of a scalar adjustment function and a term involving the noncentrality parameters. More specifically, the curvature asymptotic relative efficiency is a function of the determinants of the matrices in the noncentrality parameters, while the trace asymptotic relative efficiency is a function of the traces of the matrices in the noncentrality parameters.

Woolson and Sen [79] give an application of these efficiency criteria to the multivariate one-sample location problem. The primary purpose of this research is to apply these two measures of asymptotic relative efficiency to the multivariate longitudinal data problem. Of concern in this work is the comparison of models that utilise the Kronecker product approach to models that do not. Incorporating a test designed to test the null hypothesis that the within-subject variance-covariance matrix has a Kronecker product structure will be considered. An index that can be used to measure how far a given variance-covariance matrix departs from Kronecker product will be presented.

1.3 Motivation

This research was motivated by the analysis of longitudinal data arising from quality of life studies in cancer clinical trials. Most cancer treatments are palliative in nature and patient's quality of life is of primary concern. During the course of cancer treatment,

the quality of life of the cancer patient is closely monitored. The assessment of the effects on quality of life of different treatments in clinical trials is now regarded as an important tool in comparing the effectiveness of different treatments. Quality of life is a multi-dimensional construct comprising the physical, emotional and social well-being of patients.

Additionally, quality of life and its dimensions are not directly observable, hence the need for several items measuring the same latent variable. Quality of life is also a dynamic, time-dependent process resulting in repeated measurements over time per individual. The fact that most patients are usually very ill frequently results in quality of life data sets that are unbalanced and incomplete. The linear model for multivariate longitudinal data with a Kronecker structured covariance matrix presents a reasonable and flexible way of dealing with the complex issues associated with quality of life data.

In practical work carried out in various disciplines, the longitudinal design is very popular and usually involves collecting multiple characteristics on the subjects under study instead of a single characteristic over time. Modelling covariance structure is even more critical in this setting because of the two dimensions involved (characteristics and time). Without specifying a covariance model, $\frac{1}{2}TC(TC+1)$ covariance parameters must be estimated. Modelling the covariance structure using a Kronecker product model results in a tremendous reduction in the number of covariance parameters to be estimated. This may be especially advantageous in studies that result in highly unbalanced and/or missing data, a common problem in designs that involve following subjects over time. However, the validity of this model will depend to a large extent on the special covariance structure that it assumes.

Assuming the Kronecker product structure is valid, we wish to compare the linear model with a Kronecker structured covariance matrix with other approaches typically used for multivariate longitudinal data and discover which advantages, if any, that it offers. We will show that the model is very flexible, with applications to many kinds of longitudinal designs, and offers a rich class of covariance structures. On the other hand, if the Kronecker product structure is not valid, then we also wish to find out the negative

consequences of imposing such a structure.

1.4 Objectives of the study

1. Apply the CARE (Curvature Asymptotic Relative Efficiency) and the TARE (Trace Asymptotic Relative Efficiency), both measures of asymptotic relative efficiency developed for the multiparameter testing problem, to investigate efficiency as it relates to testing hypotheses of interest in multivariate longitudinal data.

2. Investigate the consequences of ignoring correlations that arise both within and across time in multivariate longitudinal data.

3. Incorporate a preliminary test for the null hypothesis that the within-subject variance-covariance matrix has a Kronecker product structure.

4. Assess the gain in efficiency that results from exploiting a Kronecker structured covariance matrix in testing hypotheses of interest in multivariate longitudinal data.

5. To describe an index that can be used to measure the departure of a given variance-covariance matrix from Kronecker product structure.

6. Investigate the consequences of imposing a Kronecker product covariance matrix when there is some departure from the Kronecker product structure.

1.5 Problem domain (Example)

In this section, a real multivariate longitudinal data set is discussed in order to highlight the issues that will be the focus of this dissertation. The data is kindly provided by Dr. Jure Manfreda at the Respiratory Hospital, Health Sciences Center, in Winnipeg, Manitoba. Survey data were collected yearly between 1976 and 1991, using both occupational and non-occupational surveys. The data selected do not represent any particular group and generalization of results to the Manitoba population or any segment

of it should not be attempted. The data set was created for the purpose of developing a methodological (statistical) approach. The discussion presented here will focus on FEV1 and FVC, both measures of lung function capacity. Methods of measuring lung function (spirometry) were the same.

Spirometry refers to the measurement of the forced expiratory vital capacity (FVC) and the expiratory flow rates which occur during a FVC maneuver [16]. A maneuver consists of a subject inhaling as much air as possible, then exhaling it as rapidly and completely as possible. Spirometry is used to detect chronic obstructive pulmonary disease (COPD). COPD is a term commonly used to broadly refer to individuals (usually patients) with non-specific obstructive lung disease. The incidence and prevalence of COPD has increased tremendously in recent years and it has now become a major public health problem. The high prevalence justifies efforts to detect early obstruction of airways caused by COPD. Spirometry is the first test for early detection of COPD, where "early" means before the occurrence of significance symptoms requiring the attendance of a physician. Only spirometry can detect COPD 5-10 years before the onset of significant symptoms. The earlier one can detect airway obstruction, the better the response to therapy.

The most important spirometry variable is FEV1, short for forced expiratory volume in 1 second. We can think of FEV1 as the average flow rate during the first second of the forced vital capacity maneuver. FEV1 is reduced with airflow limitation or obstruction. "Restriction" in lung disorders always means a decrease in lung volume. Spirometry provides a measure of the FVC, the volume of air that can be exhaled after a subject takes as deep a breath as possible. A reduction in the FVC measured by spirometry is consistent with restriction.

From the above discussion, we see that using spirometry, two disorders can be detected. The first disorder is *obstruction*, which refers to reduced flow rates and is detected by a reduced FEV1. The second disorder is *restriction*, which refers to reduced lung volume and detected by a reduced FVC given that obstruction of airways has been excluded. Therefore, spirometry maneuvres are best visualised by graphs that enable

one to simultaneously view flow rates and volumes produced by the maneuvres. If the maneuvers have been conducted over time on the same subjects, then there is need for a statistical procedure that can simultaneously consider flow rates and volumes measured over time.

Demographic variables collected at the beginning of the study included gender, birth date and age at which an individual started smoking. Smoking status represents a time varying covariate and was collected at each time point along with the date of test, height in inches and weight in pounds. Whether an individual was a surface worker or an underground worker represents a time invariant covariate since it remained fixed over the course of the study. The discussion presented here will focus on the potential effects of being a surface or underground worker on lung function capacity as measured by FEV1 and FVC. The data considered will be restricted to the subjects with complete data during the first 5 years of the study ($1976-1980$), resulting in 140 subjects of whom 52 were surface workers and 88 were underground workers. Issues that one must consider in applying the linear model with a Kronecker structured covariance matrix model to this type of data are now presented.

In using the Kronecker structured covariance matrix, the model incorporates two sources of correlation: the correlation that exists between FEV1 and FVC at each time point as well as the correlation that exists over time within each of FEV1 and FVC. A potentially inefficient way to analyze this data would be by ordinary least squares (unless the covariance matrix is known to be $\Sigma_o = \sigma^2 I_{TC}$). This would not only ignore the correlation between FEV1 and FVC at each time point but also the correlation within each of FEV1 and FVC over time. How inefficient this is when the underlying covariance structure between the two variables measured over time is known to take on a Kronecker product form is investigated. Mathematical expressions for the TARE and CARE will be derived and numerical results presented for some specific within-subject design matrices and covariance structures.

Suppose that the within-subject variance covariance matrix Σ_o for FEV1 and FVC is known to have the Kronecker product form, but instead one models the data using an

unstructured $TC \times TC$ matrix. How inefficient is the latter approach in this situation? The answer to this question will be sought by assessing the gain in relative efficiency that may result from taking advantage of the Kronecker product structure rather than using an unstructured $TC \times TC$ matrix. The efficiency of the test based on a completely unstructured covariance matrix relative to a test based on a Kronecker structured covariance matrix is evaluated, employing two measures of asymptotic relative efficiency. A simulation study is conducted to assess the gain in efficiency that may result from taking advantage of the Kronecker product structure.

Another question of considerable practical interest is as follows: how does one know that the variance-covariance matrix for FEV1 and FVC can be modelled as the Kronecker product of two matrices? To find an answer to this question, a preliminary test of the hypothesis that the covariance matrix has a Kronecker product structure versus the hypothesis that the covariance matrix is completely unstructured is presented. The observed significance level and power of this test will also be examined.

So far, we have considered the situation in which the Kronecker product structure is thought to be suitable for FEV1 and FVC. The consequence of ignoring this structure either by the use of ordinary least squares or by fitting a completely unstructured covariance matrix are presented as issues that need further investigation. The converse of this situation is also of interest and needs further investigation, that is, the situation in which the within-subject variance covariance matrix for FEV1 and FVC measured over time is known to deviate from the Kronecker product form. In this case, modelling the data using a completely unstructured covariance matrix may be more suitable than imposing a Kronecker structured covariance matrix. What are the consequences of imposing a Kronecker product structure in testing hypotheses of interest in multivariate longitudinal data? The answer to this question will also be sought. An index that measures how far a given covariance matrix deviates from Kronecker product is a useful measure and is introduced. A simulation study to assess the consequences of imposing a Kronecker structured covariance matrix on hypothesis testing will also be conducted.

1.6 Thesis organisation

In Chapter 2, we review the existing literature on areas that are relevant to this dissertation. This includes the linear model for correlated data and estimation thereof, models for multivariate longitudinal data and the comparison of tests both in the one parameter and multi-parameter testing problem. Chapter 3 presents a detailed review of existing results that are useful in this dissertation. These include estimation in the linear model for correlated data assuming normally distributed errors using maximum likelihood and restricted maximum likelihood estimation. Measuring test efficiency in the one parameter case and the concept of Pitman efficiency will be discussed, including an example of measuring test efficiency in the one parameter case. Test efficiency in the multi-parameter case will also be reviewed and an example illustrated using two parameters.

Chapter 4 investigates the problem of how inefficient ordinary least squares can be. The chapter begins with a detailed look at the formulation of the linear model for multivariate longitudinal data with a Kronecker structured covariance matrix. Since efficiency is defined in terms of hypothesis testing, a discussion of hypothesis testing and power in the linear model for multivariate longitudinal data is presented. Algebraic results for measuring test efficiency using the TARE and CARE are presented and numerical results for some within-subject design matrices and special covariance structures are given. Chapter 5 investigates the potential gain in test efficiency that may result from utilising the Kronecker product structure when it is appropriate. A test of the null hypothesis that the within-subject variance-covariance matrix has a Kronecker product structure is discussed. Algebraic results for measuring test efficiency are presented and numerical results from a simulation study presented.

Chapter 6 investigates the consequences of imposing a Kronecker structured covariance matrix in testing hypotheses of interest from multivariate longitudinal data when it is not appropriate. An index, referred to as the Kronecker product deviation index, which measures how far a given within-subject variance-covariance matrix departs from Kronecker product structure, is introduced and computed for a specially defined class of

matrices. Numerical results from a Monte-carlo simulation study designed to evaluate the consequences of imposing Kronecker structured covariance matrix on testing hypotheses of interest are also presented.

1.7 Simulation Study Overview

Chapters 5 and 6 involve simulating data from the multivariate normal distribution. Chapter 5 assumes a Kronecker product covariance matrix of the form $\Sigma_o = \Delta \otimes \Omega$ for the within-subject variance-covariance matrix. In this chapter, the matrix Δ is defined by parameters ρ_c and γ and the matrix Ω is defined by a parameter ρ_t. The values of the parameters considered in simulating a multivariate normal distribution are: ρ_c from -0.6 to 0.6 by 0.3; γ from 0.5 to 2.0 by 0.5 and ρ_t from 0.1 to 0.9 by 0.1. At each parameter combination, 200 simulation trials are carried out. Depending on the purpose of a simulation, 200 trials may be perceived to be small. For this study, however, interest lies in the overall relationship between the measures of asymptotic relative efficiency and the covariance parameters, and we will see that a sample of 200 at each parameter combination is sufficient to demonstrate the nature of the relationships of interest.

Chapter 6 assumes a non-Kronecker product covariance matrix for the within-subject variance-covariance matrix Σ_o. Σ_o is now defined by covariance parameters σ_{11}, σ_{22} and σ_{12} and correlation parameters ρ_1, ρ_2 and ρ_{12}. The covariance parameters σ_{11}, σ_{22} and σ_{12} are kept fixed at 4, 4 and 2 respectively. The values of the correlation parameters considered in simulating a multivariate normal distribution are: ρ_1 from 0.1 to 0.9 by 0.2; ρ_2 from 0.1 to 0.9 by 0.2 and ρ_{12} from 0.2 to 0.8 by 0.2. As in Chapter 5, 200 simulation trials are carried out at each parameter combination.

Chapter 2

Review of relevant research and theory

2.1 Introduction

The literature on the analysis of a single characteristic measured in a longitudinal design is extensive. The currently available methods cover quantitative data (continous or measured data) as well as qualitative data (binary and count data). Developments that have taken place in the last decade have also made it possible to cope with both unbalanced designs and missing data. Many of these techniques have also been incorporated into statistical software. Sections 2.1 and 2.2 will review the literature for a univariate quantitative characteristic.

The body of literature for multiple characteristics measured in a longitudinal design is rapidly expanding. As with the univariate case, earlier methods in this area were mainly "analysis of variance" based but there is now a move towards more "regression" based methods. The literature for multivariate longitudinal data is reviewed chronologically in section 2.3.

The literature on the comparison of tests is presented in section 2.4. The comparison of two tests for a given situation with the aim of evaluating their relative efficiencies

is a fundamental concept in this dissertation. The review presented is for both the single parameter and multi-parameter settings.

2.2 Linear model for longitudinal data

The models for univariate quantitative longitudinal data are based primarily on the multivariate normal distribution with the repeated observations within an individual assumed to follow a multivariate normal distribution. This means that for the T_i (say) observations on individual i, $i = 1, 2, \ldots, I$, represented by the $T_i \times 1$ vector y_i, we assume that y_i has a multivariate normal distribution with mean vector μ_i and $T_i \times T_i$ variance-covariance matrix Σ_{oi} which is unspecified. In addition, the mean structure for the repeated observations is assumed to be linear, which means that y_i arises from the linear model $y_i = X_i\beta + \epsilon_i$, where X_i is the design matrix for the i^{th} individual and ϵ_i is the vector of deviations with multivariate normal distribution with an unspecified covariance matrix Σ_{oi}. β is a vector of unknown fixed effects. Laird [33] gives several features of the multivariate normal with a linear mean structure that makes it particularly attractive for modelling continous longitudinal data. One of the features mentioned is the fact that the mean vector and covariance matrix are distinct parameters that can be modelled separately.

Ware [75] gives a straightforward discussion of linear models for longitudinal data that include modelling both the expected values of the responses and their covariance structure. The model discussed by Ware [75] is of the form given above, that is $y_i = X_i\beta + \epsilon_i$. This approach to modelling the mean function of y_i as $X_i\beta$ is more direct than the mean value function usually assumed for growth data. For example, the model considered by Rao [53] for a balanced growth-curve data is given by $E[y_i] = A\beta$ where the matrix A is constant over all units representing powers of time or orthogonal polynomials. These model is restrictive in two ways: (i) all units must have the same design in time and (ii) other covariates that are not functions of time cannot be included in the model. Grizzle and Allen [20] generalised Rao's [53] model by appending a vector of covariates,

giving $E[y_i] = A\beta x_i$, where x_i is the vector of covariate values for the i^{th} subject. The model given by Ware [75] is a further generalisation of the Grizzle and Allen [20] model.

Ware [75] also considers possible forms for the within subject variance-covariance matrix Σ_{oi}. Specifically, he discusses three models for the covariance structure: multivariate or unstructured, random effects and the first-order autoregressive model. The unstructured model is suitable when the data are relatively balanced and there are not too many missing values. It is also a reasonable model when the number of observations per unit is not too large compared to the total number of units. When the data are highly unbalanced and/or there are lots of missing data, or when the number of observations per unit is large relative to the total number of units, then structured models for the covariance structure must be considered.

Jennrich and Schluchter [25] provide a detailed discussion of modelling unbalanced and incomplete longitudinal data using structured covariance matrices. They model the expected value of the responses as a linear function of unknown regression parameters as in Ware [75]. The covariance matrix Σ_{oi} is modelled as an arbitrary function of a set of unknown covariance parameters. The covariance structures discussed include the random-effects and the first-order autoregressive models discussed in Ware [75]. Additional structures discussed include independence, compound symmetry, factor analytic, banded or general autoregressive models.

The mixed model is a useful alternative for modelling unbalanced and/or incomplete data. The mixed model approach is a further generalisation of the linear model discussed by Ware [75] and Schluchter [25]. It is formulated so that the probability distribution for the repeated measurements has the same form for every unit but the parameters of that distribution vary from one unit to unit. Laird and Ware [34] discuss a general family of random effects models.

Although the models proposed by Ware [75] and Schluchter [25] are useful, one cannot always model serial correlation. This is because their approach requires one to choose a particular covariance structure and this need not be the first-order autoregressive structure. Modelling serial correlation is important in longitudinal data, especially when

measurements are collected over extended periods of time. For this reason, Diggle [13] provides another choice of covariance structure useful for longitudinal data. The special correlation structure provides parameters for measurement error, variation between experimental units and serial correlation within units. In using this model, one must not only include parameters for measurement error and variation between experimental units, but also provide for serial correlation between measurements within a unit.

2.3 Estimation in the linear model for longitudinal data

Estimation of the parameters defining the $p \times 1$ mean parameter vector β and the within subject covariance matrix Σ_{oi} denoted by the $q \times 1$ vector ϕ has been discussed by various authors. Under the assumption of multivariate normality of the repeated observations within an individual, estimation procedures are mostly likelihood-based. Ware [75] notes that when the data are balanced and complete, closed form maximum likelihood estimators of β and θ are easily available. However, in the more typical situation involving unbalanced and/or incomplete data, closed form solutions do not exist and iterative procedures must be used.

Jennrich and Schluchter [25] consider maximum likelihood and restricted maximum likelihood estimation using the Newton-Raphson and Fisher Scoring algorithms as well as the Estimation-Maximisation (EM) algorithm. Diggle [13] discusses maximum likelihood estimation using the simplex algorithm of Nelder and Mead [45]. Laird, Lange and Stram [32] consider the use of the Estimation-Maximisation algorithm for both maximum likelihood and restricted maximum likelihood estimation.

For the mixed effects model, estimation of the mean and covariance parameters is usually accomplished using iterative procedures. Laird and Ware [34] discuss using the Estimation-Maximisation algorithm to obtain both maximum likelihood and restricted maximum likelihood estimators as well as a combination of empirical Bayes and maximum likelihood estimators. Jennrich and Sampson [24] discuss three algorithms for

maximum likelihood estimation of mean and variance components in a mixed analysis of variance model. These include the Newton-Raphson algorithm, the Fisher Scoring algorithm and the Hemmerle and Hartley algorithm. However, their work is more geared towards analysis of variance and may be of limited use in the longitudinal setting. Lindstrom and Bates [39] also consider the Newton-Raphson and Estimation-Maximisation algorithms for the random effects model using both maximum likelihood and restricted maximum likelihood estimation. Wolfinger, Tobias and Sall [77] give several algorithms for computing gaussian likelihoods or restricted likelihoods for the mixed model.

2.4 Models for multivariate longitudinal data

The literature on the analysis of several characteristics measured in a longitudinal design continues to grow steadily, more so in the past fifteen years. The analysis of such data has usually been confined to analysing each of the response variables separately. Boik [6] notes that separate analyses may be appropriate if the dependent variables are uncorrelated or if they are measures of distinct theoretical constructs. A combined analysis (or multivariate analysis) is necessary if the dependent variables are functionally related, as is usually the case. As we shall infer from the discussion that follows, a lot of focus in this area has been in the growth curve setting.

Pothoff and Roy [52] provide an extension of the usual multivariate analysis of variance model and show that it applies to many kinds of problems, including growth curve. One application of the extended model is to the situation where more than one characteristic is measured on each occasion. The situation described is as follows using the notation in their paper: m groups of animals are being measured at, say, $q\prime$ points in time, more than one characteristic associated with growth is measured at each of the $q\prime$ time points. Pothoff and Roy [52] do not impose any structure on the within-subject variance covariance matrix. An application of the extended model is demonstrated using data collected by investigators at the University of North Carolina Dental School. Measurements were made on 11 girls and 16 boys at 4 different ages (8, 10, 12 and

14). Each measurement is the distance, in mm, from the centre of the pituitary to the pteryomaxillary fissure.

Reinsel [54] also considers the longitudinal design where several characteristics are measured on each individual at each time point, assuming a multivariate random effects model for the repeated measurements. The covariance structure is the multivariate analog of the compound symmetry pattern used in the univariate case. Reinsel's model applies only to balanced data and considers estimation under a restricted covariance structure. Reinsel [54] notes that if no special assumptions are made about the structure of the covariance matrix, then we have the general model considered by Pothoff and Roy [52] among others but involving multiple measurements. Applications of the proposed model are shown using the growth curve data introduced by Pothoff and Roy [52] and medical data from the Department of Anesthesiology, University of Wisconsin-Madison, from an experiment designed to study the effects of certain anesthetics on dogs.

Wang [73] examines the relationship between the mixed-model analysis and multivariate approach to a repeated measures design with multiple measurements. In the multivariate setting, the two methods are referred to as the multivariate mixed model approach and the doubly multivariate model approach, respectively. The two approaches, like the Pothoff and Roy [52] and Reinsel [54] models, differ in the assumptions underlying the models. The multivariate approach imposes no structure on the correlation structure and represents one extreme on the spectrum characterising covariance structure parsimony. The mixed-model approach imposes a structure based on the assumption of the mixed-effects model. Thomas [63] also considers the multivariate mixed model analysis for multivariate repeated measures. This model is an extension of the univariate mixed model approach whose validity depends on a special covariance structure for the multivariate repeated meausures discussed in Reinsel [54]. Thomas [63] derives conditions for the validity of the multivariate mixed model analysis and presents a test for determining whether or not given data satisfy these conditions.

Boik [6] reviews both the multivariate mixed model approach and the doubly multivariate model approach for analysing repeated measures on multivariate responses.

Three new results concerning the multivariate mixed model are also presented. One of the results presented is that, given multivariate normality, a condition called multivariate sphericity of the covariance matrix is both necessary and sufficient for the validity of the multivariate mixed model analysis. A likelihood ratio test can be employed to test for departure from multivariate sphericity. Boik [7] compares the two models for multivariate repeated measures: the doubly multivariate model and the p-variate generalisation of Scheffe's mixed model (the multivariate mixed model). Boik [7] points out that requiring multivariate sphericity for the multivariate mixed model approach is in fact a disadvantage and that even small departures from multivariate sphericity inflate the size of multivariate mixed model tests. Boik [7] notes that the question of how to model the covariance parameters is even more critical because without specifying a covariance model, there are $\frac{1}{2}pt(p+t)$ covariance parameters to estimate where p is the number of dependent variables measured and t is the number of occasions. The model presented by Boik [7] is a special case of Reinsel's [54] multivariate random effects growth curve model. Vasdekis [67] generalises the model proposed by Reinsel [54] to the mixed effects model with an arbitrary number of random effects and considers maximum likelihood and restricted maximum likelihood estimation. Reinsel [54] considered a single random factor.

More recent developments in the analysis of multivariate longitudinal data include the work of Rochon [55], who considers bivariate repeated measures and applies a generalised estimating equations approach to relate each set of repeated measures to important explanatory variables. Zhang [81] presents several choices of structured covariance matrices for analysing multivariate repeated measures and provides a computational algorithm using Gibbs sampling. Matsuyama and Ohashi [41] develop a bivariate mixed effects model that is a generalisation of the univariate mixed model discussed by Laird and Ware [34]. Estimation is achieved via the Gibbs sampler. Sy, Taylor and Cumberland [60] present another model for multivariate repeated measures that incorporates random effects, correlated stochastic processes and measurement errors. Maximum likelihood estimation is used to obtain estimates of the fixed effects and covariance parameters. Diaz and Johnson [11] consider the situation when the patterned within subject

covariance matrix can be reduced to a diagonal form, for example, the Wiener stochastic process. Diagonalisation results in a reduction in the number of parameters to be estimated. Maximum likelihood estimates are derived for the mean vector and covariance matrices.

2.5 The comparison of tests

Suppose we have two α level tests available for testing a given set of null and alternative hypotheses. For comparing the two tests, a reasonable measure of relative efficiency is to use

$$\mathrm{RE}(\alpha, \beta, \theta_a) = \frac{n_1}{n_2},$$

where n_1 and n_2 are the minimum sample sizes required for the two tests at level α to have the same power β against a fixed alternative $\theta = \theta_a$. To study this ratio for all values of α, β and θ_a would be a very complicated study, as stated in Woolson [80], hence the restriction to asymptotic results in test comparison. An alternative to the approach described above was considered and generalised by Noether [46]. He proposed that a sequence of alternative hypotheses depending on the sample size N be chosen such that the limit of this sequence approaches the null hypothesis and simultaneously, the power is bounded away from one. By Pitmans's theorem, the null and alternative hypotheses are stated as

$$H_o : \theta = \theta_o \quad vs \quad H_a : \theta = \theta_N = \theta_o + N^{-\delta}\lambda,$$

where λ is a fixed but arbitrary constant considered to be small and $\delta > 0$. Comparing two tests using Pitman efficiency is discussed in detail in Kendall and Stuart [26] and Gibbons [19].

Now consider the multiparameter testing problem where, in the Pitman sense, the null and alternative hypotheses are given by

$$H_o : \theta = \theta_o \quad vs \quad H_a : \theta = \theta_N = \theta_o + N^{-\delta}\lambda,$$

where now $\theta^T = (\theta_1, \theta_2, \ldots, \theta_r)$ are the location parameters for r populations and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ is a fixed but arbitrary non-null vector considered to be small and $\delta > 0$. How does one now measure the relative efficiency of two tests that are available for testing the above hypotheses? Woolson [80] first considered this problem, focusing on comparing two test statistic sequences which have limiting chi-square distributions with possibly different degrees of freedom. He suggested and justified some measures of asymptotic relative efficiency that may be used in comparing the two test sequences. These measures are presented in detail in Woolson and Sen [79]. The new measures of asymptotic relative efficiency for the multiparameter testing problem are:

1. Local asymptotic relative efficiency (LARE),
2. Curvature asymptotic relative efficiency (CARE) and
3. Trace asymptotic relative efficiency (TARE).

The three measures depend on the level of significance of the test α and the degrees of freedom for the two test statistics. The CARE and TARE are "average" measures of efficiency, independent of the direction of approach of θ_N to θ_o. The CARE is a function of the ratio of the determinants of the non-centrality parameters while the TARE is a function of the ratio of the traces of the non-centrality parameters. An application of the TARE and CARE to the one-sample location problem is discussed in Woolson [78].

Chapter 3

A detailed review of relevant concepts

3.1 Estimation in the linear model for correlated data assuming normally distributed errors

The matrix formulation of the linear model for correlated data is given by:

$$\underset{TI\times 1}{y} = \underset{TI\times p}{X}\ \underset{p\times 1}{\beta} + \underset{TI\times 1}{\epsilon}, \tag{3.1}$$

with:

$$\mathrm{Cov}(y) = \Sigma = \begin{pmatrix} \Sigma_o & 0 & \dots & 0 \\ 0 & \Sigma_o & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma_o \end{pmatrix}, \tag{3.2}$$

where Σ is a block-diagonal matrix with $T \times T$ non-zero blocks Σ_o, each representing the covariance matrix of the vector of observations on a single subject. Suppose we re-parameterise Σ_o:

$$\Sigma_o = \sigma^2 V_o, \tag{3.3}$$

so that:

$$\Sigma = \sigma^2 V, \tag{3.4}$$

where σ^2 is a scale factor. Hence,

$$\Sigma = \sigma^2 V = \begin{pmatrix} \Sigma_o & 0 & \dots & 0 \\ 0 & \Sigma_o & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma_o \end{pmatrix} = \begin{pmatrix} \sigma^2 V_o & 0 & \dots & 0 \\ 0 & \sigma^2 V_o & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 V_o \end{pmatrix}. \tag{3.5}$$

Under this specifications, the linear model for correlated data treats y as a realization of a multivariate normal random vector with

$$Y \sim \text{MVN}(X\beta, \sigma^2 V). \tag{3.6}$$

3.1.1 Maximum likelihood estimation

Consider simultaneous estimation of the parameters of interest β and the covariance parameters σ^2 and V_o using the likelihood function given by

$$f_Y(y; \beta, \Sigma) = 2\pi^{-\frac{TI}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)). \tag{3.7}$$

Substituting $\sigma^2 V$ for Σ, we obtain

$$f_Y(y : \beta, \sigma^2, V) = 2\pi^{-\frac{TI}{2}} |\sigma^2 V|^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta)). \tag{3.8}$$

Noting that

$$|\Sigma|^{\frac{1}{2}} = \sqrt{|\sigma^2 V|} = \sigma^{TI} |V_o|^{\frac{I}{2}}, \tag{3.9}$$

the likelihood function using the re-parameterised form of Σ is now given by

$$L(\beta, \sigma^2, V_o) = 2\pi^{-\frac{TI}{2}} (\sigma^2)^{-\frac{TI}{2}} |V_o|^{-\frac{I}{2}} \exp(-\frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta)). \tag{3.10}$$

The log-likelihood is therefore given by

$$\begin{aligned}&\log L(\beta, \sigma^2, V_o)\\ &= -\frac{TI}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| - \frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta). \end{aligned} \tag{3.11}$$

To find the maximum likelihood estimators of β, σ^2 and V, proceed as follows:

1. Fix V_o and find the maximum likelihood estimator of β by differentiating equation (3.11) with respect to β and setting the equation to zero. It can be shown that

$$\begin{aligned}\frac{\partial \log L(\beta, \sigma^2, V_o)}{\partial \beta} &= -2X'V^{-1}y + 2X'V^{-1}X\beta \\ &\stackrel{set}{=} 0.\end{aligned} \tag{3.12}$$

   Solving the above equation, we obtain

$$\begin{aligned}X'V^{-1}y &= X'V^{-1}X\beta,\\ \hat{\beta}_{V_o} &= (X'V^{-1}X)^{-1}X'V^{-1}y.\end{aligned} \tag{3.13}$$

2. Substitute $\hat{\beta}_{V_o}$ given by equation (3.13) into equation (3.11) and obtain

$$\begin{aligned}&l(\hat{\beta}_{V_o}, \sigma^2, V_o)\\ &= -\frac{TI}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| - \frac{1}{2\sigma^2}(y - X\hat{\beta}_{V_o})'V^{-1}(y - X\hat{\beta}_{V_o})\\ &= -\frac{TI}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| - \frac{1}{2\sigma^2}\mathrm{RSS}(V_o),\end{aligned} \tag{3.14}$$

   where

$$\mathrm{RSS}(V_o) = (y - X\hat{\beta}_{V_o})'V^{-1}(y - X\hat{\beta}_{V_o}). \tag{3.15}$$

3. Now maximise equation (3.14) with respect to σ^2 for fixed V_o to obtain the maximum likelihood estimation of σ^2; that is,

$$\begin{aligned}\frac{\partial l}{\partial \sigma^2} &= -\frac{TI}{2\sigma^2} + \frac{\mathrm{RSS}(V_o)}{2\sigma^4}\\ &\stackrel{set}{=} 0.\end{aligned} \tag{3.16}$$

Solving the above equation gives

$$\hat{\sigma}^2_{V_o} = \frac{\text{RSS}(V_o)}{TI} \tag{3.17}$$

4. Substitute $\hat{\sigma}^2_{V_o}$ given by equation (3.17) into equation (3.14) to obtain a reduced log-likelihood for V_o, that is

$$\begin{aligned} &l(\hat{\beta}_{V_o}, \hat{\sigma}^2_{V_o}, V_o) \\ &\quad = -\frac{I}{2}\log|V_o| - \frac{TI}{2}\log \text{RSS}(V_o) - \frac{TI}{2}\log 2\pi + \frac{TI}{2}\log TI - \frac{TI}{2}. \end{aligned} \tag{3.18}$$

5. The reduced log-likelihood is now a function of the q unknown parameters in V_o and in simplified form is given by

$$\begin{aligned} l_r(V_o) &= l(\hat{\beta}_{V_o}, \hat{\sigma}^2_{V_o}, V_o) \\ &= -\frac{I}{2}\log|V_o| - \frac{TI}{2}\log \text{RSS}(V_o) - \frac{TI}{2}\{1 + \log 2\pi - \log TI\} \\ &= -\frac{I}{2}\log|V_o| - \frac{TI}{2}\log \text{RSS}(V_o) - \frac{TI}{2}\{1 + \log\left(\frac{2\pi}{TI}\right)\}. \end{aligned} \tag{3.19}$$

Ignoring the constants in equation (3.19) above, we have

$$\begin{aligned} l_r(V_o) &= l(\hat{\beta}_{V_o}, \hat{\sigma}^2_{V_o}, V_o) \\ &= -\frac{1}{2}\{TI\log \text{RSS}(V_o) + I\log|V_o|\}. \end{aligned} \tag{3.20}$$

6. Maximisation of equation (3.19) with respect to V_o yields $\hat{V}_o$.

7. Substitute $\hat{V}_o$ into equation (3.13) and obtain

$$\hat{\beta}_{\hat{V}_o} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y. \tag{3.21}$$

8. Substitute $\hat{V}_o$ into equation (3.17) and obtain

$$\hat{\sigma}^2_{\hat{V}_o} = \frac{\text{RSS}(\hat{V}_o)}{TI} \tag{3.22}$$

In general, maximisation of $l_r(V_o)$ given by equation (3.19) will require numerical optimisation techniques such as the Newton-Raphson algorithm. The dimensionality of the optimisation process will depend on what structure has been imposed on Σ_o.

3.1.2 Restricted maximum likelihood estimation

Unfortunately, maximum likelihood estimation as presented in the previous section produces biased estimators of variance components. For example, in the classical linear model given by

$$y = X\beta + \epsilon, \tag{3.23}$$

where X is a $n \times p$ matrix of full rank, β is $p \times 1$ vector of unknown parameters and ϵ is an $n \times 1$ normally distributed random vector with mean 0 and variance $\sigma^2 I_{n\times n}$, the maximum likelihood estimator of β is

$$\hat{\beta} = (X'X)^{-1}X'y \tag{3.24}$$

and the maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = \mathrm{SS}_{\mathrm{res}}/n, \tag{3.25}$$

which is biased. $\mathrm{SS}_{\mathrm{res}}$ is the residual sum of squares given by

$$(y - X\hat{\beta})'(y - X\hat{\beta}). \tag{3.26}$$

The unbiased estimator is given by

$$s^2 = \mathrm{SS}_{\mathrm{res}}/(n - p), \tag{3.27}$$

which in fact, is the restricted maximum likelihood estimator of σ^2 under this model. Note that p is the number of parameters in β and hence the difference between $\hat{\sigma}^2$ and s^2 is that the former has not taken into account the fact that β is also estimated while the latter does. In general, restricted maximum likelihood estimation is a modified approach which takes into account the fact that β is also estimated in estimating variance

components and essentially yields unbiased estimators. Recall from the previous section that the log-likelihood for the data $y_1, y_2, \ldots, y_I$ is

$$\begin{aligned} \log L(\beta, \phi) &= L_I(\beta, \phi) \\ &= -\frac{TI}{2}\log 2\pi - \frac{I}{2}\log|\Sigma_o| - \frac{1}{2}\sum_{i=1}^{I}(y_i - X_i\beta)'\Sigma_o^{-1}(y_i - X_i\beta), \end{aligned} \quad (3.28)$$

where ϕ is the vector of covariance parameters.

Restricted maximum likelihood estimation maximises the part of the likelihood which is invariant to β, that is, the restricted maximum likelihood estimator is defined as the maximum likelihood estimator based on a linearly transformed set of data $y^* = Ay$ such that the distribution of y^* does not depend on β. Harville [22] calls the elements of A "error contrasts". This is equivalent to saying that the restricted maximum likelihood estimator maximises the likelihood of a vector of linear combinations of the observations which are invariant to $X\beta$. If we use the data vector y to estimate ϕ, then as noted before, the maximum likelihood estimator of ϕ takes no account of the loss in degrees of freedom resulting from estimating β. Patterson and Thompson [51] proposed a modified maximum likelihood technique which does not suffer from this defect. The technique proposed consists of maximising the likelihood function associated with a specified set of $(TI - p)$ linearly independent error contrasts rather than the full likelihood function given earlier.

The linear model for correlated data (using the matrix formulation of the model) treats y as a realization of a multivariate normal random vector with

$$Y \sim \text{MVN}(X\beta, \Sigma). \quad (3.29)$$

Note that for convenience, σ^2 has been reabsorbed into V and hence we are now using Σ and not $\sigma^2 V$. Also, $\Sigma = \Sigma(\phi)$, where ϕ is a $q \times 1$ vector used to characterise Σ. The likelihood function is therefore given by

$$f_Y(y; \beta, \Sigma) = 2\pi^{-\frac{TI}{2}}|\Sigma|^{-\frac{1}{2}}\exp(-\frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)), \quad (3.30)$$

and the log-likelihood is given by

$$\begin{aligned}&\log L(\beta, \Sigma)\\ &\quad = -\frac{TI}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta).\end{aligned} \tag{3.31}$$

If β is estimated by $\hat{\beta}$ for fixed Σ, then

$$\begin{aligned}&\log L(\hat{\beta}, \Sigma)\\ &\quad = -\frac{TI}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\hat{\beta})'\Sigma^{-1}(y - X\hat{\beta}),\end{aligned} \tag{3.32}$$

and this is the reduced log-likelihood used to find the maximum likelihood estimator of Σ. So what form does the reduced log-likelihood used to find the restricted maximum likelihood estimator of Σ take, and how does it differ from equation (3.32)?

Definition 3.1.1 (Error Contrast) *A linear combination $a'y$ of the observations such that*

$$\begin{aligned}E(a'y) &= 0, \text{ that is,}\\ a'X &= 0,\end{aligned}$$

is called an error contrast.

The maximum possible number of linearly independent error contrasts in any such set is $(TI - p)$. Define the $TI \times TI$ matrix $A = I_{(TI)} - X(X'X)^{-1}X'$. Also, define the $TI \times (TI - p)$ matrix B to satisfy $A = BB'$ and $B'B = I_{(TI-p)}$. The vector $w = B'y$ provides a particular set of $(TI - p)$ linearly independent error contrasts.

Proposition 3.1.1 *$w = B'y$ is an error contrast.*

The proof is as follows:

$$\begin{aligned}E(B'y) &= B'E(y)\\ &= B'X\beta\end{aligned}$$

$$
\begin{aligned}
&= I(B'X\beta) \\
&= (B'B)(B'X\beta) \quad (\text{since } I = B'B) \\
&= (B')(BB')(X\beta) \\
&= B'AX\beta \quad (\text{since } A = BB') \\
&= B'\{I - X(X'X)^{-1}X'\}X\beta \\
&= B'\{X - X(X'X)^{-1}X'X\}\beta \\
&= B'\{X - X\}\beta \\
&= 0
\end{aligned}
$$

The likelihoods based on $w = B'y$ and $w^* = B'^*y$, where B'^* is an error contrast matrix, are proportional. This implies that the restricted maximum likelihood estimator based on $B'y$ and B'^*y are identical. In general, the likelihood function associated with any other set of $(TI - p)$ linearly independent error contrasts is proportional to that associated with w. So what is the likelihood function associated with $w = B'y$? Denote the likelihood function associated with w by $f(B'y|\phi)$, where $f(.\,|\phi)$ is the probability density function of w indexed by ϕ. We now proceed to find a convenient expression for $f(B'y|\phi)$.

For fixed ϕ, the maximum likelihood estimator of β is the generalised least squares estimator given by

$$
\begin{aligned}
\hat{\beta} &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \\
&= G'y. \qquad (3.33)
\end{aligned}
$$

From the above equation, we obtain

$$
\begin{aligned}
G &= \{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\}' \\
&= \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}. \qquad (3.34)
\end{aligned}
$$

Denoting the probability density function of $\hat{\beta} = G'y$ by $f_{\hat{\beta}}(.\,|\phi, \beta)$, it can be shown that since $\hat{\beta}$ is distributed as $N \sim (\beta, (X'\Sigma^{-1}X)^{-1})$, then

$$f_{\hat{\beta}}(G'y|\phi, \beta) =$$

$$(2\pi)^{-\frac{p}{2}}|X'\Sigma^{-1}X|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\hat{\beta}-\beta)'(X'\Sigma^{-1}X)(\hat{\beta}-\beta)\}. \tag{3.35}$$

If we now denote the probability density function of y by $f_Y(y|\phi,\beta)$, then

$$f_Y(y|\phi,\beta) = (2\pi)^{-\frac{TI}{2}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-X\beta)'(\Sigma)^{-1}(y-X\beta)\}. \tag{3.36}$$

Now define z such that

$$z = Ty = [B,G]y, \tag{3.37}$$

where T is a $TI \times TI$ matrix and consider the transformation from y to Ty. Now, $y = T^{-1}z$ and the Jacobian of the transformation is given by

$$|J| = |\frac{d\,y}{d\,z}| = |\frac{d\,(T^{-1}z)}{d\,z}| = |T^{-1}| = |T|^{-1}. \tag{3.38}$$

We now derive $|T|$.

We know that for a square matrix such as T,

$$\begin{aligned} |T'T| &= |T'||T| \\ &= |T||T| \\ &= |T|^2, \end{aligned}$$

so that

$$|T| = |T'T|^{\frac{1}{2}}. \tag{3.39}$$

Since $T = [B,G]$, by using result (3.39) we obtain

$$\begin{aligned} |T| &= |[B,G]|^{\frac{1}{2}} \\ &= |[B,G]'[B,G]|^{\frac{1}{2}} \\ &= |[B',G'][B,G]|^{\frac{1}{2}} \\ &= \begin{vmatrix} B'B & B'G \\ G'B & G'G \end{vmatrix}^{\frac{1}{2}}. \end{aligned}$$

Definition 3.1.2 (Schur Complements) *If*

$$\begin{aligned} A &= \begin{pmatrix} E & F \\ G & H \end{pmatrix} \quad \text{and} \\ S &= H - GE^{-1}F, \quad \text{then} \\ |A| &= \begin{vmatrix} E & F \\ G & H \end{vmatrix} \\ &= |E||S| \end{aligned}$$

Using the above definition of Schur complements, we have

$$\begin{aligned} |T|^2 &= \begin{vmatrix} B'B & B'G \\ G'B & G'G \end{vmatrix} \\ &= |B'B||G'G - G'B(B'B)^{-1}B'G| \\ &= |I_{(TI-p)}||G'G - G'BI_{(TI-p)}{}^{-1}B'G| \quad (\text{since} \quad B'B = I_{(TI-p)}) \\ &= |G'G - G'BB'G| \\ &= |G'G - G'AG| \quad (\text{since} \quad BB' = A) \\ &= |G'G - G'\{I - X(X'X)^{-1}X'\}G| \\ &= |G'G - G'G + G'X(X'X)^{-1}X'G| \\ &= |(X'X)^{-1}X'G| \quad (\text{since} \quad G'X = I) \\ &= |(X'X)^{-1}| \quad (\text{since} \quad X'G = I), \end{aligned}$$

hence

$$|T| = |(X'X)^{-1}|^{\frac{1}{2}} = |(X'X)|^{-\frac{1}{2}} \neq 0. \tag{3.40}$$

The probability density function of $z = Ty$ is therefore given by

$$\begin{aligned} g(z) &= |J| f_Y(T^{-1}z) \\ &= \frac{1}{|T|} f_Y(y|\phi, \beta) \end{aligned} \tag{3.41}$$

Also,

$$g(z) = f(G'y|\phi,\beta)f(B'y|\phi), \tag{3.42}$$

since $w = B'y$ and $\hat{\beta} = G'y$ are independent, which is proved by showing that the covariance between w and $\hat{\beta}$ is zero:

$$\begin{aligned}
\mathrm{cov}[\mathrm{w},\hat{\beta}] &= E\{(w - E(w))(\hat{\beta} - E(\hat{\beta}))\} \\
&= E[w(\hat{\beta} - \beta)'] \\
&= E[B'y(G'y - \beta)'] \\
&= E[B'y(y'G - \beta')] \\
&= E[B'yy'G] - E[B'y\beta'] \\
&= B'E[yy']G - E[B'y\beta'] \\
&= B'[\mathrm{var(y)} + \mathrm{E(y)E(y')}]\mathrm{G} - \mathrm{B'X}\beta\beta' \quad (\text{since} \quad \mathrm{E(y)} = \mathrm{X}\beta) \\
&= B'[\Sigma + (X\beta)(X\beta)']G \quad (\text{since} \quad B'X = 0) \\
&= B'\Sigma G + B'(X\beta)(X\beta)'G \\
&= B'\Sigma G \quad (\text{since} \quad B'X = 0) \\
&= B'\Sigma[\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}] \\
&= B'\Sigma\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \\
&= B'X(X'\Sigma^{-1}X)^{-1} \\
&= 0 \quad (\text{since} \quad B'X = 0)
\end{aligned}$$

Equating equations (3.41) and (3.42), we obtain

$$f(z) = \frac{1}{|T|}f_Y(y|\phi,\beta) = f(G'y|\phi,\beta)f(B'y|\phi), \tag{3.43}$$

and hence

$$f(B'y|\phi) = \frac{f_Y(y|\phi,\beta)}{|T|f(G'y|\phi,\beta)}. \tag{3.44}$$

The probability density function of $w = B'y$ is therefore given by:

$$\begin{aligned} & f(B'y|\phi) \\ &= \frac{equation \quad (3.36)}{|T|equation \quad (3.35)} \\ &= \frac{(2\pi)^{-\frac{Tf}{2}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-X\beta)'(\Sigma)^{-1}(y-X\beta)\}}{(2\pi)^{-\frac{p}{2}}|X'\Sigma^{-1}X|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\hat{\beta}-\beta)'(X'\Sigma^{-1}X)(\hat{\beta}-\beta)\}|(X'X)|^{-\frac{1}{2}}}. \end{aligned} \tag{3.45}$$

The following result will now be used to simplify equation (3.45):

$$\begin{aligned} & (y-X\beta)'(\Sigma)^{-1}(y-X\beta) \\ &= (y-X\hat{\beta})'(\Sigma)^{-1}(y-X\hat{\beta}) + (\hat{\beta}-\beta)'(X'\Sigma^{-1}X)(\hat{\beta}-\beta). \end{aligned} \tag{3.46}$$

The proof is as follows:

$$\begin{aligned} & (y-X\beta)'(\Sigma)^{-1}(y-X\beta) \\ &= \{y-X\beta+X\hat{\beta}-X\hat{\beta}\}'\Sigma^{-1}\{y-X\beta+X\hat{\beta}-X\hat{\beta}\} \\ &= \{y-X\hat{\beta}+X(\hat{\beta}-\beta)\}'\Sigma^{-1}\{y-X\hat{\beta}+X(\hat{\beta}-\beta)\} \\ &= \{y-X\hat{\beta}+X(\hat{\beta}-\beta)\}'\{\Sigma^{-1}(y-X\hat{\beta})+\Sigma^{-1}X(\hat{\beta}-\beta)\} \\ &= \{(y-X\hat{\beta})'+(\hat{\beta}-\beta)'X'\}\{\Sigma^{-1}(y-X\hat{\beta})+\Sigma^{-1}X(\hat{\beta}-\beta)\} \\ &= (y-X\hat{\beta})'\Sigma^{-1}(y-X\hat{\beta})+(y-X\hat{\beta})'\Sigma^{-1}X(\beta-\hat{\beta}) \\ &+ (\hat{\beta}-\beta)'X'\Sigma^{-1}(y-X\hat{\beta})+(\hat{\beta}-\beta)'(X'\Sigma^{-1}X)(\hat{\beta}-\beta) \\ &= (y-X\hat{\beta})'\Sigma^{-1}(y-X\hat{\beta})+(\hat{\beta}-\beta)'(X'\Sigma^{-1}X)(\hat{\beta}-\beta) \\ &+ 2(\hat{\beta}-\beta)'X'\Sigma^{-1}(y-X\hat{\beta}) \end{aligned} \tag{3.47}$$

Substituting $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ into the last expression of (3.47) above, we have

$$\begin{aligned} & 2(\hat{\beta}-\beta)'X'\Sigma^{-1}(y-X\hat{\beta}) \\ &= 2\{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y-\beta\}'X'\Sigma^{-1}\{y-X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y\} \\ &= 2\{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y-\beta\}'\{X'\Sigma^{-1}y-X'\Sigma^{-1}y\} \\ &= 0. \end{aligned}$$

Substituting 0 for the last expression of (3.47), we obtain result (3.46). Using result (3.46), equation (3.45) simplifies to:

$$f(B'y|\phi) = \tag{3.48}$$
$$(2\pi)^{-\frac{(TI-p)}{2}}|(X'X)|^{\frac{1}{2}}|\Sigma|^{-\frac{1}{2}}|X'\Sigma^{-1}X|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-X\hat{\beta})'\Sigma^{-1}(y-X\hat{\beta})\}.$$

The log-likelihood is therefore given by

$$\log L^*(\hat{\beta},\Sigma) =$$
$$-\frac{(TI-p)}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(y-X\hat{\beta})'\Sigma^{-1}(y-X\hat{\beta}) + \frac{1}{2}\log|X'X|$$
$$-\frac{1}{2}\log|X'\Sigma^{-1}X|. \tag{3.49}$$

and this is the reduced log-likelihood used to find the restricted maximum likelihood estimator of Σ. It is very similar to equation (3.32), the reduced log-likelihood used to find the maximum likelihood estimator of Σ. The difference between equations (3.32) and (3.49) is that the latter has a new piece given by $|X'\Sigma^{-1}X|^{-\frac{1}{2}}$ that is very closely related to var$(\hat{\beta})$. The other additional piece in equation (3.49) corresponds to the Jacobian of the transformation $|(X'X)|^{-\frac{1}{2}}$ which does not depend on any of the parameters of the model and can therefore be ignored in making inferences for β or ϕ.

