Inference for Generalized Multivariate Analysis of Variance (GMANOVA) Models and High-dimensional Extensions

INFERENCE FOR GENERALIZED MULTIVARIATE ANALYSIS OF VARIANCE (GMANOVA) MODELS AND HIGH-DIMENSIONAL EXTENSIONS

BY

SAYANTEE JANA, M.Sc., (Statistics)
McMaster University, Hamilton, Canada

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2017)	McMaster University
(Mathematics & Statistics)	Hamilton, Ontario, Canada

TITLE:	Inference for Generalized Multivariate Analysis of Variance (GMANOVA) Models and High-dimensional Extensions
AUTHOR:	Sayantee Jana M.Sc. (Statistics) University of Calcutta, Kolkata, India M.Sc., (Statistics) McMaster University, Hamilton, Canada
SUPERVISORS:	Dr. JEMILA S. HAMID Prof. NARAYANSWAMY BALAKRISHNAN
NUMBER OF PAGES:	xvii, 170

Abstract

A Growth Curve Model (GCM) is a multivariate linear model used for analyzing longitudinal data with short to moderate time series. It is a special case of Generalized Multivariate Analysis of Variance (GMANOVA) models. Analysis using the GCM involves comparison of mean growths among different groups. The classical GCM, however, possesses some limitations including distributional assumptions, assumption of identical degree of polynomials for all groups and it requires larger sample size than the number of time points. In this thesis, we relax some of the assumptions of the traditional GCM and develop appropriate inferential tools for its analysis, with the aim of reducing bias, improving precision and to gain increased power as well as overcome limitations of high-dimensionality.

Existing methods for estimating the parameters of the GCM assume that the underlying distribution for the error terms is multivariate normal. In practical problems, however, we often come across skewed data and hence estimation techniques developed under the normality assumption may not be optimal. Simulation studies conducted in this thesis, in fact, show that existing methods are sensitive to the presence of skewness in the data, where estimators are associated with increased bias and mean square error (MSE), when the normality assumption is violated. Methods appropriate for skewed distributions are, therefore, required. In this thesis, we relax the distributional assumption of the GCM and provide estimators for the mean and covariance matrices of the GCM under multivariate skew normal (MSN) distribution. An estimator for the additional skewness parameter of the MSN distribution is also provided. The estimators are derived using the expectation maximization (EM) algorithm and extensive simulations are performed to examine the performance of the estimators. Comparisons with existing estimators show that our estimators perform better than existing estimators, when the underlying distribution is multivariate skew normal. Illustration using real data set is also provided, wherein Triglyceride levels from the Framingham Heart Study is modelled over time.

The GCM assumes equal degree of polynomial for each group. Therefore, when groups means follow different shapes of polynomials, the GCM fails to accommodate this difference in one model. We consider an extension of the GCM, wherein mean responses from different groups can have different shapes, represented by polynomials of different degree. Such a model is referred to as Extended Growth Curve Model (EGCM). We extend our work on GCM to EGCM, and develop estimators for the mean and covariance matrices under MSN errors. We adopted the Restricted Expectation Maximization (REM) algorithm, which is based on the multivariate Newton-Raphson (NR) method and Lagrangian optimization. However, the multivariate NR method and hence, the existing REM algorithm are applicable to vector parameters and the parameters of interest in this study are matrices. We, therefore, extended the NR approach to matrix parameters, which consequently allowed us to extend the REM algorithm to matrix parameters. The performance of the proposed estimators were examined using extensive simulations and a motivating real data example was provided to illustrate the application of the proposed estimators.

Finally, this thesis deals with high-dimensional application of GCM. Existing methods for a GCM are developed under the assumption of small p large n ($n >> p$) and are not appropriate for analyzing high-dimensional longitudinal data, due to singularity of the sample covariance matrix. In a previous work, we used Moore-Penrose generalized inverse to overcome this challenge. However, the method has some limitations around near singularity, when $p \approx n$. In this thesis, a Bayesian framework was used to derive a test for testing the linear hypothesis on the mean parameter of the GCM, which is applicable in high-dimensional situations. Extensive simulations are performed to investigate the performance of the test statistic and establish optimality characteristics. Results show that this test performs well, under different conditions, including the near singularity zone. Sensitivity of the test to mis-specification of the parameters of the prior distribution are also examined empirically. A numerical example is provided to illustrate the usefulness of the proposed method in practical situations.

Keywords: Growth Curve Model (GCM), GMANOVA models, Bayesian methods, High-dimensional data, Longitudinal analysis, Multivariate Skew Normal distribution, Extended Growth Curve Model (EGCM), EM algorithm, Restricted EM algorithm, Matrix Newton Raphson method

Publications

1. Jana S., Balakrishnan N. and Hamid J. "Estimation of the parameters of the Growth Curve Model under Multivariate Skew Normal distribution". Submitted to *Sankhya: The Indian Journal of Statistics, Series B.* (Chapter 3)

2. Jana S., Balakrishnan N. and Hamid J. "Estimation of the parameters of the Extended Growth Curve Model under Multivariate Skew Normal distribution". Submitted to *Journal of Multivariate Analysis.* (Chapter 4)

3. Jana S., Balakrishnan N. and Hamid J. "Bayesian Growth Curve Model useful for high-dimensional longitudinal data". Submitted to *Journal of Applied Statistics.* (Chapter 5)

Acknowledgements

I would like to thank my supervisor, Dr. J. Hamid for generously supporting me during my PhD and providing ideas for the thesis. I am really grateful for the regular meetings, long discussions on the work and for the immense critical improvement of my writing. I am grateful to my co-supervisor, Prof. N. Balakrishnan for providing deep insights and suggestions to the thesis, comments on the write up and for supporting me.

I thank Prof. J. Beyene for being in my PhD supervisory committee and providing feedback and comments on the thesis and for providing me TA support. I express my gratitude to the external examiner, Dr. Martin Singull and the internal-external examiner, Dr. Gregory Pond for their comments, feedback on the thesis and their insightful questions. My deepest gratitude to Prof. Rajat Bhaduri for agreeing to chair my defence on such a short notice.

I would like to thank the Department of Mathematics and Statistics and