The following results will be used in what follows:

1. $|\Sigma|^{\frac{1}{2}} = \sqrt{|\sigma^2 V|} = \sigma^{TI}|V_o|^{\frac{I}{2}}$.

2.
$$\begin{aligned} |X'\Sigma^{-1}X|^{\frac{1}{2}} &= \sqrt{|X'(\sigma^2V)^{-1}X|} \\ &= \sqrt{|X'\sigma^{-2}V^{-1}X|} \\ &= \sigma^{-p}|X'V^{-1}X|^{\frac{1}{2}}. \end{aligned}$$

Substituting $\sigma^2 V$ for Σ, (3.48) becomes

$$f(B'y|\phi) =$$
$$(2\pi)^{-\frac{(TI-p)}{2}}|(X'X)|^{\frac{1}{2}}(\sigma^2)^{-\frac{TI}{2}}|V_o|^{-\frac{I}{2}}(\sigma^2)^{\frac{p}{2}}|X'V^{-1}X|^{-\frac{1}{2}} \times$$
$$\exp\{-\frac{1}{2\sigma^2}(y-X\hat{\beta})'V^{-1}(y-X\hat{\beta})\}, \tag{3.50}$$

and the log-likelihood given by equation (3.49) becomes

$$\begin{aligned}\log L^*(\beta, \sigma^2, V_o) = \\ -\frac{(TI-p)}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| + \frac{p}{2}\log\sigma^2 - \frac{1}{2}\log|X'V^{-1}X| \\ -\frac{1}{2\sigma^2}(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}).\end{aligned} \tag{3.51}$$

For given V_o, $\hat{\beta}$ is written as

$$\tilde{\beta}_{V_o} = (X'V^{-1}X)^{-1}X'V^{-1}y. \tag{3.52}$$

To find the restricted maximum likelihood estimators of σ^2 and V, use the following iterative process:

1. Substitute $\tilde{\beta}_{V_o}$ given by equation (3.52) into equation (3.51) and obtain

$$\begin{aligned}l^*(\tilde{\beta}_{V_o}, \sigma^2, V_o) = \\ -\frac{TI-p}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| + \frac{p}{2}\log\sigma^2 - \frac{1}{2}\log|X'V^{-1}X| \\ -\frac{1}{2\sigma^2}(y - X\hat{\beta}_{V_o})'V^{-1}(y - X\hat{\beta}_{V_o}) \\ = -\frac{TI-p}{2}\log 2\pi - \frac{TI}{2}\log\sigma^2 - \frac{I}{2}\log|V_o| + \frac{p}{2}\log\sigma^2 - \frac{1}{2}\log|X'V^{-1}X| \\ -\frac{1}{2\sigma^2}\mathrm{RSS}(V_o),\end{aligned} \tag{3.53}$$

   where

$$\mathrm{RSS}(V_o) = (y - X\tilde{\beta}_{V_o})'V^{-1}(y - X\tilde{\beta}_{V_o}). \tag{3.54}$$

2. Now maximise equation (3.53) with respect to σ^2 for fixed V_o to obtain the restricted maximum likelihood estimator of σ^2; that is,

$$\begin{aligned}\frac{\partial l^*}{\partial\sigma^2} &= -\frac{TI}{2\sigma^2} + \frac{p}{2\sigma^2} - \frac{\mathrm{RSS}(V_o)}{2\sigma^4} \\ &\stackrel{set}{=} 0.\end{aligned} \tag{3.55}$$

   Solving the above equation gives

$$\tilde{\sigma}^2_{V_o} = \frac{\mathrm{RSS}(V_o)}{TI-p} \tag{3.56}$$

3. Substitute $\tilde{\sigma}^2_{V_o}$ given by equation (3.56) into equation (3.53) to obtain a reduced log-likelihood for V_o, that is:

$$\begin{aligned} &l(\tilde{\beta}_{V_o}, \tilde{\sigma}^2_{V_o}, V_o) = \\ &\quad -\frac{I}{2}\log|V_o| - \frac{(TI-p)}{2}\log \mathrm{RSS}(V_o) - \frac{(TI-p)}{2}\log 2\pi + \frac{(TI-p)}{2}\log(TI-p) \\ &\quad -\frac{(TI-P)}{2} - \frac{1}{2}\log|X'V^{-1}X|. \end{aligned} \tag{3.57}$$

4. The reduced log-likelihood is now a function of the q unknown parameters in V_o and in simplified form is given by

$$\begin{aligned} &l^*_r(V_o) \\ &= l(\tilde{\beta}_{V_o}, \tilde{\sigma}^2_{V_o}, V_o) \\ &= -\frac{I}{2}\log|V_o| - \frac{1}{2}\log|X'V^{-1}X| - \frac{(TI-p)}{2}\log \mathrm{RSS}(V_o) \\ &- \frac{(TI-P)}{2}\{1 + \log 2\pi - \log(TI-p)\} \\ &= -\frac{I}{2}\log|V_o| - \frac{1}{2}\log|X'V^{-1}X| - \frac{(TI-p)}{2}\log \mathrm{RSS}(V_o) \\ &- \frac{(TI-p)}{2}\{1 + \log\left(\frac{2\pi}{TI-p}\right)\}. \end{aligned} \tag{3.58}$$

Ignoring the constants in equation (3.58) above, we have

$$\begin{aligned} l^*_r(V_o) &= l(\tilde{\beta}_{V_o}, \tilde{\sigma}^2_{V_o}, V_o) \\ &= -\frac{1}{2}\{(TI-p)\log \mathrm{RSS}(V_o) + I\log|V_o|\} - \frac{1}{2}\log|X'V^{-1}X|. \end{aligned} \tag{3.59}$$

5. Maximisation of equation (3.58) with respect to V_o yields $\tilde{V}_o$.

6. Substitute $\tilde{V}_o$ into equation (3.52) to obtain

$$\tilde{\beta}_{\tilde{V}_o} = (X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-1}y. \tag{3.60}$$

7. Substitute $\tilde{V}_o$ into equation (3.56) to obtain:

$$\tilde{\sigma}^2_{\tilde{V}_o} = \frac{\mathrm{RSS}(\tilde{V}_o)}{TI-p} \tag{3.61}$$

As with maximum likelihood estimation, maximisation of $l_r^*(V_o)$ given by equation (3.58) requires numerical optimisation techniques such as the Newton-Raphson algorithm. Once again, the dimensionality of the optimisation process will depend on what structure has been imposed on Σ_o. The estimation procedure is greatly simplified by the block diagonal structure for V, particularly in terms of evaluating its inverse and determinant.

3.2 Measuring test efficiency in the one parameter case: Pitman efficiency

The problem of comparing two tests for a given situation with the aim of evaluating their relative efficiency is reviewed. This is important because of the need to evaluate the loss of efficiency incurred in using any other test apart from the optimum one. Following Kendall and Stuart [26], the simplest way of comparing two tests for a given null hypothesis against a given alternative for fixed sample size is by direct examination of their power functions. The following definition of the relative efficiency of two tests is taken from Kendall and Stuart [26]:

Definition 3.2.1 *If an "efficient" test (that is, the most powerful in the class considered) of size α requires it to be based on n_1 observations to attain a certain power, and a second size α test requires n_2 observations to attain the same power against the same alternative, then the relative efficiency of the second test in attaining that power against that alternative is n_1/n_2.*

A similar definition can be found in Gibbons [19] where it is referred to as power efficiency. The relative efficiency or power of two tests is therefore a function of three arguments:

1. the size α of the tests,

2. the distance between the value of the parameter under the null hypothesis and its value under the alternative, and

3. the sample size n_1 required by the efficient test.

Hence we do not have, by this definition, a single summary measure of the relative efficiency of one test to another. The efficiency varies as the arguments listed above change. The need to achieve a single measure of efficiency is the driving force behind restriction to asymptotic results in evaluating test efficiency.

The following definition of the asymptotic relative efficiency of two tests is based on Gibbons [19]:

Definition 3.2.2 *Let A and B be two consistent tests of a null hypothesis H_o and alternative hypothesis H_a at significance level α. The asymptotic relative efficiency of test A relative to test B is the limiting value of the ratio n_b/n_a, where n_a is the number of observations required for the power of test A to equal the power of test B based on n_b observations while simultaneously, $n_b \to \infty$ and $H_a \to H_o$.*

Gibbons [19] notes that in many applications, the above ratio is the same for all choices of α so that the ARE is a single number.

The ARE of two tests can also be obtained by applying Pitman's theorem. Pitman efficiency was considered and generalised by Noether [46]. The general review presented here is based primarily on Kendall and Stuart [26] and Gibbons [19]. Suppose we have two consistent size α tests T_n and T_n^* for testing the hypothesis set

$$H_o : \theta = \theta_o \quad vs \quad H_a : \theta > \theta_o.$$

The first test rejects H_o if $t_n \geq t_{n,\alpha}$ while the second test rejects H_o if $t_n^* \geq t_{n,\alpha}^*$, where t_n and t_n^* are chosen such that:

$$P(T_n \geq t_{n,\alpha} \,|\, \theta = \theta_o) = \alpha$$

and

$$P(T_n^* \geq t_{n,\alpha}^* \,|\, \theta = \theta_o) = \alpha$$

respectively. The sequence of alternative hypotheses considered is such that θ approaches the value tested, θ_o, with increasing sample size. If T_n and T_n^* satisfy regularity conditions outlined in Noether [46] and Kendall and Stuart [26], then the ARE of T relative to T^* is

$$ARE(T, T^*) = \lim_{n\to\infty} \left[\frac{dE(T_n)/d\theta\,|_{\theta=\theta_o}}{dE(T_n^*)/d\theta\,|_{\theta=\theta_o}}\right]^2 \frac{\sigma^2(T_n^*)\,|_{\theta=\theta_o}}{\sigma^2(T_n)\,|_{\theta=\theta_o}}, \tag{3.62}$$

or simply,

$$ARE(T, T^*) = \lim_{n\to\infty} \frac{e(T_n)}{e(T_n^*)}, \tag{3.63}$$

where $e(T_n)$ is the *efficacy* of the test statistic T_n when used to test the hypothesis $\theta = \theta_o$ and

$$e(T_n) = \frac{[dE(T_n)/d\theta]^2\,|_{\theta=\theta_o}}{\sigma^2(T_n)\,|_{\theta=\theta_o}}. \tag{3.64}$$

3.3 An example of measuring test efficiency in the one parameter case

Suppose we wish to test the hypothesis that the mean μ of a normal population with known variance (taken to be equal to 1), is 0 versus that it's greater than zero, that is:

$$H_o : \mu = 0 \quad vs \quad H_a : \mu > 0. \tag{3.65}$$

We will consider two tests for testing the above null hypothesis and use them to illustrate the ideas of measuring test efficiency in the one parameter case. The first test is the usual z test based on the mean and the second test is the sign test. In this section, the asymptotic relative efficiency of the sign test relative to the usual test is found using the methods discussed above.

For testing the above null hypothesis, the power function of the usual test is given by

$$Power_{usual} = \Phi(\sqrt{n}\mu - z_\alpha) \tag{3.66}$$

while the approximate power function for the sign test is given by

$$Power_{sign} = \Phi\left\{\frac{\sqrt{n}(\Phi(\mu) - 0.5) - 0.5z_\alpha}{\sqrt{\Phi(\mu)(1-\Phi(\mu))}}\right\} \tag{3.67}$$

Figures 3.1 and 3.2 show the above power functions for varying values of n as a function of μ. Figure 3.3 shows the two power functions together at $n = 50$, with the usual test

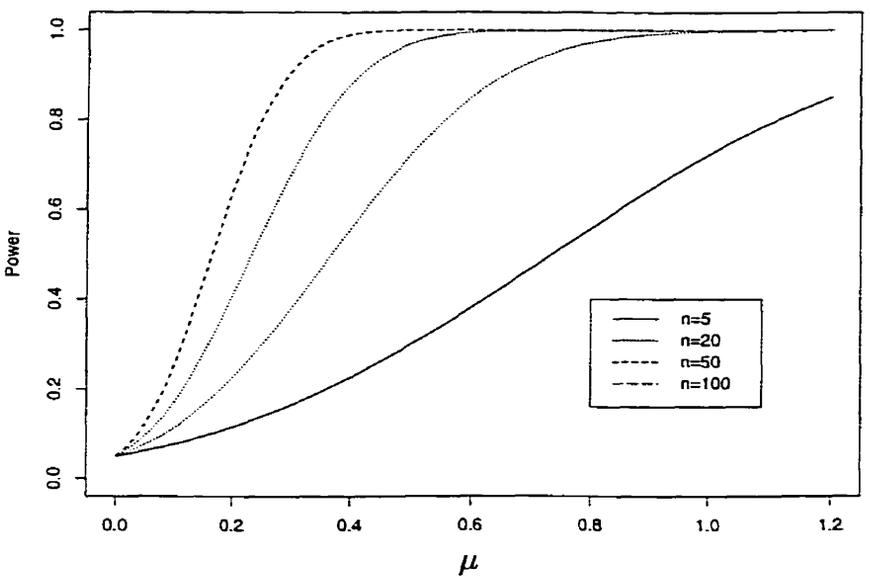

Figure 3.1: Power function based on the the usual z test for the mean μ of a normal population at $\alpha = 0.05$.

clearly having higher power than the sign test in the neighbourhood of H_o. Re-arranging the above power functions, expressions for the sample sizes needed to achieve a given

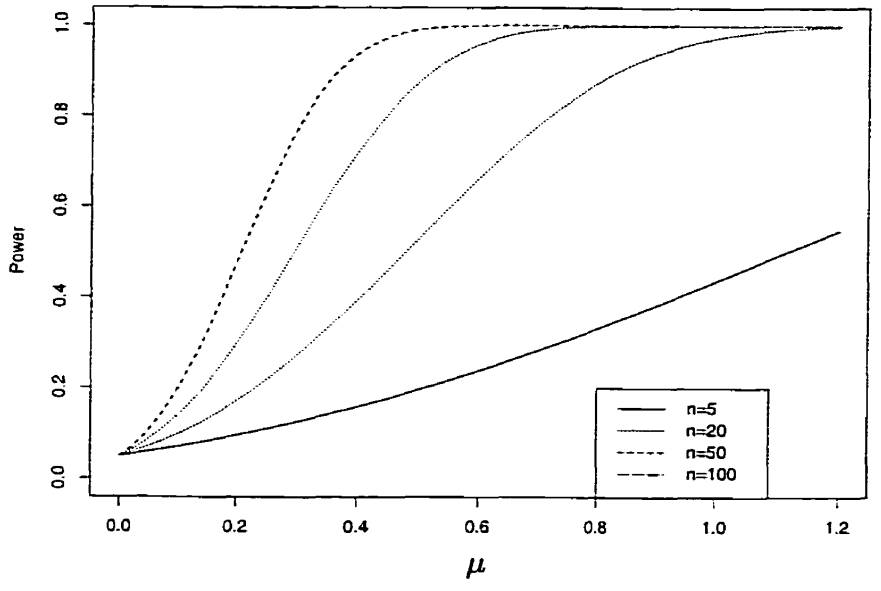

Figure 3.2: Power function based on the sign test for the mean μ of a normal population at $\alpha = 0.05$.

power for the usual test and the sign test are

$$n_{usual} = \left\{ \frac{\Phi^{-1}(1-\beta) + z_\alpha}{\mu} \right\}^2 \tag{3.68}$$

and

$$n_{sign} = \left\{ \frac{\Phi^{-1}(1-\beta)\sqrt{\Phi(\mu)(1-\Phi(\mu))} + 0.5 z_\alpha}{\Phi(\mu) - 0.5} \right\}^2, \tag{3.69}$$

respectively. Figure 3.4 shows the ratio n_{usual}/n_{sign} for varying values of power. Note that as H_a approaches H_o, the ratio appears to be somewhere between 0.63 and 0.65. As the power increases, the ratio approaches this value even faster even when H_a is far from H_o. The asymptotic relative efficiency of the sign test relative to the usual test is the limiting value of n_{usual}/n_{sign}, where n_{sign} is the number of observations required for the power of the sign test to equal the power of the usual test based on n_{usual} observations

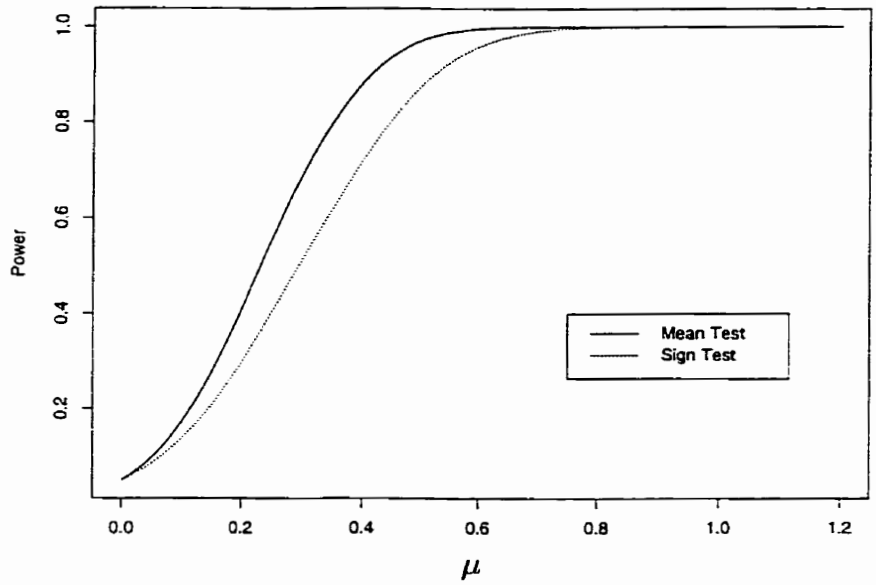

Figure 3.3: Power functions based on the usual z test and sign test for the mean μ of a normal population at $\alpha = 0.05$ and $n = 50$.

while simultaneously $n_{usual} \to \infty$ and $H_a \to H_o$. We have that

$$\begin{aligned}
(ARE)^{\frac{1}{2}} &= \lim_{\mu \to 0} \left\{ \frac{\Phi^{-1}(1-\beta) + z_\alpha}{\Phi^{-1}(1-\beta)\sqrt{\Phi(\mu)(1-\Phi(\mu))} + 0.5 z_\alpha} \right\} \left\{ \frac{\Phi(\mu) - 0.5}{\mu} \right\} \\
&= 2 \lim_{\mu \to 0} \left\{ \frac{\Phi(\mu) - 0.5}{\mu} \right\} \\
&= 2 \lim_{\mu \to 0} \frac{d}{d\mu} \left\{ \frac{\Phi(\mu) - 0.5}{\mu} \right\} \\
&= 2 \lim_{\mu \to 0} \frac{\phi(\mu)}{1} \\
&= 2 \lim_{\mu \to 0} \frac{1}{\sqrt{2\pi}} e^{\frac{-\mu^2}{2}} \\
&= 2 \lim_{\mu \to 0} \frac{1}{\sqrt{2\pi}} (1) \\
&= \sqrt{\frac{2}{\pi}}.
\end{aligned}$$

Overall:

$$ARE = \lim_{\mu \to 0} \frac{n_{usual}}{n_{sign}} = \left\{ \sqrt{\frac{2}{\pi}} \right\}^2 = \frac{2}{\pi} = 0.637. \tag{3.70}$$

Alternatively, the ARE of the two tests can be obtained by expanding their power functions in Taylor series at $\mu = 0$ and equating the two series so that the two tests have equal power against the same alternative. For the usual test, expanding the power function in a Taylor series gives:

$$\begin{aligned} &\Phi(\sqrt{n}\mu - z_\alpha) \\ &\quad = \Phi(-z_\alpha) + \mu\phi(z_\alpha)\sqrt{n} \\ &\quad = \alpha + \mu\phi(z_\alpha)\sqrt{n}. \end{aligned} \tag{3.71}$$

For the sign test, expanding the power function in a Taylor series gives:

$$\begin{aligned} &\Phi\left\{ \frac{\sqrt{n}(\Phi(\mu) - 0.5) - 0.5z_\alpha}{\sqrt{\Phi(\mu)(1 - \Phi(\mu))}} \right\} \\ &\quad = \Phi(-z_\alpha) + \mu \frac{1}{\sqrt{2\pi}} \phi(z_\alpha) \frac{\sqrt{n}}{0.5} \\ &\quad = \alpha + \mu\phi(z_\alpha)\sqrt{\frac{2n}{\pi}}. \end{aligned} \tag{3.72}$$

Figures 3.5 and 3.6 display the power functions, and their respective Taylor series approximations, of the usual test and sign test for varying sample sizes. Observe that the Taylor series approximations get closer to the power functions with increasing sample size.

Equating the two approximations so that the two tests have the same power against the same alternative, we have

$$\begin{aligned} &\alpha + \mu\phi(z_\alpha)\sqrt{n_{usual}} \\ &\quad = \alpha + \mu\phi(z_\alpha)\sqrt{\frac{2n_{sign}}{\pi}}, \end{aligned}$$

which simplifies to

$$\frac{n_{usual}}{n_{sign}} = \frac{2}{\pi}, \tag{3.73}$$

the same result obtained previously.

The ARE of the sign test relative to the usual test could also have been obtained by applying Pitman's theorem. For testing

$$H_o : \mu = 0 \quad vs \quad H_a : \mu > 0 \tag{3.74}$$

using the mean (usual test), the test statistic is

$$T^* = \frac{\bar{x}}{1/\sqrt{n}} = \sqrt{n}\bar{x}.$$

For large n:

$$\mathrm{E(T^*)} = \sqrt{\mathrm{n}}\mu$$

and

$$\mathrm{var(T^*)} = \mathrm{n\ var(\bar{x})} = 1.$$

Also,

$$\frac{d}{d\mu}\mathrm{E(T^*)}\,|_{\mu=0} = \sqrt{\mathrm{n}}.$$

Hence, the efficacy of T^* is given by

$$\mathrm{e(T^*)} = \mathrm{n}. \tag{3.75}$$

For testing the same hypothesis using the sign test, we re-write H_o and H_a in terms of the population median M:

$$H_o : M = \theta = 0 \quad vs \quad H_a : M = \theta > 0 \tag{3.76}$$

since the mean and median coincide for the normal distribution. The test statistic is T, the number of positive observations. T follows the binomial distribution with parameters

n and p where $p = P(X > 0)$, the probability of observing a positive observation. Note that $p = 0.5$ under H_o. Hence,

$$\mathrm{E(T)} = \mathrm{np}$$

and

$$\mathrm{var(T)} = \mathrm{np(1-p)}.$$

Note that when $\theta = 0$, $p = 1/2$ and hence:

$$\mathrm{var(T)}\mid_{\theta=0} = \frac{\mathrm{n}}{4}.$$

Now,

$$\begin{aligned}
&\frac{d}{d\theta}\mathrm{E(T)}\mid_{\theta=0} \\
&\quad = \frac{d\,[np]\mid_{\theta=0}}{d\theta} \\
&\quad = n\frac{d}{d\theta}\left[\Phi(\theta)\right]\mid_{\theta=0} \\
&\quad = n\phi(\theta)\mid_{\theta=0} \\
&\quad = \frac{n}{\sqrt{2\pi}}.
\end{aligned}$$

The efficacy of T is given by

$$\mathrm{e(T)} = \frac{2\mathrm{n}}{\pi}. \tag{3.77}$$

The ARE of the sign test relative to the usual test is therefore given by the ratio of equation (3.77) to equation (3.75)

$$\mathrm{ARE(T, T^*)} = \frac{2\mathrm{n}/\pi}{\mathrm{n}} = \frac{2}{\pi}. \tag{3.78}$$

3.4 Measuring test efficiency in the multi-parameter case

In the multi-parameter setting, consider testing null and alternative hypotheses of the form

$$H_o : \theta = \theta_o \quad vs \quad H_a : \theta = \theta_o + N^{-\delta}\lambda$$

where $\theta^T = (\theta_1, \theta_2, \ldots, \theta_r)$ are the location parameters for r populations and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ is a fixed but arbitrary non-null vector considered to be small and $\delta > 0$. For testing the above null hypothesis, Woolson [80] considered the problem of comparing two test statistic sequences which have limiting chi-square distributions with possibly different degrees of freedom. He suggested and justified some measures of asymptotic relative efficiency that may be used in comparing the two test sequences. Of special interest was developing measures of ARE independent of λ. Woolson and Sen [79] discuss in detail these measures of ARE. The measures developed can all be justified on the behavior of the power function in the neighbourhood of the null hypothesis.

The test statistics considered were quadratic forms $Q_N^{(i)}$, where $Q_N^{(i)} \sim \chi^2(t_i; \Delta_i)$ under $H_a, i = 1, 2$. In calculating measures of ARE, they used an adjustment factor $R(t_1, t_2, \alpha)$, where

$$R(t_1, t_2, \alpha) = \frac{P\left\{\chi^2_{(t_2+2)} > \chi^2_{t_2,\alpha}\right\} - \alpha}{P\left\{\chi^2_{(t_1+2)} > \chi^2_{t_1,\alpha}\right\} - \alpha},$$

and α is the level of the test.

We shall consider two of the measures, the curvature asymptotic relative efficiency (CARE) and the trace asymptotic relative efficiency (TARE). The motivation for choosing the CARE and TARE is that they are both average measures of ARE, independent of λ. This review of the two measures is based primarily on Woolson [80] and Woolson and Sen [79]. The test ϕ_1 is based on the test statistic $Q_N^{(1)}$ and the test ϕ_2 is based on the test statistic $Q_N^{(2)}$.

The CARE of ϕ_2 relative to ϕ_1 is given by

$$\text{CARE}(\phi_2/\phi_1) = [\text{R}(\text{t}_1, \text{t}_2, \alpha)]^{1/2\delta} \left\{ \frac{|\text{D}_2'\Sigma_2^{-1}\text{D}_2|}{|\text{D}_1'\Sigma_1^{-1}\text{D}_1|} \right\}^{1/2q\delta}, \tag{3.79}$$

with D_1 and D_2 being related to the non-centrality parameters of tests ϕ_1 and ϕ_2 respectively. The CARE works out well when at least one of t_1, t_2 is $\geq q$. Geometrical considerations show that if CARE > 1, then the power function of ϕ_2 has faster average local growth at the null point than does the power function of ϕ_1.

The TARE of ϕ_2 relative to ϕ_1 is given by

$$\text{TARE}(\phi_2/\phi_1) = [\text{R}(\text{t}_1, \text{t}_2, \alpha)]^{1/2\delta} \left\{ \frac{\text{tr}(\text{D}_2'\Sigma_2^{-1}\text{D}_2)}{\text{tr}(\text{D}_1'\Sigma_1^{-1}\text{D}_1)} \right\}^{1/2\delta}. \tag{3.80}$$

The TARE is valid for all values of t_1, t_2 and q and hence the range of applicability of the TARE is wider than that of the CARE. If TARE > 1, then the test ϕ_2 is more optimum locally by virtue of its greater average power over the family of spheres.

3.5 An example of measuring test efficiency in the multi-parameter case using two parameters

The example presented here is discussed in Woolson [78], which compares the usual Hotelling's test statistic for the one-sample location problem to the nonparametric rank scores statistics utilizing the CARE and TARE as the modes of comparison. Since the two tests (at the same level α) have the same degrees of freedom, the criteria of comparison are scalar functions of the matrices in the non-centrality parameters. The CARE is a function of the determinants of these matrices while the TARE is a function of the traces of these matrices.

Let X_i $(i = 1, 2, \ldots, N)$ be N independent random vectors from the bivariate normal distribution with mean vector μ and variance-covariance matrix Σ. Consider testing the hypothesis set

$$H_o : \mu = \mu_o \quad vs \quad H_a : \mu \neq \mu_o, \tag{3.81}$$

where μ under H_a is given by $\mu_o + \frac{\lambda}{\sqrt{N}}$. Using Hotelling's test statistic, the null hypothesis is rejected if:

$$N(\bar{x} - \mu_o)'\Sigma^{-1}(\bar{x} - \mu_o) > \chi_2^2(\alpha), \tag{3.82}$$

where $\chi_2^2(\alpha)$ represents the upper 100α percentage point of the chi-square distribution with two degrees of freedom. The power of this test under H_a is given by:

$$\begin{aligned} power &= Pr\left\{N(\bar{x} - \mu_o)'\Sigma^{-1}(\bar{x} - \mu_o) > \chi_2^2(\alpha)\right\} \\ &= Pr\left\{\chi_2'^2\left(N(\mu - \mu_o)'\Sigma^{-1}(\mu - \mu_o)\right) > \chi_2^2(\alpha)\right\} \\ &= 1 - Pr\left\{\chi_2'^2\left(N(\mu - \mu_o)'\Sigma^{-1}(\mu - \mu_o)\right) \leq \chi_2^2(\alpha)\right\}. \end{aligned} \tag{3.83}$$

As an illustration, the above power function is plotted at $\alpha = 0.05$ for $N = 5, 20, 50, 100$, with

$$\mu_o = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and

$$\Sigma = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

See Figures 3.7 to 3.10. The non-centrality parameter is given by $N(\mu - \mu_o)'\Sigma^{-1}(\mu - \mu_o)$ which simplifies to $\lambda'\Sigma^{-1}\lambda$.

Let ϕ_1 be the test based on the usual Hotelling statistic and let ϕ_2 be the test based on the non-parametric rank score statistic. The non-centrality parameters in the two test statistics are $\lambda'\Sigma^{-1}\lambda$ and $\lambda' T^{-1}\lambda$ respectively. From the definitions of the CARE and TARE given previously, we have

$$\text{CARE}(\phi_2/\phi_1) = \left\{\frac{|\Sigma|}{|\mathrm{T}|}\right\}^{1/2} \tag{3.84}$$

and

$$\text{TARE}(\phi_2/\phi_1) = \frac{\mathrm{tr}\mathrm{T}^{-1}}{\mathrm{tr}\Sigma^{-1}}. \tag{3.85}$$

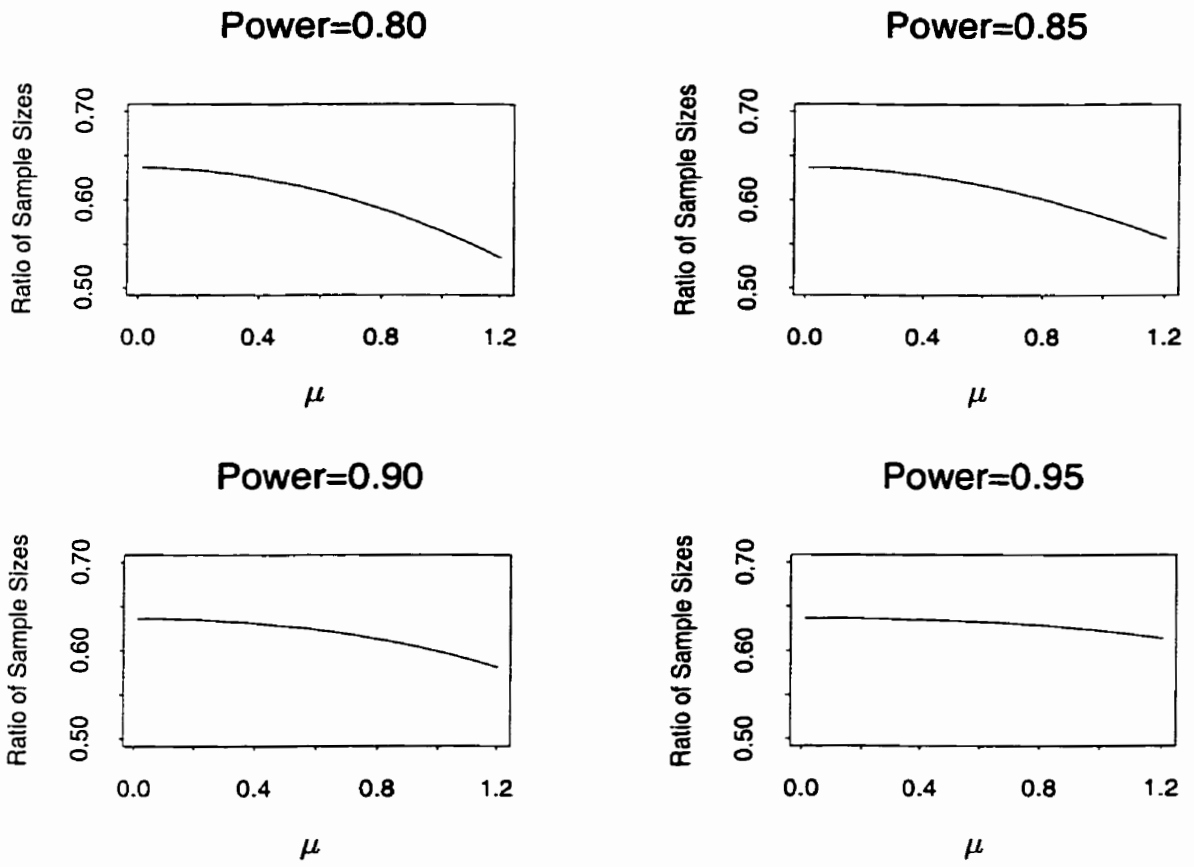

Figure 3.4: Ratio of sample sizes (n_{usual}/n_{sign}) required to achieve a given power at $\alpha = 0.05$ for the test of the mean μ of a normal population based on the usual z test and sign test.

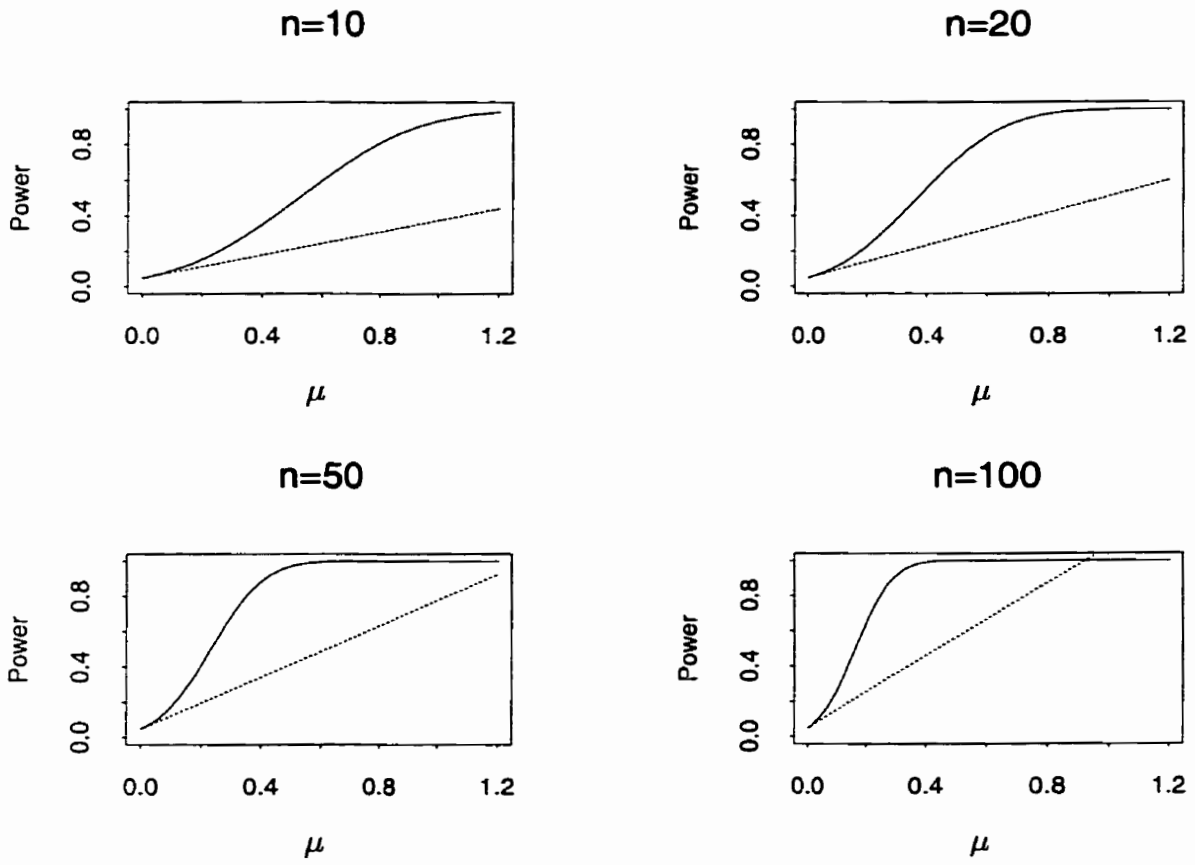

Figure 3.5: Power function (solid line) and first-order Taylor series expansion of the power function (dotted line) at $\mu = 0$ based on the usual z test for the mean μ of a normal population at $\alpha = 0.05$.

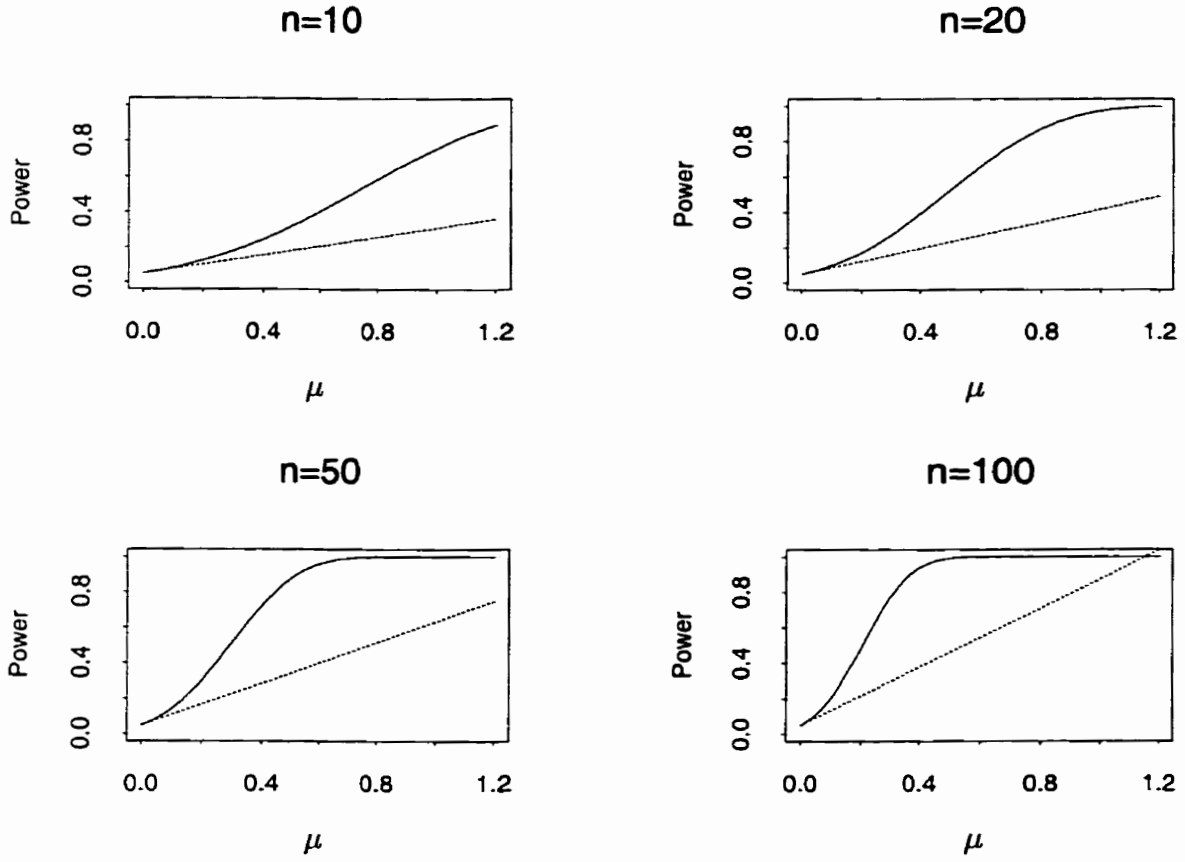

Figure 3.6: Power function (solid line) and first-order Taylor series expansion of the power function (dotted line) at $\mu = 0$ based on the sign test for the mean μ of a normal population at $\alpha = 0.05$.

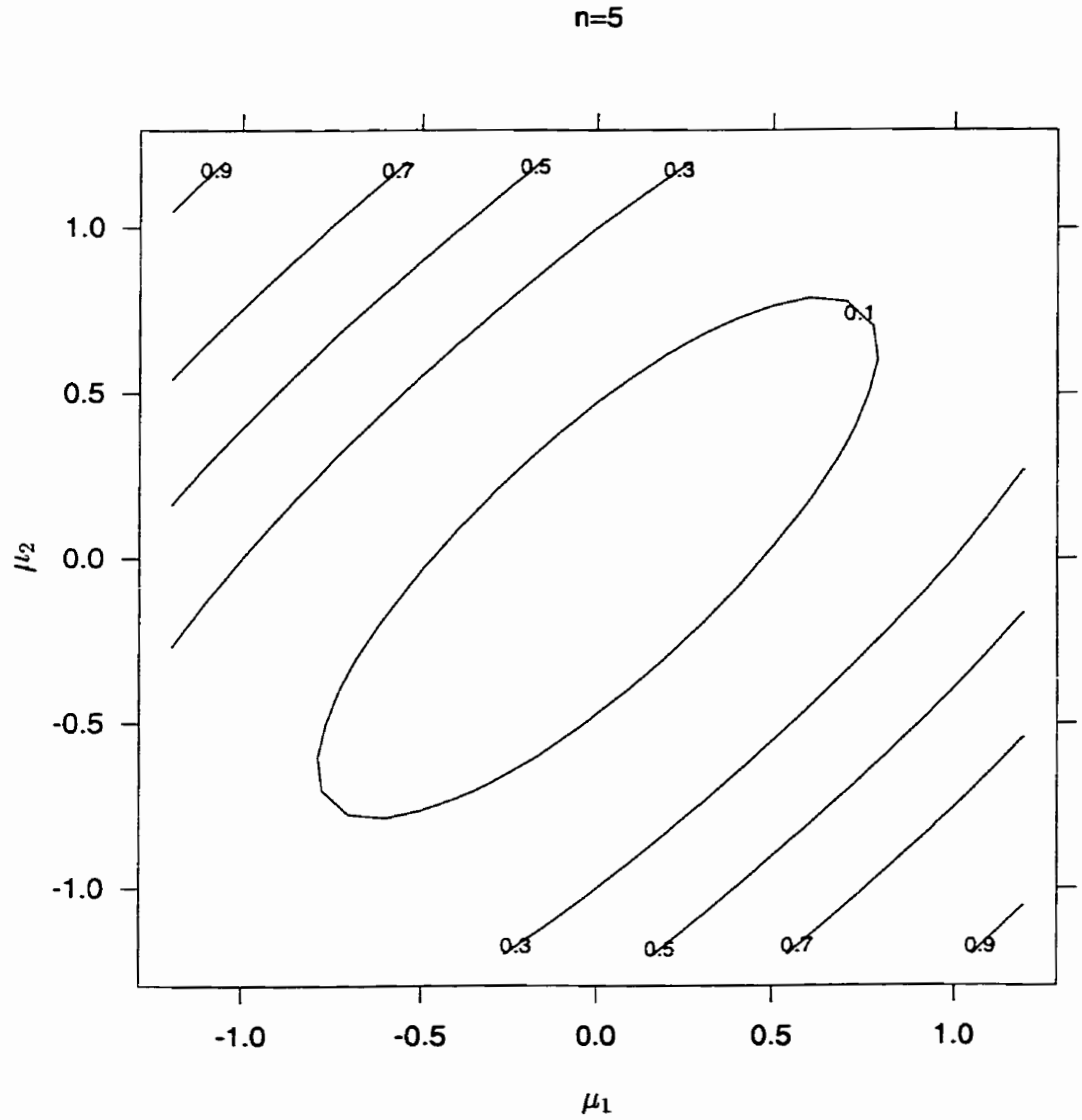

Figure 3.7: Power function based on Hotelling's test statistic for the bivariate one sample location problem (n=5).

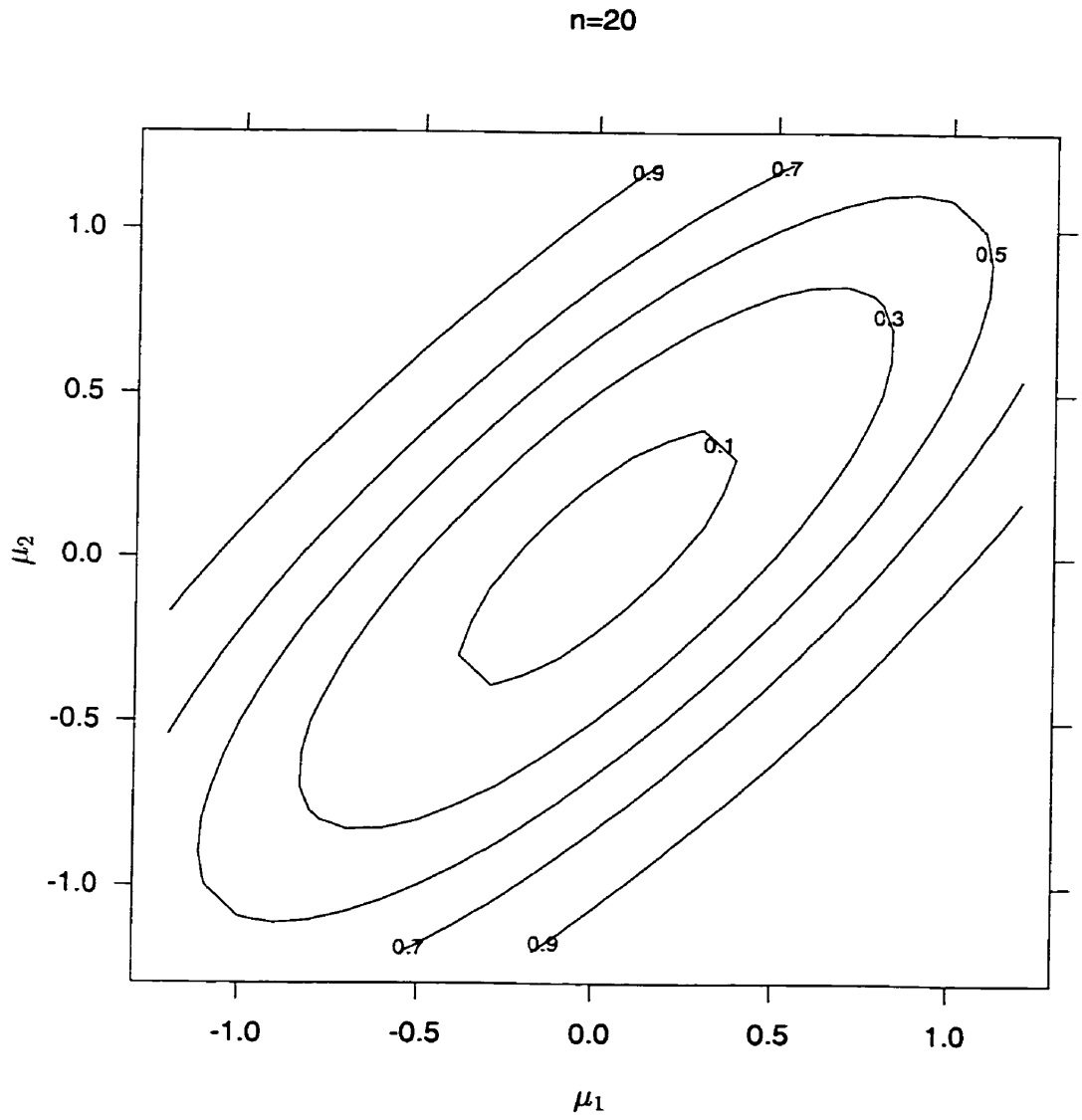

Figure 3.8: Power function based on Hotelling's test statistic for the bivariate one sample location problem (n=20).

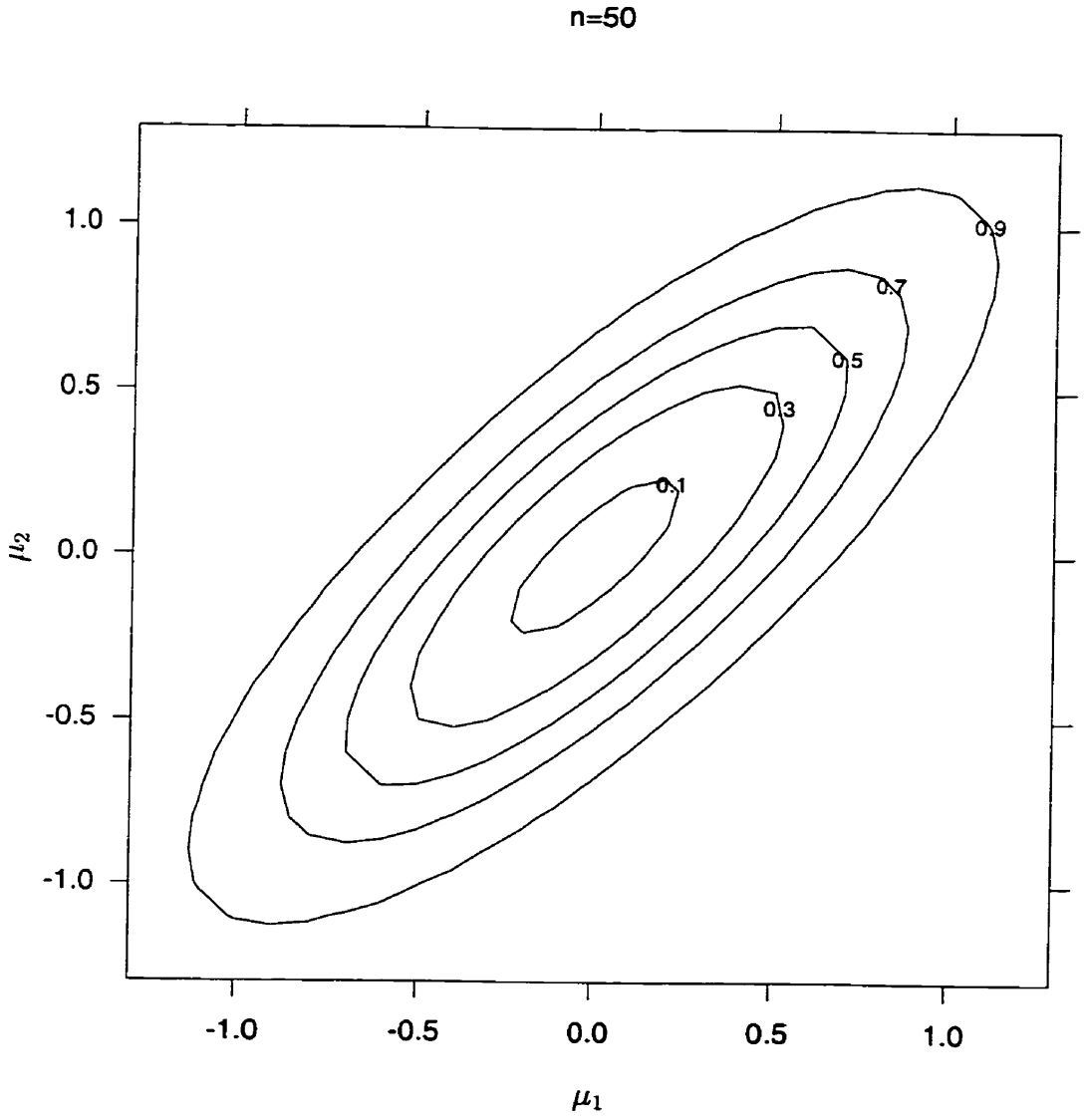

Figure 3.9: Power function based on Hotelling's test statistic for the bivariate one sample location problem (n=50).

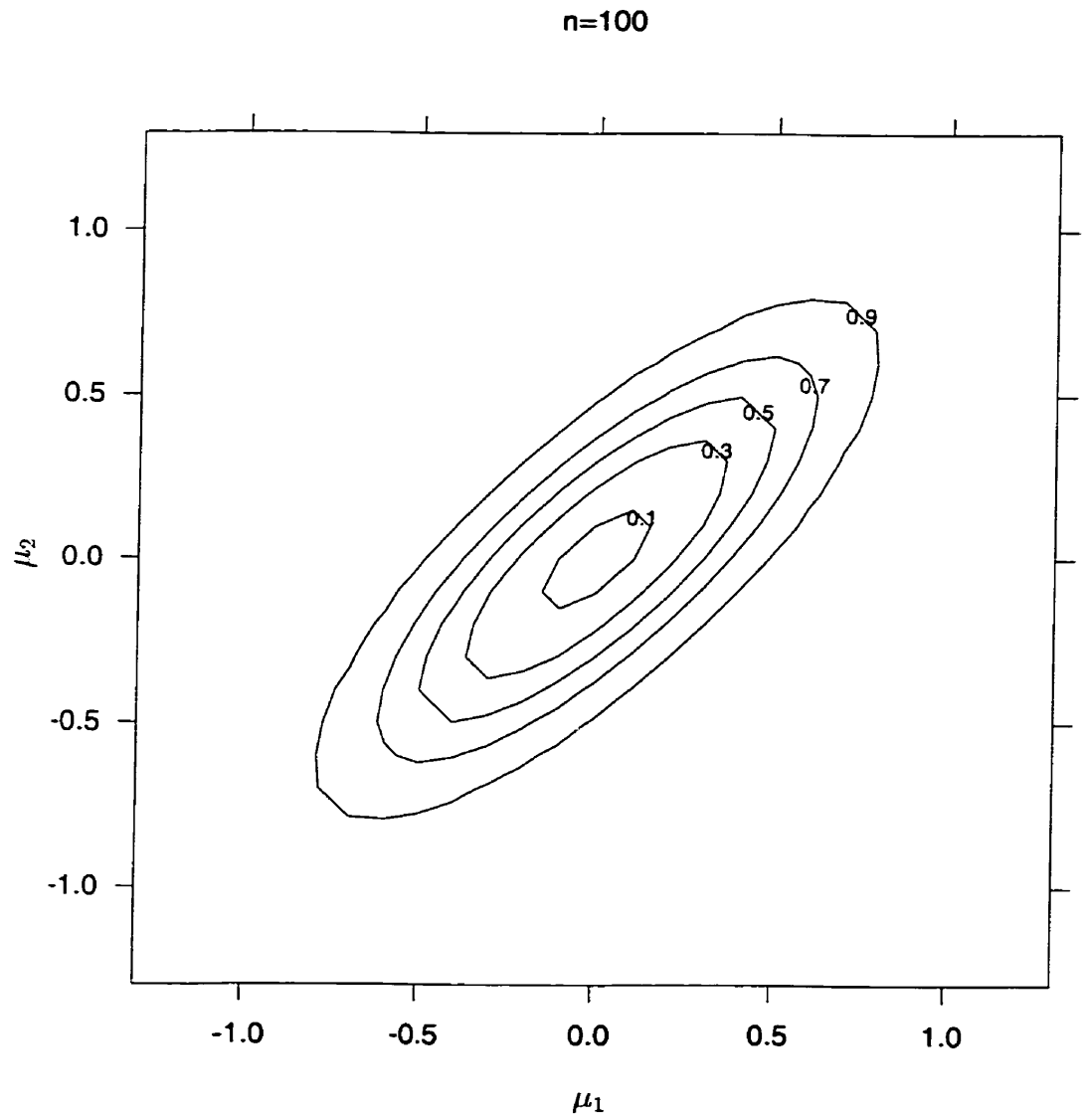

Figure 3.10: Power function based on Hotelling's test statistic for the bivariate one sample location problem (n=100).

Chapter 4

Efficiency of ordinary least squares for multivariate longitudinal data with Kronecker product covariance matrices

4.1 Introduction

In a multivariate longitudinal design, several characteristics of interest are measured on each experimental unit over time. One approach to analysing the resulting data is to use a linear model with a Kronecker structured covariance matrix. The model can be written as

$$y = (\Theta \otimes I_C \otimes X)\lambda + e, \tag{4.1}$$

where:

$\Theta \longrightarrow I \times G$ between subject design matrix for I subjects in G treatment groups;

$I_C \longrightarrow C \times C$ identity matrix;

$X \longrightarrow T \times p$ within subject design matrix;

$\lambda \longrightarrow pCG \times 1$ vector of unknown parameters; and

$e \longrightarrow$ error vector with $\Sigma = \text{cov(e)} = \mathrm{I_I} \otimes \Sigma_o$.

The parameter of interest is λ and we now consider two ways of testing hypothesis concerning λ. One approach is to use a test statistic that is a function of the ordinary least squares estimator of λ, which will be fully efficient only if $\Sigma_o = \sigma^2 I_{TC}$. For multivariate longitudinal data, this is an unlikely situation since we expect responses on the same subject to be correlated. Correlation arises from the multiple characteristics measured at each time point as well as the same characteristic being measured over time. Therefore, it is more realistic to assume that

$$\begin{aligned} \Sigma &= \text{cov(e)} = \mathrm{I_I} \otimes \Sigma_o \\ &= I_I \otimes \Delta \otimes \Omega, \end{aligned}$$

where

$I_I \longrightarrow$ the $I \times I$ identity matrix;

$\Sigma_o \longrightarrow CT \times CT$ within-subject covariance matrix;

$\Delta \longrightarrow C \times C$ covariance matrix for C dependent variables; and

$\Omega \longrightarrow T \times T$ covariance matrix for T repeated measures on each dependent variable.

Σ_o depends on γ_1 and γ_2, the parameter vectors for Δ and Ω respectively.

The present chapter investigates the problem of how inefficient ordinary least squares may become. Efficiency is defined in terms of testing hypotheses that are of interest in a given problem. Since correlation arises in two ways in this setting, we expect that ignoring these two sources of correlation and proceeding to do inference on the resulting data based on ordinary least squares will result in hypotheses of interest being inefficiently tested. As stated in Matthews [42], who considered a similar problem for cross-over designs, efficiency depends on the design in question, the choice of which is determined by many factors, some of which may not be statistical. The formulation of the model is reviewed in detail in section 4.2. Hypothesis testing and power are discussed in section 4.3. Estimation of model parameters is discussed in section 4.4. Algebraic results for efficiency evaluation using the TARE and CARE are presented in section 4.5.

Numerical results which assume some special structures for Δ and Ω are presented in section 4.6. Finally, the chapter closes with a discussion in section 4.7.

4.2 Formulation of the model

Suppose we have a sample of I individuals or experimental units that have been selected for a longitudinal study. C responses are obtained for each individual, indexed by $i = 1, 2, \ldots, I$ at the same set of T time points, indexed by $t = 1, 2, \ldots, T$. The T time points need not be equally spaced. Let y_{cti} represent the measurement of the c^{th} characteristic at occasion t on individual i for $c = 1, 2, \ldots, C$; $t = 1, 2, \ldots, T$; $i = 1, 2, \ldots, I$. The data may be represented as follows:

	1				Characteristics	C			
Individuals	1	2	...	T	Time	1	2	...	T
1	y_{111}	y_{121}	$\cdots$	y_{1T1}	$\cdots$	y_{C11}	y_{C21}	$\cdots$	y_{CT1}
2	y_{112}	y_{122}	$\cdots$	y_{1T2}	$\cdots$	y_{C12}	y_{C22}	$\cdots$	y_{CT2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	y_{11I}	y_{12I}	$\cdots$	y_{1TI}	$\cdots$	y_{C1I}	y_{C2I}	$\cdots$	y_{CTI}

Y_i is a $T \times C$ matrix representing C characteristics (columns) measured on T occasions (rows):

$$\underset{T \times C}{Y_i} = \left(y_{1i},\ y_{2i},\ \ldots,\ y_{ci},\ \ldots,\ y_{Ci} \right) = \begin{pmatrix} y_{11i} & y_{21i} & \cdots & y_{c1i} & \cdots & y_{C1i} \\ y_{12i} & y_{22i} & \cdots & y_{c2i} & \cdots & y_{C2i} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ y_{1ti} & y_{2ti} & \cdots & y_{cti} & \cdots & y_{Cti} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ y_{1Ti} & y_{2Ti} & \cdots & y_{cTi} & \cdots & y_{CTi} \end{pmatrix} . \quad (4.2)$$

Note that if $C = 1$, then only the first column in the matrix Y_i remains. This represents the $T \times 1$ vector of repeated observations on individual i used in the linear model for

correlated data. Associated with the $1 \times C$ vector y_{ti} (the t^{th} row of Y_i) is a $p \times 1$ vector of covariates or explanatory variables that is given by

$$\underset{p \times 1}{x_{ti}} = \begin{pmatrix} x_{ti1} \\ x_{ti2} \\ \vdots \\ x_{tik} \\ \vdots \\ x_{tip} \end{pmatrix}. \tag{4.3}$$

Consequently, X_i is a $T \times p$ design matrix for the i^{th} individual as shown below:

$$\underset{T \times p}{X_i} = \begin{pmatrix} x_{1i1} & x_{1i2} & \dots & x_{1ik} & \dots & x_{1ip} \\ x_{2i1} & x_{2i2} & \dots & x_{2ik} & \dots & x_{2ip} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{ti1} & x_{ti2} & \dots & x_{tik} & \dots & x_{tip} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{Ti1} & x_{Ti2} & \dots & x_{Tik} & \dots & x_{Tip} \end{pmatrix}. \tag{4.4}$$

The matrix X_i is obtained by making x_{ti} a row vector (that is, taking its transpose). The rows in X_i correspond to the different times or occasions of measurement and the columns correspond to the different covariates. This representation of the design matrix allows for time-varying covariates (for example, age), time-invariant covariates (for example, treatment group) as well as covariates that are functions of time.

4.2.1 Modelling the expected values

From section 4.2, y_{cti} represents the measurement of the c^{th} characteristic at occasion t on individual i for $c = 1, 2, \dots, C$; $t = 1, 2, \dots, T$; $i = 1, 2, \dots, I$. Hence $y_{ci} = (y_{c1i}, y_{c2i}, \dots, y_{cTi})'$ represents the $T \times 1$ vector of characteristic c measured over T occasions on individual i. Assume a model for y_{ci} of the form

$$y_{ci} = X_i b_i + \epsilon_i \tag{4.5}$$

where X_i is a $T \times p$ design matrix and b_i is a $p \times 1$ vector of unknown parameters. Letting

$$\underset{T \times C}{Y_i} = \left(\begin{array}{ccccccc} y_{1i}, & y_{2i}, & \ldots, & y_{ci}, & \ldots, & y_{Ci} \end{array} \right),$$

we have

$$Y_i = X_i \mathcal{B}_i + \mathcal{E}_i \tag{4.6}$$

for $i = 1, 2, \ldots, I$ where Y_i is the $T \times C$ matrix of observations and X_i is the $T \times p$ design matrix for the i^{th} individual. In equation (4.6) above, $\mathcal{B}_i$ is a $p \times C$ matrix of unknown parameters and $\mathcal{E}_i$ is the matrix of errors for the i^{th} individual. Using the vec operator such that $y_i = \text{vec}(\text{Y}_\text{i})$, we can express equation (4.6) as

$$y_i = (I_C \otimes X_i)\beta_i + \epsilon_i, \quad i = 1, 2, \ldots, I, \tag{4.7}$$

where $\epsilon_i = \text{vec}(\mathcal{E}_\text{i})$, $\beta_i = \text{vec}(\mathcal{B}_\text{i})$ and we have used the result that vec(PQR) = (R′ ⊗ P)vec(Q). For I individuals, we have:

$$\begin{aligned} y_1 &= (I_C \otimes X_1)\beta_1 + \epsilon_1 \\ y_2 &= (I_C \otimes X_2)\beta_2 + \epsilon_2 \\ \vdots &= \vdots \\ y_I &= (I_C \otimes X_I)\beta_I + \epsilon_I. \end{aligned}$$

The design matrix for the i^{th} individual is given by $(I_C \otimes X_i)$. This specification of the design matrix has been used by Matsuyama and Ohashi [41] for bivariate response repeated measures data.

To better illustrate the model and without loss of generality, suppose two response variables are measured on each individual at each of T occasions. Let y_{1i} and y_{2i} be the $T \times 1$ response vectors on the i^{th} individual for $i = 1, 2, \ldots, I$ and set $y_i = (y'_{1i}, y'_{2i})'$. Consider the model for a single individual given by equation (4.7) which can be written as (for two response variables):

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \begin{pmatrix} X_i & 0 \\ 0 & X_i \end{pmatrix} \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \tag{4.8}$$

where

$\beta_{ci} \longrightarrow p \times 1$ individual vector of unknown fixed effects for the c^{th} response

$\epsilon_{ci} \longrightarrow T \times 1$ within subject random error vector for the c^{th} response. Making the notation more compact, let

$$\beta_i = (\beta'_{1i}, \beta'_{2i})',$$

$$\epsilon_i = (\epsilon'_{1i}, \epsilon'_{2i})'.$$

Then

$$y_i = X_i^* \beta_i + \epsilon_i, \tag{4.9}$$

where

$\beta_i \longrightarrow 2p \times 1$ vector of unknown fixed effects,

$\epsilon_i \longrightarrow 2T \times 1$ within individual random error vector, and

$$X_i^* = \begin{pmatrix} X_i & 0 \\ 0 & X_i \end{pmatrix}.$$

With reference to equation (4.7), the complete system of equations can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_I \end{bmatrix} = \begin{bmatrix} X_1^* & 0 & \cdots & 0 \\ 0 & X_2^* & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & X_I^* \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_I \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_I \end{bmatrix}, \tag{4.10}$$

or more compactly as

$$y = X\beta + e, \tag{4.11}$$

with

$$\Sigma = \mathrm{cov(y)} = \mathrm{I_I} \otimes \Sigma_\mathrm{o}. \tag{4.12}$$

4.2.2 Modelling the covariance structure

Σ_o is the $CT \times CT$ within-subject variance-covariance matrix. It follows that if Σ_o is left completely unstructured, it is defined by $CT(CT+1)/2$ parameters. In some cases however, CT is too large relative to I and in this situation, some structure should be imposed on Σ_o. Consider modelling the $CT \times CT$ covariance matrix Σ_o using the Kronecker product of the covariance structures Δ and Ω such that $\Sigma_o = \Delta \otimes \Omega$. Since y_i is arranged by characteristic and by time within characteristic, the covariance between the outcome variables is specified by the $C \times C$ matrix Δ, whereas the covariance among the repeated measures for a given outcome variable is specified by the $T \times T$ matrix Ω.

If no restrictions are placed on Δ and Ω, then they are defined by $C(C+1)/2$ and $T(T+1)/2$ parameters respectively. Restrictions may be placed on Δ and/or Ω to ensure identifiability of all parameters. Since Δ represents the covariance matrix between the C outcome variables, it is left unstructured. Also, since Ω represents the covariance matrix for the T repeated measurements on any characteristic, it can be modelled parsimoniously, for example, using the compound symmetry and first-order autoregressive structures. Restrictions are placed on Ω to ensure identifiability of all parameters. We denote the parameter vectors for Δ and Ω by γ_1 and γ_2, respectively, and let $\kappa = [\gamma_1', \gamma_2']'$.

This approach to directly modelling the dependence of variables which exists not only within but also across time in the form of a Kronecker product of covariance matrices has been considered by various authors, including Zhang [81] for bivariate longitudinal data. Rochon [55] proposes this model as one method by which the evolving relationship between two sets of repeated measures is taken into consideration. Note that Rochon [55] suggests using $\Sigma_o = \Omega \otimes \Delta$, where Ω is the $T \times T$ covariance matrix among the repeated measures and Δ is the 2×2 covariance matrix for the pair of outcome variables. For our purposes, the order of these two matrices has been reversed. Verbyla and Cullis [69] use the Kronecker structured covariance matrix to analyse repeated measures data when an additional level of dependence exists. Galecki [18] parametrically models the covariance structure for repeated measures specified by more than one repeated factor using the

Kronecker product of underlying factor specific covariance profiles. Note that if in the model proposed by Zhang [81] one assumes that the diagonal elements of the within subject covariance matrix Σ_o are equal, then this model will be equivalent to the model proposed by Galecki [18].

For bivariate longitudinal data and again without loss of generality, the matrices Δ and Ω are given below. Now,

$$\Sigma_o = \Delta \otimes \Omega \tag{4.13}$$

where

$$\Delta = cov(y_{1ti}, y_{2ti}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

and

$$\Omega = cov(y_{ci}).$$

In later sections, Δ is reparameterised as follows (to facilitate interpretation):

$$\Delta = \begin{pmatrix} \sigma_{11} & \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & b \\ b & \gamma \end{pmatrix},$$

with σ_{11} set to one. Using this reparametrisation, γ now represents the ratio of the variances for the two characteristics and $\rho_c = \text{corr}(y_{1ti}, y_{2ti})$ is the correlation between the two characteristics at any given time.

How can one justify using a Kronecker structured covariance matrix for multivariate longitudinal data? For univariate longitudinal data, there are several choices of structured covariance structures that one can utilise. This is discussed in detail in Ware [75], Jennrich and Schluchter [25] and Diggle, Liang and Zeger [14]. Alternatively, one can introduce random effects in the model as presented by Laird and Ware [34]. The question we will address is how can one model Σ_o in the case of multivariate longitudinal data.

One possibility is to model the covariance structure as the Kronecker product of an unstructured covariance matrix (accounting for the covariance among the characteristics) with some time series covariance matrix (accounting for the covariance across time) as presented in section 4.2.2. This model is based primarily on Galecki's [18] model proposed for two or more repeated factors. Modelling data spanned by two or more repeated factors as discussed in Galecki [18] is indeed quite different from modelling multivariate longitudinal data. A similar model is presented in Zhang [81].

Suppose we imagine that the longitudinal data on an individual is spanned by two "factors", characteristic (factor A) and time (factor B), with levels $c = 1, 2, \ldots, C$ and $t = 1, 2, \ldots, T$ respectively. The complete set of observations on an individual consists of CT measurements, with variances and covariances collected in a $CT \times CT$ within subject covariance matrix Σ_o. Initially, we focus attention on the $T \times T$ submatrices of marginal distributions of measurements taken over time (Factor B) for every characteristic separately. We can assume for the purpose of modelling that the marginal covariance matrices associated with factor B (time) are equivalent at every level of Factor A (characteristic). We now shift our focus to the $C \times C$ submatrices of marginal distributions of the different characteristics (factor A) for every time point separately. Again, it is acceptable for modelling purposes to assume that the covariance matrices associated with factor A (characteristics) are equivalent at every time point (factor B).

As discussed in Galecki [18], one way to model a covariance matrix with repeated measures in more than one dimension is to use the Kronecker product of "factor" specific covariance profiles with the underlying assumption that the marginal profile for a given "factor" is invariant for every level of the other "factor". Our two dimensions here are characteristics and time and hence the $CT \times CT$ within subject covariance matrix Σ_o can be expressed as the Kronecker product of the marginal covariance matrices of the "factors": characteristics and time. For bivariate longitudinal data for example, this means that $\Delta = cov(y_{1ti}, y_{2ti})$ is constant with respect to both t (time) and i (individual) and $\Omega = cov(y_{ci})$ is constant with respect to both c (characteristics) and i (individual). Table 4.1 gives four examples of covariance structures based on an unstructured Δ and

several choices of Ω for $C = 2$ and $T = 3$. In Table 4.1, "un" denotes unstructured; "sim" denotes simple; "ar(1)" denotes first-order autoregressive and "cs" denotes compound symmetry.

Table 4.1: Examples of within-subject covariance models ($C = 2$ and $T = 3$).

Model #	Structure	Form ($\Delta \otimes \Omega$)	# Parameters
1	UN⊗SIM	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	3
2	UN⊗UN	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho^*_{12} & \rho^*_{13} \\ \rho^*_{12} & \rho^*_{22} & \rho^*_{23} \\ \rho^*_{13} & \rho^*_{23} & \rho^*_{33} \end{bmatrix}$	8
3	UN⊗AR(1)	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$	4
4	UN⊗CS	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$	4

If Σ_o is left completely unstructured in this case, it is defined by 21 parameters. Using the Kronecker structured covariance matrix results in a tremendous reduction in the number of parameters as illustrated by model 2. A further reduction in the number of parameters defining Σ_o is achieved by imposing a structure on the $T \times T$ covariance matrix Ω among the repeated measures as illustrated by models 1, 2 and 3. Note that the number of parameters in Ω does not increase with T for models 1,3 and 4. The interpretation of model 3, for example, is as follows:

1. Covariance matrix of the marginal distribution of the 2 characteristics y_{1ti} and y_{2ti} has an unstructured covariance matrix and is the same for all 3 levels of the other factor (time).

2. Covariance matrix of the marginal distribution of the 3 measurements y_{C1i}, y_{C2i} and y_{C3i} is first-order autoregressive and is the same for both levels of the other factor (characteristic).

4.2.3 Identifiability of Σ_o

Since Σ_o is expressed as the Kronecker product of two matrices, then the issue of identifiability has to be addressed. As discussed in Galecki [18], nonidentifiability arises from the fact that if $\Delta \otimes \Omega$ is equal to the overall within-subject covariance matrix Σ_o, then there exists a continuum of other pairs of covariance matrices, for example, $\delta * \Delta$ and Ω/δ for $\delta > 0$ which give the same Kronecker product. Consequently, we cannot identify Δ and Ω. To avoid this nonidentifiability, the matrix Ω is rescaled so that the upper left element of this matrix is equal to 1. This is reflected in the form of Ω given in Table 4.1.

4.2.4 A simplification of the model

Suppose that in equation 4.6 in section 4.2.1 we have $X_i = X$ so that X is a $T \times p$ matrix of known constants identical from one one individual to the next. For instance, we may think of X as containing the values of p functions of time at T time points. Then the model for Y_i becomes:

$$Y_i = X\mathcal{B}_i + \mathcal{E}_i \tag{4.14}$$

and consequently, equation (4.7) becomes:

$$y_i = (I_C \otimes X)\beta_i + \epsilon_i, \quad i = 1, 2, \ldots, I. \tag{4.15}$$

For I individuals, we suppose that

$$[\beta_1, \beta_2, \ldots, \beta_I] = \Lambda\Theta' \tag{4.16}$$

where Λ is a $pC \times G$ matrix of unknown parameters and $\Theta' = (\theta_1, \theta_2, \ldots, \theta_I)$ is a $G \times I$ matrix of known constants of full rank $G < I$, corresponding to the design matrix for I subjects in G treatment groups. In this context, X corresponds to the within subject design matrix and Θ corresponds to the between subject design matrix. Letting $Y = \{y_1, y_2, \ldots, y_I\}$, $E = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_I\}$, we have:

$$Y = (I_C \otimes X)\Lambda\Theta' + E. \tag{4.17}$$

The columns of E given by given by $\epsilon_1, \epsilon_2, \ldots, \epsilon_I$ are assumed to be independently distributed as $N \sim (0, \Sigma_o)$, where Σ_o is the $CT \times CT$ Kronecker structured covariance matrix so that $\Sigma_o = \Delta \times \Omega$ as discussed in section 4.2.2. Applying the vec operator to the above model, we obtain

$$y = (\Theta \otimes I_C \otimes X)\lambda + e, \tag{4.18}$$

where $y = \text{vec(Y)}$, $\lambda = \text{vec}(\Lambda)$ and $e = \text{vec(E)}$, with

$$\begin{aligned} \Sigma &= \text{cov(y)} = \mathrm{I_I} \otimes \Sigma_\mathrm{o} \\ &= I_I \otimes \Delta \otimes \Omega. \end{aligned} \tag{4.19}$$

4.2.5 Application to bivariate growth curve data

Consider the model given by equation (4.18). We now consider a specific application of this model, restricting attention to bivariate longitudinal data for individuals in G groups. Let the growth curves for the first and second characteristics for an individual in the g^{th} group for $g = 1, 2, \ldots, G$ be polynomials in time of degree $(p-1)$. Then the expected value of the measurement at time t for characteristic 1 in group g is given by

$$\beta_{g0} + \beta_{g1}\, t + \beta_{g2}\, t^2 + \ldots + \beta_{g,p-1} t^{p-1}, \tag{4.20}$$

and the expected value of the measurement at time t for characteristic 2 is given by

$$\alpha_{g0} + \alpha_{g1}\, t + \alpha_{g2}\, t^2 + \ldots + \alpha_{g,p-1} t^{p-1}. \tag{4.21}$$

The observation matrix Y is arranged so that each column represents $2T$ measurements on an individual, arranged by characteristic and by time within characteristic. For every unit, we have $2T$ observations with a $2T \times 2T$ covariance matrix Σ_o. We expect the T observations on characteristic 1 to be correlated among themselves and the T observations on characteristic 2 to be correlated among themselves. This correlation is captured by the matrix Ω. Additionally, we expect the observations on characteristic 1 to be correlated with the observations on characteristic 2. This correlation is captured by the matrix Δ. Suppose we have n_g individuals in group g for $g = 1, 2, \ldots, G$. Then the matrices Θ, Λ and X are defined as follows:

$$\Theta = \begin{pmatrix} \mathbf{1}_{n_1} & 0 & \ldots & 0 \\ 0 & \mathbf{1}_{n_2} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \mathbf{1}_{n_G} \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \beta_{10} & \beta_{20} & \ldots & \beta_{G0} \\ \beta_{11} & \beta_{21} & \ldots & \beta_{G1} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1,p-1} & \beta_{2,p-1} & \ldots & \beta_{G,p-1} \\ \alpha_{10} & \alpha_{20} & \ldots & \alpha_{G0} \\ \alpha_{11} & \alpha_{21} & \ldots & \alpha_{G1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{1,p-1} & \alpha_{2,p-1} & \ldots & \alpha_{G,p-1} \end{pmatrix},$$

and

$$X = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^{p-1} \\ 1 & t_2 & t_2^2 & \ldots & t_2^{p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_T & t_T^2 & \ldots & t_T^{p-1} \end{pmatrix}.$$

In matrix Θ, $\mathbf{1}_{n_g}$ denotes a $n_g \times 1$ vector of unities. If $p = 2$ and $G = 2$, then Λ is a 4×2 matrix as given below:

$$\Lambda = \begin{pmatrix} \beta_{10} & \beta_{20} \\ \beta_{11} & \beta_{21} \\ \alpha_{10} & \alpha_{20} \\ \alpha_{11} & \alpha_{21} \end{pmatrix},$$

where β_{gp} and α_{gp} represent the coefficients for the first and second characteristic respectively for g^{th} group. Row 1 and row 3 consist of the intercept effects of groups 1 and 2 of the first and second characteristics, respectively. Row 2 and row 4 consists of the slope effects of groups 1 and 2 of the first and second characteristics, respectively. Consequently,

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \\ \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

4.2.6 Application to bivariate repeated measures ANOVA data

Now suppose $I = n_1 + n_2 + \ldots + n_G$ subjects in G treatment groups are measured repeatedly on two response variables of interest under T different experimental conditions. The T conditions represent the T levels of a factor of interest, with each subject being observed under all T levels of this factor. The goals of such a design include quantifying differences in the experimental conditions as well as between the groups. The observation matrix Y is arranged so that each column represents $2T$ measurements on an individual,

arranged by characteristic and by condition within characteristic. For n_g individuals in group g for $g = 1, 2, \ldots, G$, the matrices Θ, Λ and X are defined as follows:

$$\Theta = \begin{pmatrix} \mathbf{1}_{n_1} & 0 & \ldots & 0 \\ 0 & \mathbf{1}_{n_2} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \mathbf{1}_{n_G} \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \beta_{11} & \beta_{21} & \ldots & \beta_{G1} \\ \beta_{12} & \beta_{22} & \ldots & \beta_{G2} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1T} & \beta_{2T} & \ldots & \beta_{GT} \\ \alpha_{11} & \alpha_{21} & \ldots & \alpha_{G1} \\ \alpha_{12} & \alpha_{22} & \ldots & \alpha_{G2} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{1T} & \alpha_{2T} & \ldots & \alpha_{GT} \end{pmatrix},$$

and

$$X = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

For three conditions $T = 3$ and two groups $G = 2$, Λ is a 6×2 matrix as given below:

$$\Lambda = \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \\ \alpha_{13} & \alpha_{23} \end{pmatrix},$$

and consequently

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \alpha_{21} \\ \alpha_{22} \\ \alpha_{23} \end{pmatrix}.$$

4.3 Hypothesis testing

Consider testing

$$H_o : Q\lambda = 0 \quad \text{Vs.} \quad \text{H}_\text{a} : \text{Q}\lambda \neq 0, \tag{4.22}$$

where Q is a $r \times (pCG)$ matrix of rank $\leq pCG$. In the Pitman sense, the alternative hypothesis can be written as:

$$H_a : Q\lambda = \theta_I = I^{-\delta}\omega \tag{4.23}$$

so that $\lim_{I\to\infty} \theta_I = 0$. The estimate of $Q\lambda$ is $Q\hat{\lambda}$ and can be obtained either by ordinary least squares or by generalised least squares. Asymptotically, $Q\hat{\lambda}$ is a $r \times 1$ multivariate normal random variable with mean vector $Q\lambda$ and variance-covariance matrix QVQ' where $V = \text{var}(\hat{\lambda})$. From Corollary 2.3.4 on page 64 in Myers and Milton [44], we know that

$$(Q\hat{\lambda})' (QVQ')^{-1} (Q\hat{\lambda}) \tag{4.24}$$

follows a non-central χ^2 distribution with r degrees of freedom and non-centrality parameter

$$\frac{1}{2}(Q\lambda)'\left(QVQ'\right)^{-1}(Q\lambda). \tag{4.25}$$

Lemma 4.3.1 *To test the hypotheses*

$$H_o : Q\lambda = 0 \quad \text{Vs.} \quad \mathrm{H_a} : \mathrm{Q}\lambda \neq 0,$$

compute the test statistic:

$$T = (Q\hat{\lambda})'\left(QVQ'\right)^{-1}(Q\hat{\lambda}), \tag{4.26}$$

and compare it to a χ^2 distribution with r degrees of freedom.

4.3.1 Hypotheses testing: Some examples

Example 4.3.1 *Consider the application discussed in section 4.2.5. To test the overall hypothesis of parallelism, that is*

$$H_o : \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix},$$

compute the test statistic

$$T = (Q\hat{\lambda})'\left(QVQ'\right)^{-1}(Q\hat{\lambda})$$

and compare it to a χ^2 distribution with 4 degrees of freedom, where

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Rejecting H_o implies that the overall hypothesis of parallelism does not hold at significance level α. Note that under H_o,

$$Q\lambda = \begin{pmatrix} \beta_{10} - \beta_{20} \\ \beta_{11} - \beta_{21} \\ \alpha_{10} - \alpha_{20} \\ \alpha_{11} - \alpha_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Example 4.3.2 *Now consider the application discussed in section 4.2.6. To test the hypothesis that the mean vectors for the two treatment groups are equal, that is,*

$$H_o : \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \end{pmatrix} = \begin{pmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \alpha_{21} \\ \alpha_{22} \\ \alpha_{23} \end{pmatrix},$$

compute the test statistic

$$T = (Q\hat{\lambda})' (QVQ')^{-1} (Q\hat{\lambda})$$

and compare it to a χ^2 distribution with 6 degrees of freedom, where

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Rejecting H_o implies that the equality of means for the two groups does not hold at

significance level α. Note that under H_o,

$$Q\lambda = \begin{pmatrix} \beta_{11} - \beta_{21} \\ \beta_{12} - \beta_{22} \\ \beta_{13} - \beta_{23} \\ \alpha_{11} - \alpha_{21} \\ \alpha_{12} - \alpha_{22} \\ \alpha_{13} - \alpha_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

4.3.2 Power Discussion

Consider the hypothesis set discussed in section 4.3. At significance level α, the null hypothesis is rejected if

$$T = (Q\hat{\lambda})' (QVQ')^{-1} (Q\hat{\lambda}) > \chi_r^2(\alpha).$$

Lemma 4.3.2 *The power of the test when $Q\lambda \neq 0$ is given by*

$$\begin{aligned} Power &= Pr\left\{(Q\hat{\lambda})' (QVQ')^{-1} (Q\hat{\lambda}) > \chi_r^2(\alpha)\right\} \\ &= Pr\left\{\chi_r'^2\left(\frac{1}{2}(Q\lambda)' (QVQ')^{-1} (Q\lambda)\right) > \chi_r^2(\alpha)\right\} \\ &= 1 - Pr\left\{\chi_r'^2\left(\frac{1}{2}(Q\lambda)' (QVQ')^{-1} (Q\lambda)\right) < \chi_r^2(\alpha)\right\}, \end{aligned} \tag{4.27}$$

where $\chi_r^2(\alpha)$ represents the upper 100α percentage point of the chi-square distribution with r degrees of freedom.

For fixed α and given Q, the power can be evaluated when V is known.

4.4 Estimation of model parameters and associated variances

The parameters in equation (4.18) can be estimated by ordinary least squares, but if $\text{cov(e)} \neq \sigma^2 \text{I}_{\text{TC}}$, then this method will not in general be fully efficient. A fully efficient

method is generalised least squares based on the variance covariance matrix (4.19). This is only possible when the parameter vectors γ_1 and γ_2 defining Δ and Ω, respectively, are known, as is assumed in this chapter. The practical alternative is generalised least squares based on the matrix (4.19) evaluated at the estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of the parameter vectors γ_1 and γ_2. This is the subject of subsequent chapters.

While generalised least squares is almost always the most efficient method of analysis, ignorance of γ_1 and γ_2 makes it an unattainable ideal. Also, sampling variation in the estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$ also means that the generalised least squares will not always be more efficient than ordinary least squares. The efficiency of an analysis using ordinary least squares, relative to that obtained using generalised least squares, can be evaluated if the true values of γ_1 and γ_2 are assumed to be known.

Lemma 4.4.1 *The ordinary least squares estimator of $\hat{\lambda}$ is given by*

$$\hat{\lambda}_{ols} = (\Theta'\Theta)^{-1}\Theta' \otimes I_C \otimes (X'X)^{-1}X'y, \tag{4.28}$$

while the generalised least squares estimator is given by

$$\hat{\lambda}_{gls} = (\Theta'\Theta)^{-1}\Theta' \otimes I_C \otimes (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{4.29}$$

Lemma 4.4.2 *The variance of the ordinary least squares estimator is given by*

$$\mathrm{var}(\hat{\lambda}_{\mathrm{ols}}) = (\Theta'\Theta)^{-1} \otimes \Delta \otimes (\mathrm{X'X})^{-1}\mathrm{X'\Omega X}(\mathrm{X'X})^{-1}. \tag{4.30}$$

If one proceeds to assume that the correlation matrix is given by $\Sigma_o = \sigma^2 I_{TC} = \sigma^2 I_C \otimes I_T$, then the variance of the ordinary least squares estimator in this case would be taken to be

$$\mathrm{var}_{\mathrm{incorrect}}(\hat{\lambda}_{\mathrm{ols}}) = \sigma^2(\Theta'\Theta)^{-1} \otimes \mathrm{I_C} \otimes (\mathrm{X'X})^{-1}. \tag{4.31}$$

The variance of the generalised least squares estimator is given by

$$\mathrm{var}(\hat{\lambda}_{\mathrm{gls}}) = (\Theta'\Theta)^{-1} \otimes \Delta \otimes (\mathrm{X'\Omega^{-1}X})^{-1}. \tag{4.32}$$

See Reinsel [54] to understand how the estimators and the variance of these estimators were derived. Equations (4.30), (4.31) and (4.32) will be used in section 4.5 to assess the efficiency of the test based on the ordinary least squares estimator relative to the test based on the generalised least squares estimator for a range of possible values of the parameter vectors γ_1 and γ_2. The main focus will be on using the TARE and CARE to compare the efficiency of using equation (4.31) relative to equation (4.30). The efficiency of equation (4.30) relative to equation (4.32) is also considered.

4.5 Evaluating efficiency using TARE and CARE

Consider testing the hypothesis discussed in section 4.3 using the test statistic T given by equation (4.26). The power function based on this test is given by equation (4.27). Let ϕ_3 be the test based on the generalised least squares procedure which utilises the correct covariance structure (assumed to be Kronecker structured) in both the estimator and the estimator of the variance. Let ϕ_1 and ϕ_2 be tests based on the ordinary least squares procedure. The test ϕ_2 is "correct" in that it utilises the correct covariance structure in the variance of the estimator. The test ϕ_1 is "incorrect" in that it ignores the covariance structure in the variance of the estimator.

4.5.1 Efficiency of ϕ_2 relative to ϕ_1

For the test ϕ_2, the matrix V is given by equation (4.31) while for the test ϕ_1, it is given by equation (4.30). The non-centrality parameters in the power function (4.27) for the tests ϕ_2 and ϕ_1 are given by

$$\frac{1}{2}(Q\lambda)'\left\{Q\left(\sigma^2(\Theta'\Theta)^{-1}\otimes I_C\otimes(X'X)^{-1}\right)Q'\right\}^{-1}(Q\lambda) \tag{4.33}$$

and

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)Q'\right\}^{-1}(Q\lambda), \tag{4.34}$$

respectively.

Lemma 4.5.1 *The TARE and CARE of ϕ_2 with respect to ϕ_1 are given by*

$$\text{TARE}(\phi_2/\phi_1) = \frac{\text{tr}\left\{\text{Q}\left(\sigma^2(\Theta'\Theta)^{-1}\otimes \text{I}_\text{C}\otimes(\text{X}'\text{X})^{-1}\right)\text{Q}'\right\}^{-1}}{\text{tr}\left\{\text{Q}\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(\text{X}'\text{X})^{-1}\text{X}'\Omega\text{X}(\text{X}'\text{X})^{-1}\right)\text{Q}'\right\}^{-1}} \tag{4.35}$$

and

$$\begin{aligned}\text{CARE}(\phi_2/\phi_1) &= \left\{\frac{\left|\left\{Q\left(\sigma^2(\Theta'\Theta)^{-1}\otimes I_C\otimes(X'X)^{-1}\right)Q'\right\}^{-1}\right|}{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)Q'\right\}^{-1}\right|}\right\}^{\frac{1}{r}} \\ &= \left\{\frac{\left|Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)Q'\right|}{\left|Q\left(\sigma^2(\Theta'\Theta)^{-1}\otimes I_C\otimes(X'X)^{-1}\right)Q'\right|}\right\}^{\frac{1}{r}}, \end{aligned} \tag{4.36}$$

respectively.

4.5.2 Efficiency of ϕ_1 relative to ϕ_3

The non-centrality parameter for the test ϕ_1 is given by equation (4.34). For the test ϕ_3, the matrix V is given by equation (4.32). The non-centrality parameter in the power function (4.27) for the test ϕ_3 is therefore given by

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right\}^{-1}(Q\lambda). \tag{4.37}$$

Lemma 4.5.2 *The TARE and CARE of ϕ_1 with respect to ϕ_3 are given by*

$$\text{TARE}(\phi_1/\phi_3) = \frac{\text{tr}\left\{\text{Q}\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(\text{X}'\text{X})^{-1}\text{X}'\Omega\text{X}(\text{X}'\text{X})^{-1}\right)\text{Q}'\right\}^{-1}}{\text{tr}\left\{\text{Q}\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(\text{X}'\Omega^{-1}\text{X})^{-1}\right)\text{Q}'\right\}^{-1}} \tag{4.38}$$

and

$$\begin{aligned}\text{CARE}(\phi_1/\phi_3) &= \left\{\frac{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)Q'\right\}^{-1}\right|}{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right\}^{-1}\right|}\right\}^{\frac{1}{r}} \\ &= \left\{\frac{\left|Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right|}{\left|Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'X)^{-1}X'\Omega X(X'X)^{-1}\right)Q'\right|}\right\}^{\frac{1}{r}}, \end{aligned} \tag{4.39}$$

respectively.

4.6 Numerical results for special covariance structures

The TARE and CARE for evaluating the efficiency of ϕ_2 with respect to ϕ_1 are given by equations (4.35) and (4.36) respectively. Also, the TARE and CARE for evaluating the efficiency of ϕ_1 with respect to ϕ_3 are given by equations (4.38) and (4.39) respectively. These measures of asymptotic relative efficiency are computed with the following quantities being manipulated in the computations:

1. total sample size I;
2. within subject design matrix X;
3. the matrix Q, which is dependent on the hypothesis of interest;
4. for the covariance matrix Δ, the degree of correlation among the dependent variables (ρ_c) as well as the ratio of the variability of the various characteristics (γ);
5. the covariance matrix Ω, specified to be either compound symmetry or first order autoregressive;
6. the parameter ρ_t in Ω, representing correlation between any two measurements on a given characteristic on the same subject when it is specified to be compound symmetry. When Ω is specified to be first-order autoregressive, ρ_t is the correlation between successive observations on a given characteristic on the same subject.

The number of characteristics is kept constant at $C = 2$ and the number of repeated observations per characteristic is also kept constant at $T = 3$.

Consider the model discussed in section 4.2.5 and the test of hypotheses discussed in example 4.3.1. For $G = 2$, the matrices X, Θ and Q are defined as follows:

$$X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix},$$

$$\Theta = \begin{pmatrix} \mathbf{1}_{n_1} & 0 \\ 0 & \mathbf{1}_{n_2} \end{pmatrix},$$

with $n_1 = n_2 = 30, 60, 90$ and

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

The matrices Δ and Ω are modeled following models 3 and 4 in Table 4.1. This means that Δ is completely unstructured:

$$\Delta = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

reparameterised as

$$\Delta = \begin{pmatrix} \sigma_{11} & \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{22} \end{pmatrix},$$

with σ_{11} set to one. In contrast, Ω is either first-order autoregressive,

$$\Omega = \begin{pmatrix} 1 & \rho_t & \rho_t^2 \\ \rho_t & 1 & \rho_t \\ \rho_t^2 & \rho_t & 1 \end{pmatrix},$$

or compound symmetry

$$\Omega = \begin{pmatrix} 1 & \rho_t & \rho_t \\ \rho_t & 1 & \rho_t \\ \rho_t & \rho_t & 1 \end{pmatrix}.$$

The parameter values used for ρ_c, ρ_t and γ in the covariance matrices Δ and Ω are given in Table 4.2.

Table 4.2: Values of parameters defining Δ and Ω used in computing the TARE and CARE.

Parameter	**Values**
ρ_c	-0.9 *to* 0.9 *by* 0.1
ρ_t	0.1 *to* 0.9 *by* 0.1
γ	0.5 *to* 2.5 *by* 0.5

This gives a total of 855 parameter combinations which are varied enough to represent parameters that may arise in practice.

Now consider the model discussed in section 4.2.6 and the test of hypotheses discussed in example 4.3.2. For $G = 2$, the matrices X, Θ and Q are defined as follows:

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\Theta = \begin{pmatrix} \mathbf{1}_{n_1} & 0 \\ 0 & \mathbf{1}_{n_2} \end{pmatrix},$$

with $n_1 = n_2 = 30, 60, 90$ and

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

The covariance matrices Δ and Ω are modelled following the growth curve example discussed above. In the graphs that follow, the results from fitting the two models are presented obtained from evaluating the efficiency of test ϕ_2 relative to ϕ_1. The results from evaluating the efficiency of test ϕ_1 relative to ϕ_3 did not yield very interesting results in either model.

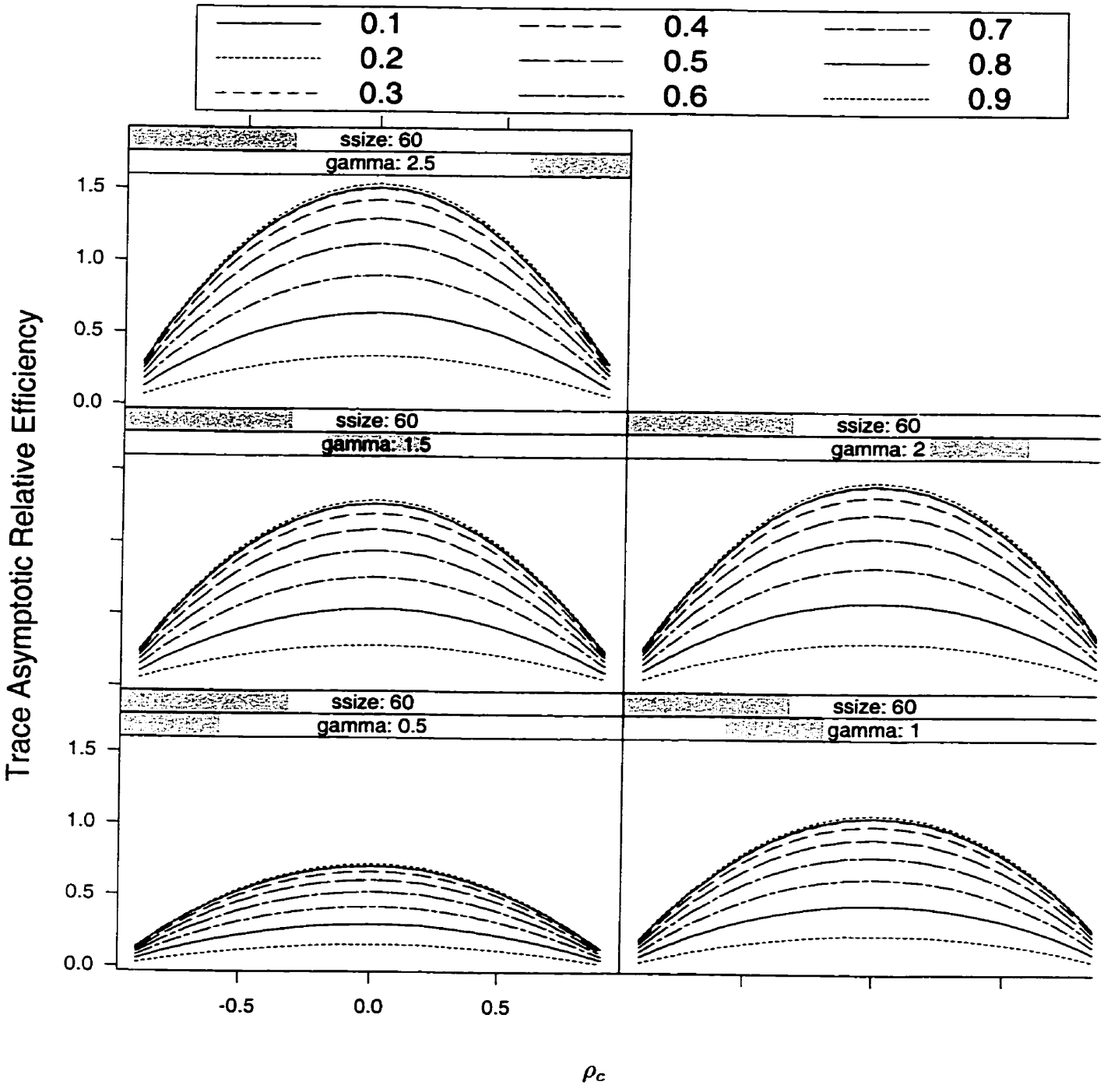

Figure 4.1: Values of TARE for ignoring correlation in the growth curve model: compound symmetry pattern for Ω.

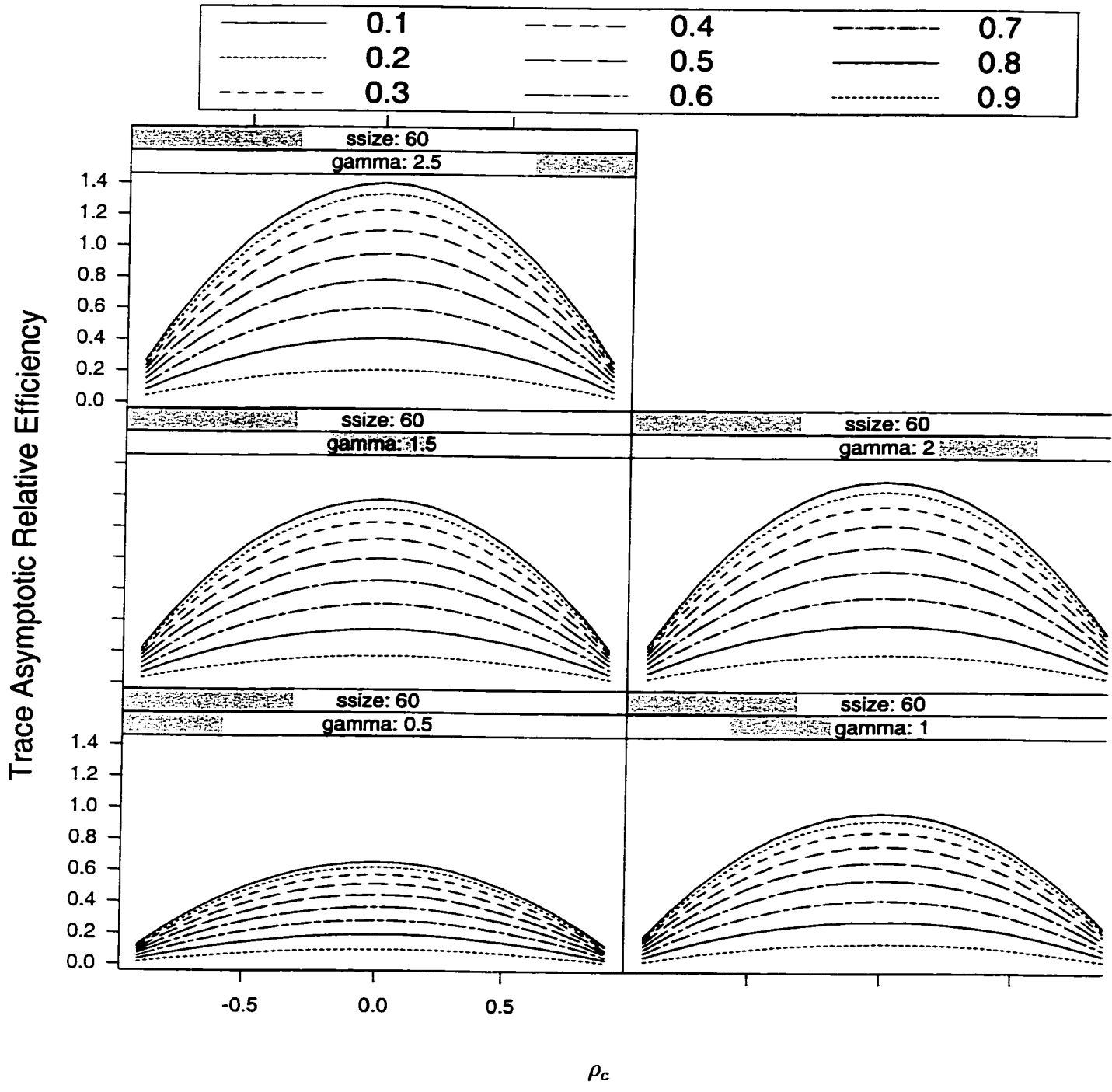

Figure 4.2: Values of TARE for ignoring correlation in the repeated measures analysis of variance model: compound symmetry pattern for Ω.

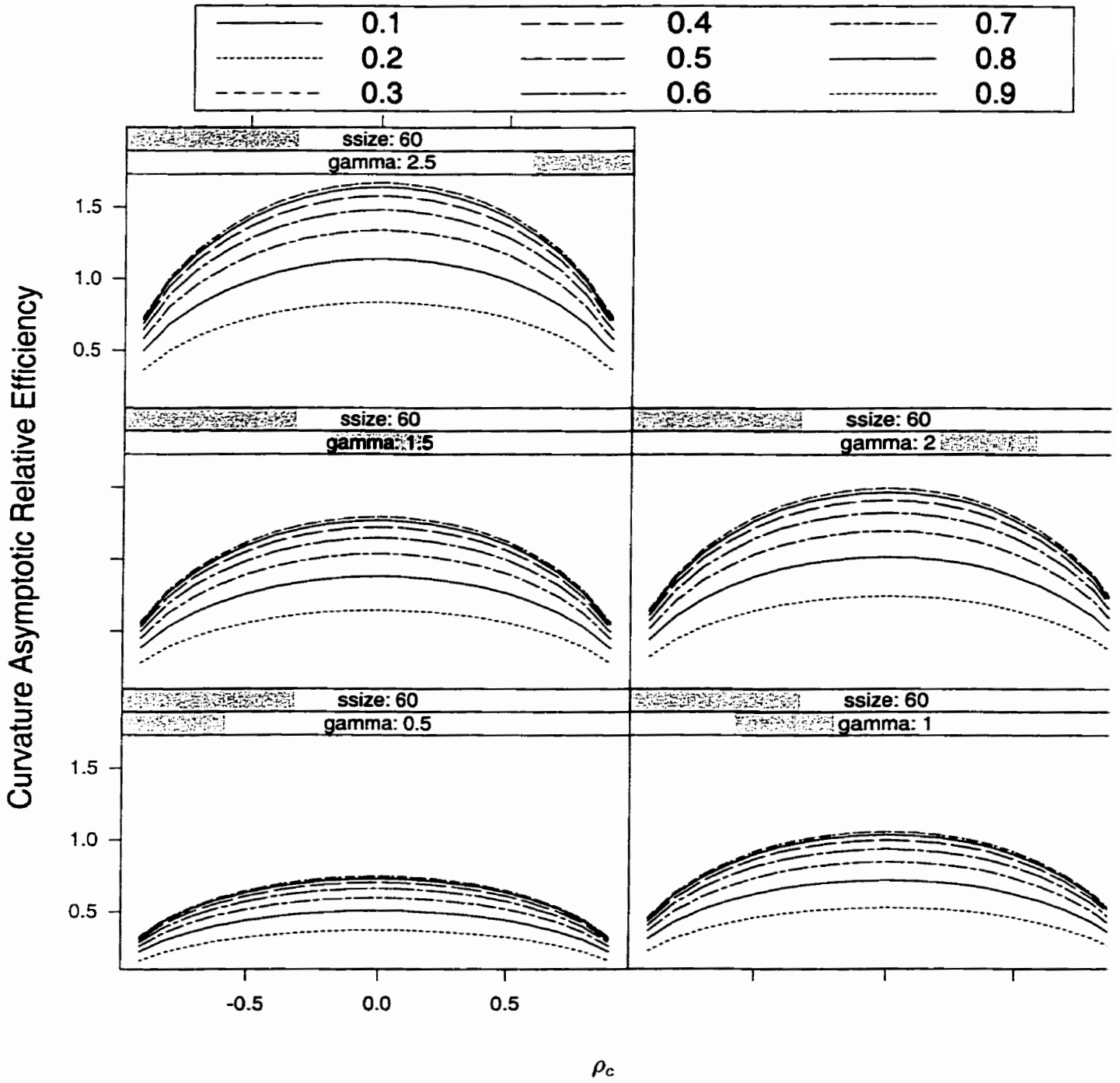

Figure 4.3: Values of CARE for ignoring correlation in the growth curve model: compound symmetry pattern for Ω.

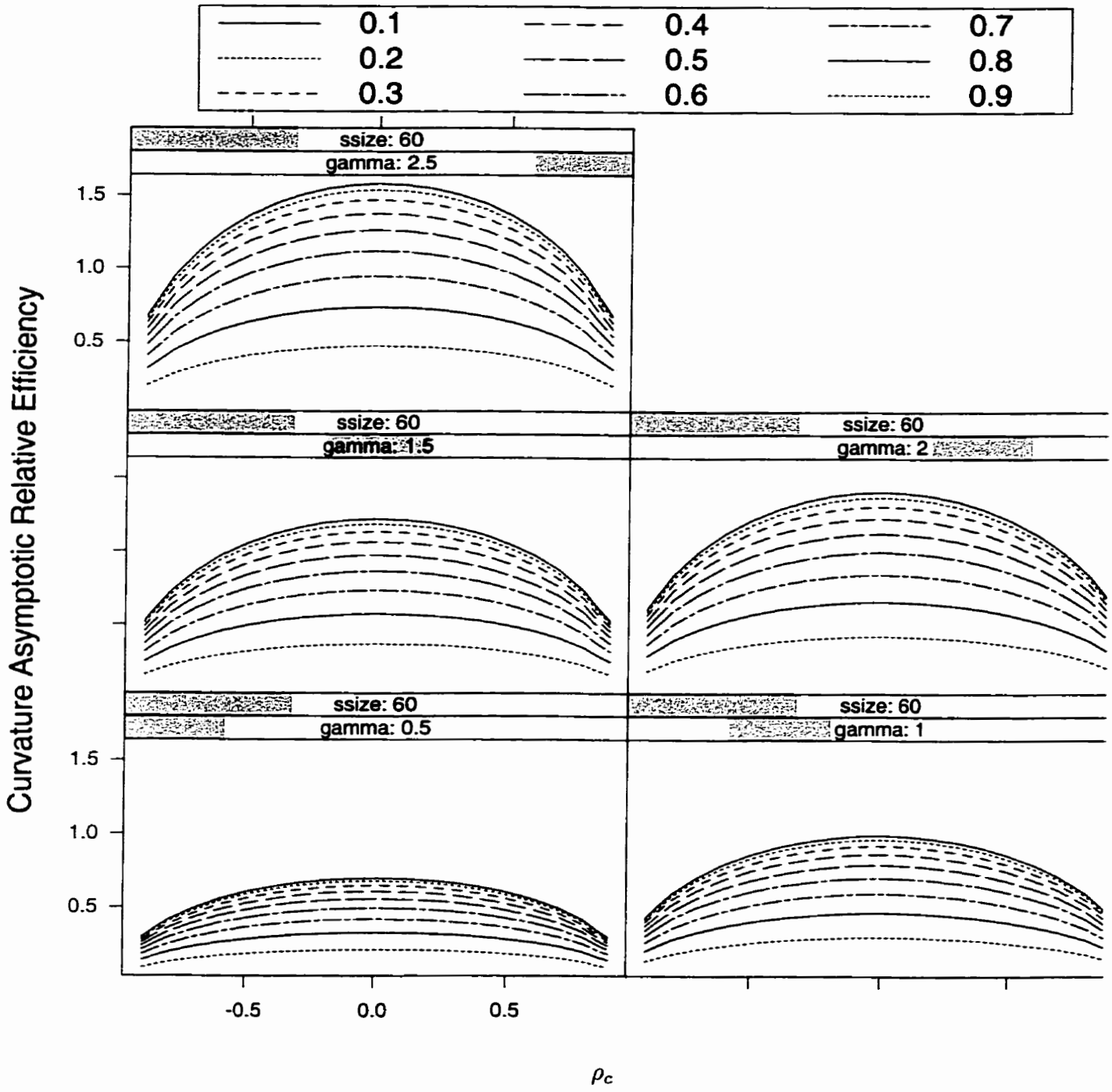

Figure 4.4: Values of CARE for ignoring correlation in the repeated measures analysis of variance model: compound symmetry pattern for Ω.

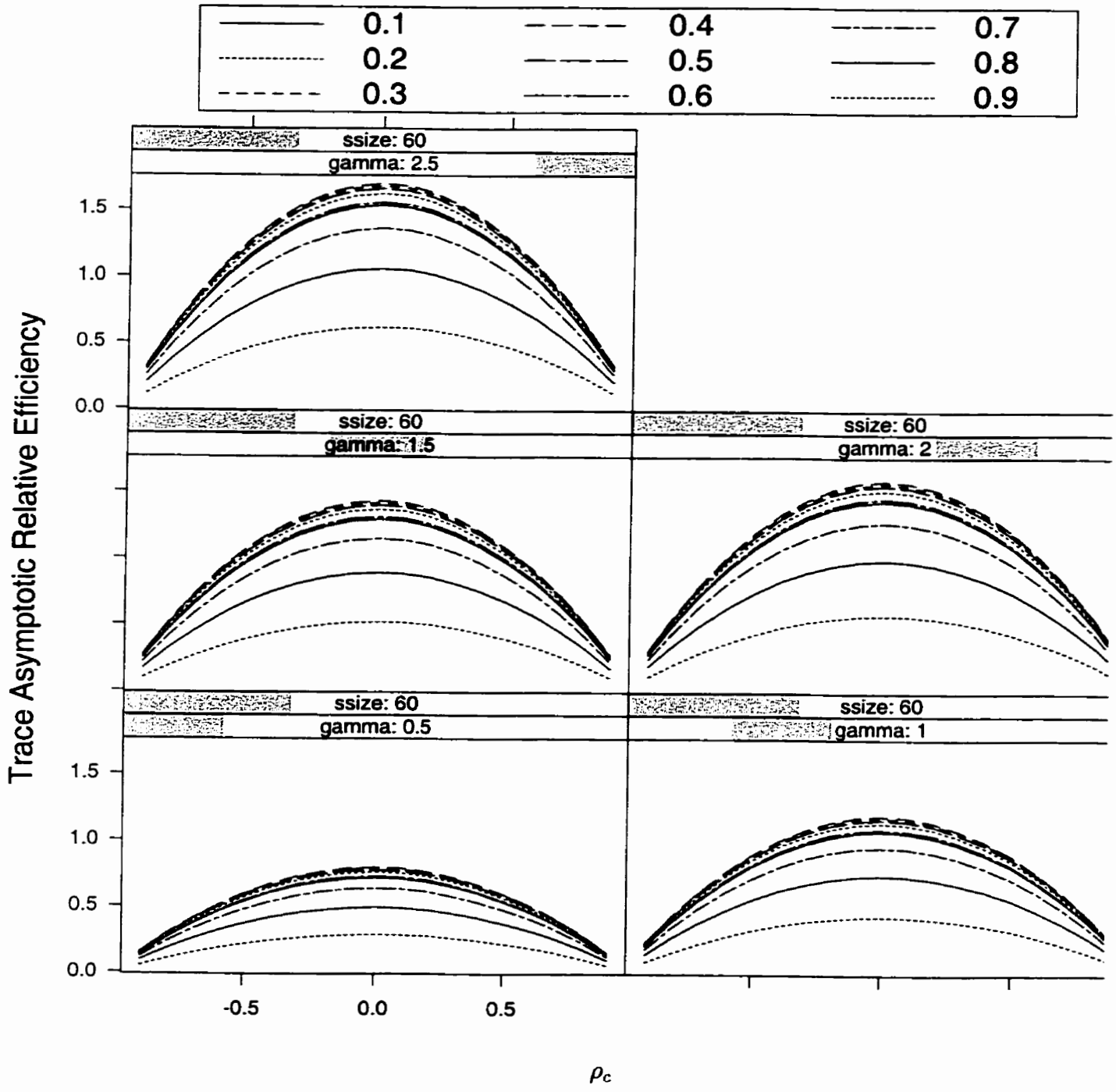

Figure 4.5: Values of TARE for ignoring correlation in the growth curve model: first-order autoregressive pattern for Ω.

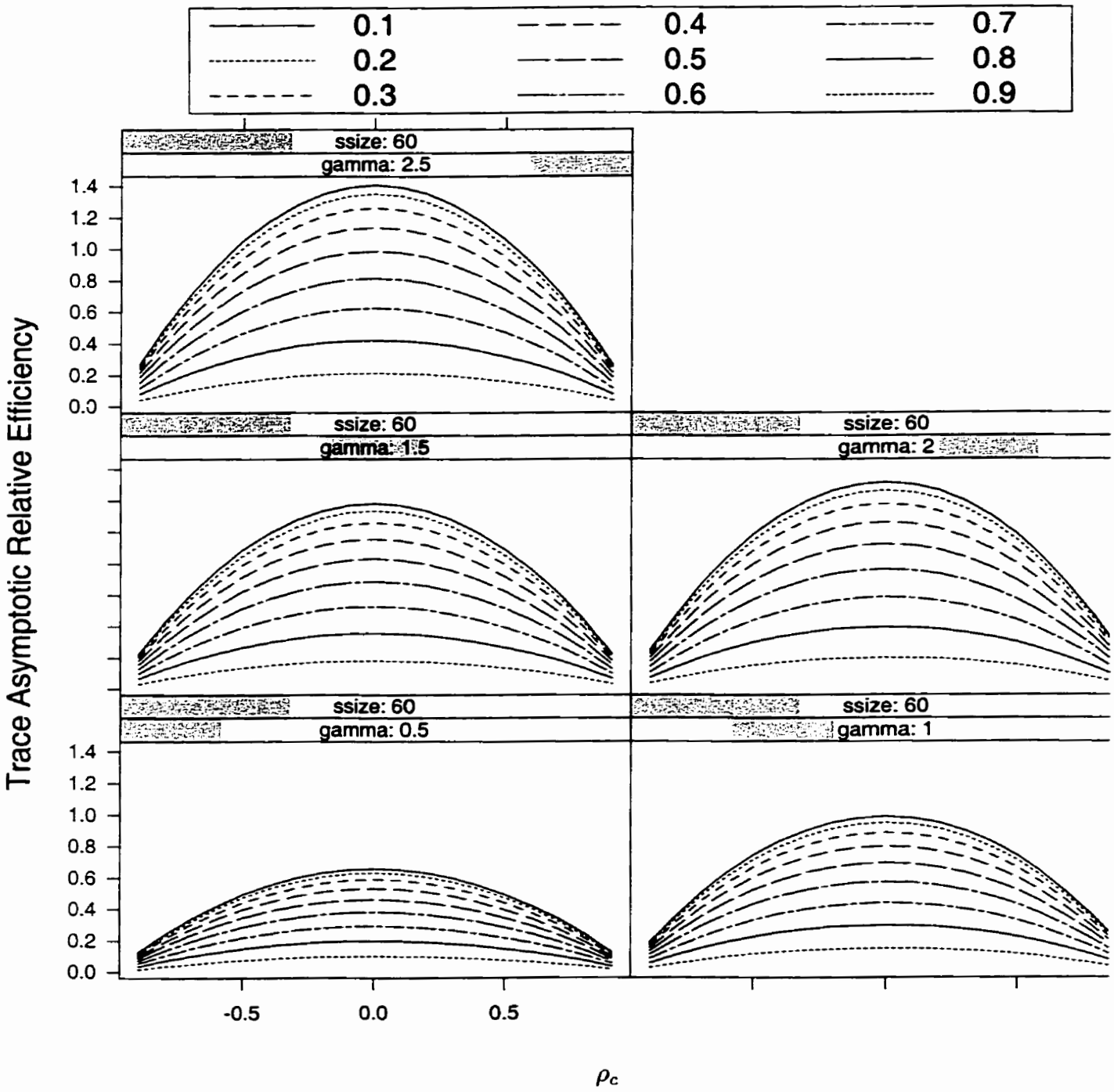

Figure 4.6: Values of TARE for ignoring correlation in the repeated measures analysis of variance model: first-order autoregressive pattern for Ω.

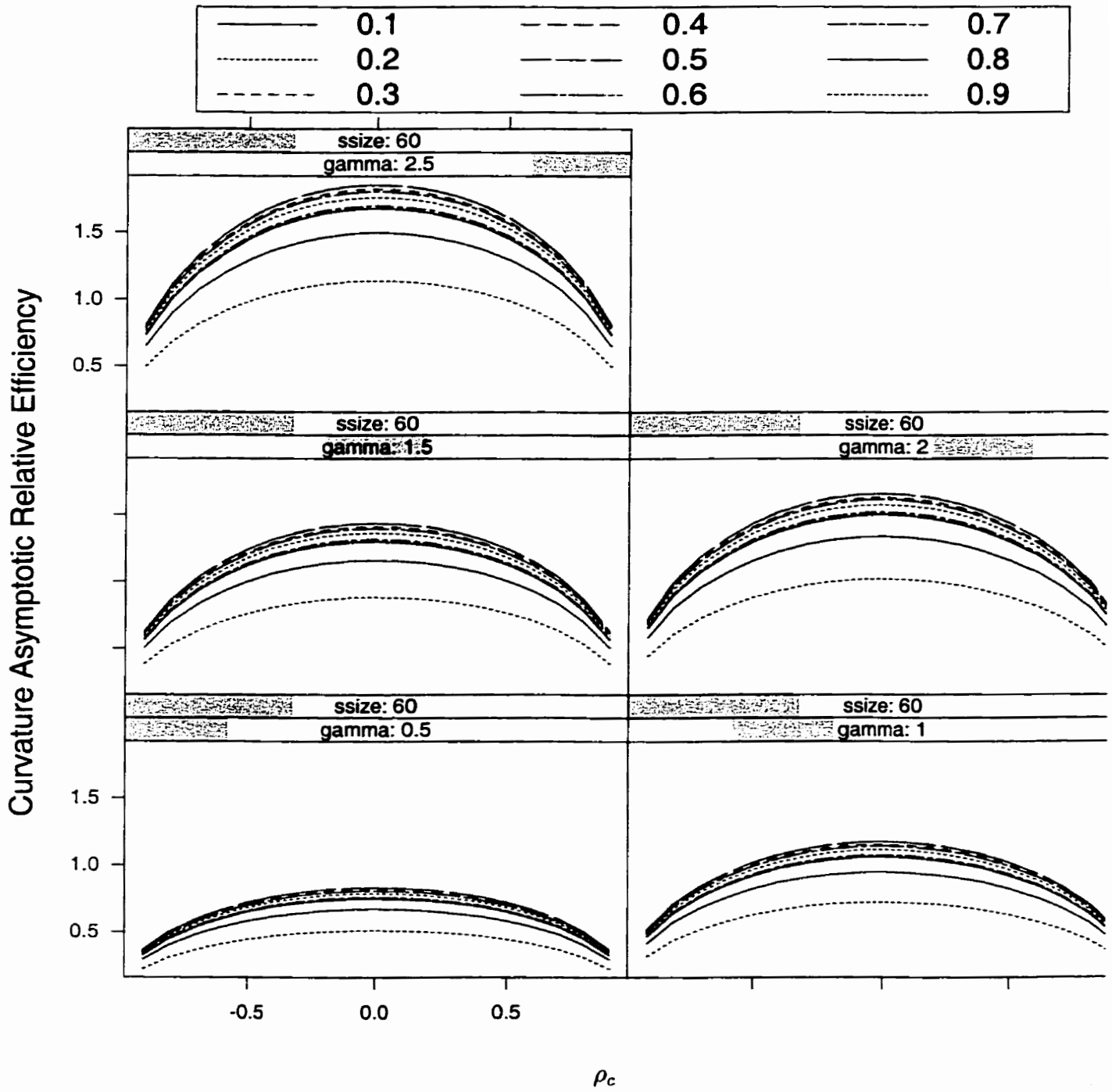

Figure 4.7: Values of CARE for ignoring correlation in the growth curve model: first-order autoregressive pattern for Ω.

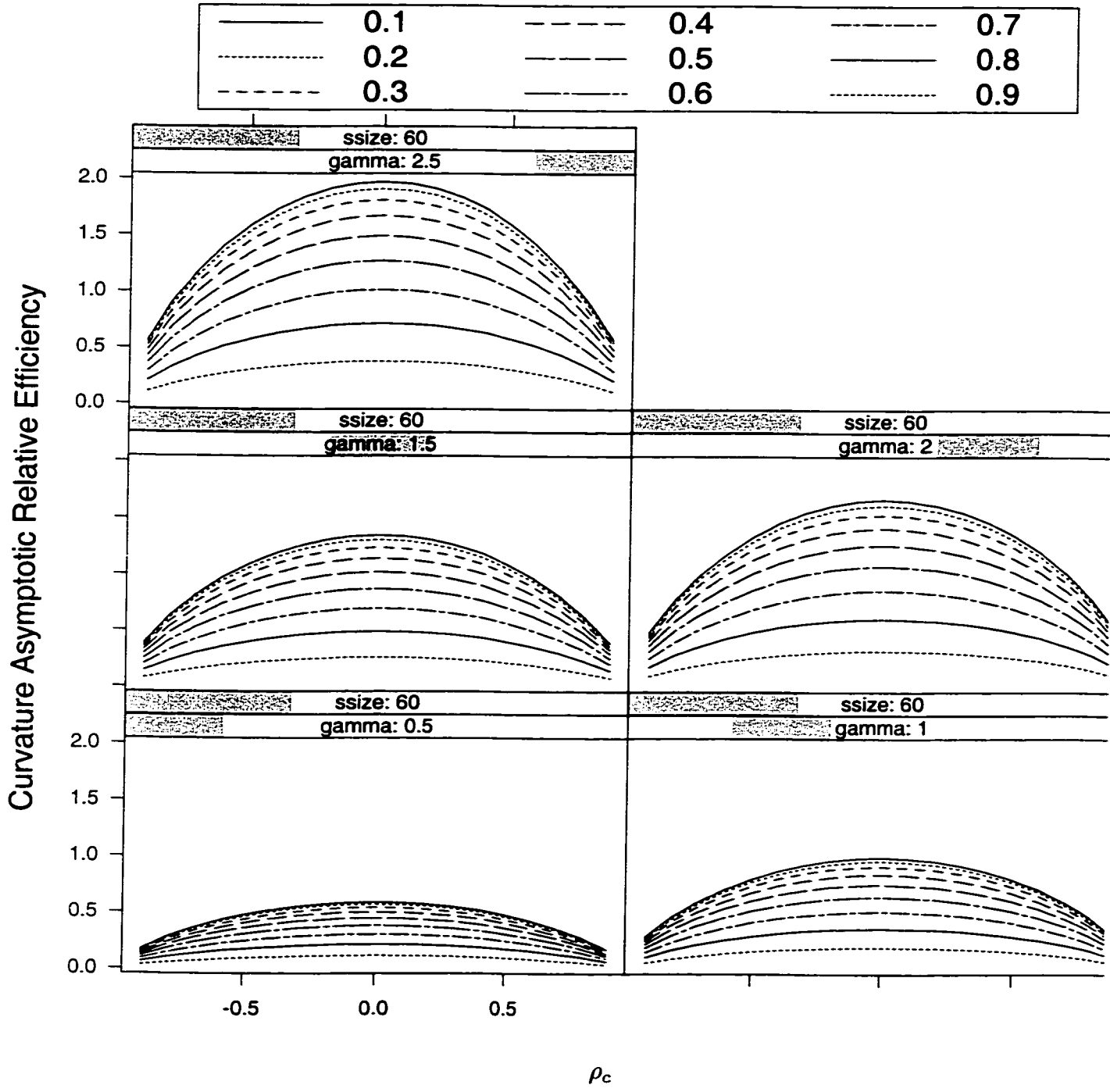

Figure 4.8: Values of CARE for ignoring correlation in the repeated measures analysis of variance model: first-order autoregressive pattern for Ω.

4.7 Discussion

The results displayed in Figures 4.1 to 4.8 can be summarised as follows:

1. First, we observe that overall, the results for the two designs investigated (growth curve and repeated measures ANOVA) are quite similar. For example, Figure 4.1 and Figure 4.2 are quite similar. This is especially true when Ω takes on the compound symmetry covariance pattern.

2. For both the compound symmetry model and the first-order autoregressive model, the graphs clearly show that the covariance parameters ρ_t, ρ_c and γ have a pronounced impact on both the TARE and CARE. It appears that the loss of efficiency, when it occurs, from ignoring correlation and the degree of the loss of efficiency are both functions of the covariance parameters γ, ρ_c and ρ_t. Overall, it appears that the efficiency of test ϕ_2 relative to test ϕ_1 is poor for high absolute values of ρ_t and ρ_c and low values of γ.

3. To gain a better understanding of the results displayed in the plots, consider a single graph in the multi-panel display. First, we observe that for a given value of ρ_t, the plot is symmetric about $\rho_c = 0$. Again, if we consider a single value of ρ_t (a single curve in the plot), the largest efficiency is observed at $\rho_c = 0$. The efficiency of test ϕ_2 relative to test ϕ_1 is worst at high negative and high positive values of ρ_c (note the shape of the curves as ρ_c moves from -0.9 to $+0.9$). The efficiency is clearly decreasing as ρ_c approaches -1 and $+1$.

4. Now we examine the effect of the parameter ρ_t. Overall, the efficiency of ϕ_2 relative to ϕ_1 is low for high values of ρ_t (values of ρ_t closer to 1). Observe that the higher the value of ρ_t, the lower the efficiency of ϕ_2 relative to ϕ_1. In fact, as ρ_t approaches 0, the maximum efficiency gets larger.

5. A question of considerable interest is: are all the plots (panels) on a given figure the same? To answer this question, we now shift focus from a single plot to the five plots or panels displayed on each figure. The five plots on any given figure display

the effect of γ (the ratio of the variances of the two characteristics) for a given design and covariance structure. The panels are clearly not the same. Overall, efficiency increases as a function of γ. For example, consider a single value of ρ_t in each plot (any single curve) and observe what happens to this curve as we move from plot to plot corresponding to different values of γ. As γ increases, so does the efficiency of ϕ_2 relative to ϕ_1. Observe that when γ is greater than one, test ϕ_2 appears to be more efficient than test ϕ_1 for a restricted range of ρ_c and low values of ρ_t. The test ϕ_2 is substantially inefficient relative to the test ϕ_1 when γ is less than or equal to one as demonstrated by the two bottom panels in each figure. The efficiency is especially poor for large values of ρ_t. The maximum efficiency achieved for each design and covariance structure also changes as γ increases.

Overall, the results are as expected. For the designs and covariance structures considered, a loss of efficiency is shown to occur from ignoring the two sources of correlation in testing hypotheses of interest. The loss is greatest when (i) the correlation between the characteristics is high and (ii) when the correlation between longitudinal measurements on a given characteristic is high. Ignoring these two correlations when they have high values is statistically less efficient than if one appropriately accounts for these correlations.

Chapter 5

Increasing efficiency from multivariate longitudinal data by using a Kronecker product to model the covariance structure

5.1 Introduction

A question of considerable practical interest, and the focus of the present chapter, concerns the potential gain in efficiency that would result from exploiting the Kronecker product covariance structure. As in the previous chapter, efficiency is defined in terms of testing hypotheses that are of interest in a given problem. Assessing the gain in efficiency that results from using a Kronecker structured covariance matrix will be accomplished by evaluating the efficiency of a test based on a completely unstructured covariance matrix relative to one based on a Kronecker structured covariance matrix. The measures of efficiency used are the TARE and CARE. Results obtained from the investigation will enable us to make general statements about the usefulness of utilising the Kronecker product structure when it exists in hypotheses testing for multivariate longitudinal data.

Additionally, we will be able to state what parameter ranges signify more serious consequences, if any, in ignoring the Kronecker product structure.

A second goal of the present chapter is to describe a preliminary test of H_o : $\Sigma_o = \Delta \otimes \Omega$ versus H_a : $\Sigma_o = \Sigma_a$, where Σ_a is an arbitrary covariance matrix in the analysis of multivariate longitudinal data. Incorporating the test in practical work will be useful in that it will provide protection against doing the wrong thing and increase efficiency if we do the right thing. Section 5.2 presents the model to be used. Section 5.3 discusses likelihood estimation of model parameters. Section 5.4 introduces the test for the Kronecker product pattern. Section 5.5 discusses hypothesis testing and power. Several examples are presented in section 5.6. Evaluating efficiency using TARE and CARE is given in section 5.7. Section 5.8 presents a Monte-carlo simulation study used to assess the gain in efficiency that results from using a Kronecker structured covariance matrix in hypothesis testing. The chapter closes with results obtained from the Monte-carlo simulation study and a general discussion in section 5.9.

5.2 Model specifications

The model presented section 4.2 is assumed in this chapter. Additionally, it is assumed that

$$y \sim MVN(\mu, \Sigma),$$

where

$$\mu = (\Theta \otimes I_C \otimes X)\lambda$$

and

$$\begin{aligned} \Sigma &= \text{cov(y)} = \mathrm{I_I} \otimes \Sigma_o \\ &= I_I \otimes \Delta \otimes \Omega. \end{aligned}$$

Suppose, for example, that we have two characteristics measured on each of three occasions for each subject. Then Δ is given by

$$\Delta = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

and is reparameterised as follows (to facilitate interpretation):

$$\Delta = \begin{pmatrix} \sigma_{11} & \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ \rho_c\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & b \\ b & \gamma \end{pmatrix},$$

with σ_{11} set to one. If Ω is assumed to be first-order autoregressive, then:

$$\Omega = \begin{pmatrix} 1 & \rho_t & \rho_t^2 \\ \rho_t & 1 & \rho_t \\ \rho_t^2 & \rho_t & 1 \end{pmatrix},$$

with the upper left element set to 1 to avoid nonidentifiability of Σ_o. In this case, parameter vectors γ_1 and γ_2 defining Δ and Ω, respectively, are given by $\gamma_1 = (\gamma, \rho_c)$ and $\gamma_2 = (\rho_t)$. γ represents the ratio of the variances for the two characteristics and ρ_c is the correlation between them. All covariance parameters are assumed to be unknown and must be estimated from the data using maximum likelihood or restricted maximum likelihood estimation.

5.3 Likelihood estimation of model parameters

Since y is assumed to be

$$y \sim MVN\left\{(\Theta \otimes I_C \otimes X)\lambda, I_I \otimes \Delta \otimes \Omega\right\}, \tag{5.1}$$

then parameters of interest can be estimated using maximum likelihood estimation or restricted maximum likelihood estimation. To find the likelihood estimators of λ, Δ and Ω, the likelihood function (and hence the log-likelihood function) is derived in terms of λ, Δ and Ω.

Lemma 5.3.1 *If $\Delta = \Delta_o$ and $\Omega = \Omega_o$, the maximum likelihood estimate or the restricted maximum likelihood estimate of λ is given by*

$$\hat{\lambda} = (\Theta'\Theta)^{-1}\Theta' \otimes I_C \otimes (X'\Omega_o^{-1}X)^{-1}X'\Omega_o^{-1}y. \tag{5.2}$$

Refer back to equation (4.29). Equation (5.2) is substituted back into the log-likelihood function and the resulting equation, a function of Δ_o and Ω_o, is maximised with respect to γ_1 and γ_2. Maximisation yields $\hat{\Delta}$ and $\hat{\Omega}$, the maximum likelihood or restricted maximum likelihood estimates of Δ_o and Ω_o respectively.

Lemma 5.3.2 *Substituting the likelihood estimates back into equation (5.2), we get:*

$$\hat{\lambda}_{\Delta\times\Omega} = (\Theta'\Theta)^{-1}\Theta' \otimes I_C \otimes (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y, \tag{5.3}$$

with

$$\hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega}) = (\Theta'\Theta)^{-1} \otimes \hat{\Delta} \otimes (X'\hat{\Omega}^{-1}X)^{-1}. \tag{5.4}$$

Lemma 5.3.3 *If λ is estimated ignoring the Kronecker product structure by using an arbitrary completely unstructured $CT \times CT$ matrix, then*

$$\hat{\lambda}_{un} = (\Theta'\Theta)^{-1}\Theta' \otimes (X^{*\prime}\hat{\Sigma}_o^{-1}X^*)^{-1}X^{*\prime}\hat{\Sigma}_o^{-1}y, \tag{5.5}$$

with

$$\hat{\text{var}}(\hat{\lambda}_{un}) = (\Theta'\Theta)^{-1} \otimes (X^{*\prime}\hat{\Sigma}_o^{-1}X^*)^{-1}, \tag{5.6}$$

where $X^* = I_C \otimes X$ and $\hat{\Sigma}_o$ is the maximum likelihood or restricted maximum likelihood estimate of Σ_o, an arbitrary $CT \times CT$ matrix.

5.4 Testing for the Kronecker product pattern

We can test the null hypothesis that the $CT \times CT$ covariance matrix Σ_o has the Kronecker product structure $\Sigma_o = \Delta \otimes \Omega$ using the likelihood ratio test statistic:

$$\text{LR} = \left\{ \frac{|\hat{\Sigma}_o|}{|\hat{\Delta}|^T |\hat{\Omega}|^C} \right\}^{\frac{-I}{2}}, \tag{5.7}$$

where Σ_o has an arbitrary pattern and $\hat{\Sigma}_o$ is its maximum likelihood or restricted maximum likelihood estimate. $\hat{\Delta}$ and $\hat{\Omega}$ are the maximum likelihood or restricted maximum likelihood estimates of Δ and Ω assuming Σ_o is a Kronecker structured covariance matrix. Under the null hypothesis, the quantity $2\log_e(LR)$ has an asymptotic Chi-square distribution with

$$\frac{TC(TC+1)}{2} - \frac{C(C+1)+2}{2} \tag{5.8}$$

degrees of freedom when Ω is assumed to be compound symmetry or first-order autoregressive. The test described here is similar to the test discussed in Diaz and Johnson [11] for testing for the Wiener stochastic process pattern in the covariance matrix of multivariate repeated measures data.

5.5 Hypothesis testing and power

Hypothesis testing concerning the parameter vector λ is based on the result (5.3) which, in conjunction with (5.1), implies that

$$\hat{\lambda} \sim MVN\left\{\lambda, (\Theta'\Theta)^{-1} \otimes \Delta \otimes (X'\Omega^{-1}X)^{-1}\right\}. \tag{5.9}$$

We assume (5.9) continues to hold, to a good approximation, if we replace γ_1 and γ_2, the parameter vectors defining Δ and Ω respectively, with their likelihood estimates. This gives

$$\hat{\lambda} \sim MVN\left\{\lambda, \hat{V} = \hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega})\right\} \tag{5.10}$$

where $\hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega})$ is given by equation (5.4). If a completely unstructured covariance matrix has been used, then $\hat{\lambda}$ is given by (5.5) and

$$\hat{\lambda} \sim MVN\left\{\lambda, \hat{V} = \hat{\text{var}}(\hat{\lambda}_{un})\right\}, \tag{5.11}$$

where $\hat{\text{var}}(\hat{\lambda}_{un})$ is given by equation (5.6). Consider testing the hypotheses discussed in section 4.3. In this case, the estimate of $Q\lambda$ is $Q\hat{\lambda}$, with $\hat{\lambda}$ given either by equation 5.2 if the parameters defining Δ and Ω are known or by equations 5.3 and 5.5 if the covariance parameters have to be estimated. Since

$$\hat{\lambda} \sim MVN(\lambda, \hat{V}), \tag{5.12}$$

then

$$Q\hat{\lambda} \sim MVN\left\{Q\lambda, Q\hat{V}Q'\right\}, \tag{5.13}$$

and

$$(Q\hat{\lambda})'\left(Q\hat{V}Q'\right)^{-1}(Q\hat{\lambda}) \tag{5.14}$$

follows a non-central χ^2 distribution with r degrees of freedom and non-centrality parameter

$$\frac{1}{2}(Q\lambda)'\left(Q\hat{V}Q'\right)^{-1}(Q\lambda). \tag{5.15}$$

Lemma 5.5.1 *To test the hypotheses*

$$H_o : Q\lambda = 0 \quad \text{Vs.} \quad \mathrm{H_a} : \mathrm{Q}\lambda \neq 0,$$

compute the test statistic

$$T^* = (Q\hat{\lambda})'\left(Q\hat{V}Q'\right)^{-1}(Q\hat{\lambda}), \tag{5.16}$$

and compare it to a χ^2 distribution with r degrees of freedom.

Following section 4.3.2, the null hypothesis $H_o : Q\lambda = 0$ is rejected at level α if

$$T^* = (Q\hat{\lambda})' (Q\hat{V}Q')^{-1} (Q\hat{\lambda}) > \chi^2_r(\alpha).$$

Lemma 5.5.2 *The power of the test under the alternative hypothesis* $H_a : Q\lambda \neq 0$ *is given by*

$$Power = 1 - Pr\left\{\chi'^2_r\left(\frac{1}{2}(Q\lambda)' (Q\hat{V}Q')^{-1} (Q\lambda)\right) < \chi^2_r(\alpha)\right\}, \tag{5.17}$$

where $\chi^2_r(\alpha)$ *represents the upper* 100α *percentage point of the (central) chi-square distribution with* r *degrees of freedom.*

For fixed α and given Q, the power can be evaluated once $\hat{V}$ has been obtained from the data using maximum likelihood or restricted maximum likelihood estimation.

5.6 Examples

In this section, we present three examples to illustrate the application of the linear model for multivariate longitudinal data with a Kronecker structured covariance matrix. The examples will also serve to illustrate multivariate longitudinal designs that frequently occur in practice. The first example considers a growth curve setting while the second example considers a repeated measures analysis of variance problem. The third example re-visits the data introduced in Chapter 1. One of the major roles of these examples is to provide structure to simulations that will be carried out later in this chapter and also in the next chapter.

The first example concerns 18 patients randomized to two treatment groups in order to evaluate the changes in vertical position on the mandible. Three variables, called SOr-Me (mm), ANS-Me (mm), and Pal-Me (degrees) were measured at three time points during therapy. The data has been discussed previously in Timm [65], Thomas [63] and Boik [6]. Timm [65] gives the mean plots of the data for each group and variable, which suggests that the growth curves for the three variables are approximately linear.

A question of practical interest concerns whether the growth curves for the two groups are parallel for one or more variables. For our purposes, we will consider only the last two variables, that is, ANS-Me and Pal-Mp angle.

The model fit to the two variables is

$$y = (\Theta \otimes I_C \otimes X)\lambda + e, \tag{5.18}$$

with

$$\Theta = \begin{pmatrix} \mathbf{1}_9 & 0 \\ 0 & \mathbf{1}_9 \end{pmatrix},$$

$$I_C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \\ \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

The error vector e is assumed to have covariance matrix Σ given by $\Sigma = \text{cov(e)} = \text{I}_\text{I} \otimes \Sigma_\text{o} = \text{I}_\text{I} \otimes \Delta \otimes \Omega$, and Ω is assumed to be first-order autoregressive. The estimates

Table 5.1: Estimated regression coefficients for the growth curve example.

Group	Characteristic	Parameter	Estimate	Standard Error
1	1	β_{10}	64.411	1.782
1	1	β_{11}	1.194	0.2998
1	2	α_{10}	24.893	0.589
1	2	α_{11}	0.222	0.267
2	1	β_{20}	65.726	1.782
2	1	β_{21}	1.444	0.2998
2	2	α_{20}	24.146	1.589
2	2	α_{21}	0.144	0.267

Table 5.2: Estimated covariance parameters for the growth curve example.

Covariance Matrix	Parameter	Estimate
Δ	σ_{11}	29.374
	σ_{12}	1.257
	σ_{22}	23.378
Ω Auto-Regressive(1)	ρ	0.972

obtained for the regression coefficients, obtained by maximum likelihood estimation, are given in Table 5.1. The covariance parameter estimates are given in Table 5.2.

To test for parallelism simultaneously for both variables, that is

$$H_o : \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix},$$

we compute the test statistic

$$T = (Q\hat{\lambda})' (Q\hat{V}Q')^{-1} (Q\hat{\lambda})$$

and compare it to a χ^2 distribution with 4 degrees of freedom where:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

We obtain $T = 0.803$ with a p-value of 0.938. From this, we conclude that the two treatments do not differ significantly with respect to their linear growth curves. This is in agreement with conclusions drawn by others who have looked at this data.

The second example considers a repeated measures analysis of variance setting, with measurements taken under what we can think of as three experimental conditions. The data is discussed in Hand and Crowder [21]. The data relates to patients who suffer from panic attacks (group 1) and the control set who do not suffer from panic attacks (group 2). 11 repeated measures are recorded on 3 variables: the first variable is the score on an anxiety scale, increasing from 0 to 8; the second is CO_2 expiration; and the third is pulse rate. The three variables are recorded together at times $4, 6, 8, 10, 11, 14, 16, 17, 18, 19$ and 23 minutes. Times $4, 11, 14, 19$ and 23 are rest times. Times $6, 8$ and 10 are times at which subjects are spoken to on the topic about which they are anxious. Times $6, 17$ and 18 are times at which subjects are asked to hyperventilate. There was missing data for some of the subjects on the response variables.

As mentioned in Hand and Crowder [21], the scope for relating anxiety scores to the explanatory variables is wide. Their analysis focused on the effects of "group" and "circumstances" (rest, spoken to, hyperventilate), on CO_2 expiration. We will also focus on the effects of "group" and "circumstances", applying the linear model for multivariate longitudinal data with Kronecker structured covariance matrix. The data is treated as a two sample bivariate repeated measures ANOVA, with the subjects in two groups measured repeatedly under three different conditions (rest, spoken to, hyperventilate). Times $4, 10$ and 16 are chosen to correspond to each of the three conditions respectively. The two variables considered are anxiety scores and CO_2 expiration.

The model fit to the two variables is:

$$y = (\Theta \otimes I_C \otimes X)\lambda + e \tag{5.19}$$

with:

$$\Theta = \begin{pmatrix} \mathbf{1}_{21} & 0 \\ 0 & \mathbf{1}_{19} \end{pmatrix},$$

$$I_C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \alpha_{21} \\ \alpha_{22} \\ \alpha_{23} \end{pmatrix}.$$

The error vector e is assumed to have covariance matrix Σ given by $\Sigma = \text{cov(e)} = \text{I}_\text{I} \otimes \Sigma_\text{o} = \text{I}_\text{I} \otimes \Delta \otimes \Omega$, and Ω is assumed to be compound symmetry. The estimates obtained for

Table 5.3: Estimated regression coefficients for the repeated measures analysis of variance example.

Group	Characteristic	Parameter	Estimate	Standard Error
1	1	β_{11}	3.667	0.414
1	1	β_{12}	6.000	0.414
1	1	β_{13}	5.571	0.414
1	2	α_{11}	32.190	0.836
1	2	α_{12}	30.095	0.836
1	2	α_{13}	19.190	0.836
2	1	β_{21}	1.000	0.435
2	1	β_{22}	4.684	0.435
2	1	β_{23}	2.842	0.435
2	2	α_{21}	35.947	0.879
2	2	α_{22}	34.053	0.879
2	2	α_{23}	19.579	0.879

Table 5.4: Estimated covariance parameters for the repeated measures analysis of variance example.

Covariance Matrix	Parameter	Estimate
Δ	σ_{11}	3.600
	σ_{12}	-0.752
	σ_{22}	14.665
Ω	ρ	0.395
Compound Symmetry		

the regression coefficients, using maximum likelihood estimation, are given in Table 5.3. The covariance parameter estimates are given in Table 5.4.

To test the hypothesis that the mean vectors for the two groups (panic=yes,

panic=no) are equal simultaneously for both variables, that is:

$$H_o: \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \end{pmatrix} = \begin{pmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \alpha_{21} \\ \alpha_{22} \\ \alpha_{23} \end{pmatrix},$$

compute the test statistic:

$$T = (Q\hat{\lambda})' (Q\hat{V}Q')^{-1} (Q\hat{\lambda})$$

and compare it to a χ^2 distribution with 6 degrees of freedom where:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

We obtain $T = 44.912$ with a p-value of 4.87×10^{-8}. From this, we conclude that the two groups do differ significantly with respect to anxiety scores or CO_2 expiration or both.

The third example re-visits the data introduced in Chapter 1. In treating the data as a growth curve problem, the model fit to FEV1 and FVC is:

$$y = (\Theta \otimes I_C \otimes X)\lambda + e \tag{5.20}$$

with:

$$\Theta = \begin{pmatrix} \mathbf{1}_{52} & 0 \\ 0 & \mathbf{1}_{88} \end{pmatrix},$$

$$I_C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix},$$

and

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \\ \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

The error vector e is assumed to have covariance matrix Σ given by $\Sigma = \text{cov(e)} = \text{I}_\text{I} \otimes \Sigma_\text{o} = \text{I}_\text{I} \otimes \Delta \otimes \Omega$, and Ω is assumed to be first-order autoregressive. The estimates obtained for the regression coefficients, obtained by maximum likelihood estimation, are given in Table 5.5. The covariance parameter estimates are given in Table 5.6.

To test for parallelism simultaneously for both variables, that is:

$$H_o : \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix},$$

we compute the test statistic:

$$T = (Q\hat{\lambda})' (Q\hat{V}Q')^{-1} (Q\hat{\lambda})$$

Table 5.5: Estimated regression coefficients from fitting a growth curve model to FEV1 and FVC.

Group	Variable	Parameter	Estimate	Standard Error
SW	fev1	β_{10}	5.1226	0.1298
SW	fev1	β_{11}	0.4058	0.08493
SW	fvc	α_{10}	5.6547	0.1206
SW	fvc	α_{11}	0.3744	0.07890
UW	fev1	β_{20}	5.4422	0.09287
UW	fev1	β_{21}	0.3150	0.06075
UW	fvc	α_{20}	5.7548	0.08627
UW	fvc	α_{21}	0.3162	0.05644

Table 5.6: Estimated covariance parameters from fitting a growth curve model to FEV1 and FVC.

Covariance Matrix	Parameter	Estimate
Δ	σ_{11}	2.8413
	σ_{12}	2.2662
	σ_{22}	2.4518
Ω Auto-Regressive(1)	ρ	0.1712

and compare it to a χ^2 distribution with 4 degrees of freedom where:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

We obtain $T = 8.995$ with a p-value of 0.061. From this, we conclude that at $\alpha = 0.1$, surface and underground workers differ significantly with respect to one or both of their lung function capacities. Suppose the data is now treated as a repeated measures analysis

of variance problem with the error vector e is assumed to have covariance matrix Σ given by $\Sigma = \text{cov(e)} = \mathrm{I_I} \otimes \Sigma_o = \mathrm{I_I} \otimes \Delta \otimes \Omega$, and Ω is assumed to be compound symmetry. In testing the hypothesis that the mean vectors for the two groups (worker=surface, worker=underground) are equal simultaneously for both variables, we obtain $T = 20.073$ with a p-value of 0.029. From this, we conclude that the two groups do differ significantly with respect to FEV1 or FVC or both.

The test for the Kronecker product pattern described in section 5.4 is also applied to this data. Using the first model $(un \otimes ar(1))$, we obtain $\chi^2 = 870.821$ and using the second model $(un \otimes cs)$, we obtain $\chi^2 = 830.661$. The null hypothesis that the within-subject variance-covariance matrix has a Kronecker product pattern is therefore rejected in both cases, indicating that for these data, the unstructured covariance matrix may be more suitable.

5.7 Evaluating efficiency using TARE and CARE

Consider testing the hypothesis discussed in section 5.5 using the test statistic T^* given by equation (5.16). The power function based on this test is given by equation (5.17). Let ϕ_1^* be the test based on the likelihood procedure which specifies the correct covariance structure and ϕ_2^* be the test based on the likelihood procedure which ignores the Kronecker product structure by specifying a completely unstructured covariance matrix. To evaluate the efficiency of ϕ_2^* relative to ϕ_1^*, the TARE and CARE of ϕ_2^* relative to ϕ_1^* are computed. The TARE and CARE in this specific case are discussed in the next two sections.

5.7.1 Efficiency of ϕ_2^* relative to ϕ_1^*

If the covariance parameters defining Δ and Ω are known, then for the test ϕ_1^*, the covariance matrix V is given by $(\Theta'\Theta)^{-1} \otimes \Delta \otimes (X'\Omega^{-1}X)^{-1}$. The non-centrality

parameters in the power function (5.17) for the tests ϕ_2^* and ϕ_1^* are therefore given by

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\Sigma_o^{-1}\,X^*)^{-1}\right)Q'\right\}^{-1}(Q\lambda) \tag{5.21}$$

and

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right\}^{-1}(Q\lambda), \tag{5.22}$$

respectively.

Lemma 5.7.1 *The TARE and CARE of ϕ_2^* with respect to ϕ_1^* are given by:*

$$\mathrm{TARE}(\phi_2^*/\phi_1^*)=\frac{\mathrm{tr}\left\{\mathrm{Q}\left((\Theta'\Theta)^{-1}\otimes(\mathrm{X}^{*\prime}\Sigma_o^{-1}\,\mathrm{X}^*)^{-1}\right)\mathrm{Q}'\right\}^{-1}}{\mathrm{tr}\left\{\mathrm{Q}\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}\right)\mathrm{Q}'\right\}^{-1}} \tag{5.23}$$

and

$$\begin{aligned}\mathrm{CARE}(\phi_2^*/\phi_1^*) &= \left\{\frac{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\Sigma_o^{-1}\,X^*)^{-1}\right)Q'\right\}^{-1}\right|}{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right\}^{-1}\right|}\right\}^{\frac{1}{r}}\\ &= \left\{\frac{\left|Q\left((\Theta'\Theta)^{-1}\otimes\Delta\otimes(X'\Omega^{-1}X)^{-1}\right)Q'\right|}{\left|Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\Sigma_o^{-1}\,X^*)^{-1}\right)Q'\right|}\right\}^{\frac{1}{r}},\end{aligned} \tag{5.24}$$

respectively.

The quantities (5.23) and (5.24) are estimated in the simulation study later in this chapter.

5.7.2 Estimated efficiency of ϕ_2^* relative to ϕ_1^*

For the test ϕ_2^*, the estimated covariance matrix $\hat{V}$ is given by equation (5.6) while for the test ϕ_1^*, it is given by equation (5.4). The non-centrality parameters in the power function (5.17) for the tests ϕ_2^* and ϕ_1^* are estimated by

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\hat{\Sigma}_o^{-1}\,X^*)^{-1}\right)Q'\right\}^{-1}(Q\lambda) \tag{5.25}$$

and

$$\frac{1}{2}(Q\lambda)'\left\{Q\left((\Theta'\Theta)^{-1}\otimes\hat{\Delta}\otimes(X'\hat{\Omega}^{-1}X)^{-1}\right)Q'\right\}^{-1}(Q\lambda), \tag{5.26}$$

respectively.

Lemma 5.7.2 *The estimated TARE and CARE of ϕ_2^* with respect to ϕ_1^* are given by*

$$\mathrm{TARE}_{\mathrm{est}}(\phi_2^*/\phi_1^*) = \frac{\mathrm{tr}\left\{\mathrm{Q}\left((\Theta'\Theta)^{-1}\otimes(\mathrm{X}^{*\prime}\hat{\Sigma}_o^{-1}\,\mathrm{X}^*)^{-1}\right)\mathrm{Q}'\right\}^{-1}}{\mathrm{tr}\left\{\mathrm{Q}\left((\Theta'\Theta)^{-1}\otimes\hat{\Delta}\otimes(\mathrm{X}'\hat{\Omega}^{-1}\mathrm{X})^{-1}\right)\mathrm{Q}'\right\}^{-1}} \tag{5.27}$$

and

$$\begin{aligned}\mathrm{CARE}_{\mathrm{est}}(\phi_2^*/\phi_1^*) &= \left\{\frac{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\hat{\Sigma}_o^{-1}\,X^*)^{-1}\right)Q'\right\}^{-1}\right|}{\left|\left\{Q\left((\Theta'\Theta)^{-1}\otimes\hat{\Delta}\otimes(X'\hat{\Omega}^{-1}X)^{-1}\right)Q'\right\}^{-1}\right|}\right\}^{\frac{1}{r}} \\ &= \left\{\frac{\left|Q\left((\Theta'\Theta)^{-1}\otimes\hat{\Delta}\otimes(X'\hat{\Omega}^{-1}X)^{-1}\right)Q'\right|}{\left|Q\left((\Theta'\Theta)^{-1}\otimes(X^{*\prime}\hat{\Sigma}_o^{-1}\,X^*)^{-1}\right)Q'\right|}\right\}^{\frac{1}{r}},\end{aligned} \tag{5.28}$$

respectively.

The quantities (5.27) and (5.28) are used to evaluate the efficiency of test ϕ_2^* relative to test ϕ_1^* when the covariance parameters defining Δ and Ω are unknown and estimated from the data by fitting a Kronecker product covariance structure.

5.8 A Monte-carlo study

5.8.1 Data generation

Multivariate normal data with $\mu = (\Theta\otimes I_C\otimes X)\lambda$ and $\Sigma = I_I\otimes\Sigma_o = I_I\otimes\Delta\otimes\Omega$ is generated for two characteristics $C = 2$ and three time points $T = 3$. Multivariate normal data were generated using the Cholesky root of the variance-covariance matrix.

To illustrate this, let X be $N_p(\mu, \Sigma)$ and Y be $N_p(0, I)$. Y is generated by p repeated calls to a univariate normal generator and X is obtained from the transformation $X = LY + \mu$ such that $LL' = \Sigma$.

The vector μ is specified from the design specifications and results of a previous study. The study was conducted by Dr. Tom Zullo in the school of Dental Medicine at the University of Pittsburgh and is discussed in Timm [65], Thomas [63] and Boik [6] among others. The study concerned the relative effectiveness of two orthopaedic adjustments of the mandible. Nine subjects were assigned to each of two orthopaedic treatments, called activator treatments. On each of three occasions, three dependent variables were observed which, in combination, reflected the position and size of the mandible. Mean plots of the data for each group and variable revealed that the growth curves of the three variables were approximately linear. Timm [65] fit a quadratic regression model to the data. In the study described here, μ is specified by ignoring the quadratic terms and using data only for the first two variables.

The matrices Δ and Ω are specified according to model 3 in Table 4.1. The parameter values used for ρ_c, ρ_t and γ in the covariance matrices Δ and Ω are given in Table 5.7. This gives a total of 180 parameter combinations. The range of parameters

Table 5.7: Values of parameters defining Δ and Ω used in the Monte-carlo simulation study.

Parameter	**Values**
ρ_c	-0.6 *to* 0.6 *by* 0.3
ρ_t	0.1 *to* 0.9 *by* 0.1
γ	1 *to* 2 *by* 0.5

and parameter combinations considered are varied enough to represent parameters that may arise in practice. Computational problems were encountered for values of ρ_t very close to 0 and 1 and hence the range of values considered for ρ_t is restricted to lie between 0.1 and 0.9, inclusive. For each set of parameters considered, 200 simulation trials were carried out.

5.8.2 Model fit and quantities of interest

The data generated by each trial is analyzed using a linear model for multivariate longitudinal data. More specifically, a linear growth curve model is fit to each of the two response variables in each of the two groups. Evaluation of the efficiency of test ϕ_2^* relative to test ϕ_1^* depends on whether the covariance parameters defining Δ and Ω are known or not. If they are known, then $\hat{\lambda}$ is given by (5.2) and var$(\hat{\lambda})$ is given by $(\Theta'\Theta)^{-1} \otimes \Delta \otimes (X'\Omega^{-1}X)^{-1}$. To evaluate the efficiency of test ϕ_2^* relative to test ϕ_1^* following section 5.7.1, an unstructured covariance matrix is fit to the simulated data and $\hat{\lambda}_{un}$ given by equation (5.5) and its covariance matrix $\hat{\text{var}}(\hat{\lambda}_{un})$ given by equation (5.6) are computed. The TARE and CARE given by equations 5.23 and 5.24 are then evaluated. In simplified form, we evaluate

$$\text{TARE}(\phi_2^*/\phi_1^*) = \frac{\text{tr}\left\{\text{Qvar}(\hat{\lambda}_{\text{un}})\text{Q}'\right\}^{-1}}{\text{tr}\left\{\text{Qvar}(\hat{\lambda}_{\Delta\times\Omega})\text{Q}'\right\}^{-1}} \tag{5.29}$$

and

$$\text{CARE}(\phi_2^*/\phi_1^*) = \frac{\left|\text{Qvar}(\hat{\lambda}_{\Delta\times\Omega})\text{Q}'\right|}{\left|\text{Qvar}(\hat{\lambda}_{\text{un}})\text{Q}'\right|}. \tag{5.30}$$

In practical situations, the covariance parameters are unknown and must be estimated from the data. The efficiency of test ϕ_2^* relative to test ϕ_1^* is evaluated by fitting two different covariance models to the data:

1. unstructured covariance matrix;

2. Kronecker product covariance matrix (the true model).

Using the Kronecker product covariance matrix means computing $\hat{\lambda}_{\Delta\otimes\Omega}$ given by equation (5.3) and its covariance matrix $\hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega})$ given by equation (5.4). Ignoring the Kronecker structure and using an unstructured covariance matrix means computing $\hat{\lambda}_{un}$ given by equation (5.5) and its covariance matrix $\hat{\text{var}}(\hat{\lambda}_{un})$ given by equation (5.6).

The quantities discussed in section 5.7.2 for evaluating the efficiency of ϕ_2^* (test based on a likelihood procedure that ignores the Kronecker product covariance structure) relative to ϕ_1^* (test based on a likelihood procedure that specifies the Kronecker product covariance structure) are computed from the generated data. Specifically, the quantities are given by equation (5.27) for the TARE and equation (5.28) for the CARE. In simplified form, we evaluate

$$\text{TARE}_{\text{est}}(\phi_2^*/\phi_1^*) = \frac{\text{tr}\left\{\text{Q}\hat{\text{var}}(\hat{\lambda}_{\text{un}})\text{Q}'\right\}^{-1}}{\text{tr}\left\{\text{Q}\hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega})\text{Q}'\right\}^{-1}} \tag{5.31}$$

and

$$\text{CARE}_{\text{est}}(\phi_2^*/\phi_1^*) = \frac{\left|\text{Q}\hat{\text{var}}(\hat{\lambda}_{\Delta\times\Omega})\text{Q}'\right|}{\left|\text{Q}\hat{\text{var}}(\hat{\lambda}_{\text{un}})\text{Q}'\right|} \tag{5.32}$$

where Q reflects a hypothesis of interest. The results are summarised in the tables and graphs that follow.

5.9 Results and discussion

Tables 5.8 to 5.16 give the efficiency, as measured using the TARE and CARE, of test ϕ_2^* relative to test ϕ_1^* for the covariance parameter values given in Table 5.7 obtained from the simulation study. For clarity, the results are presented separately for each value of ρ_t and are cross-classified by the values of ρ_c and γ.

The parameters ρ_c and γ do not appear to have a significant impact on the TARE and CARE as can be clearly seen from Table 5.8 to Table 5.16. We observe that the values of the TARE and CARE do not change very much as ρ_c progresses from -0.6 to 0.6 and as γ progresses from 0.5 to 2. As a result, the values within a given table are quite close. It appears that the parameter which impacts efficiency most profoundly is ρ_t. The values of the TARE and CARE change as we move from table to table. The CARE gives higher values for all parameter combinations considered. The TARE and CARE are closer for

Table 5.8: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.1$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care
-0.6	0.955	0.924	0.930	0.965	0.925	0.931	0.97	0.926	0.932	0.97	0.926	0.932	3.86	0.925	0.931
-0.3	1	0.947	0.974	1	0.943	0.967	1	0.944	0.967	1	0.944	0.967	4	0.945	0.969
0	1	0.950	0.974	1	0.946	0.967	1	0.947	0.967	1	0.947	0.967	4	0.947	0.969
0.3	1	0.956	0.985	1	0.949	0.973	1	0.950	0.973	1	0.951	0.973	4	0.951	0.976
0.6	1	0.961	1.002	1	0.958	0.991	1	0.959	0.991	1	0.959	0.991	4	0.959	0.994
Total	4.955	0.948	0.973	4.965	0.944	0.966	4.97	0.945	0.966	4.97	0.946	0.966	19.86	0.946	0.968

Table 5.9: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.2$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.819	0.850	1	0.819	0.850	1	0.819	0.850	1	0.820	0.850	4	0.819	0.85
-0.3	1	0.820	0.850	1	0.821	0.850	1	0.821	0.850	1	0.822	0.850	4	0.821	0.85
0	1	0.821	0.850	1	0.822	0.850	1	0.823	0.850	1	0.823	0.850	4	0.822	0.85
0.3	1	0.821	0.850	1	0.822	0.850	1	0.823	0.850	1	0.824	0.850	4	0.823	0.85
0.6	1	0.820	0.850	1	0.822	0.850	1	0.823	0.850	1	0.823	0.850	4	0.822	0.85
Total	5	0.820	0.850	5	0.821	0.850	5	0.822	0.850	5	0.822	0.850	20	0.821	0.85

Table 5.10: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.3$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.742	0.796	1	0.742	0.796	1	0.742	0.796	1	0.742	0.796	4	0.742	0.79
-0.3	1	0.742	0.796	1	0.743	0.796	1	0.744	0.796	1	0.744	0.796	4	0.743	0.79
0	1	0.743	0.796	1	0.744	0.796	1	0.744	0.796	1	0.745	0.796	4	0.744	0.79
0.3	1	0.743	0.796	1	0.744	0.796	1	0.745	0.796	1	0.745	0.796	4	0.744	0.79
0.6	1	0.743	0.796	1	0.744	0.796	1	0.744	0.796	1	0.745	0.796	4	0.744	0.79
Total	5	0.743	0.796	5	0.743	0.796	5	0.744	0.796	5	0.744	0.796	20	0.743	0.79

Table 5.11: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.4$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.691	0.769	1	0.691	0.769	1	0.692	0.769	1	0.692	0.769	4	0.691	0.76
-0.3	1	0.692	0.769	1	0.692	0.769	1	0.692	0.769	1	0.693	0.769	4	0.692	0.76
0	1	0.692	0.769	1	0.692	0.769	1	0.693	0.769	1	0.693	0.769	4	0.693	0.76
0.3	1	0.692	0.769	1	0.693	0.769	1	0.693	0.769	1	0.693	0.769	4	0.693	0.76
0.6	1	0.692	0.769	1	0.692	0.769	1	0.693	0.769	1	0.693	0.769	4	0.692	0.76
Total	5	0.692	0.769	5	0.692	0.769	5	0.692	0.769	5	0.693	0.769	20	0.692	0.76

Table 5.12: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.5$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	4	0.667	0.77
-0.3	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	1	0.668	0.772	4	0.667	0.77
0	1	0.667	0.772	1	0.667	0.772	1	0.668	0.772	1	0.668	0.772	4	0.667	0.77
0.3	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	4	0.667	0.77
0.6	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	1	0.667	0.772	4	0.667	0.77
Total	5	0.667	0.772	5	0.667	0.772	5	0.667	0.772	5	0.667	0.772	20	0.667	0.77

Table 5.13: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.6$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.674	0.813	1	0.674	0.813	1	0.674	0.813	1	0.674	0.813	4	0.674	0.81
-0.3	1	0.674	0.813	1	0.674	0.813	1	0.674	0.813	1	0.674	0.813	4	0.674	0.81
0	1	0.673	0.813	1	0.674	0.813	1	0.674	0.813	1	0.674	0.813	4	0.674	0.81
0.3	1	0.673	0.813	1	0.673	0.813	1	0.673	0.813	1	0.673	0.813	4	0.673	0.81
0.6	1	0.673	0.813	1	0.673	0.813	1	0.673	0.813	1	0.673	0.813	4	0.673	0.81
Total	5	0.673	0.813	5	0.673	0.813	5	0.674	0.813	5	0.674	0.813	20	0.673	0.81

Table 5.14: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.7$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.730	0.921	1	0.730	0.921	1	0.730	0.921	1	0.731	0.921	4	0.730	0.92
-0.3	1	0.729	0.921	1	0.730	0.921	1	0.730	0.921	1	0.730	0.921	4	0.730	0.92
0	1	0.729	0.921	1	0.729	0.921	1	0.729	0.921	1	0.729	0.921	4	0.729	0.92
0.3	1	0.728	0.921	1	0.728	0.921	1	0.728	0.921	1	0.728	0.921	4	0.728	0.92
0.6	1	0.728	0.921	1	0.727	0.921	1	0.727	0.921	1	0.727	0.921	4	0.727	0.92
Total	5	0.729	0.921	5	0.729	0.921	5	0.729	0.921	5	0.729	0.921	20	0.729	0.92

Table 5.15: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.8$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	0.893	1.180	1	0.894	1.180	1	0.894	1.180	1	0.895	1.180	4	0.894	1.18
-0.3	1	0.892	1.180	1	0.893	1.180	1	0.893	1.180	1	0.893	1.180	4	0.893	1.18
0	1	0.891	1.180	1	0.891	1.180	1	0.891	1.180	1	0.891	1.180	4	0.891	1.18
0.3	1	0.890	1.180	1	0.889	1.180	1	0.889	1.180	1	0.889	1.180	4	0.889	1.18
0.6	1	0.889	1.180	1	0.888	1.180	1	0.888	1.180	1	0.887	1.180	4	0.888	1.18
Total	5	0.891	1.180	5	0.891	1.180	5	0.891	1.180	5	0.891	1.180	20	0.891	1.18

Table 5.16: TARE and CARE for the test based on a unstructured model relative to one based on a Kronecker product model cross-classified by ρ_c and γ: $\rho_t = 0.9$.

γ	0.5			1			1.5			2			Total		
ρ_c	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Care	$\frac{N}{200}$	Tare	Car
-0.6	1	1.429	1.982	1	1.464	2.029	1	1.465	2.029	1	1.466	2.029	4	1.456	2.01
-0.3	1	1.420	1.969	1	1.463	2.029	1	1.464	2.029	1	1.464	2.029	4	1.452	2.01
0	1	1.420	1.974	1	1.459	2.029	1	1.460	2.029	1	1.460	2.029	4	1.450	2.01
0.3	1	1.410	1.961	1	1.456	2.029	1	1.456	2.029	1	1.455	2.029	4	1.444	2.01
0.6	1	1.423	1.982	1	1.453	2.029	1	1.452	2.029	1	1.451	2.029	4	1.445	2.01
Total	5	1.420	1.974	5	1.459	2.029	5	1.459	2.029	5	1.459	2.029	20	1.449	2.01

small values of ρ_t with an increasing difference as ρ_t approaches 0.9. At $\rho_t = 0.8$, the CARE exceeds 1 and at $\rho_t = 0.9$, both measures exceed 1. This is somewhat surprising and is worth further investigation. To understand how the parameter ρ_t affects efficiency, a graphical representation of the results is shown in Figures 5.1 to 5.4. A quadratic curve has been fit to the data. A loss of efficiency occurs for values of ρ_t from 0.1 to about 0.8. The degree of the loss of efficiency depends on the value of ρ_t. Efficiency drops as we move from $\rho_t = 0.1$ to $\rho_t = 0.5$ and then begins to rise again. There appears to be no loss of efficiency for high values of ρ_t (above 0.8). Efficiency is worst for mid values of ρ_t.

For the design and within-subject variance-covariance matrices considered, the results presented in the Tables 5.8 to 5.16 and Figures 5.1 and 5.4 demonstrate the usefulness of utilising the Kronecker product covariance structure for multivariate longitudinal data. If one ignores the underlying Kronecker product covariance structure, a potential loss of efficiency will occur in testing hypotheses that are of interest. The parameters γ and ρ_c defining the covariance matrix Δ do not appear to impact the efficiency very significantly. However, the parameter ρ_t defining the covariance matrix Ω appears to have a large impact on efficiency.

5.10 Evaluating the performance of the test for the Kronecker product pattern in the covariance matrix

This section investigates the performance of the test described in section 5.4 for testing the null hypothesis that the $CT \times CT$ covariance matrix Σ_o has the Kronecker product structure $\Delta \times \Omega$. The null and alternative hypotheses are given respectively by :

$$H_o : \Sigma_o = \Delta \otimes \Omega$$

versus

$$H_a : \Sigma_o = \Sigma_a \quad (\textit{Arbitrary}).$$

As mentioned in the introduction to this chapter, there are two reasons why this is a very useful test in practical work involving multivariate longitudinal data. First, the test will provide one with protection against doing the wrong thing in terms of basing inference on an incorrect covariance matrix. Secondly, it will give one increased efficiency if one does the right thing. The test is based on computing the likelihood ratio test statistic given by:

$$\mathrm{LR} = \left\{ \frac{|\hat{\Sigma}_o|}{|\hat{\Delta}|^T |\hat{\Omega}|^C} \right\}^{\frac{-f}{2}} \tag{5.33}$$

where Σ_o has an arbitrary pattern and $\hat{\Sigma}_o$ is its maximum likelihood or restricted maximum likelihood estimate. $\hat{\Delta}$ and $\hat{\Omega}$ are the maximum likelihood or restricted maximum likelihood estimates of Δ and Ω assuming Σ_o has a Kronecker structured covariance matrix. Under the null hypothesis, 2In(LR) has an asymptotic chi-square distribution with

$$\frac{TC(TC+1)}{2} - \frac{C(C+1)+2}{2} \tag{5.34}$$

degrees of freedom when Ω is assumed to be compound symmetry or first-order autoregressive. The idea of incorporating the test is that the choice of parameter estimates contained in the vector $\hat{\lambda}$ to be used in testing hypotheses of interest will depend on the results of this test. If the p-value obtained is less than or equal to α, where α is the fixed significance level of the test, then $\hat{\lambda}_{un}$ will be used in subsequent analysis since the null hypothesis for the Kronecker product pattern will have been rejected. Otherwise, $\hat{\lambda}_{\Delta\times\Omega}$ will be used.

The data in the Monte-carlo study discussed in section 5.8 was generated using the covariance matrix $\Sigma_o = \Delta \times \Omega$ and varying the parameters in Δ and Ω. Hence, we can evaluate the performance of the test for the Kronecker product pattern by finding the empirical Type I error rates for this test for a given α. This is accomplished for each parameter combination considered by counting the number of times (out of the total number of simulation trials per parameter combination) that the null hypothesis $H_o : \Sigma_o = \Delta \times \Omega$ is rejected. This translates to the number of times $\hat{\lambda}_{un}$ is chosen over

$\hat{\lambda}_{\Delta\times\Omega}$ is subsequent analysis. We then observe what happens to the Type I error rates as α goes from 0.05 to 0.95 in steps of 0.05 for the different parameter combinations considered. The results are presented in Table 5.17, from which we observe that the performance of the test described in section 5.4 is very good, with empirical Type I error rates being very close to α.

Table 5.17: Type I error rates (per 100 tests) for the test for the Kronecker product covariance structure cross-classified by α and ρ_t.

	ρ_t								
α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	5.86	5.50	5.00	5.50	6.00	5.70	6.05	6.00	7.00
0.1	11.88	11.85	11.38	11.50	12.00	12.50	12.00	12.00	12.00
0.15	18.60	18.00	18.15	17.50	18.50	18.50	18.00	18.50	17.05
0.2	23.08	22.50	23.00	23.00	23.00	23.50	23.68	23.50	23.95
0.25	28.90	29.00	28.00	28.50	28.50	29.05	30.00	30.00	30.90
0.3	34.94	35.42	34.50	33.50	32.50	33.00	33.50	34.00	33.50
0.35	37.46	38.40	38.42	38.00	38.00	37.50	37.00	37.00	36.50
0.4	41.97	41.60	41.82	42.50	41.50	41.00	40.95	39.50	40.10
0.45	47.05	46.50	46.00	46.50	46.18	45.50	44.72	45.50	46.48
0.5	53.50	52.45	51.78	51.00	50.50	52.00	51.50	51.50	53.02
0.55	59.06	58.00	57.00	56.50	57.00	57.00	56.62	57.50	58.00
0.6	62.43	62.40	61.50	61.50	60.50	61.00	61.00	62.42	63.10
0.65	68.67	66.40	66.30	66.50	66.50	64.97	66.00	66.00	67.65
0.7	72.40	72.00	72.00	72.00	72.50	73.50	71.60	71.00	71.50
0.75	76.91	75.50	76.50	76.50	77.50	77.50	77.50	77.50	76.55
0.8	78.95	79.00	79.50	80.00	80.00	80.50	81.00	80.50	80.05
0.85	84.89	85.00	85.50	85.08	85.50	86.00	86.00	84.45	83.50
0.9	91.44	91.50	90.50	90.50	90.50	90.50	90.00	89.50	89.50
0.95	95.44	95.50	95.50	96.00	96.00	94.50	94.50	95.50	96.00

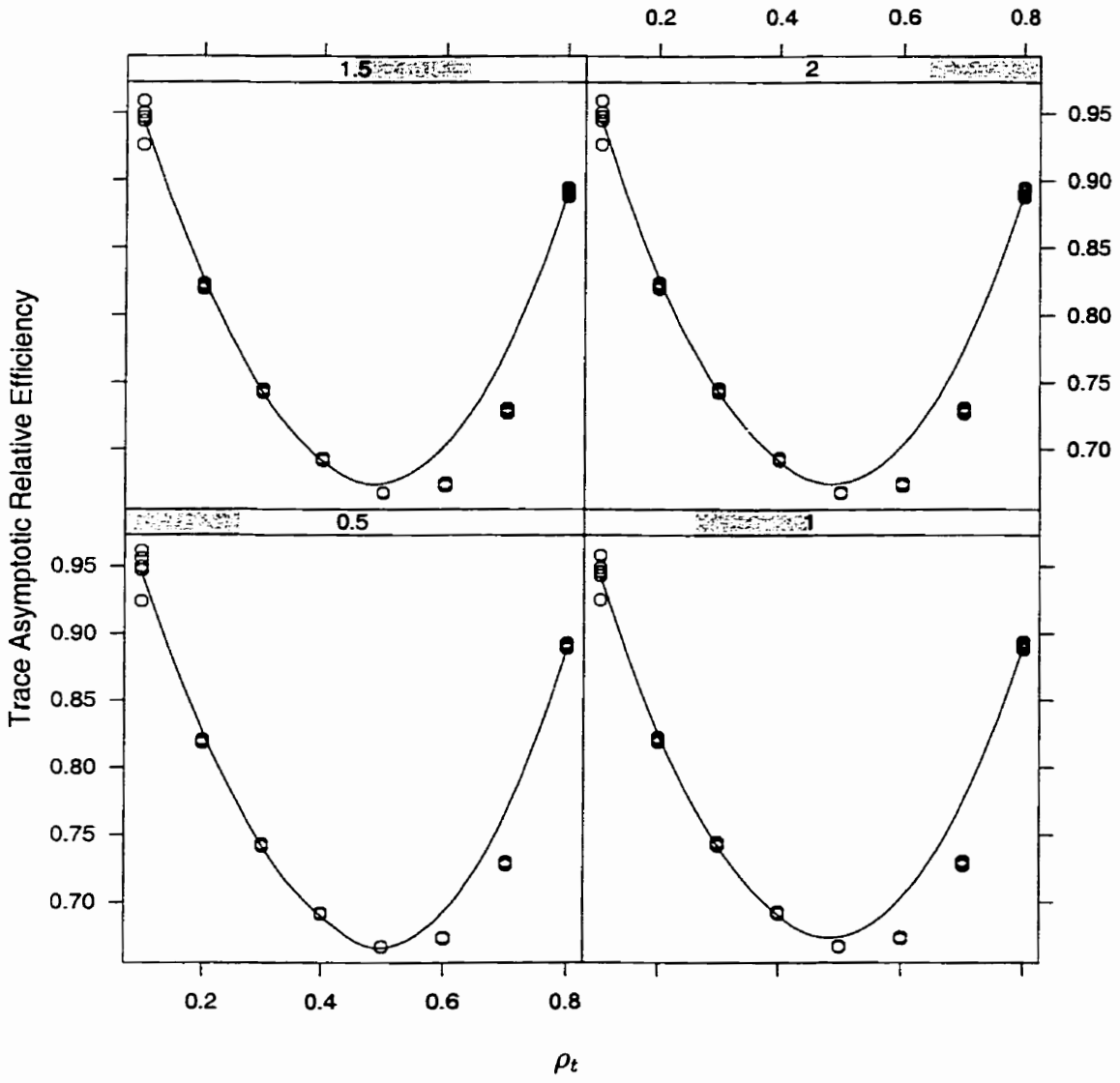

Figure 5.1: TARE (averaged over 200 simulations) of the test based on a unstructured within-subject variance-covariance matrix relative to one based on the Kronecker product model for varying values of γ.

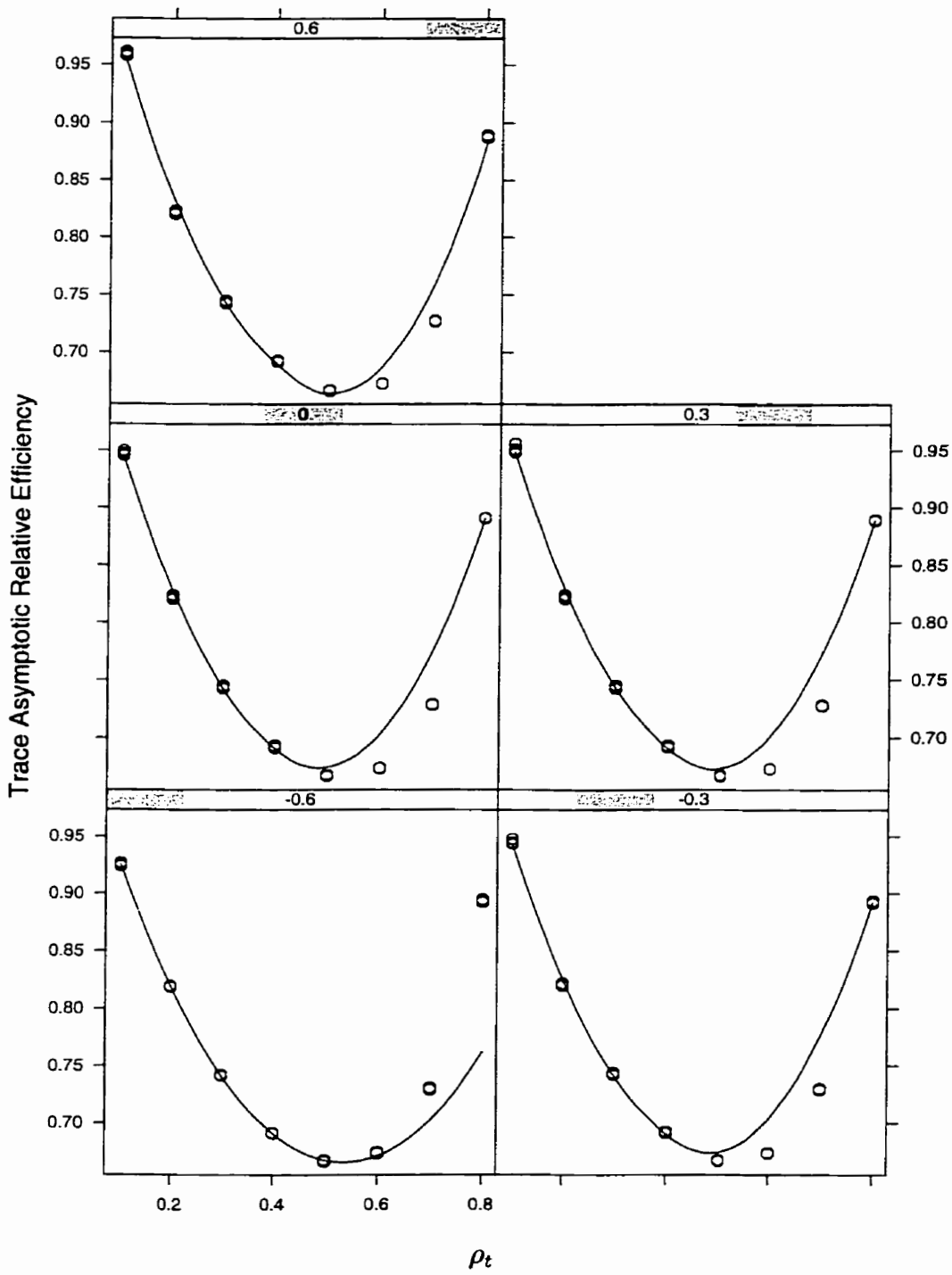

Figure 5.2: TARE (averaged over 200 simulations) of the test based on a unstructured within-subject variance-covariance matrix relative to one based on the Kronecker product model for varying values of ρ_c.

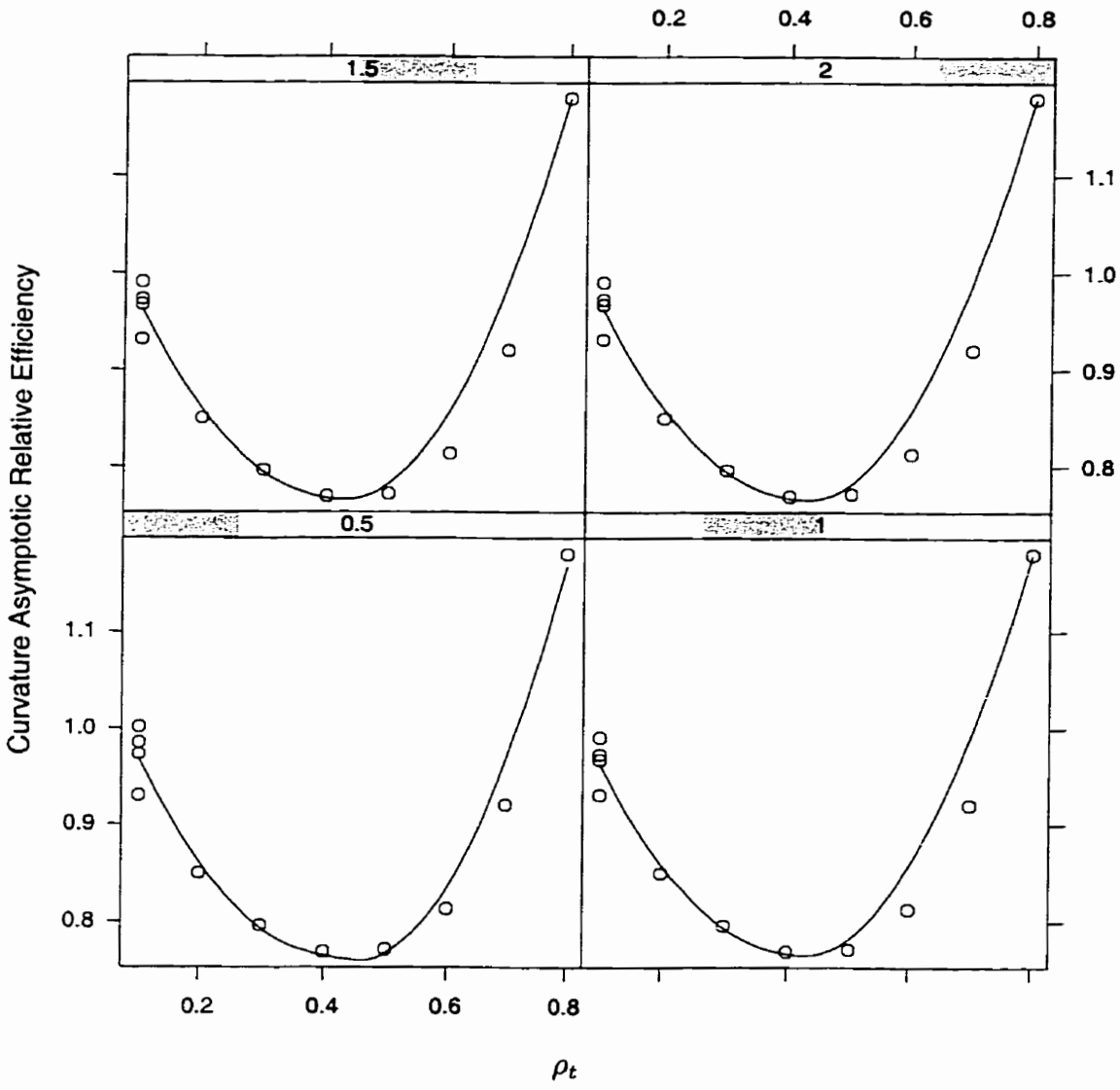

Figure 5.3: CARE (averaged over 200 simulations) of the test based on a unstructured within-subject variance-covariance matrix relative to one based on the Kronecker product model for varying values of γ.

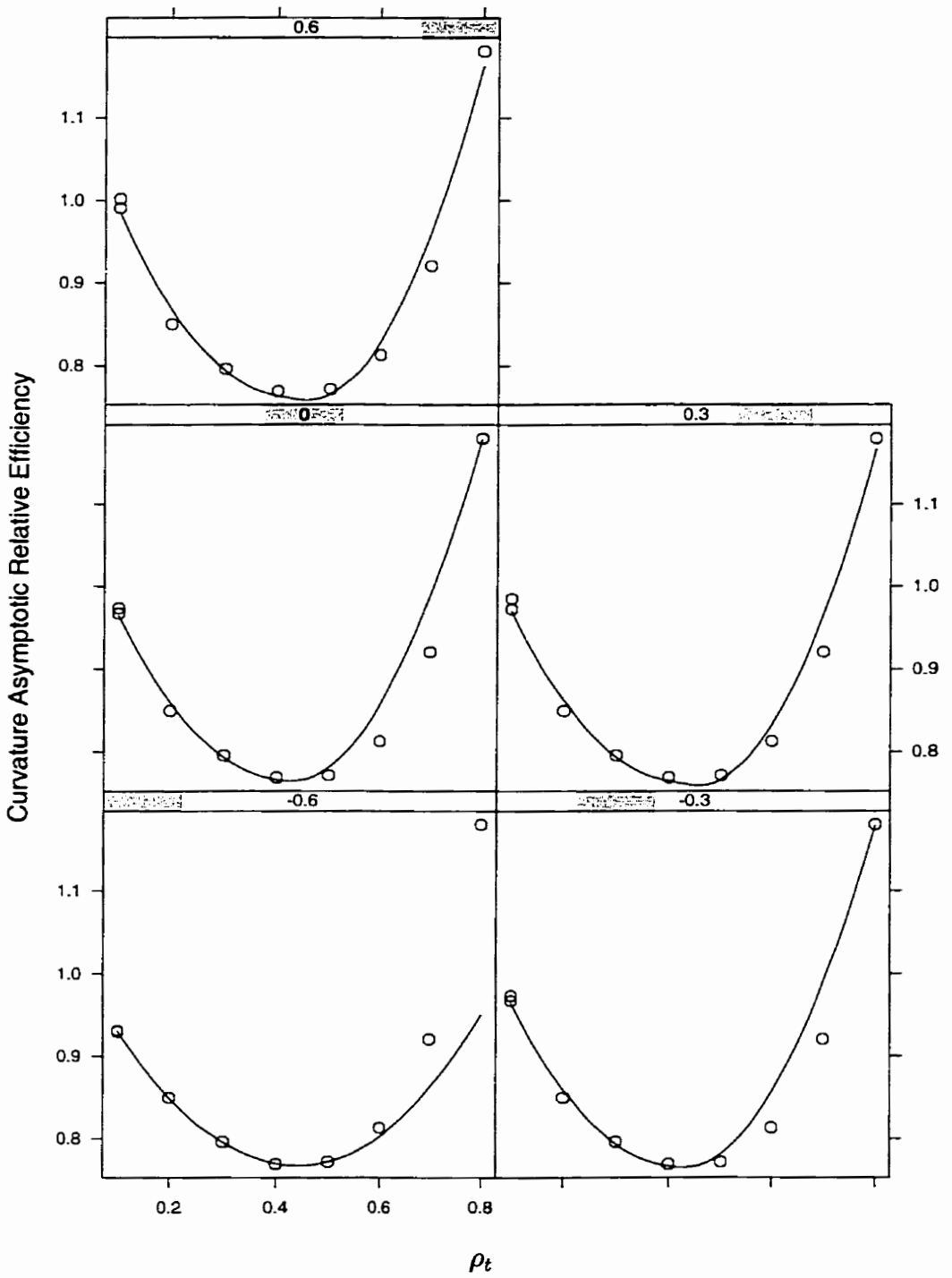

Figure 5.4: CARE (averaged over 200 simulations) of the test based on a unstructured within-subject variance-covariance matrix relative to one based on the Kronecker product model for varying values of ρ_c.

Chapter 6

The effect of covariance structure on hypothesis testing in multivariate longitudinal data

6.1 Introduction

In Chapter 5, we investigated the gain from utilising a Kronecker structured covariance matrix for multivariate longitudinal data. Using the TARE and CARE, we evaluated the efficiency of a test based on a completely unstructured covariance matrix relative to one based on a Kronecker structured covariance matrix. For the designs and covariance structures considered, the results demonstrated that if one ignores the underlying Kronecker product covariance structure, a potential loss of efficiency could occur in testing hypotheses of interest. The degree of the loss of efficiency was determined to a large extent by the parameters defining the matrices Δ and Ω. The parameter ρ_t in Ω had the greatest effect on efficiency.

In this chapter, we investigate the converse of the situation considered in Chapter 5. Specifically, the loss from imposing a Kronecker structured covariance matrix in testing hypotheses of interest in multivariate longitudinal data is investigated. To achieve this,

the concept of non-Kronecker product covariance matrices is introduced and a class of matrices that is non-Kronecker product defined. The class of matrices are specified in a way that makes them easy to interpret. An index, referred to as the Kronecker product deviation index, is introduced. It is used to quantify how far a given covariance matrix departs from Kronecker product structure. To assess the consequences of imposing a Kronecker product covariance matrix, hypotheses of interest are tested using two models. The first model is based on a Kronecker product covariance matrix and the second model is based on a non-Kronecker product covariance matrix. The impact of the Kronecker product deviation index on the results of hypothesis testing using the two models are carefully studied.

Results obtained from the investigation will enable us to make general statements about the consequences of imposing Kronecker product structure when it is not appropriate in testing hypotheses of interest for multivariate longitudinal data. Additionally, we will be able to state what parameter ranges signify more serious consequences, if any, in imposing the Kronecker product structure. Section 6.2 gives an alternative and more general formulation for the within-subject variance-covariance matrix Σ_o. Section 6.3 discusses how to measure departure from the Kronecker product structure. Section 6.4 presents a Monte-carlo simulation study designed to investigate the impact of the Kronecker product deviation index on testing hypotheses of interest. The chapter closes with results and a general discussion in section 6.5.

6.2 An alternative formulation for Σ_o

A detailed discussion on modelling the covariance matrix for multivariate longitudinal data in the form of a Kronecker product is discussed in detail in section 4.2.2. A justification for using the model is also given. Two of the structures that are most commonly used are reproduced in Table 6.1. In these structures, Δ represents the covariance matrix for the characteristics at any time point and Ω represents the correlation matrix for any of the characteristics over time. To model the within-subject covariance matrix

Table 6.1: Examples of within-subject covariance models ($C = 2$ and $T = 3$).

Structure	**Form** ($\Delta \otimes \Omega$)	**# Parameters**
UN⊗AR(1)	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$	4
UN⊗CS	$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$	4

Σ_o as the Kronecker product of two matrices Δ and Ω , the assumption is made that Δ does not change with time and Ω is the same for all characteristics. This assumption will work in many situations but will be unrealistic in some cases. For example, if one is modelling both diastolic and systolic blood pressure over time in a group of patients diagnosed with high blood pressure, then it is reasonable to assume that Ω is the same for both of these characteristics. However, if one is modelling two distinctly different characteristics that have been measured over time, then it may be unrealistic to assume that their correlation structures over time are the same. An alternative formulation for Σ_o must therefore be considered.

There are several options one may consider, one of which is to drop the assumption that the covariance matrix that models the T repeated measurements on a given characteristic is the same for all characteristics. For C characteristics each measured on T occasions, consider the following formulation for Σ_o:

$$\Sigma_o = \begin{pmatrix} \sigma_{11}\Omega_1 & \sigma_{12}\Omega_{12} & \cdots & \sigma_{1C}\Omega_{1C} \\ \sigma_{21}\Omega_{21} & \sigma_{22}\Omega_2 & \cdots & \sigma_{2C}\Omega_{2C} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{C1}\Omega_{C1} & \sigma_{C2}\Omega_{C2} & \cdots & \sigma_{CC}\Omega_C \end{pmatrix},$$

where Ω_c, for $c = 1, 2, \ldots, C$, represents the $T \times T$ correlation matrices for characteristics $1, 2, \ldots, C$ measured on T occasions and Ω_{ck} for $c = 1, 2, \ldots, C$ and $k = 1, 2, \ldots, C$ represents the correlation matrices between the pairs of characteristics over time. The matrices

Ω_c and Ω_{ck} for $c = 1, 2, \ldots, C$ and $k = 1, 2, \ldots, C$ may be left completely unstructured or may be structured in some way. Possible structures include compound symmetry and first-order autoregressive. Without loss of generality, consider two characteristics each measured on three occasions and assume a first-order autoregressive structure over time. The correlation matrix for the first characteristic is given by

$$\Omega_1 = \begin{pmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{pmatrix},$$

and for the second characteristic by

$$\Omega_2 = \begin{pmatrix} 1 & \rho_2 & \rho_2^2 \\ \rho_2 & 1 & \rho_2 \\ \rho_2^2 & \rho_2 & 1 \end{pmatrix}.$$

Also,

$$\Omega_{12} = \begin{pmatrix} 1 & \rho_{12} & \rho_{12}^2 \\ \rho_{12} & 1 & \rho_{12} \\ \rho_{12}^2 & \rho_{12} & 1 \end{pmatrix},$$

representing the correlation between the two characteristics over time. The overall within-subject covariance matrix Σ_o is therefore given by:

$$\Sigma_o = \begin{pmatrix} \sigma_{11} \begin{pmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{pmatrix} & \sigma_{12} \begin{pmatrix} 1 & \rho_{12} & \rho_{12}^2 \\ \rho_{12} & 1 & \rho_{12} \\ \rho_{12}^2 & \rho_{12} & 1 \end{pmatrix} \\ \sigma_{12} \begin{pmatrix} 1 & \rho_{12} & \rho_{12}^2 \\ \rho_{12} & 1 & \rho_{12} \\ \rho_{12}^2 & \rho_{12} & 1 \end{pmatrix} & \sigma_{22} \begin{pmatrix} 1 & \rho_2 & \rho_2^2 \\ \rho_2 & 1 & \rho_2 \\ \rho_2^2 & \rho_2 & 1 \end{pmatrix} \end{pmatrix}$$

which cannot be expressed as the exact Kronecker product of two matrices except in the special case when the parameters ρ_1, ρ_2 and ρ_{12} are all equal to each other. The matrix Σ_o is defined by 6 parameters, namely: σ_{11}, σ_{12} and σ_{22}, ρ_1 (defining Ω_1), ρ_2 defining (Ω_2) and ρ_{12} defining (Ω_{12}).

6.3 Measuring departure from Kronecker product

An important goal of the present chapter is to define an index that gives an indication of how well a given $TC \times TC$ variance-covariance matrix for C characteristics measured on T occasions can be expressed as the Kronecker product of a $C \times C$ matrix and a $T \times T$ matrix. The index would enable one to decide when to base inference for multivariate longitudinal data on a model with a Kronecker structured covariance matrix. This section describes the index that will be used which is based on Verhees and Wansbeek [71].

Verhees and Wansbeek [71] describe a multimode direct product model for covariance structure analysis. They justify the model by stating that in the psychometric literature, there is evidence that the modes in multimode data interact multiplicatively. They also state that a basic expression of this idea is that a covariance matrix may then be written as the repeated Kronecker product of k, say, parameter matrices, where k is the number of modes. This is, in fact, the covariance matrix that has been central to the work done in this dissertation, specifically as it applies to multivariate longitudinal data with $k = 2$ to reflect the two dimensions (characteristics and time). Verhees and Wansbeek [71] call this model the "factorial covariance structure". For this model, they give an integrated treatment of maximum likelihood, weighted least squares and unweighted least squares estimators. In this section, we focus on the unweighted least squares estimator. We pay particular attention to the modified unweighted least squares estimator which is non-iteratively computable. To avoid confusion and to be consistent with the notation in Verhees and Wansbeek [71], the following equivalences should be kept in mind. The notation on the right is the notation used in this dissertation and the notation on the left is the notation used in Verhees and Wansbeek [71].

1. $\Omega = \Sigma_o$
2. $\Sigma_1 = \Delta$
3. $\Sigma_2 = \Omega$

4. $n_1 = C \quad n_2 = T$.

Consider the data vector for a single subject i, $i = 1, 2, \ldots, I$, in multivariate longitudinal data with $C = n_1$ characteristics measured on $T = n_2$ occasions. Let y_{i2} represent the $T \times 1$ vector representing the data for a given characteristic measured on T occasions and y_{i1} represent the $C \times 1$ vector representing the data for C characteristics measured at each time point. The first index (characteristic) is the slower running index and the second index (time) is the faster running index. The total number of observations on a given subject is $n = n_1 \times n_2 = C \times T$ which are given by the vector y_i of length $n_1 \times n_2 = C \times T$. The variability in the observations is summarised in their $n \times n = CT \times CT$ sample covariance matrix S. The covariance matrix from which y_i is drawn is given by Σ_o, also $n \times n = CT \times CT$. Σ_o is said to have a factorial covariance structure when it has the form:

$$\Sigma_o = \Delta \otimes \Omega \tag{6.1}$$

where Δ and Ω are symmetric positive-definite matrices of order $C \times C$ and $T \times T$ respectively. We consider estimation of the matrices Δ and Ω in equation (6.1). Three criteria are available for estimating the parameters in Δ and Ω. The three criteria are maximum likelihood, weighted least squares and unweighted least squares. The three criteria are given by equations 6.2, 6.3 and 6.4 respectively.

$$\min_{\theta}(\mathrm{In}|\Sigma_o| + \mathrm{tr}S\Sigma_o^{-1}) \tag{6.2}$$

$$\min_{\theta} \mathrm{tr}((S - \Sigma_o)S^{-1})^2 \tag{6.3}$$

$$\min_{\theta} \mathrm{tr}(S - \Sigma_o)^2 \tag{6.4}$$

The parameter vector θ contains the parameters in Δ and Ω. The three criteria can be summarised as:

$$\min \ \mathrm{tr}((S - \Sigma_o)W^{-1})^2, \tag{6.5}$$

with $W = \hat{\Sigma}_o$ (ML), $W = S$ (WLS) or $W = I_n$ (ULS).

6.3.1 Notation

The following notation will be useful:

1. C_i is a $n \times n$ commutation matrix that changes the running order of the observations in the vector y in such a way that $C_i y$ has the i^{th} index fastest.

2. $n^i = \prod_{j \neq i}^{k} n_j = \frac{n}{n_i}$

3. W_i, S_i and Σ_{oi} are the permuted versions of W, S and Σ_o respectively. For example, considering the general situation with k dimensions, $\Omega_i = \Sigma^i \otimes \Sigma_i$ where the $n^i \times n^i$ matrix Σ^i is given by:

$$\Sigma^i = \Sigma_1 \otimes \cdots \otimes \Sigma_{i-1} \otimes \Sigma_{i+1} \otimes \cdots \otimes \Sigma_k. \tag{6.6}$$

   The vectorized matrices are given by $\sigma_i = \text{vec } \Sigma_i$ and $\sigma^i = \text{vec } \Sigma^i$.

4. $\tilde{s}_i = \text{vec } \tilde{S}_i$ where $\tilde{S}_i$ comes from stacking each of the $(n^i)^2$ blocks of order $n_i \times n_i$ of S_i in a vector according to the vec operator, and placing these vectors together next to each other as columns of the matrix $\tilde{S}_i$.

6.3.2 Estimation

Following Verhees and Wansbeek [71], the estimator $\hat{\sigma}_i$ of σ_i is given by

$$\hat{\sigma}_i = [\hat{X}_i'[\hat{W}_i \pi \hat{W}_i]^{-1} \hat{X}_i]^{-1} \hat{X}_i'[\hat{W}_i \pi \hat{W}_i]^{-1} \tilde{s}_i, \tag{6.7}$$

where $W_i = \hat{\Omega}_i$ (ML), $W_i = S_i$ (WLS) or $W_i = I_n$ (ULS), and hats on X_i and W_i indicate their possible dependence on unknown parameters that also have to be estimated. Also, $\hat{W}_i \pi \hat{W}_i = B_i(\hat{W}_i \otimes \hat{W}_i)B_i'$, where B_i is a permutation matrix. Elaboration of equation (6.7) gives three distinct estimators for the three estimation criteria. All three estimators require an iterative procedure. Fortunately, a modification of the unweighted least squares estimator is possible that allows for non-iterative estimator.

The unweighted least squares estimator for σ_i is

$$\hat{\sigma}_i = (\hat{\sigma}^{i\prime}\hat{\sigma}^i)^{-1}\tilde{S}_i\hat{\sigma}^i, \tag{6.8}$$

and the unweighted least squares estimator for σ^i is

$$\hat{\sigma}^i = (\hat{\sigma}_i'\hat{\sigma}_i)^{-1}\tilde{S}_i'\hat{\sigma}_i. \tag{6.9}$$

Substituting equation (6.9) into equation (6.8) yields

$$(\tilde{S}_i\tilde{S}_i' - \hat{\lambda}I)\hat{\sigma}_i = 0, \tag{6.10}$$

where $\hat{\lambda}$ is defined as

$$\hat{\lambda} = \hat{\sigma}_i'\hat{\sigma}_i\hat{\sigma}^{i\prime}\hat{\sigma}^i = \hat{\sigma}^{i\prime}\hat{\sigma}^i, \tag{6.11}$$

imposing the normalisation that $\hat{\sigma}_i'\hat{\sigma}_i = 1$. This gives the very important result that $\hat{\sigma}_i$ is an eigenvector of the $n_i^2 \times n_i^2$ matrix $\tilde{S}_i\tilde{S}_i'$. The optimum value of the modified unweighted least squares criterion is shown to be $a - \hat{\lambda}$, where $a = \text{tr}\ (\text{S}_i^2)$. Therefore, in order to render this minimal, the largest eigenvalue in equation (6.10) should be chosen. Verhees and Wansbeek [71] proved that there exists a non-iterative unweighted least squares estimator for $\hat{\sigma}_i$. This estimator is consistent but not asymptotically efficient. When $k = 2$, as in our case with multivariate longitudinal data, $\hat{\sigma}_1$ can be seen to be the first left singular vector of $\tilde{S}_1$ and $\hat{\sigma}_2$ to be the first right singular vector of this same matrix $\tilde{S}_1$. Based on this, we define an index that measures how far a given variance-covariance matrix is from Kronecker product:

Definition 6.3.1 *Let Σ be a $n \times n$ variance-covariance matrix. The Kronecker product deviation index of Σ denoted by $\delta(\Sigma)$ is defined as:*

$$\delta(\Sigma) = \min_{\Delta,\Omega} \text{tr}(\Sigma - (\Delta \otimes \Omega))^2. \tag{6.12}$$

Definition 6.3.2 *Let Σ be a $n \times n$ variance-covariance matrix. A modified definition of the Kronecker product deviation index of Σ denoted by $\delta^*(\Sigma)$ is given by:*

$$\delta^*(\Sigma) = \frac{\delta(\Sigma)}{|\Sigma|^{2/n}}. \tag{6.13}$$

The second definition ensures that the Kronecker product deviation index is invariant under scale change. In this thesis, the matrices considered are all of comparable size in terms of their determinants, hence the first definition of the Kronecker product deviation index is used.

6.3.3 An example

To illustrate the modified unweighted least squares estimator that is the solution of an eigenvalue equation, consider the following matrix:

$$\Sigma_o = \begin{pmatrix} 4 & 3.2 & 2.56 & 2 & 0.4 & 0.08 \\ 3.2 & 4 & 3.2 & 0.4 & 2 & 0.4 \\ 2.56 & 3.2 & 4 & 0.08 & 0.4 & 2 \\ 2 & 0.4 & 0.08 & 4 & 1.6 & 0.64 \\ 0.4 & 2 & 0.4 & 1.6 & 4 & 1.6 \\ 0.08 & 0.4 & 2 & 0.64 & 1.6 & 4 \end{pmatrix}.$$

The modified unweighted least squares estimates of σ_1 and σ_2 are found to be:

$$\hat{\sigma}_1 = \begin{pmatrix} 0.745 \\ 0.246 \\ 0.246 \\ 0.568 \end{pmatrix},$$

and

$$\hat{\sigma}_2 = \begin{pmatrix} 0.47 \\ 0.263 \\ 0.174 \\ 0.263 \\ 0.47 \\ 0.263 \\ 0.174 \\ 0.263 \\ 0.47 \end{pmatrix}.$$

Hence,

$$\hat{\Delta} = \begin{pmatrix} 0.745 & 0.246 \\ 0.246 & 0.568 \end{pmatrix},$$

and

$$\hat{\Omega} = \begin{pmatrix} 0.47 & 0.263 & 0.174 \\ 0.263 & 0.47 & 0.263 \\ 0.174 & 0.263 & 0.47 \end{pmatrix}.$$

Applying Definition 6.3.1, the value of the Kronecker product deviation index is found to be 10.154.

The values of the Kronecker product deviation index obtained by applying Definition 6.3.1 to the class of matrices introduced in section 6.2 are displayed in the histogram in Figure 6.1. The values in the histogram are obtained by specifying values for the correlation parameters as given in Table 6.2. The parameters σ_{11}, σ_{22} and σ_{12} and are kept constant at 4, 4 and 2 respectively. Note that the distribution of these values is strongly skewed to the right. Summary statistics for these values are also given in Table 6.3.

The results are also given for each value of ρ_{12} in Table 6.4 to Table 6.7 where they have been cross-classified by the parameters ρ_1 and ρ_2. The value of the criterion in the

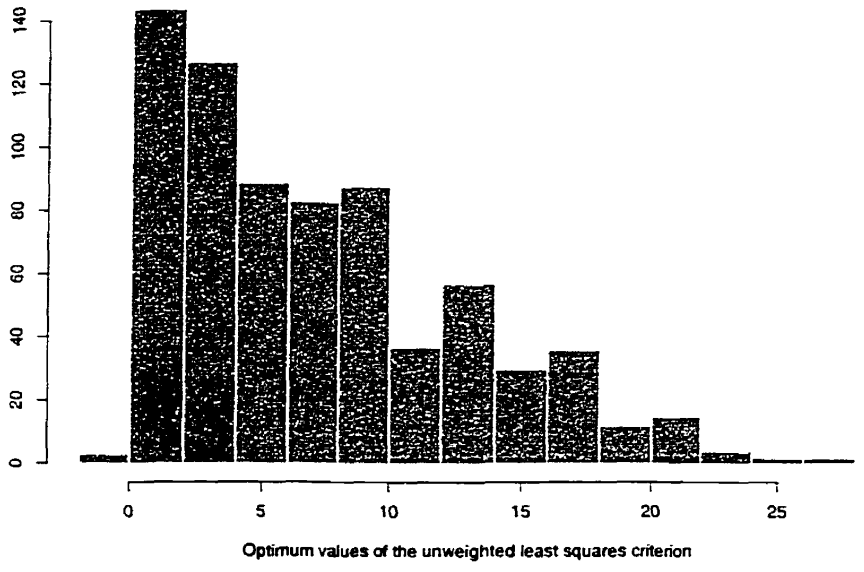

Figure 6.1: Histogram of the values of the Kronecker product deviation index.

Table 6.2: Values of correlation parameters used in computing the Kronecker product deviation index.

Parameter	**Values**
ρ_1	0.1 *to* 0.9 *by* 0.1
ρ_2	0.1 *to* 0.9 *by* 0.1
ρ_{12}	0.1 *to* 0.9 *by* 0.1

optimum is 0 when the values of the parameters ρ_1, ρ_2 and ρ_{12} are equal. Also note that the tables are symmetric along $\rho_1 = \rho_2$.

Examining each of the four tables closely, we focus on the cell where the value of the Kronecker product deviation index is 0. This is the cell for which the values of ρ_1, ρ_2 and ρ_{12} are equal. Observe that the values in the immediate vicinity of this cell are small and increase quickly as we move outwards away from this cell in all directions. This intuitively makes sense since moving away from this cell (which we can think of as

Table 6.3: Overall summary statistics for the Kronecker product deviation index.

Min.	0
1st Qu.	2.555
Median	6.024
Mean	7.233
3rd Qu.	10.73
Max.	26.54

Table 6.4: Values of the Kronecker product deviation index cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.2$.

	ρ_2								
ρ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.26	0.39	1.31	3.08	5.74	9.30	13.74	18.95	24.69
0.2	0.39	0.00	0.40	1.64	3.76	6.79	10.73	15.49	20.88
0.3	1.31	0.40	0.26	0.94	2.48	4.94	8.31	12.54	17.50
0.4	3.08	1.64	0.94	1.02	1.95	3.78	6.52	10.15	14.58
0.5	5.74	3.76	2.48	1.95	2.23	3.38	5.43	8.38	12.18
0.6	9.30	6.79	4.94	3.78	3.38	3.80	5.11	7.31	10.38
0.7	13.74	10.73	8.31	6.52	5.43	5.11	5.62	7.01	9.29
0.8	18.95	15.49	12.54	10.15	8.38	7.31	7.01	7.56	8.98
0.9	24.69	20.88	17.50	14.58	12.18	10.38	9.29	8.98	9.51

Table 6.5: Values of the Kronecker product deviation index cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.4$.

	ρ_2								
ρ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2.50	2.06	2.37	3.49	5.47	8.33	12.06	16.59	21.74
0.2	2.06	1.12	0.95	1.59	3.08	5.48	8.78	12.93	17.81
0.3	2.37	0.95	0.28	0.41	1.39	3.27	6.08	9.79	14.29
0.4	3.49	1.59	0.41	0.00	0.43	1.75	4.00	7.19	11.23
0.5	5.47	3.08	1.39	0.43	0.27	0.99	2.63	5.20	8.68
0.6	8.33	5.48	3.27	1.75	0.99	1.06	2.02	3.92	6.75
0.7	12.06	8.78	6.08	4.00	2.63	2.02	2.28	3.44	5.53
0.8	16.59	12.93	9.79	7.19	5.20	3.92	3.44	3.82	5.12
0.9	21.74	17.81	14.29	11.23	8.68	6.75	5.53	5.12	5.58

Table 6.6: Values of the Kronecker product deviation index cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.6$.

	ρ_2								
ρ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	7.52	6.42	6.02	6.37	7.50	9.46	12.25	15.83	20.08
0.2	6.42	4.88	4.04	3.95	4.67	6.24	8.68	11.97	16.03
0.3	6.02	4.04	2.76	2.22	2.49	3.63	5.66	8.59	12.36
0.4	6.37	3.95	2.22	1.22	1.02	1.67	3.23	5.72	9.11
0.5	7.50	4.67	2.49	1.02	0.30	0.43	1.46	3.44	6.35
0.6	9.46	6.24	3.63	1.67	0.43	0.00	0.45	1.85	4.19
0.7	12.25	8.68	5.66	3.23	1.46	0.45	0.29	1.03	2.73
0.8	15.83	11.97	8.59	5.72	3.44	1.85	1.03	1.09	2.08
0.9	20.08	16.03	12.36	9.11	6.35	4.19	2.73	2.08	2.32

Table 6.7: Values of the Kronecker product deviation index cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.8$.

	ρ_2								
ρ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	16.04	14.19	12.95	12.36	12.46	13.30	14.90	17.25	20.29
0.2	14.19	11.96	10.35	9.41	9.17	9.71	11.05	13.21	16.14
0.3	12.95	10.35	8.38	7.06	6.47	6.66	7.69	9.58	12.31
0.4	12.36	9.41	7.06	5.37	4.39	4.19	4.85	6.41	8.85
0.5	12.46	9.17	6.47	4.39	3.00	2.38	2.62	3.76	5.83
0.6	13.30	9.71	6.66	4.19	2.38	1.31	1.07	1.74	3.34
0.7	14.90	11.05	7.69	4.85	2.62	1.07	0.32	0.45	1.52
0.8	17.25	13.21	9.58	6.41	3.76	1.74	0.45	0.00	0.47
0.9	20.29	16.14	12.31	8.85	5.83	3.34	1.52	0.47	0.29

the center of the table) means that the values of the parameters ρ_1 and ρ_2 are moving further away from the value of ρ_{12}. The departure of ρ_1 and ρ_2 further and further away from ρ_{12} implies that we are getting further away from the Kronecker product structure and hence the values of the Kronecker product deviation index are getting larger.

6.4 A Monte-Carlo study

In this section, a Monte-carlo simulation study was undertaken to evaluate the impact of the Kronecker product deviation index on testing hypotheses of interest in multivariate longitudinal data. The evaluation is done both under the null and alternative hypotheses. The test based on imposing a Kronecker product covariance matrix is compared to a test based on a non-Kronecker product covariance matrix. These tests are investigated for a multivariate longitudinal design consisting of data from two groups of subjects measured on three different occasions on two characteristics.

6.4.1 Data generation

Multivariate normal data with $\mu = (\Theta \otimes I_C \otimes X)\lambda$ and $\Sigma = I_I \otimes \Sigma_o = I_I \otimes \Sigma_a$, where $\Sigma_o = \Sigma_a$ is as defined in section 6.2, is generated for two characteristics $C = 2$ and three time points $T = 3$. As discussed in Chapters 4 and 5, for testing the hypotheses

$$H_o : Q\lambda = 0 \quad \text{Vs.} \quad \mathrm{H_a} : \mathrm{Q}\lambda \neq 0, \tag{6.14}$$

we compute the test statistic

$$T^* = (Q\hat{\lambda})' \left(Q\hat{V}Q'\right)^{-1} (Q\hat{\lambda}), \tag{6.15}$$

and compare it to a χ^2 distribution with r degrees of freedom. The null hypothesis $H_o : Q\lambda = 0$ is rejected at level α if

$$T^* = (Q\hat{\lambda})' \left(Q\hat{V}Q'\right)^{-1} (Q\hat{\lambda}) > \chi^2_r(\alpha).$$

In the Monte-carlo simulation study conducted here, the vector μ is specified in two different ways. To assess the impact of the Kronecker product deviation index under the null hypothesis, the vector μ is specified so that the null hypothesis is true. This means specifying μ so that $Q\lambda = 0$. To assess the impact of the Kronecker product deviation index under the alternative hypothesis, μ is specified so that the null hypothesis is not true. This means specifying μ so that $Q\lambda \neq 0$. To be more specific, the vector μ is specified for a bivariate growth curve data problem where one is interested in the overall hypothesis of parallelism. Under H_o, we have

$$H_o : \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

If H_o is true, then

$$Q\lambda = \begin{pmatrix} \beta_{10} - \beta_{20} \\ \beta_{11} - \beta_{21} \\ \alpha_{10} - \alpha_{20} \\ \alpha_{11} - \alpha_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

If H_o is not true, then $Q\lambda \neq 0$.

Under the null hypothesis, we specify μ so that:

$$\begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix},$$

where the intercepts and slopes for the two groups are equal for both characteristics.

Under the alternative hypothesis, μ is specified so that:

$$\begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} \neq \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

For the sake of simplicity, we specify equal intercepts ($\beta_{10} = \beta_{20}, \alpha_{10} = \alpha_{20}$) but different slopes. For the slopes, we let $\beta_{21} = k\beta_{11}$ and $\alpha_{21} = k\alpha_{11}$ and specify the constant k so that the alternative hypothesis represents just a slight departure from the null hypothesis.

The specific values in μ are specified following modified results of the study described in section 5.8. The study concerned the relative effectiveness of two orthopaedic adjustments of the mandible. Nine subjects were assigned to each of two orthopaedic treatments, called activator treatments. On each of three occasions, three dependent variables were observed. The three dependent variables, in combination, reflected the position and size of the mandible. Mean plots of the data for each group and variable revealed that the growth curves of the three variables were at least linear. Timm [65] fit a quadratic regression model to the data. In the study described here, the parameters for group 1 are specified by ignoring the quadratic terms and using the data from the first group on the first two variables. The parameters for group 2 are specified as a function of group 1 parameters, that is:

$$\begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ k\beta_{11} \\ \alpha_{10} \\ k\alpha_{11} \end{pmatrix},$$

and k is specified so that group 2 slopes for both characteristics are 20% greater than group 1 slopes. Note that when $k = 1$, $Q\lambda = 0$ and the data is therefore generated assuming H_o is true. All other values of k, that is, $k > 1$ signify some departure from H_o and hence $Q\lambda \neq 0$. In the simulation study done here, only one alternative is considered which is the case where group 2 slopes are 20% greater than the group 1 slopes. The

justification is that the main focus here is to understand the effect of the Kronecker product deviation index on hypothesis testing.

The within-subject variance-covariance matrix is specified to be non-Kronecker product following the specification outlined in section 6.2. To evaluate how far each of the covariance matrices departs from the Kronecker product form, the index of departure discussed in section 6.3 is computed prior to data generation. The further the index is from 0, the further the given covariance matrix is from Kronecker product. The parameter values used in specifying the within-subject variance covariance matrix are given in Table 6.8. This gives a total of 100 parameter combinations. The range of parameters

Table 6.8: Values of parameters defining Ω used in the Monte-carlo simulation study.

Parameter	**Values**
ρ_1	0.1 *to* 0.9 *by* 0.2
ρ_2	0.1 *to* 0.9 *by* 0.2
ρ_{12}	0.2 *to* 0.8 *by* 0.2

and parameter combinations considered is varied enough to represent parameter combinations that may arise in practice. Computational problems are encountered for values of ρ_1, ρ_2 and ρ_{12} very close to 0 and 1 and hence the range of values considered for these parameters is restricted to lie between 0.1 and 0.9 inclusive. The parameters σ_{11} and σ_{22} are kept constant. We did not see the need to vary these parameters since they had minimal effects on any of the quantities of interest investigated in Chapter 5. For each set of parameter combinations, 200 simulation trials are carried out.

6.4.2 Model fit and quantities of interest

The data generated by each trial is analyzed using a linear model for multivariate longitudinal data. More specifically, a linear growth curve model is fit to each of the two response variables in each of the two groups. The model fit to the two variables is

$$y = (\Theta \otimes I_C \otimes X)\lambda + e, \tag{6.16}$$

with

$$\Theta = \begin{pmatrix} \mathbf{1}_9 & 0 \\ 0 & \mathbf{1}_9 \end{pmatrix},$$

$$I_C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and

$$\lambda = \text{vec}(\Lambda) = \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \\ \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix}.$$

To test for parallelism simultaneously for both variables, that is:

$$H_o : \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \alpha_{20} \\ \alpha_{21} \end{pmatrix},$$

we compute the test statistic

$$T = (Q\hat{\lambda})' (Q\hat{V}Q')^{-1} (Q\hat{\lambda})$$

and compare it to a χ^2 distribution with 4 degrees of freedom, where

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Two different covariance models are fit to the data:

1. a non-Kronecker product covariance matrix;

2. a Kronecker product covariance matrix.

Each trial yields two test statistics and the p-values for the two tests are easily obtained from the χ^2 distribution with 4 degrees of freedom. Note that for each trial, we obtain two test statistics and corresponding p-values as a result of fitting two different covariance models to the data. For each of the two models fit, we can evaluate observed significance level by counting the proportion of times (out of the total number of trials) that the null hypothesis $Q\lambda = 0$ is falsely rejected. This is the number of times the p-value obtained is less than or equal to the fixed significance level α out of the total number of trials carried out. When $k > 1$, $Q\lambda \neq 0$ and the null hypothesis is not true. In a similar way, we can find the power for the two models fit by counting the proportion of times (out of the total number of trials) that the null hypothesis $Q\lambda = 0$ is correctly rejected. This is the number of times the p-value obtained is less than or equal to the fixed significance level α out of the total number of trials carried out. The primary focus, however, will be on the impact of the Kronecker product deviation index on hypothesis testing.

6.5 Results and discussion

Under the null hypothesis, data was generated to simulate a two treatment bivariate growth curve model in which the mean vectors for the groups on the two variables

measured on three occasions are equal. Under the alternative hypothesis, data was generated to simulate a two treatment bivariate growth curve model in which the mean vectors for the groups on the two variables measured on three occasions are not equal. Specifically, the intercepts for the two characteristics were set equal in both groups, but the slopes for both characteristics in group 2 exceeded those of group 1 by 20%. This represents a slight departure from the null hypothesis. The hypothesis of interest in both cases was the overall hypothesis of parallelism between the two groups which was tested by fitting two different covariance models to the data. Tables 6.9 to 6.12 give the values of observed significance level obtained while Tables 6.13 to 6.16 give the achieved power.

Table 6.9: Values of the observed significance level for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.2$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	7.00	11.50	8.50	12.00	8.00	12.00	9.00	11.00	5.50	11.50
0.3	7.00	11.50	9.50	12.00	10.50	11.00	9.00	6.50	7.50	12.00
0.5	6.50	10.50	10.50	10.50	11.50	7.50	8.50	7.50	7.00	12.00
0.7	7.00	11.50	11.00	8.50	9.00	8.50	7.50	10.00	7.50	11.50
0.9	5.00	11.50	7.00	10.50	7.54	10.05	6.53	10.55	6.00	11.50

The histograms in Figure 6.2 show the distribution of the p-values obtained from testing the null hypothesis of parallelism between the two groups under the two covariance models. Under the null hypothesis, the distribution of the p-values is expected to be close to uniform when the hypothesis of parallelism between two groups is tested using an unstructured covariance matrix. Some slight deviation from the uniform distribution may be expected when the same hypothesis is tested using a Kronecker product covariance

Table 6.10: Values of the observed significance level for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.4$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	7.14	12.86	6.00	12.00	7.50	12.00	8.00	11.00	6.50	8.50
0.3	5.00	11.50	7.50	11.50	8.50	12.00	8.50	11.00	7.00	10.00
0.5	5.50	11.50	6.50	11.50	10.00	11.50	10.50	8.50	7.07	11.62
0.7	6.00	11.00	7.50	11.00	11.50	10.00	10.50	8.00	8.08	11.62
0.9	5.50	11.50	6.03	10.05	8.08	10.10	8.04	11.56	7.04	11.56

Table 6.11: Values of the observed significance level for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.6$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	9.80	14.71	5.50	11.50	4.50	12.00	5.50	12.00	6.50	11.50
0.3	6.50	11.50	6.50	11.50	6.50	12.00	7.50	12.50	9.05	11.56
0.5	5.50	11.50	6.00	11.50	8.50	12.50	8.50	12.50	9.23	10.77
0.7	3.50	11.50	6.00	11.50	8.00	12.00	11.00	11.50	8.60	11.83
0.9	5.00	11.50	5.53	10.55	9.18	9.69	9.28	10.82	9.28	11.34

Table 6.12: Values of the observed significance level for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.8$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	7.04	11.56	6.50	11.50	6.00	12.00	4.00	12.00	5.00	11.50
0.3	7.00	11.50	7.50	11.50	7.00	12.00	5.00	12.50	7.00	11.50
0.5	5.50	11.50	7.50	11.50	6.50	12.50	7.00	12.50	8.06	10.75
0.7	5.00	12.00	6.00	12.50	9.00	12.50	9.50	12.50	9.46	12.84
0.9	4.00	11.50	6.50	11.00	8.81	11.40	10.91	10.91		

Table 6.13: Values of empirical power for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.2$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	20.00	27.00	20.50	26.00	20.00	24.00	22.50	23.50	40.00	99.50
0.3	24.50	32.00	31.00	32.00	31.50	29.00	29.50	25.50	29.50	85.00
0.5	40.00	51.50	47.00	51.50	55.00	49.00	49.50	53.00	38.00	40.00
0.7	92.00	97.50	96.50	98.50	97.00	99.50	66.50	100.00	49.00	57.50
0.9	100.00	100.00	100.00	100.00	98.49	100.00	66.83	99.50	63.50	80.00

Table 6.14: Values of empirical power for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.4$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	19.29	30.00	20.50	28.50	19.50	25.00	22.50	28.00	76.50	100.00
0.3	21.50	31.50	27.00	31.50	30.00	29.00	30.00	28.50	95.50	100.00
0.5	28.50	46.50	41.00	47.00	46.50	47.50	52.00	42.50	46.46	85.86
0.7	71.00	93.00	86.50	93.00	96.00	95.00	99.50	100.00	65.15	63.64
0.9	100.00	100.00	100.00	100.00	100.00	100.00	99.50	100.00	86.93	88.94

Table 6.15: Values of empirical power for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.6$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	16.67	33.33	16.00	30.50	17.00	28.00	22.50	32.50	78.50	99.50
0.3	16.50	32.00	23.50	32.00	25.50	33.00	31.50	34.50	94.97	99.50
0.5	22.00	44.00	33.50	44.00	42.00	44.00	49.50	42.00	100.00	99.49
0.7	53.50	82.00	67.50	83.00	77.50	83.00	88.50	84.00	80.65	98.92
0.9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.48

Table 6.16: Values of empirical power for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis cross-classified by ρ_1 and ρ_2: $\rho_{12} = 0.8$.

ρ_1	**0.1**		**0.3**		**0.5**		**0.7**		**0.9**	
ρ_2	KP	UN	KP	UN	KP	UN	KP	UN	KP	UN
0.1	13.57	32.66	15.00	33.50	16.50	34.00	25.00	40.50	91.00	99.00
0.3	14.50	36.50	18.00	37.00	24.50	38.00	32.50	44.00	94.50	99.00
0.5	22.00	46.50	29.00	46.00	38.50	48.00	48.00	50.00	95.70	97.85
0.7	49.50	79.50	59.50	78.50	71.50	78.50	78.50	76.50	95.95	93.92
0.9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		

matrix. Quantile plots of the p-values when the null hypothesis is true, based on the uniform distribution for both the unstructured and Kronecker product covariance models, are given in Figure 6.3. It appears that the distribution of the p-values from the two tests do not depart too much from the uniform distribution.

To better understand the role of the Kronecker product deviation index, the p-values obtained from the two tests are plotted conditioning on intervals of the Kronecker product deviation index. Figure 6.4 shows the results where the endpoints of the Kronecker product deviation index intervals are chosen so as to make the counts of points in the intervals as nearly equal as possible. Figure 6.5 shows the same results where the endpoints of the Kronecker product deviation index intervals are chosen so as to make the intervals to be of equal width. The p-values in both plots are given when the alternative hypothesis is true. Figures 6.6 and 6.7 show the results under the null hypothesis. Figures 6.4 and 6.5 clearly show that the variability in the p-values obtained increases as the Kronecker product deviation index increases. Another important observation is that the dark cloud of points is shifting towards the horizontal axis as the Kronecker product

deviation index increases. This implies that overall, while the the p-values of the test using the unstructured covariance matrix get smaller as the Kronecker product deviation index increases, the p-values of the test using the Kronecker product covariance matrix are getting larger. Figure 6.8 further clarifies this point with a general upward trend in the scatter of points confirmed by the fitted least squares regression line. Overall, the figure shows that the differences in power observed under the two tests is increasing as the Kronecker product deviation index increases. The effect on the observed significance level of the test is not as pronounced even though we do observe an increase in the variability as the Kronecker product deviation index increases.

Another outcome of interest in this study was the power of the test for the null hypothesis that the within subject variance covariance matrix has a Kronecker product structure described in section 5.4. Figure 6.9 shows the relationship between the Kronecker product deviation index and the p-values obtained from testing this hypothesis. From the figure, we note that when the Kronecker product index is large, the p-value for the test of the null hypothesis that the within subject variance covariance matrix has a Kronecker product structure is small, indicating that the test is, for the most part, doing the right thing and is able to detect departure from Kronecker product structure.

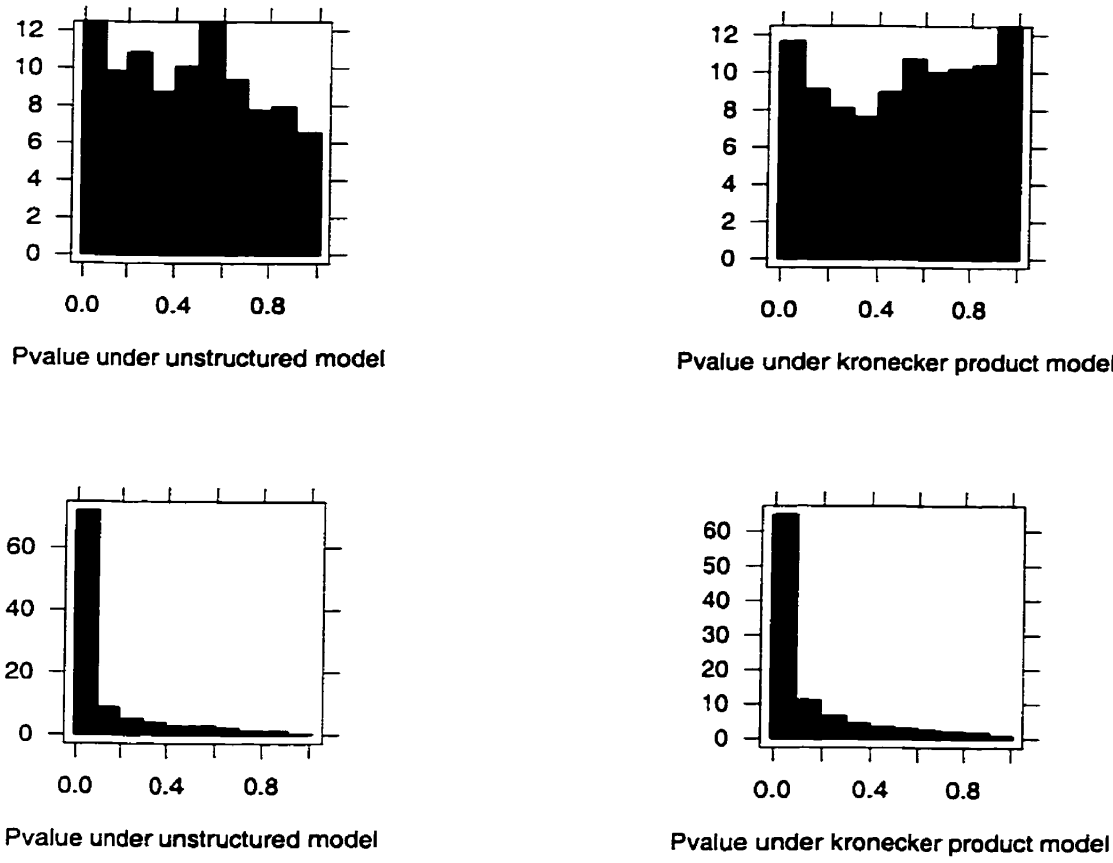

Figure 6.2: Histograms of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis (top figures) and alternative hypothesis (bottom figures).

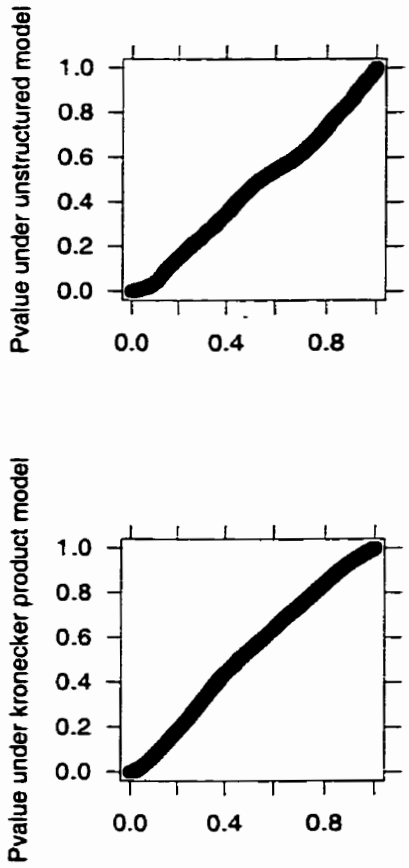

Figure 6.3: Quantile plots of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis based on the uniform distribution.

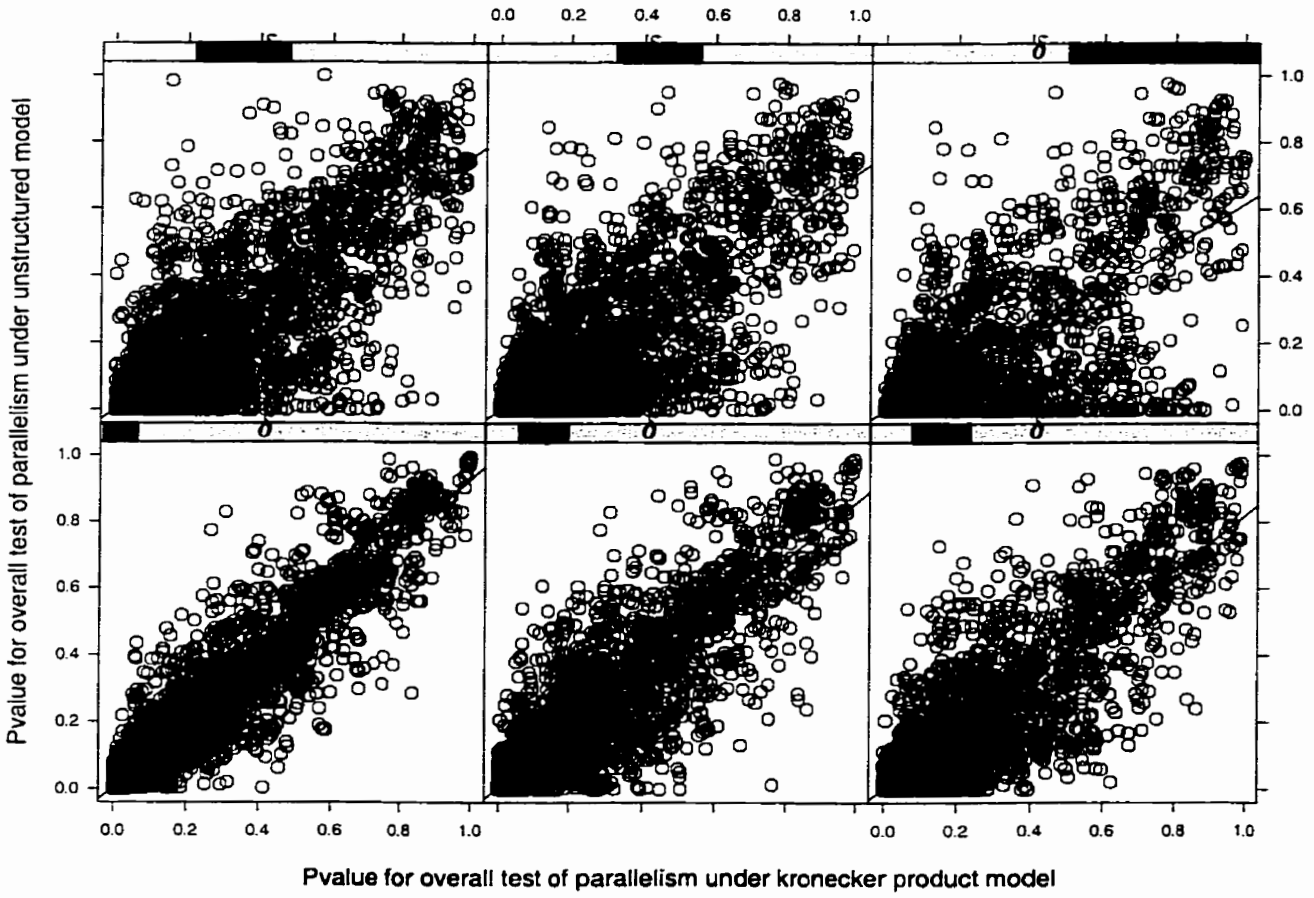

Figure 6.4: Scatter plots and fitted least squares regression lines of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis. Plots are conditioned on intervals of the Kronecker product deviation index (counts of points in the intervals nearly equal).

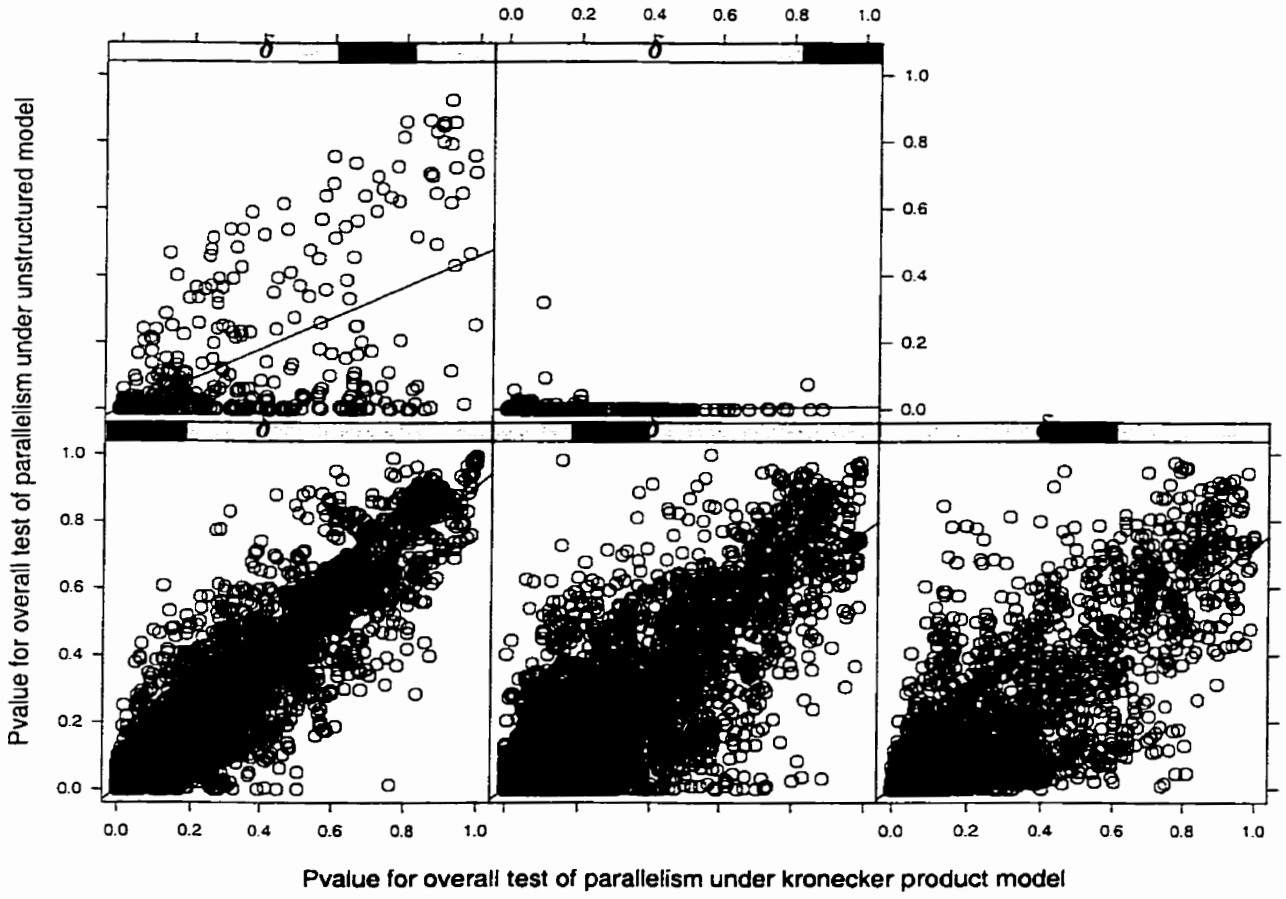

Figure 6.5: Scatter plots and fitted least squares regression lines of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis. Plots are conditioned on intervals of the Kronecker product deviation index (intervals of equal width).

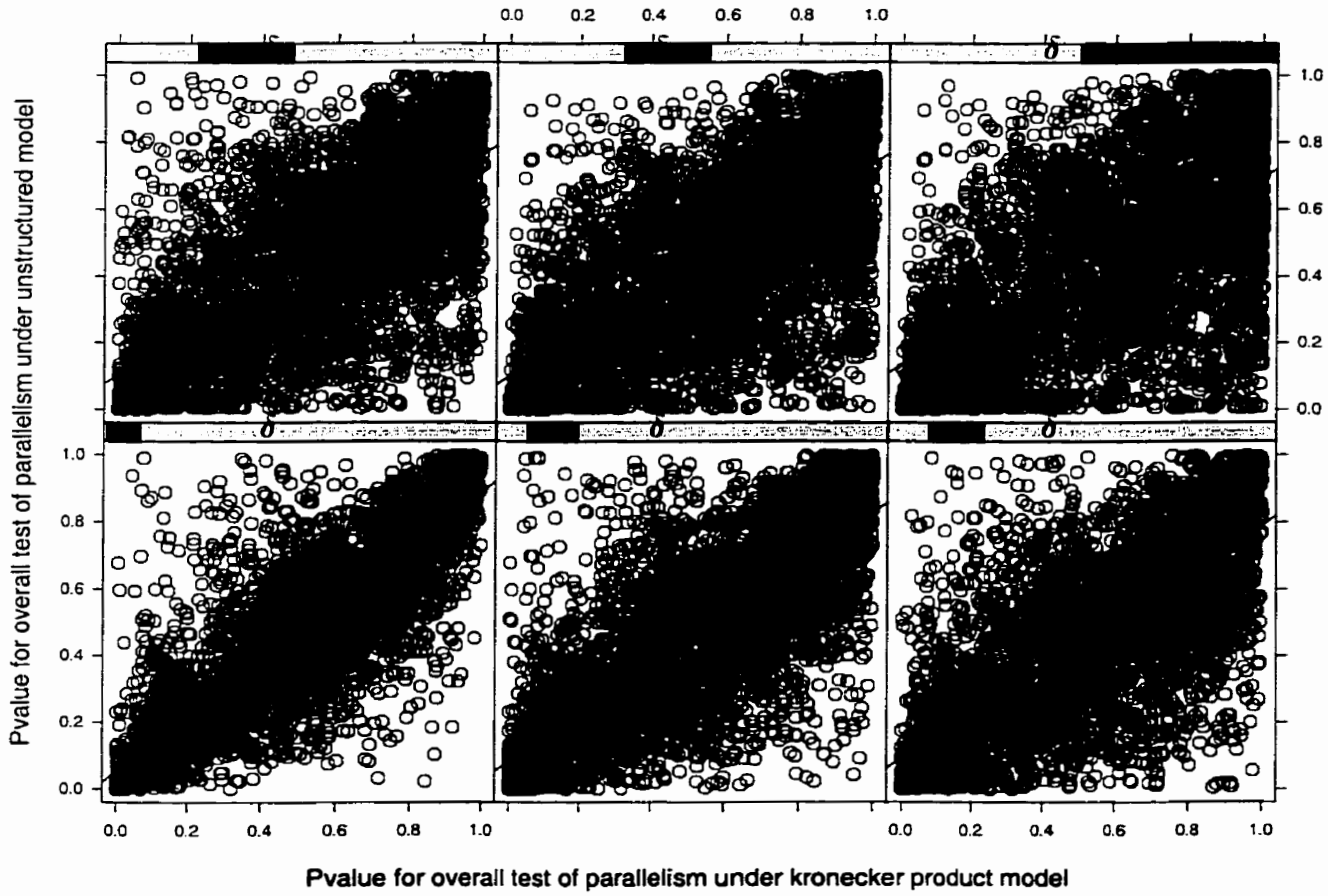

Figure 6.6: Scatter plots and fitted least squares regression lines of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis. Plots are conditioned on intervals of the Kronecker product deviation index (counts of points in the intervals nearly equal).

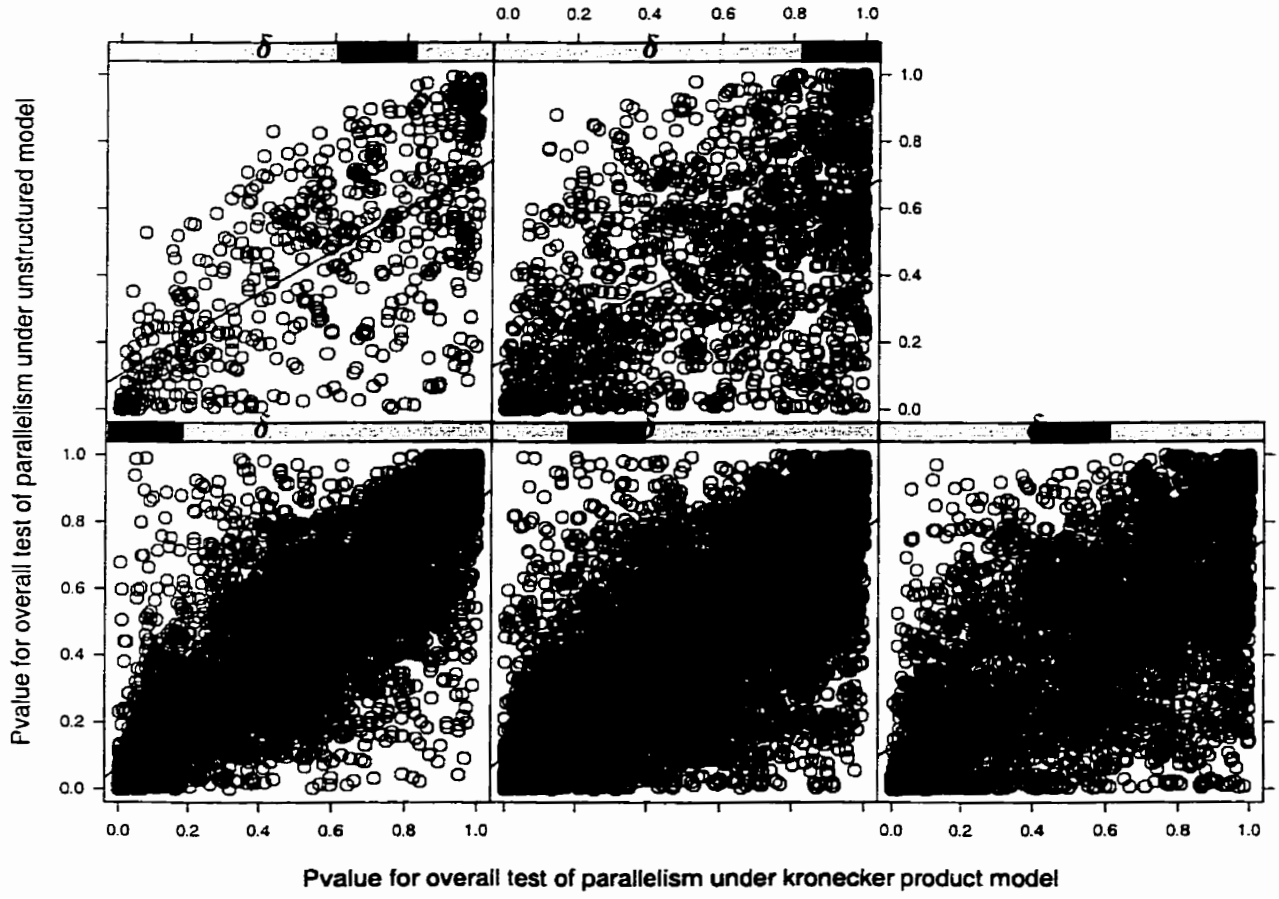

Figure 6.7: Scatter plots and fitted least squares regression lines of the p-values for the test of overall hypothesis of parallelism between two groups in a growth curve model under the null hypothesis. Plots are conditioned on intervals of the Kronecker product deviation index (intervals of equal width).

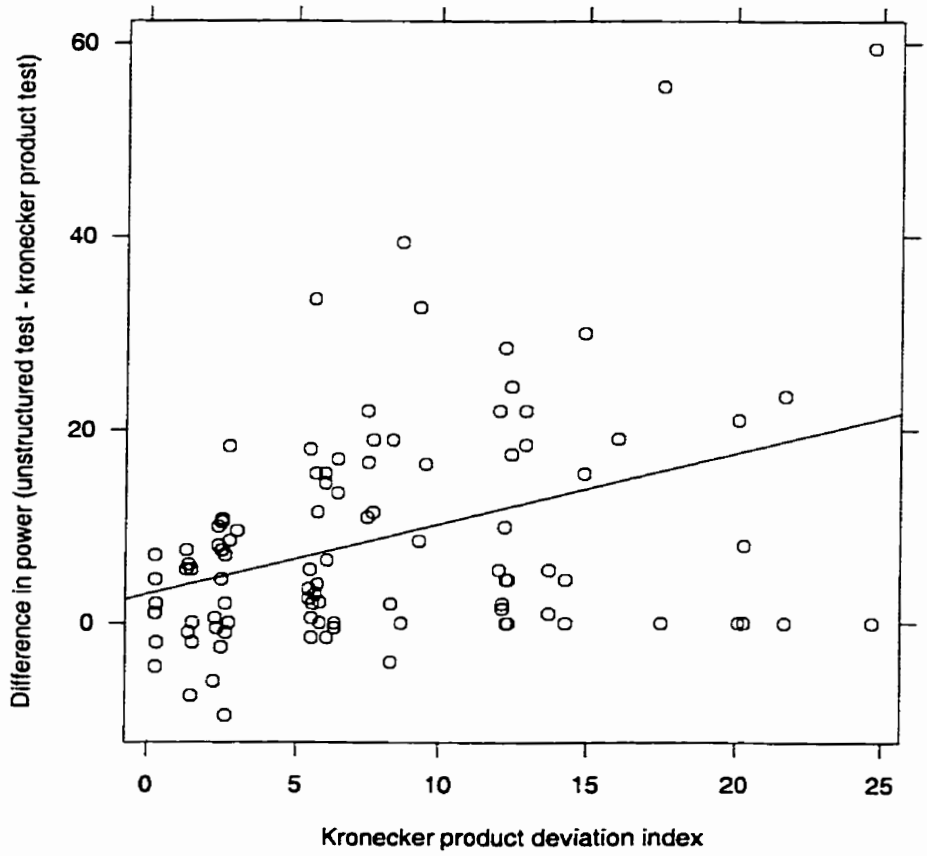

Figure 6.8: Scatter plot and fitted least squares regression line of the difference in power (power of test based on a unstructured covariance matrix - power of test based on a Kronecker product covariance matrix) for the test of overall hypothesis of parallelism between two groups in a growth curve model under the alternative hypothesis.

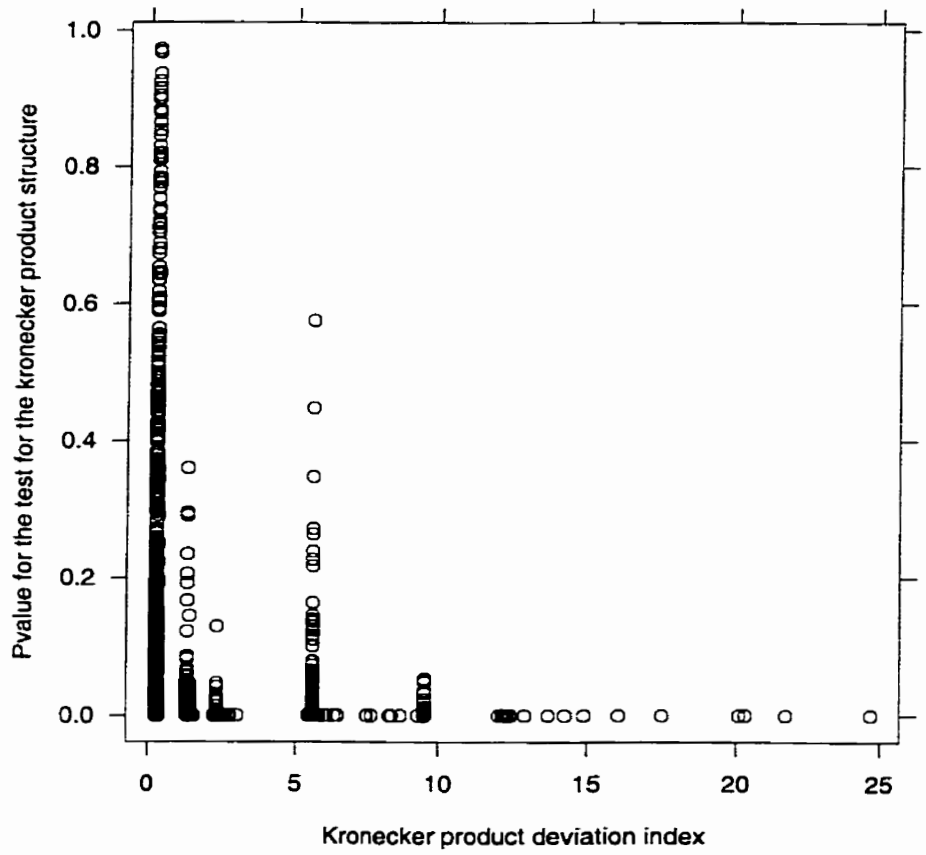

Figure 6.9: Scatter plot of the Kronecker product deviation index versus the p-values for the test of the null hypothesis that the within-subject variance-covariance matrix has a Kronecker product structure.

Chapter 7

Summary and further research

7.1 General discussion

In this dissertation, we have investigated efficiency in the linear model for multivariate longitudinal data with a Kronecker structured covariance matrix. The distinguishing characteristic of this model is that it requires the within-subject variance-covariance matrix to be specified as the Kronecker product of two matrices that reflect the two dimensions underlying multivariate longitudinal data, namely characteristic and time. The same model has been used to model covariance structure when two or more repeated factors are present in a given study as discussed by Galecki [18].

Some advantages of using this model for multivariate longitudinal data include clear and meaningful interpretation in terms of the contribution of the characteristic and time dimensions to the overall within-subject variance-covariance matrix. Under different settings, efficiency was evaluated by deriving the trace asymptotic relative efficiency (TARE) and curvature asymptotic relative efficiency (CARE), both measures of asymptotic relative efficiency. Both measures can be applied to compare competing test statistics which have limiting non-central chi-square distributions through a suitable Pitman alternative.

One approach commonly used to analyse data from multivariate longitudinal de-

signs is ordinary least squares. If different subjects are being measured at different times, this might be a reasonable approach. However, if the same subjects being measured over time, it is more realistic to assume that the measurements within a subject are correlated. Modelling the covariance matrix using the Kronecker product of two matrices is one way to capture this correlation. The efficiency of a test procedure that ignores correlation relative to a test that models the covariance matrix as the Kronecker product of two matrices (assumed to be the true structure) was evaluated using the TARE and CARE.

Numerical results were presented for two designs (growth curve and repeated measures analysis of variance) and two covariance structures for the matrix that models the repeated measures on a given characteristic (compound symmetry and first-order autoregressive). The covariance parameters ρ_t, ρ_c and γ were found to have a pronounced effect on both measures of asymptotic relative efficiency. The degree of the loss of efficiency was clearly demonstrated to be a function of these covariance parameters. For the designs and covariance matrices considered, the results indicate that the efficiency of a test procedure that ignores correlation relative to a test that models the covariance matrix as the Kronecker product of two matrices (assumed to be the true structure) is worse for high values of ρ_t and ρ_c and low values of γ.

Another issue investigated, and considered to be of considerable practical interest, was the potential gain in efficiency that would result from testing hypotheses of interest using a test that utilised the Kronecker product structure. The efficiency of a test procedure that ignores the Kronecker product structure relative to one that models the covariance matrix as the Kronecker product of two matrices was evaluated using the TARE and CARE. Expressions for the TARE and CARE are derived. To estimate efficiency, a Monte-carlo simulation study was conducted. The design in the simulation study was specified to correspond to a two group bivariate growth curve setting and the covariance matrix was specified to be the Kronecker product of a unstructured covariance matrix and a first-order autoregressive matrix. The parameters defining the two matrices were varied to represent various parameter combinations that are likely to arise in practise.

Once again, efficiency is shown to be a function of the covariance parameters ρ_t, ρ_c and γ. For the design and covariance parameters considered, the parameter ρ_t defining Ω was found to have the greatest impact on efficiency. In practical work, one would also need to know if the Kronecker product structure is suitable or not. In this regard, a test of the null hypothesis that the within-subject variance-covariance matrix has a Kronecker product structure was also presented. Since in the simulation study the within-subject variance-covariance matrix was specified to be the Kronecker product of a unstructured covariance matrix and a first-order autoregressive matrix, the performance of the test was evaluated as a by-product of the simulations. The type I error rates of the test of Kronecker product structure were found to be very close to the nominal values.

The validity of the model considered so far depends largely on the special covariance structure that it assumes. If the Kronecker product model is not suitable in a given situation, then there should be some consequences if hypotheses of interest are tested under the Kronecker product structure. The consequences of imposing the special covariance structure were also investigated. A class of matrices with some degree of departure from the Kronecker product model is introduced. An index, called the Kronecker product deviation index, is used to quantify how far a given variance-covariance matrix departs from Kronecker product. The index is described and evaluated for the class of matrices introduced.

A Monte-carlo simulation study using this class of covariance matrices was used to compare the impact of the Kronecker product deviation index on a test based on imposing a Kronecker product structure relative to a test based on a unstructured covariance matrix. The null hypothesis of interest was that of overall parallelism in a two group bivariate growth curve design. The results obtained indicate that the greatest negative consequence from imposing a Kronecker product model in testing hypotheses of interest occurred when there was moderate departure from the null hypothesis. For the parameter combinations considered, the power of the test that imposed the Kronecker product model was consistently lower. Also, the difference the power between the two tests was found to increase as the Kronecker product deviation index increased.

7.2 Limitations

1. The within-subject design considered is the same for all subjects. This is too restrictive and a more general specification that allows different subjects to have different designs should be considered.

2. Study considered two within-subject designs (growth curve and repeated measures ANOVA). Other designs should be considered.

3. Study only looked at covariates that are time invariant, for example, the treatment group that an individual is assigned to. It would be useful to consider time varying covariates as well, for example, characteristics of the subject that change with time and that may have an effect on the response of interest.

4. Two covariance matrices were used for Ω, namely, compound symmetry and first-order autoregressive. Other types of matrices that can be used to model Ω, including unstructured and simple, should be considered.

7.3 Further research

1. Most of the work in this dissertation has focused on two within-subject design matrices: the growth curve design and the repeated measures analysis of variance design. Another design that is equally important and needs to be investigated is the crossover design. In a crossover trial, the entire study period is first divided into say p experimental phases. A wash-out period is usually allowed between the phases. The design also specifies a number of different treatment sequences and outlines the order in which the treatments are to be administered for each treatment sequence. Subjects are then randomly assigned to the different treatment sequences. The design is commonly used in areas such as agriculture and medicine. If multiple characteristics are being measured on each study subject, then the model with a Kronecker product covariance matrix can be used. The efficiency of this model needs to be investigated for the crossover design.

2. In Chapter 4, the efficiency of a test procedure that ignores correlation relative one that models the covariance matrix as the Kronecker product of two matrices (assumed to be the true structure) was evaluated using the trace asymptotic relative efficiency and curvature asymptotic relative efficiency. Mathematical expression for the trace asymptotic relative efficiency and curvature asymptotic relative efficiency were derived and shown to be functions of various quantities. For different within-subject designs and different covariance matrices, it would be useful to establish bounds on both the trace asymptotic relative efficiency and curvature asymptotic relative efficiency. For example, if the within-subject design corresponds to a growth curve setting and the within-subject variance-covariance matrix is the Kronecker product of a unstructured covariance matrix and a first-order autoregressive matrix, then bounds on both the TARE and CARE should be derived.

3. The greatest benefit from using the Kronecker product model for multivariate longitudinal data may be in the presence of unbalanced and/or missing data, a common problem in designs that involve long or short term follow up of subjects. Further simulation work is required to investigate this. To investigate the benefit of using the Kronecker product model in the presence of missing data, for example, data can be simulated with a Kronecker product covariance matrix for the within-subject variance-covariance matrix. Observations can then be systematically deleted at different rates in specific patterns. The trace asymptotic relative efficiency and curvature asymptotic relative efficiency from testing hypothesis of interest can then be evaluated for a test based on a Kronecker product model relative to other tests, for example, a test based on a unstructured covariance matrix.

4. Another potential benefit from using the Kronecker product model for multivariate longitudinal data may be in cases where a given study involves only a small number of subjects. It is expected that the effects of mis-specified covariance structures on testing hypotheses of interest may not be substantial if the number of subjects I in the study is large. However, this is likely not to be the case when only a small number of subjects have been enrolled or are available for the study. Hence,

the trace asymptotic relative efficiency and curvature asymptotic relative efficiency need to be evaluated with small and moderate sample sizes.

5. The response vector for a given subject $y_i, i = 1, \ldots, I$ has been represented so far with the time index running faster than the characteristics index. For example, for a single subject with 2 characteristics measured on 3 occasions, the response vector is given by:

$$y_i = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix}. \tag{7.1}$$

with the first index representing characteristics and the second index representing time. Using this formulation, the within-subject variance-covariance matrix has been expressed as $\Sigma_o = \Delta \otimes \Omega$. The covariance between the outcome variables is specified by the $C \times C$ matrix Δ whereas the covariance among the repeated measures for a given outcome variable is specified by the $T \times T$ matrix Ω. If the order of the two dimensions is reversed, we now have:

$$y_i = \begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}. \tag{7.2}$$

This formulation of y_i may have some advantages and needs to be considered, especially in terms of the within-subject variance-covariance matrix Σ_o. For example, in defining a class of matrices that depart from Kronecker product, the correlation between the characteristics can be specified to change over time. Without loss of

generality, for two characteristics measured on three occasions, let ρ_1, ρ_2 and ρ_3 be the correlations between characteristic 1 and characteristic 2 at times 1, 2 and 3 respectively. Allowing ρ_1, ρ_2 and ρ_3 to take on different values removes the rather restrictive assumption that the correlation between the characteristics is constant over time. This seems more practical and can also be easily interpreted. In extreme situations, we can model $\rho_1 > 0$, $\rho_2 = 0$ and $\rho_3 < 0$. This alternative formulation needs to be investigated both in terms of evaluating efficiency in different settings using the trace asymptotic relative efficiency and curvature asymptotic relative efficiency and in assessing the consequences of imposing the Kronecker product structure on testing hypothses of interest.

6. The model considered in this dissertation applies to multivariate longitudinal data when the measurements are assumed to be multivariate normal. In many practical situations, however, this assumption will not hold. In particular, when the responses are discrete or represent count data, different methodology must be used. When a single outcome variable is being recorded over time, generalised estimating equations provide one practical means for dealing with discete or count data. The approach to be taken when multiple outcome variables that are discrete or represent count data needs to be investigated. In particular, a way of modelling the covariance matrix that is similar to the Kronecker product approach for continous data should be sought.

Bibliography

[1] B. Abraham and Ch. E. Minder. A time series model with random coefficients. *Communications in Statistics, Theory and Methods*, 11(12):1381–1391, 1982.

[2] Blair M. Anderson, T. W. Anderson, and Ingram Olkin. Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *The Annals of Statistics*, 14(2):405–417, 1986.

[3] Adelchi Azzalini. Growth curve analysis for patterned covariance matrices. In M. L. Puri and J. P. Vilaplana, editors, *New perspectives in theoretical and applied statistics*, pages 61–74. John Wiley and Sons Inc., 1987.

[4] Peter Bloomfield and Geoffrey S. Watson. The inefficiency of least squares. *Biometrika*, 62(1):121–128, 1975.

[5] R. D. Bock. *Multivariate statistical methods in behavioral research.* McGraw-Hill Series in Psychology. McGraw-Hill Inc., New-York, 1975.

[6] Robert R. Boik. The mixed model for multivariate repeated measures: Validity conditions and an approximate test. *Psychometrika*, 53:469–486, 1988.

[7] Robert R. Boik. Scheffe's mixed model for multivariate repeated measures: A relative efficiency evaluation. *Communications in Statistics, Theory and Methods*, 20:1233–1255, 1991.

[8] Vernon M. Chinchilli and Walter H. Carter Jr. A likelihood ratio test for a patterned covariance matrix in a multivariate growth-curve model. *Biometrics*, 40:151–156, 1984.

[9] D. R. Cox, R. Fitzpatrick, A. E. Fletcher, S. M. Gore, D. J. Spiegelhalter, and D. R. Jones. Quality of life assessment: Can we keep it simple? *Journal of the Royal Statistical Society, Series A*, 155:353–393, 1992.

[10] Martin J. Crowder and David J. Hand. *Analysis of Repeated Measurements*, volume 41 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, London, 1990.

[11] Carroll J. Diaz and William D. Johnson. An F-test for multivariate repteated measures data with the Wiener stochastic process pattern in the covariance matrix. *Communications in Statistics, Theory and Methods*, 27(2):275–289, 1998.

[12] Terry E. Dielman and Roger C. Pfaffenberger. Efficiency of ordinary least squares for linear models with autocorrelataion. *Journal of the American Statistical Association*, 84(405):248–248, 1989.

[13] Peter J. Diggle. An approach to the analysis of repeated measurements. *Biometrics*, 44:959–971, 1988.

[14] Peter J. Diggle, Kung-Yee Liang, and Scott L. Zeger. *Analysis of Longitudinal Data*, volume 13 of *Oxford Statistical Science Series*. Oxford University Press, London, 1994.

[15] Dorothy D. Dunlop. Regression for longitudinal data: A bridge from least squares. *The American Statistician*, 48(4):299–303, 1994.

[16] Paul L. Enright and Robert E. Hyatt. *Office Spirometry: A practical guide to the selection and use of spirometers*. Lea & Febiger, Philadelphia, 1987.

[17] Garrett M. Fitzmaurice, Nan M. Laird, and Andrea G. Rotnizky. Regression models for discrete longitudinal responses. *Statistical Science*, 8(3):284–309, 1993.

[18] Andrzej T. Galecki. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics, Theory and Methods*, 23(11):3105–3119, 1994.

[19] Jean Dickinson Gibbons. *Nonparametric statistical inference*, volume 65 of *Statistics, textbooks and monographs*. Marcell Dekker, Inc., New York, second edition, 1985.

[20] James E. Grizzle and David M. Allen. Analysis of growth and dose response curves. *Biometrics*, 25:357–381, 1969.

[21] David Hand and Martin Crowder. *Practical Longitudinal Data Analysis*. Texts in Statistical Science. Chapman and Hall, London, 1996.

[22] David A. Harville. Bayesian inference for variance components using only error contrasts. *Biometrika*, 61:383–385, 1974.

[23] P. Hopwood, R. J. Stephens, and D. Machin. Approaches to the analysis of quality of life data: experience gained from a Medical Research Council Lung Cancer Working Party palliative chemotherapy trial. *Quality of Life Research*, 3:339–352, 1994.

[24] R. I. Jennrich and P. F. Sampson. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18(1):11–17, 1976.

[25] Robert I. Jennrich and Mark D. Schluchter. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42:805–820, 1986.

[26] M. Kendall and A. Stuart. *Inference and relationship*, volume 2 of *The advanced theory of statistics*. Charles Griffin and company ltd., london, fourth edition, 1979.

[27] M. G. Kenward. The use of fitted higher-order polynomial coefficients as covariates in the analysis of growth curves. *Biometrics*, 41:19–28, 1985.

[28] Gary G. Koch, Janet D. Elashoff, and Ingrid A. Amara. Repeated measurements - design and analysis. In Samuel Kotz and Norman L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 8, pages 46–73. John Wiley and Sons Inc., 1982.

[29] Walter Kramer. Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. *Journal of the American Statistical Association*, 75(372):1005–1009, 1980.

[30] W. J. Krzanowski and F. H. C. Marriott. *Multivariate Analysis Part 2: Classification, Covariance Structures and Repeated Measurements*, volume 2 of *Kendall's Library of Statistics*. Arnold, London, 1995.

[31] Nan Laird. Longitudinal data analysis, 1997. Draft course notes presented by the author at the NSF/CBMS Longitudinal Data Analysis Conference at the University of Missouri-Columbia.

[32] Nan Laird, Nicholas Lange, and Daniel Stram. Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, 82(397):97–105, 1987.

[33] Nan M. Laird. Topics in likelihood-based methods for longitudinal data analysis. *Statistica Sinica*, 1:33–50, 1991.

[34] Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

[35] Nicholas Lange and Nan M. Laird. The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association*, 84(405):241–247, 1989.

[36] Jack C. Lee. Prediction and estimation of growth curves with special covariance structure. *Journal of the American Statistical Association*, 83(402):432–440, 1988.

[37] Martin L. Lesser, Nina E. Kohn, Barbara A. Napolitano, and Savita Pahwa. The FU-PLOT: A graphical method for visualising the timing of Follow-Up in longitudinal studies. *The American Statistician*, 49(2):139–143, 1995.

[38] J. K. Lindsey. *Models for Repeated Measurements*, volume 10 of *Oxford Statistical Science Series*. Oxford University Press Inc., New York, 1993.

[39] Mary J. Lindstrom and Douglas M. Bates. Newton-Raphson and EM algorithms for mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022, 1988.

[40] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*. Probability and Mathematical Statistics. Academic Press, London, 1979.

[41] Yutuka Matsuyama and Yasuo Ohashi. Mixed models for bivariate repeated measures data using Gibbs sampling. *Statistics in Medicine*, 16:1587–1601, 1997.

[42] J. N. S. Matthews. The analysis of data from crossover deigns: The efficiency of ordinary least squares. *Biometrics*, 46:689–696, 1990.

[43] Robert A. Mclean, William L. Sanders, and Walter W. Stroup. A unified approach to mixed linear models. *The American Statistician*, 45(1):54–64, 1991.

[44] Raymond H. Myers and Janet S. Milton. *A first course in the theory of linear statistical models*. The Duxbury advanced series in statistics and decision sciences. PWS-KENT Publishing Company, Boston, 1991.

[45] J. A. Nelder and R. Mead. A simplex method for function minimisation. *Computing Journal*, 7:303–313, 1965.

[46] Gottfried E. Noether. On a theorem of pitman. *Annals of Mathematical Statistics*, 26:64–68, 1955.

[47] M. Olshewski and M. Schumacher. Statistical analysis of quality of life in cancer clinical trials. *Statistics in Medicine*, 9:749–763, 1990.

[48] Richard W. Park. Efficient estimation of a system of regression equations when the disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62:500–509, 1967.

[49] Taesung Park and Robert F. Woolson. Generalised multivariate models for longitudinal data. *Communications in Statistics, Simulations*, 21(4):925–946, 1992.

[50] H. I. Patel. Analysis of repeated measures designs with changing covariates in clinical trials. *Biometrika*, 73(3):707–715, 1986.

[51] H. D. Patterson and R. Thompson. Recovery of interblock information when block sizes are unequal. *Biometrika*, 58:545–554, 1971.

[52] Richard F. Pothoff and S. N. Roy. A generalised multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51(3 and 4):313–326, 1964.

[53] C. R. Rao. Some problems involving linear hypothesis in multivariate analysis. *Biometrika*, 46:49–58, 1959.

[54] Greg Reinsel. Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*, 77(377):190–195, 1982.

[55] James Rochon. Analyzing bivariate repeated measures for discrete and continous outcome variables. *Biometrics*, 52:740–750, 1996.

[56] M. Schumacker, M. Olschewski, and G. Schulgen. Assessment of quality of life in clinical trials. *Statistics in Medicine*, 10:1915–1930, 1991.

[57] Shayle R. Searle. *Matrix algebra useful for statistics.* Wiley series in probability and mathematical statistics. John Wiley and Sons Inc., New York, 1982.

[58] Burton Singer. Longitudinal data analysis. In Samuel Kotz and Norman L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 5, pages 142–155. John Wiley and Sons Inc., 1982.

[59] Judith D. Singer. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4):323–355, 1998.

[60] J. P. Sy, J. M. G. Taylor, and W. G. Cumberland. A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics*, 53:542–555, 1997.

[61] P. K. Tandon. Applications of global statistics in analysing quality of life data. *Statistics in Medicine*, 9:819–827, 1990.

[62] C. W. Therrien and K. Fukunaga. Properties of separable covariance matrices and their associated Gaussian random processes. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 6(5):652–656, 1984.

[63] Roland D. Thomas. Univariate repeated measures techniques applied to multivariate data. *Psychometrika*, 48:451–464, 1983.

[64] Neil H. Timm. *Multivariate analysis with applications in education and psychology.* Wadsworth Publishing Company Inc., Belmont, California, 1975.

[65] Neil H. Timm. Multivariate analysis of repeated measurements. In P. R. Krishnaiah, editor, *Handbook of statistics, Analysis of variance*, volume 1, pages 41–87. North Holland publishing company, 1980.

[66] Kao-Tai Tsai and James A. Koziol. Score and Wald tests for the multivariate growth curve model with missing data and a patterned covariance matrix. *Communications in Statistics, Theory and Methods*, 22(2):311–317, 1993.

[67] V. G. S. Vasdekis. *An investigation of certain methods in the analysis of growth curves.* PhD thesis, University of Oxford, England, 1993. Unpublished Doctor of Philosophy thesis.

[68] A. P. Verbyla. Analysis of repeated measures with changing covariates. *Biometrika*, 75(1):172–174, 1988.

[69] A. P. Verbyla and B. R. Cullis. The analysis of multistratum and spatially correlated repeated measures data. *Biometrics*, 48:1015–1032, 1992.

[70] A. P. Verbyla and W. N. Venables. An extension of the growth curve model. *Biometrika*, 75(1):129–138, 1988.

[71] J. Verhees and T. J. Wansbeek. An multimode direct product model for covariance structure analysis. *Journal of Mathematical and Statistical Psychology*, 43:231–240, 1990.

[72] Edward F. Vonesh and Randy L. Carter. Efficient inference for random coefficient growth curve models with unbalanced data. *Biometrics*, 43:617–628, 1987.

[73] Ming C. Wang. On the analysis of multivariate repeated measures designs. *Communications in Statistics, Theory and Methods*, 12:1647–1659, 1983.

[74] James H. Ware. Growth curves. In Samuel Kotz and Norman L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 3, pages 539–542. John Wiley and Sons Inc., 1982.

[75] James H. Ware. Linear models for the analysis of longitudinal studies. *The American Statistician*, 39(2):95–101, 1985.

[76] Russ Wolfinger. A tutorial on mixed models. Technical report, SAS Institute Inc., SAS Campus Drive, Cary NC, 27513 USA, October 1992.

[77] Russ Wolfinger, Randy Tobias, and John Sall. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM Journal of Scientific Computing*, 15(6):1294–1310, 1994.

[78] Robert F. Woolson. Application of an efficiency criterion to the multivariate one-sample location problem. *Sankhya: The Indian Journal of Statistics, Series B*, 38:290–293, 1976.

[79] Robert F. Woolson and Pranab Kumar Sen. Asymptotic comparison of a class of multivariate multiparameter tests. *Communications in Statistics*, 3(9):813–828, 1974.

[80] Robert Francis Woolson. *Some alternative measures of asymptotic relative efficiency for the multiparameter testing problem with application to the growth curve problem.* PhD thesis, University of North Carolina, Chapel Hill, 1974. Doctor of Philosophy thesis.

[81] Xiang Zhang. *Multivariate longitudinal data analysis with a family of covariance matrices.* PhD thesis, University of California, Los Angeles, 1996. Unpublished Doctor of Philosophy thesis.

[82] A.H. Zwinderman. The measurement of change of quality of life in clinical trials. *Statistics in Medicine*, 9:931–942, 1990.

Appendix A

Chapter 4 Computer Programs

A.1 Computing the TARE and CARE for a growth curve design and compound symmetry structure for Ω

```
%macro chap4;

proc iml;

* Specify the within-subject design matrix X and find
  its transpose;

X = {1 -1,
     1  0,
     1  1};
Xp = t(X);

* Specify the matrix Q corresponding to a hypothesis of
  interest and find its transpose;

Q = {1 0 0 0 -1 0 0 0,
     0 1 0 0 0 -1 0 0,
     0 0 1 0 0 0 -1 0,
     0 0 0 1 0 0 0 -1};
Qp  = t(Q);

* Specify the between-subject design matrix Theta that
  depends on the total sample size I and the number of
  groups G and find its transpose;

gsize = &n*30;
tsize = gsize*2;
g1 = J(gsize, 1, 1);
g2 = J(gsize, 1, 1);
theta = block(g1,g2);
thetap = t(theta);

* Specify the identity matrix needed for OLS estimation
  whose dimension depends on the number of characteristics;
```

```
I2={1 0,
    0 1};

* Create V-C matrix based on the Kronecker product of a
  unstructured matrix delta and a compound symmetry matrix
  or first order autoregressive matrix omega;

gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
rhoc2 = rhoc*rtgamma;
diag = 1;
invrhot2 = 1 - rhot2;

* Defining Delta;

delta = J(2,2,0);
delta[1,1] = diag;
delta[1,2] = rhoc2;
delta[2,1] = rhoc2;
delta[2,2] = gamma;


* Defining Compound Symmetry Omega;

omega = J(3,3,0);
omega[1,1] = diag;
omega[1,2] = rhot;
omega[1,3] = rhot;
omega[2,1] = rhot;
omega[2,2] = diag;
omega[2,3] = rhot;
omega[3,1] = rhot;
omega[3,2] = rhot;
omega[3,3] = diag;

* Defining AR(1) Omega;

omega2 = J(3,3,0);
omega2[1,1] = diag;
omega2[1,2] = rhot;
omega2[1,3] = rhot2;
omega2[2,1] = rhot;
omega2[2,2] = diag;
omega2[2,3] = rhot;
omega2[3,1] = rhot2;
omega2[3,2] = rhot;
omega2[3,3] = diag;

* Quantities needed to compute TARE and CARE;

qua1 = thetap * theta;
qua1inv = inv(qua1);
qua3 = Xp * X;
```

```
qua3inv = inv(qua3);
qua4 = Xp * omega;
qua5 = qua4 * X;
qua6 = qua3inv * qua5 * qua3inv;

* Computing the TARE21 and CARE21;

nume = qua1inv @ I2 @ qua3inv;
nume2 = Q * nume * Qp;
nume3 = inv(nume2);
nume4 = det(nume2);
deno = qua1inv @ delta @ qua6;
deno2 = Q * deno * Qp;
deno3 = inv(deno2);
deno4 = det(deno2);
tare21 = trace(nume3) / trace(deno3);
care21 = (deno4 / nume4)##0.25;

* Additional Quantities needed to find TARE13 and CARE13;

qua7 = inv(omega);
qua8 = Xp * qua7 * X;
qua8inv  = inv(qua8);

* Computing the TARE13 and CARE13;

deno5 = qua1inv @delta @ qua8inv;
deno6 = Q * deno5 * Qp;
deno7 = inv(deno6);
deno8 = det(deno6);
tare13 = trace(deno3) / trace(deno7);
care13 = (deno8 / deno4)##0.25;

* Sending results to an external text file;

file 'chap4a.text';
put tare21 +1 care21 +1 tare13 +1 care13;

* Setting printing options and printing the parameters
  used and results obtained;

* options linesize=96 pagesize=54 nocenter nodate nonumber;
* title1;
* print tsize gamma rhoc rhot tare21 care21 tare13 care13;
* quit;

%mend chap4;

* Inputting prameters and results obtained into a SAS
  data set;

%macro accum;

data cscov2;
  infile 'chap4a.text';
  input tare21 care21 tare13 care13;
```

```
run;

data cscov3;
ssize = &n*60;
gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
run;

data cscov;
  set cscov3;
  set cscov2;
run;

proc append base=chap4a data=cscov;
run;

proc datasets nolist;
   delete cscov cscov2 cscov3;
run;
quit;

%mend accum;

%macro para;
%do n=1 %to 3;
  %do r=1 %to 5;
    %do p=0 %to 18;
      %do q=1 %to 9;
%chap4
%accum
    %end;
   %end;
 %end;
%end;
%mend para;

%para;
run;

* Saving results as a permanent SAS data set;

libname wam 'wambugu';
run;

proc datasets library=work;
   copy out=wam memtype=data;
   select chap4a;
run;
quit;

proc contents data=wam.chap4a position;
run;
```

A.2 Computing the TARE and CARE for a growth curve design and first-order autoregressive structure for Ω

```
%macro chap4;

proc iml;

* Specify the within-subject design matrix X and find its
  transpose;

X = {1 -1,
     1  0,
     1  1};
Xp = t(X);

* Specify the matrix Q corresponding to a hypothesis of
  interest and find its transpose;

Q = {1 0 0 0 -1 0 0 0,
     0 1 0 0 0 -1 0 0,
     0 0 1 0 0 0 -1 0,
     0 0 0 1 0 0 0 -1};
Qp  = t(Q);

* Specify the between-subject design matrix Theta that
  depends on the total sample size I and the number of
  groups G and find its transpose;

gsize = &n*30;
tsize = gsize*2;
g1 = J(gsize, 1, 1);
g2 = J(gsize, 1, 1);
theta = block(g1,g2);
thetap = t(theta);

* Specify the identity matrix needed for OLS estimation
  whose dimension depends on the number of characteristics;

I2={1 0,
    0 1};

* Create V-C matrix based on the Kronecker product of a
  unstructured matrix delta and a compound symmetry matrix
  or first order autoregressive matrix omega;

gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
rhoc2 = rhoc*rtgamma;
diag = 1;
invrhot2 = 1 - rhot2;

* Defining Delta;
```

```
delta = J(2,2,0);
delta[1,1] = diag;
delta[1,2] = rhoc2;
delta[2,1] = rhoc2;
delta[2,2] = gamma;


* Defining Compound Symmetry Omega;

omega = J(3,3,0);
omega[1,1] = diag;
omega[1,2] = rhot;
omega[1,3] = rhot;
omega[2,1] = rhot;
omega[2,2] = diag;
omega[2,3] = rhot;
omega[3,1] = rhot;
omega[3,2] = rhot;
omega[3,3] = diag;

* Defining AR(1) Omega;

omega2 = J(3,3,0);
omega2[1,1] = diag;
omega2[1,2] = rhot;
omega2[1,3] = rhot2;
omega2[2,1] = rhot;
omega2[2,2] = diag;
omega2[2,3] = rhot;
omega2[3,1] = rhot2;
omega2[3,2] = rhot;
omega2[3,3] = diag;

* Quantities needed to compute TARE and CARE;

qua1 = thetap * theta;
qua1inv = inv(qua1);
qua3 = Xp * X;
qua3inv = inv(qua3);
qua4 = Xp * omega2;
qua5 = qua4 * X;
qua6 = qua3inv * qua5 * qua3inv;

* Computing the TARE21 and CARE21;

nume = qua1inv @ I2 @ qua3inv;
nume2 = Q * nume * Qp;
nume3 = inv(nume2);
nume4 = det(nume2);
deno = qua1inv @ delta @ qua6;
deno2 = Q * deno * Qp;
deno3 = inv(deno2);
deno4 = det(deno2);
tare21 = trace(nume3) / trace(deno3);
care21 = (deno4 / nume4)##0.25;
```

```
* Additional Quantities needed to find TARE13 and CARE13;

qua7 = inv(omega2);
qua8 = Xp * qua7 * X;
qua8inv  = inv(qua8);

* Computing the TARE13 and CARE13;

deno5 = qua1inv @delta @ qua8inv;
deno6 = Q * deno5 * Qp;
deno7 = inv(deno6);
deno8 = det(deno6);
tare13 = trace(deno3) / trace(deno7);
care13 = (deno8 / deno4)##0.25;

* Sending results to an external text file;

file 'chap4b.text';
put tare21 +1 care21 +1 tare13 +1 care13;

* Setting printing options and printing the parameters
  used and results obtained;

* options linesize=96 pagesize=54 nocenter nodate nonumber;
* title1;
* print tsize gamma rhoc rhot tare21 care21 tare13 care13;
* quit;

%mend chap4;

* Inputting prameters and results obtained into a SAS data
  set;

%macro accum;

data arcov2;
  infile 'chap4b.text';
  input tare21 care21 tare13 care13;
run;

data arcov3;
ssize = &n*60;
gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
run;

data arcov;
  set arcov3;
  set arcov2;
run;

proc append base=chap4b data=arcov;
run;
```

```
proc datasets nolist;
   delete arcov arcov2 arcov3;
run;
quit;

%mend accum;

%macro para;
%do n=1 %to 3;
  %do r=1 %to 5;
    %do p=0 %to 18;
      %do q=1 %to 9;
%chap4
%accum
    %end;
   %end;
 %end;
%end;
%mend para;

%para;
run;

* Saving results as a permanent SAS data set;

libname wam 'wambugu';
run;

proc datasets library=work;
   copy out=wam memtype=data;
   select chap4b;
run;
quit;

proc contents data=wam.chap4b position;
run;
```

A.3 **Computing the** TARE and CARE for a repeated measures analysis of variance design and compound symmetry structure for Ω

```
%macro chap4;

proc iml;

* Specify the within-subject design matrix X and find its
  transpose;

X = {1 0 0,
     0 1 0,
     0 0 1};
Xp = t(X);

* Specify the matrix Q corresponding to a hypothesis of
 interest and find its transpose;

Q = {1 0 0 0 0 0 -1 0 0 0 0 0,
     0 1 0 0 0 0  0 -1 0 0 0 0,
     0 0 1 0 0 0  0 0 -1 0 0 0,
     0 0 0 1 0 0  0 0 0 -1 0 0,
     0 0 0 0 1 0  0 0 0 0 -1 0,
     0 0 0 0 0 1  0 0 0 0 0 -1};
Qp  = t(Q);

* Specify the between-subject design matrix Theta that
  depends on the total sample size I and the number of
  groups G and find its transpose;

gsize = &n*30;
tsize = gsize*2;
g1 = J(gsize, 1, 1);
g2 = J(gsize, 1, 1);
theta = block(g1,g2);
thetap = t(theta);

* Specify the identity matrix needed for OLS estimation
  whose dimension depends on the number of characteristics;

I2={1 0,
    0 1};

* Create V-C matrix based on the Kronecker product of a
  unstructured matrix delta and a compound symmetry matrix
  or first order autoregressive matrix omega;

gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
rhoc2 = rhoc*rtgamma;
```

```
diag = 1;
invrhot2 = 1 - rhot2;

* Defining Delta;

delta = J(2,2,0);
delta[1,1] = diag;
delta[1,2] = rhoc2;
delta[2,1] = rhoc2;
delta[2,2] = gamma;


* Defining Compound Symmetry Omega;

omega = J(3,3,0);
omega[1,1] = diag;
omega[1,2] = rhot;
omega[1,3] = rhot;
omega[2,1] = rhot;
omega[2,2] = diag;
omega[2,3] = rhot;
omega[3,1] = rhot;
omega[3,2] = rhot;
omega[3,3] = diag;

* Defining AR(1) Omega;

omega2 = J(3,3,0);
omega2[1,1] = diag;
omega2[1,2] = rhot;
omega2[1,3] = rhot2;
omega2[2,1] = rhot;
omega2[2,2] = diag;
omega2[2,3] = rhot;
omega2[3,1] = rhot2;
omega2[3,2] = rhot;
omega2[3,3] = diag;

* Quantities needed to compute TARE and CARE;

qua1 = thetap * theta;
qua1inv = inv(qua1);
qua3 = Xp * X;
qua3inv = inv(qua3);
qua4 = Xp * omega;
qua5 = qua4 * X;
qua6 = qua3inv * qua5 * qua3inv;

* Computing the TARE21 and CARE21;

nume = qua1inv @ I2 @ qua3inv;
nume2 = Q * nume * Qp;
nume3 = inv(nume2);
nume4 = det(nume2);
deno = qua1inv @ delta @ qua6;
deno2 = Q * deno * Qp;
```

```
deno3 = inv(deno2);
deno4 = det(deno2);
tare21 = trace(nume3) / trace(deno3);
care21 = (deno4 / nume4)##0.17;

* Additional Quantities needed to find TARE13 and CARE13;

qua7 = inv(omega);
qua8 = Xp * qua7 * X;
qua8inv  = inv(qua8);

* Computing the TARE13 and CARE13;

deno5 = qua1inv @delta @ qua8inv;
deno6 = Q * deno5 * Qp;
deno7 = inv(deno6);
deno8 = det(deno6);
tare13 = trace(deno3) / trace(deno7);
care13 = (deno8 / deno4)##0.17;

* Sending results to an external text file;

file 'chap4d.text';
put tare21 +1 care21 +1 tare13 +1 care13;

* Setting printing options and printing the parameters
  used and results obtained;

  options linesize=96 pagesize=54 nocenter nodate nonumber;
* title1;
* print tsize gamma rhoc rhot tare21 care21 tare13 care13;
* quit;

%mend chap4;

* Inputting prameters and results obtained into a SAS
  data set;

%macro accum;

data cscov2;
  infile 'chap4d.text';
  input tare21 care21 tare13 care13;
run;

data cscov3;
ssize = &n*60;
gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
run;

data cscov;
  set cscov3;
  set cscov2;
run;
```

```
proc append base=chap4d data=cscov;
run;

proc datasets nolist;
   delete cscov cscov2 cscov3;
run;
quit;

%mend accum;

%macro para;
%do n=1 %to 3;
  %do r=1 %to 5;
    %do p=0 %to 18;
      %do q=1 %to 9;
%chap4
%accum
    %end;
   %end;
 %end;
%end;
%mend para;

%para;
run;

* Saving results as a permanent SAS data set;

libname wam 'wambugu';
run;

proc datasets library=work;
   copy out=wam memtype=data;
   select chap4d;
run;
quit;

proc contents data=wam.chap4d position;
run;
```

A.4 **Computing the** TARE and CARE for a repeated measures analysis of variance design and first-order autoregressive structure for Ω

```
%macro chap4;

proc iml;

* Specify the within-subject design matrix X and find its
  transpose;

X = {1 0 0,
     0 1 0,
     0 0 1};
Xp = t(X);

* Specify the matrix Q corresponding to a hypothesis of
  interest and find its transpose;

Q = {1 0 0 0 0 0 -1 0 0 0 0 0,
     0 1 0 0 0 0  0 -1 0 0 0 0,
     0 0 1 0 0 0  0 0 -1 0 0 0,
     0 0 0 1 0 0  0 0 0 -1 0 0,
     0 0 0 0 1 0  0 0 0 0 -1 0,
     0 0 0 0 0 1  0 0 0 0 0 -1};
Qp  = t(Q);

* Specify the between-subject design matrix Theta that
  depends on the total sample size I and the number of
  groups G and find its transpose;

gsize = &n*30;
tsize = gsize*2;
g1 = J(gsize, 1, 1);
g2 = J(gsize, 1, 1);
theta = block(g1,g2);
thetap = t(theta);

* Specify the identity matrix needed for OLS estimation
  whose dimension depends on the number of characteristics;

I2={1 0,
    0 1};

* Create V-C matrix based on the Kronecker product of a
  unstructured matrix delta and a compound symmetry matrix
  or first order autoregressive matrix omega;

gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
```

```
rhoc2 = rhoc*rtgamma;
diag = 1;
invrhot2 = 1 - rhot2;

* Defining Delta;

delta = J(2,2,0);
delta[1,1] = diag;
delta[1,2] = rhoc2;
delta[2,1] = rhoc2;
delta[2,2] = gamma;


* Defining Compound Symmetry Omega;

omega = J(3,3,0);
omega[1,1] = diag;
omega[1,2] = rhot;
omega[1,3] = rhot;
omega[2,1] = rhot;
omega[2,2] = diag;
omega[2,3] = rhot;
omega[3,1] = rhot;
omega[3,2] = rhot;
omega[3,3] = diag;

* Defining AR(1) Omega;

omega2 = J(3,3,0);
omega2[1,1] = diag;
omega2[1,2] = rhot;
omega2[1,3] = rhot2;
omega2[2,1] = rhot;
omega2[2,2] = diag;
omega2[2,3] = rhot;
omega2[3,1] = rhot2;
omega2[3,2] = rhot;
omega2[3,3] = diag;

* Quantities needed to compute TARE and CARE;

qua1 = thetap * theta;
qua1inv = inv(qua1);
qua3 = Xp * X;
qua3inv = inv(qua3);
qua4 = Xp * omega2;
qua5 = qua4 * X;
qua6 = qua3inv * qua5 * qua3inv;

* Computing the TARE21 and CARE21;

nume = qua1inv @ I2 @ qua3inv;
nume2 = Q * nume * Qp;
nume3 = inv(nume2);
nume4 = det(nume2);
deno = qua1inv @ delta @ qua6;
```

```
deno2 = Q * deno * Qp;
deno3 = inv(deno2);
deno4 = det(deno2);
tare21 = trace(nume3) / trace(deno3);
care21 = (deno4 / nume4)##0.25;

* Additional Quantities needed to find TARE13 and CARE13;

qua7 = inv(omega2);
qua8 = Xp * qua7 * X;
qua8inv  = inv(qua8);

* Computing the TARE13 and CARE13;

deno5 = qua1inv @delta @ qua8inv;
deno6 = Q * deno5 * Qp;
deno7 = inv(deno6);
deno8 = det(deno6);
tare13 = trace(deno3) / trace(deno7);
care13 = (deno8 / deno4)##0.25;

* Sending results to an external text file;

file 'chap4e.text';
put tare21 +1 care21 +1 tare13 +1 care13;

* Setting printing options and printing the parameters
  used and results obtained;

  options linesize=96 pagesize=54 nocenter nodate nonumber;
* title1;
* print tsize gamma rhoc rhot tare21 care21 tare13 care13;
* quit;

%mend chap4;

* Inputting prameters and results obtained into a SAS
  data set;

%macro accum;

data arcov2;
  infile 'chap4e.text';
  input tare21 care21 tare13 care13;
run;

data arcov3;
ssize = &n*60;
gamma=&r*0.5;
rhoc = (&p - 9)*0.1;
rhot = &q*0.1;
run;

data arcov;
  set arcov3;
  set arcov2;
```

```
run;

proc append base=chap4e data=arcov;
run;

proc datasets nolist;
   delete arcov arcov2 arcov3;
run;
quit;

%mend accum;

%macro para;
%do n=1 %to 3;
  %do r=1 %to 5;
    %do p=0 %to 18;
      %do q=1 %to 9;
%chap4
%accum
    %end;
   %end;
 %end;
%end;
%mend para;

%para;
run;

* Saving results as a permanent SAS data set;

libname wam 'wambugu';
run;

proc datasets library=work;
   copy out=wam memtype=data;
   select chap4e;
run;
quit;

proc contents data=wam.chap4e position;
run;
```

Appendix B

Chapter 5 Computer Programs

B.1 Simulation program to compute TARE and CARE

```
* Referencing the library where simulation results will be
  saved as a permanent SAS data set;

libname syl 'chapter5';
run;

* Ensuring program runs as it would in interactive mode;

options nosyntaxcheck;

* Suppress printing of the PROC MIXED output;

%global _PRINT_;
%let _PRINT_ = OFF;

* The simulation program starts here;

%macro simulate;

%do j=1 %to 200; * j = the simulation
                   index;

* One rep of the simulation starts here*;

data sbp(replace=yes);

* Generate 6 independent standard normal random variables
  for two groups;

do i = 1 to 50; * where 50 is the desired sample size in
    each group,i is the subject index;

    x1 = rannor(647 + i + &j*2);
x2 = rannor(372 + i + &j*2);
x3 = rannor(425 + i + &j*2);
x4 = rannor(162 + i + &j*2);
x5 = rannor(528 + i + &j*2);
```

```
x6 = rannor(289 + i + &j*2);
x7 = rannor(467 + i + &j*2);
x8 = rannor(732 + i + &j*2);
x9 = rannor(245 + i + &j*2);
x10 = rannor(612 + i + &j*2);
x11 = rannor(258 + i + &j*2);
x12 = rannor(829 + i + &j*2);
output;
end;
run;

* Convert generated data set into matrices;

proc iml;
use sbp;
read all var {x1 x2 x3 x4 x5 x6} into x1;
read all var {x7 x8 x9 x10 x11 x12} into x2;
tx1 = t(x1);
tx2 = t(x2);

* Create V-C matrix based on the Kronecker product and
  obtain lower triangular choleski factorization;

rhoc = (&p - 3)*0.3;
rhot = &q*0.1;
gamma=&r*0.5;
w = (&p - 1)*9 + &q;
z = (w - 1)*4 + &r;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
rhoc2 = rhoc*rtgamma;
diag=1;
A = J(2,2,0);
A[1,1] = diag;
A[1,2] = rhoc2;
A[2,1] = rhoc2;
A[2,2] = gamma;
B = J(3,3,0);
B[1,1] = diag;
B[1,2] = rhot;
B[1,3] = rhot2;
B[2,1] = rhot;
B[2,2] = diag;
B[2,3] = rhot;
B[3,1] = rhot2;
B[3,2] = rhot;
B[3,3] = diag;
sigma = A @ B;
lsigma=t(root(sigma));

* Enter the mean vector for group 1;

meanv1 = {112,
          116,
          120,
          63.5,
```

```
         63,
         62.5};
mu1=repeat(meanv1,1,50);

* Enter the mean vector for group 2;

meanv2 = {113,
         118,
         123,
         59,
         62,
         65};
mu2=repeat(meanv2,1,50);

* Create 6 correlated variables on each individual in
  group 1 and save as a SAS data set;

y1=lsigma*tx1+mu1;
ty1=t(y1);
varnames='sbp1':'sbp6';
create sbpmod1 from ty1 [colname=varnames];
append from ty1;

* Create 6 correlated variables on each individual in
  group 2 and save as a SAS data set;

y2=lsigma*tx2+mu2;
ty2=t(y2);
varnames='sbp1':'sbp6';
create sbpmod2 from ty2 [colname=varnames];
append from ty2;
quit;

*Combine the two data sets;

data sbpmod;
  set sbpmod1 sbpmod2;
run;

* Create variable to use in transposing the created data;

data nested;
   do i = 1 to 100;
     output;
   end;
run;

* Merge the two data sets;

data all;
  merge nested sbpmod;
run;

* Transpose data into format required by PROC MIXED;

proc transpose data=all out=allt;
```

```
  by i;
  var sbp1 sbp2 sbp3 sbp4 sbp5 sbp6;
run;

* Create data set on I individuals based on C
  characteristics measured T times;

data mult;
    do person = 1 to 100;
       do chara = 1 to 2;
         do time = 1 to 3;
            output;
         end;
       end;
    end;
run;

* Create design matrix to fit a linear growth-curve model
  to the two response variables;

proc iml;
A1=J(50,1,1);
A2=J(50,1,1);
theta=block(A1,A2);
thetap=t(theta);
X = { 1 -1,
      1 0,
      1 1};
Xp=t(X);
I2={1 0,
    0 1};
D = theta@I2@X;
varnames='x1':'x8';
create des from D [colname=varnames];
append from D;
quit;

* Create the final data set to be used by the Mixed
  procedure;

data actual;
   merge allt mult des;
run;

* Model fitting and estimating parameters of interest;

/* Completely Unstructured V-C Matrix */
proc mixed data=actual;
  class person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8/ covb noint s;
  repeated / type=un subject=person r=1 ri=1;
  make 'SolutionF' out=seun&j;
  make 'R' out=cmun&j;
  make 'COVB' out=emun&j;
run;
quit;
```

```
/* Unstructured by AR(1) V-C Matrix */
proc mixed data=actual;
  class chara time person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8 / covb noint s;
  repeated chara time / type=un@ar(1) subject=person
      r=1 ri=1;
  make 'SolutionF' out=sear&j;
  make 'R' out=cmar&j;
  make 'COVB' out=emar&j;
run;
quit;

* Convert generated data sets into matrices and compute
  the likelihood ratio test statistic, chi-square
  statistic, p-value and the measures of asymptotic
  relative efficiency;

proc iml;
rhoc = (&p - 3)*0.3;
rhot = &q*0.1;
gamma=&r*0.5;
rhot2 = rhot**2;
rtgamma = sqrt(gamma);
rhoc2 = rhoc*rtgamma;
diag=1;
A = J(2,2,0);
A[1,1] = diag;
A[1,2] = rhoc2;
A[2,1] = rhoc2;
A[2,2] = gamma;
B = J(3,3,0);
B[1,1] = diag;
B[1,2] = rhot;
B[1,3] = rhot2;
B[2,1] = rhot;
B[2,2] = diag;
B[2,3] = rhot;
B[3,1] = rhot2;
B[3,2] = rhot;
B[3,3] = diag;
A1=J(50,1,1);
A2=J(50,1,1);
theta=block(A1,A2);
thetap=t(theta);
X = { 1 -1,
      1 0,
      1 1};
Xp=t(X);
use cmun&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into arbi&j;
num&j = det(arbi&j);
use cmar&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into kscm&j;
deno&j = det(kscm&j);
lr&j = J(1,1,0);
```

```
lr&j = (num&j/deno&j)##-50;
chi&j = 2#log(lr&j);
pval&j = 1 - probchi(chi&j,17);
Q = {1 0 0 0 -1 0 0 0,
     0 1 0 0 0 -1 0 0,
     0 0 1 0 0 0 -1 0,
     0 0 0 1 0 0 0 -1};
qprime = t(Q);

* Code to find TARE and CARE based on optimal
  variance-covariance matrix;

qua1 = thetap*theta;
qua1inv = inv(qua1);
qua4 = Xp * B;
qua5 = qua4 * X;
qua6 = inv(qua5);
qua7 = qua1inv @ A @ qua6;
real = Q * qua7 * qprime;
realinv = inv(real);
realt = trace(realinv);
realc = det(real);
use emun&j var{_col1 _col2 _col3 _col4 _col5 _col6
    _col7 _col8};
read all var _num_ into vh2&j;
use emar&j var{_col1 _col2 _col3 _col4 _col5 _col6
    _col7 _col8};
read all var _num_ into vh1&j;
qua2&j = Q*vh2&j*qprime;
qua1&j = Q*vh1&j*qprime;
tar2&j = inv(qua2&j);
tar1&j = inv(qua1&j);
tare&j = trace(tar2&j)/trace(tar1&j);
atar&j = trace(tar2&j)/realt;
car2&j = det(qua2&j);
car1&j = det(qua1&j);
care&j = (car1&j/car2&j)##0.25;
acar&j = (realc/car2&j)##0.25;
lrt&j = J(1,7,0);
lrt&j[1,1] = lr&j;
lrt&j[1,2] = chi&j;
lrt&j[1,3] = pval&j;
lrt&j[1,4] = tare&j;
lrt&j[1,5] = care&j;
lrt&j[1,6] = atar&j;
lrt&j[1,7] = acar&j;
lrtm&j = repeat(lrt&j,8,1);
para&j = J(8,1,0);
para&j[1,1] = 116;
para&j[2,1] = 4;
para&j[3,1] = 63;
para&j[4,1] = -0.5 ;
para&j[5,1] = 118;
para&j[6,1] = 5;
para&j[7,1] = 62;
para&j[8,1] = 3;
```

```
lmod&j = lrtm&j || para&j;
varnames = 'hypo1':'hypo8';
create unar&j from lmod&j [colname=varnames];
append from lmod&j;
quit;


* Merging the results from the different models.
  Note: The merged data sets will be empty if one of the
  data sets is empty as a result of the model failing
  to converge;

data std&j;
   merge seun&j(drop=_df_ _t_ _pt_ rename=(_est_=est_un)
       rename=(_se_=se_un))
         sear&j(drop=_df_ _t_ _pt_ rename=(_est_=est_ar)
       rename=(_se_=se_ar))
         unar&j(rename=(hypo1=lrtest) rename=(hypo2=chisq)
       rename=(hypo3=pvalue) rename=(hypo4=Tare)
       rename=(hypo5=Care) rename=(hypo6=atare)
       rename=(hypo7=acare) rename=(hypo8=trueval));
   attrib simu length=$8;
   simu="Sim &j ";
output;

proc datasets nolist force;
  delete sbp sbpmod1 sbpmod2 sbpmod nested all allt mult
      des actual sear&j seun&j unar&j;
  append base=std new=std&j;
run;
quit;

%end;
%mend simulate;

%macro combine;

data std;
   set std;
rhoc = (&p - 3)*0.3;
rhot = &q*0.1;
gamma=&r*0.5;
w = (&p - 1)*9 + &q;
z = (w - 1)*4 + &r;
run;

proc append base=syl.results2 data=std;
run;

proc datasets nolist;
   delete std;
run;

%mend combine;

%macro krone;
```

```
%do p=1 %to 5;
  %do q=1 %to 9;
    %do r=1 %to 4;
%simulate
%combine
   %end;
  %end;
%end;
%mend krone;

%krone;
run;
```

Appendix C

Chapter 6 Computer Programs

C.1 Computing the Kronecker product deviation index

```
* This program computes the umweighted least squares
  estimator under the assumption of a "factorial
  covariance structure" or kronecker product covariance
  structure for the within-subject covariance matrix
  as described in Verhees and Wansbeek (1990) when k=2;

* The program also computes the value of the criterion
  that is minimised in finding  the umweighted least
  squares estimator;

%macro chap6a;

proc iml;

* Specifying the covariance parameters;

rho1 = &p*0.1;
rho2 = &q*0.1;
rho3 = &r*0.1;
diag = 1;
rho1m = rho1**2;
rho2m = rho2**2;
rho3m = rho3**2;

* Specifying the matrices that make up the within-subject
  covariance matrix Sigmao;

A = J(3,3,0);
A[1,1] = diag;
A[1,2] = rho1;
A[1,3] = rho1m;
A[2,1] = rho1;
A[2,2] = diag;
A[2,3] = rho1;
A[3,1] = rho1m;
```

```
A[3,2] = rho1;
A[3,3] = diag;
Amod = A#4;

B = J(3,3,0);
B[1,1] = diag;
B[1,2] = rho3;
B[1,3] = rho3m;
B[2,1] = rho3;
B[2,2] = diag;
B[2,3] = rho3;
B[3,1] = rho3m;
B[3,2] = rho3;
B[3,3] = diag;
Bmod = B#2;

C = J(3,3,0);
C[1,1] = diag;
C[1,2] = rho2;
C[1,3] = rho2m;
C[2,1] = rho2;
C[2,2] = diag;
C[2,3] = rho2;
C[3,1] = rho2m;
C[3,2] = rho2;
C[3,3] = diag;
Cmod = C#4;

* Specifying Sigmao;

AB = Amod//Bmod;
BC = Bmod//Cmod;
Sigmao = AB||BC;

* Computing the umweighted least squares estimator
  following Verhees and Wansbeek (1990) and finding
  the value of the criterion in the optimum;

C1 = {1 0 0 0 0 0,
      0 0 0 1 0 0,
 0 1 0 0 0 0,
 0 0 0 0 1 0,
 0 0 1 0 0 0,
 0 0 0 0 0 1};
S = Sigmao;
S1 = C1 * S * t(C1);
a1 = trace(S1**2);
* print a1;
S11 = shape(S1[1:2, 1:2],4,1);
S12 = shape(S1[3:4, 1:2],4,1);
S13 = shape(S1[5:6, 1:2],4,1);
S14 = shape(S1[1:2, 3:4],4,1);
S15 = shape(S1[3:4, 3:4],4,1);
S16 = shape(S1[5:6, 3:4],4,1);
S17 = shape(S1[1:2, 5:6],4,1);
S18 = shape(S1[3:4, 5:6],4,1);
```

```
S19 = shape(S1[5:6, 5:6],4,1);
* print S1;
* print S11 S12 S13 S14 S15 S16 S17 S18 S19;
S1tilde = S11||S12||S13||S14||S15||S16||S17||S18||S19;
* print S1tilde;
realmat1 = S1tilde * t(S1tilde);
eval1 = eigval(realmat1);
evec1 = eigvec(realmat1);
* print eval1;
* print evec1;
lambdah = eval1[1,1];
* print lambdah;
S2 = S;
a2 = trace(S2**2);
* print a2;
S21 = shape(S2[1:3, 1:3],9,1);
S22 = shape(S2[4:6, 1:3],9,1);
S23 = shape(S2[1:3, 4:6],9,1);
S24 = shape(S2[4:6, 4:6],9,1);
* print S2;
* print S21 S22 S23 S24;
S2tilde = S21||S22||S23||S24;
* print S2tilde;
realmat2 = S2tilde * t(S2tilde);
eval2 = eigval(realmat2);
evec2 = eigvec(realmat2);
* print eval2;
* print evec2;
call svd(u1,q1,v1,S1tilde);
* print u1 q1 v1;
delta = u1[1:4,1];
* print delta;
call svd(u2,q2,v2,S2tilde);
* print u2 q2 v2;
omega = u2[1:9,1];
* print omega;
criterio = a1 - lambdah;
*print criterio;

* Sending results to an external file;

file 'chap63a.text';
put criterio;

%mend chap6a;

* Inputting parameters and results obtained into a SAS
  data set;

%macro accum;

data verhees2;
  infile 'chap63a.text';
  input criterio;
  attrib criterio format=6.3;
run;
```

```
data verhees3;
  rho1 = &p*0.1;
  rho2 = &q*0.1;
  rho3 = &r*0.1;
run;

data verhees;
   set verhees3;
   set verhees2;
run;

proc append base=chap63a data=verhees;
run;

proc datasets nolist;
   delete verhees verhees2 verhees3;
run;

quit;

%mend accum;

%macro para;
%do p=1 %to 9;
  %do q=1 %to 9;
    %do r=1 %to 9;
%chap6a
%accum
    %end;
  %end;
%end;
%mend para;

%para;
run;

* Saving results as a permanent SAS data set;

libname kpd 'chapter6';
run;

proc datasets library=work;
   copy out=kpd memtype=data;
   select chap63a;
run;
quit;

proc contents data=kpd.chap63a position;
run;

proc print data=kpd.chap63a;
run;
```

C.2 Simulation program to evaluate the impact of the Kronecker product deviation index under the null hypothesis

```
* Referencing the library where simulation results will be
   saved as a permanent SAS data set;

libname kpd 'chapter6';
run;

* Ensuring program runs in batch mode as it would in
  interactive mode;

options nosyntaxcheck nosource;

* Suppress printing of the PROC MIXED output;

%global _PRINT_;
%let _PRINT_ = OFF;

* The simulation program starts here;

%macro simulate;

%do j=1 %to 200;

* j = the simulation index;

* One rep of the simulation starts here;

data sbp (replace=yes);

* Generate 6 independent standard normal random variables
  for two groups;

do i = 1 to 50; * where 50 is the desired sample size in each
                  group,
                  i is the subject index;
x1 = rannor(647 + i + &j*2);
x2 = rannor(372 + i + &j*2);
x3 = rannor(425 + i + &j*2);
x4 = rannor(162 + i + &j*2);
x5 = rannor(528 + i + &j*2);
x6 = rannor(289 + i + &j*2);
x7 = rannor(467 + i + &j*2);
x8 = rannor(732 + i + &j*2);
x9 = rannor(245 + i + &j*2);
x10 = rannor(612 + i + &j*2);
x11 = rannor(258 + i + &j*2);
x12 = rannor(829 + i + &j*2);
output;
end;
run;
```

```
* Convert generated data set into matrices;

proc iml;
use sbp;
read all var {x1 x2 x3 x4 x5 x6} into x1;
read all var {x7 x8 x9 x10 x11 x12} into x2;
tx1 = t(x1);
tx2 = t(x2);

* Create the within subject V-C matrix;

* Specifying the covariance parameters;

rho1 = &p*0.1;
rho2 = &q*0.1;
rho3 = &r*0.1;
diag = 1;
rho1m = rho1**2;
rho2m = rho2**2;
rho3m = rho3**2;

* Specifying the matrices that make up the within-subject
  covariance
matrix Sigmao;

A = J(3,3,0);
A[1,1] = diag;
A[1,2] = rho1;
A[1,3] = rho1m;
A[2,1] = rho1;
A[2,2] = diag;
A[2,3] = rho1;
A[3,1] = rho1m;
A[3,2] = rho1;
A[3,3] = diag;
Amod = A#4;

B = J(3,3,0);
B[1,1] = diag;
B[1,2] = rho3;
B[1,3] = rho3m;
B[2,1] = rho3;
B[2,2] = diag;
B[2,3] = rho3;
B[3,1] = rho3m;
B[3,2] = rho3;
B[3,3] = diag;
Bmod = B#2;

C = J(3,3,0);
C[1,1] = diag;
C[1,2] = rho2;
C[1,3] = rho2m;
C[2,1] = rho2;
C[2,2] = diag;
C[2,3] = rho2;
```

```
C[3,1] = rho2m;
C[3,2] = rho2;
C[3,3] = diag;
Cmod = C#4;

* Specifying Sigmao (lsigma in this program);

AB = Amod//Bmod;
BC = Bmod//Cmod;
lsigma = AB||BC;

* Enter the mean vector for group 1;

meanv1 = {113,
          118,
          123,
          59,
          62,
          65};
mu1=repeat(meanv1,1,50);

* Enter the mean vector for group 2;

meanv2 = {113,
          118,
          123,
          59,
          62,
          65};
mu2=repeat(meanv2,1,50);

* Create 6 correlated variables on each individual in group 1
  and save as a SAS data set;

y1=lsigma*tx1+mu1;
ty1=t(y1);
varnames='sbp1':'sbp6';
create sbpmod1 from ty1 [colname=varnames];
append from ty1;

* Create 6 correlated variables on each individual in group 2
  and save as a SAS data set;

y2=lsigma*tx2+mu2;
ty2=t(y2);
varnames='sbp1':'sbp6';
create sbpmod2 from ty2 [colname=varnames];
append from ty2;
quit;

*Combine the two data sets;

data sbpmod;
  set sbpmod1 sbpmod2;
run;
```

```
* Create variable to use in transposing the created data;

data nested;
   do i = 1 to 100;
     output;
   end;
run;

* Merge the two data sets;

data all;
  merge nested sbpmod;
run;

* Transpose data into format required by PROC MIXED;

proc transpose data=all out=allt;
  by i;
  var sbp1 sbp2 sbp3 sbp4 sbp5 sbp6;
run;

* Create data set on I individuals based on C characteristics
  measured T times;

data mult;
    do person = 1 to 100;
       do chara = 1 to 2;
         do time = 1 to 3;
            output;
         end;
       end;
    end;
run;

* Create design matrix to fit a linear growth-curve model to the
  two response variables;

proc iml;
A1=J(50,1,1);
A2=J(50,1,1);
theta=block(A1,A2);
thetap=t(theta);
X = { 1 -1,
      1 0,
      1 1};
Xp=t(X);
I2={1 0,
    0 1};
D = theta@I2@X;
varnames='x1':'x8';
create des from D [colname=varnames];
append from D;
quit;

* Create the final data set to be used by the Mixed procedure;
```

```
data actual;
   merge allt mult des;
run;

* Model fitting and estimating parameters of interest;

/* Completely Unstructured V-C Matrix */
proc mixed data=actual;
  class person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8/ covb noint s;
  repeated / type=un subject=person r=1 ri=1;
  make 'SolutionF' out=seun&j;
  make 'R' out=cmun&j;
  make 'COVB' out=emun&j;
run;
quit;

/* Unstructured by AR(1) V-C Matrix */
proc mixed data=actual;
  class chara time person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8 / covb noint s;
  repeated chara time / type=un@ar(1) subject=person r=1 ri=1;
  make 'SolutionF' out=sear&j;
  make 'R' out=cmar&j;
  make 'COVB' out=emar&j;
run;
quit;


* Convert generated data sets into matrices and compute the
  likelihood ratio test statistic, chi-square statistic and
  p-value;

proc iml;
use cmun&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into arbi&j;
num&j = det(arbi&j);
use cmar&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into kscm&j;
deno&j = det(kscm&j);
lr&j = J(1,1,0);
lr&j = (num&j/deno&j)##-50;
chi&j = 2#log(lr&j);
pval&j = 1 - probchi(chi&j,17);
Q = {1 0 0 0 -1 0 0 0,
     0 1 0 0 0 -1 0 0,
     0 0 1 0 0 0 -1 0,
     0 0 0 1 0 0 0 -1};
qprime = t(Q);

* Computing the test statistics abd p-values for testing the
  hypothesis of overall parallelism;;

use emun&j var{_col1 _col2 _col3 _col4 _col5 _col6 _col7 _col8};
read all var _num_ into vh2&j;
use emar&j var{_col1 _col2 _col3 _col4 _col5 _col6 _col7 _col8};
```

```
read all var _num_ into vh1&j;
qua2&j = Q*vh2&j*qprime;
qua1&j = Q*vh1&j*qprime;

use seun&j var{_est_};
read all var _num_ into lam2&j;
use sear&j var{_est_};
read all var _num_ into lam1&j;
qua3&j = Q*lam1&j;
qua4&j = Q*lam2&j;
qua5&j = t(qua3&j);
qua6&j = t(qua4&j);
tar&j = qua5&j*inv(qua1&j)*qua3&j;
tun&j = qua6&j*inv(qua2&j)*qua4&j;
pvar&j = 1 - probchi(tar&j, 4);
pvun&j = 1 - probchi(tun&j, 4);

lrt&j = J(1,7,0);
lrt&j[1,1] = lr&j;
lrt&j[1,2] = chi&j;
lrt&j[1,3] = pval&j;
lrt&j[1,4] = tar&j;
lrt&j[1,5] = tun&j;
lrt&j[1,6] = pvar&j;
lrt&j[1,7] = pvun&j;
lrtm&j = repeat(lrt&j,8,1);
para&j = J(8,1,0);
para&j[1,1] = 118;
para&j[2,1] = 5;
para&j[3,1] = 62;
para&j[4,1] = 3 ;
para&j[5,1] = 118;
para&j[6,1] = 5;
para&j[7,1] = 62;
para&j[8,1] = 3;
lmod&j = lrtm&j || para&j;
varnames = 'hypo1':'hypo8';
create unar&j from lmod&j [colname=varnames];
append from lmod&j;
quit;

* Merging the results from the different models.
  Note: The merged data sets will be empty if one of the data
  sets is empty as a result of the model failing to converge;

data std&j;
  length _effect_ $8 est_un se_un est_ar se_ar lrtest chisq
      pvalue T_AR T_UN PVA_AR PVA_UN trueval 8 simu $8;
run;

data std&j;
   set std&j(obs=0);
run;

data std&j;
    attrib _effect_ length=$8;
```

```
    merge seun&j(drop=_df_ _t_ _pt_ rename=(_est_=est_un)
        rename=(_se_=se_un))
         sear&j(drop=_df_ _t_ _pt_ rename=(_est_=est_ar)
        rename=(_se_=se_ar))
         unar&j(rename=(hypo1=lrtest) rename=(hypo2=chisq)
         rename=(hypo3=pvalue) rename=(hypo4=T_AR)
        rename=(hypo5=T_UN)
         rename=(hypo6=PVA_AR) rename=(hypo7=PVA_UN)
        rename=(hypo8=trueval));
   attrib simu length=$8;
   simu="Sim &j ";
output;

proc datasets nolist;
  delete sbp sbpmod1 sbpmod2 sbpmod nested all allt mult des
      actual sear&j
  seun&j unar&j;
  append base=std new=std&j;
run;
quit;

%end;
%mend simulate;

%macro combine;

data std;
   set std;
rho1 = &p*0.1;
rho2 = &q*0.1;
rho3 = &r*0.1;
run;

proc append base=kpd.dec18 data=std(where=(_effect_='X1'));
run;

proc datasets nolist;
   delete std;
run;

%mend combine;

%macro krone;
%do p=1 %to 9 %by 2;
  %do q=1 %to 9 %by 2;
    %do r=2 %to 8 %by 2;
%simulate
%combine
   %end;
  %end;
%end;
%mend krone;

%krone;
run;
```

C.3 Simulation program to evaluate the impact of the Kronecker product deviation index under the alternative hypothesis

```
* Referencing the library where simulation results will be
   saved as a permanent SAS data set;

libname kpd 'chapter6';
run;

* Ensuring program runs in batch mode as it would in
  interactive mode;

options nosyntaxcheck nosource;

* Suppress printing of the PROC MIXED output;

%global _PRINT_;
%let _PRINT_ = OFF;

* The simulation program starts here;

%macro simulate;

%do j=1 %to 200;

* j = the simulation index;

* One rep of the simulation starts here;

data sbp (replace=yes);

* Generate 6 independent standard normal random variables for
  two groups;

do i = 1 to 50; * where 50 is the desired sample size in each
                  group,
                  i is the subject index;
x1 = rannor(647 + i + &j*2);
x2 = rannor(372 + i + &j*2);
x3 = rannor(425 + i + &j*2);
x4 = rannor(162 + i + &j*2);
x5 = rannor(528 + i + &j*2);
x6 = rannor(289 + i + &j*2);
x7 = rannor(467 + i + &j*2);
x8 = rannor(732 + i + &j*2);
x9 = rannor(245 + i + &j*2);
x10 = rannor(612 + i + &j*2);
x11 = rannor(258 + i + &j*2);
x12 = rannor(829 + i + &j*2);
output;
end;
run;
```

```
* Convert generated data set into matrices;
proc iml;
use sbp;
read all var {x1 x2 x3 x4 x5 x6} into x1;
read all var {x7 x8 x9 x10 x11 x12} into x2;
tx1 = t(x1);
tx2 = t(x2);

* Create the within subject V-C matrix;

* Specifying the covariance parameters;

rho1 = &p*0.1;
rho2 = &q*0.1;
rho3 = &r*0.1;
diag = 1;
rho1m = rho1**2;
rho2m = rho2**2;
rho3m = rho3**2;

* Specifying the matrices that make up the within-subject
  covariance
matrix Sigmao;

A = J(3,3,0);
A[1,1] = diag;
A[1,2] = rho1;
A[1,3] = rho1m;
A[2,1] = rho1;
A[2,2] = diag;
A[2,3] = rho1;
A[3,1] = rho1m;
A[3,2] = rho1;
A[3,3] = diag;
Amod = A#4;

B = J(3,3,0);
B[1,1] = diag;
B[1,2] = rho3;
B[1,3] = rho3m;
B[2,1] = rho3;
B[2,2] = diag;
B[2,3] = rho3;
B[3,1] = rho3m;
B[3,2] = rho3;
B[3,3] = diag;
Bmod = B#2;

C = J(3,3,0);
C[1,1] = diag;
C[1,2] = rho2;
C[1,3] = rho2m;
C[2,1] = rho2;
C[2,2] = diag;
C[2,3] = rho2;
C[3,1] = rho2m;
```

```
C[3,2] = rho2;
C[3,3] = diag;
Cmod = C#4;

* Specifying Sigmao (lsigma in this program);

AB = Amod//Bmod;
BC = Bmod//Cmod;
lsigma = AB||BC;

* Enter the mean vector for group 1;

meanv1 = {113,
          118,
          123,
          59,
          62,
          65};
mu1=repeat(meanv1,1,50);

* Enter the mean vector for group 2;

meanv2 = {112,
          118,
          124,
          58.4,
          62,
          65.6};
mu2=repeat(meanv2,1,50);

* Create 6 correlated variables on each individual in group 1
  and save as a SAS data set;

y1=lsigma*tx1+mu1;
ty1=t(y1);
varnames='sbp1':'sbp6';
create sbpmod1 from ty1 [colname=varnames];
append from ty1;

* Create 6 correlated variables on each individual in group 2
  and save as a SAS data set;

y2=lsigma*tx2+mu2;
ty2=t(y2);
varnames='sbp1':'sbp6';
create sbpmod2 from ty2 [colname=varnames];
append from ty2;
quit;

*Combine the two data sets;

data sbpmod;
  set sbpmod1 sbpmod2;
run;

* Create variable to use in transposing the created data;
```

```
data nested;
   do i = 1 to 100;
     output;
   end;
run;

* Merge the two data sets;

data all;
  merge nested sbpmod;
run;

* Transpose data into format required by PROC MIXED;

proc transpose data=all out=allt;
  by i;
  var sbp1 sbp2 sbp3 sbp4 sbp5 sbp6;
run;

* Create data set on I individuals based on C characteristics
  measured T times;

data mult;
    do person = 1 to 100;
       do chara = 1 to 2;
         do time = 1 to 3;
            output;
         end;
       end;
    end;
run;

* Create design matrix to fit a linear growth-curve model to the
  two response variables;

proc iml;
A1=J(50,1,1);
A2=J(50,1,1);
theta=block(A1,A2);
thetap=t(theta);
X = { 1 -1,
      1 0,
      1 1};
Xp=t(X);
I2={1 0,
    0 1};
D = theta@I2@X;
varnames='x1':'x8';
create des from D [colname=varnames];
append from D;
quit;

* Create the final data set to be used by the Mixed procedure;

data actual;
```

```
   merge allt mult des;
run;

* Model fitting and estimating parameters of interest;

/* Completely Unstructured V-C Matrix */
proc mixed data=actual;
  class person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8/ covb noint s;
  repeated / type=un subject=person r=1 ri=1;
  make 'SolutionF' out=seun&j;
  make 'R' out=cmun&j;
  make 'COVB' out=emun&j;
run;
quit;

/* Unstructured by AR(1) V-C Matrix */
proc mixed data=actual;
  class chara time person;
  model col1 = x1 x2 x3 x4 x5 x6 x7 x8 / covb noint s;
  repeated chara time / type=un@ar(1) subject=person r=1 ri=1;
  make 'SolutionF' out=sear&j;
  make 'R' out=cmar&j;
  make 'COVB' out=emar&j;
run;
quit;

* Convert generated data sets into matrices and compute the
  likelihood ratio test statistic, chi-square statistic and
  p-value;

proc iml;
use cmun&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into arbi&j;
num&j = det(arbi&j);
use cmar&j var{col1 col2 col3 col4 col5 col6};
read all var _num_ into kscm&j;
deno&j = det(kscm&j);
lr&j = J(1,1,0);
lr&j = (num&j/deno&j)##-50;
chi&j = 2#log(lr&j);
pval&j = 1 - probchi(chi&j,17);
Q = {1 0 0 0 -1 0 0 0,
     0 1 0 0 0 -1 0 0,
     0 0 1 0 0 0 -1 0,
     0 0 0 1 0 0 0 -1};
qprime = t(Q);

* Computing the test statistics and p-values for testing the
  hypothesis of overall parallelism;;

use emun&j var{_col1 _col2 _col3 _col4 _col5 _col6 _col7 _col8};
read all var _num_ into vh2&j;
use emar&j var{_col1 _col2 _col3 _col4 _col5 _col6 _col7 _col8};
read all var _num_ into vh1&j;
qua2&j = Q*vh2&j*qprime;
```

```
qua1&j = Q*vh1&j*qprime;

use seun&j var{_est_};
read all var _num_ into lam2&j;
use sear&j var{_est_};
read all var _num_ into lam1&j;
qua3&j = Q*lam1&j;
qua4&j = Q*lam2&j;
qua5&j = t(qua3&j);
qua6&j = t(qua4&j);
tar&j = qua5&j*inv(qua1&j)*qua3&j;
tun&j = qua6&j*inv(qua2&j)*qua4&j;
pvar&j = 1 - probchi(tar&j, 4);
pvun&j = 1 - probchi(tun&j, 4);

lrt&j = J(1,7,0);
lrt&j[1,1] = lr&j;
lrt&j[1,2] = chi&j;
lrt&j[1,3] = pval&j;
lrt&j[1,4] = tar&j;
lrt&j[1,5] = tun&j;
lrt&j[1,6] = pvar&j;
lrt&j[1,7] = pvun&j;
lrtm&j = repeat(lrt&j,8,1);
para&j = J(8,1,0);
para&j[1,1] = 118;
para&j[2,1] = 5;
para&j[3,1] = 62;
para&j[4,1] = 3 ;
para&j[5,1] = 118;
para&j[6,1] = 6;
para&j[7,1] = 62;
para&j[8,1] = 3.6;
lmod&j = lrtm&j || para&j;
varnames = 'hypo1':'hypo8';
create unar&j from lmod&j [colname=varnames];
append from lmod&j;
quit;

* Merging the results from the different models.
  Note: The merged data sets will be empty if one of the data
  sets is empty as a result of the model failing to converge;

data std&j;
  length _effect_ $8 est_un se_un est_ar se_ar lrtest chisq
      pvalue T_AR T_UN PVA_AR PVA_UN trueval 8 simu $8;
run;

data std&j;
   set std&j(obs=0);
run;

data std&j;
    attrib _effect_ length=$8;
    merge seun&j(drop=_df_ _t_ _pt_ rename=(_est_=est_un)
        rename=(_se_=se_un))
```

```
          sear&j(drop=_df_ _t_ _pt_ rename=(_est_=est_ar)
         rename=(_se_=se_ar))
          unar&j(rename=(hypo1=lrtest) rename=(hypo2=chisq)
          rename=(hypo3=pvalue) rename=(hypo4=T_AR)
         rename=(hypo5=T_UN)
          rename=(hypo6=PVA_AR) rename=(hypo7=PVA_UN)
         rename=(hypo8=trueval));
    attrib simu length=$8;
    simu="Sim &j ";
output;

proc datasets nolist;
  delete sbp sbpmod1 sbpmod2 sbpmod nested all allt mult des
      actual sear&j seun&j unar&j;
  append base=std new=std&j;
run;
quit;

%end;
%mend simulate;

%macro combine;

data std;
   set std;
rho1 = &p*0.1;
rho2 = &q*0.1;
rho3 = &r*0.1;
run;

proc append base=kpd.jan05 data=std(where=(_effect_='X1'));
run;

proc datasets nolist;
   delete std;
run;

%mend combine;

%macro krone;
%do p=1 %to 9 %by 2;
  %do q=1 %to 9 %by 2;
    %do r=2 %to 8 %by 2;
%simulate
%combine
   %end;
  %end;
%end;
%mend krone;

%krone;
run;
```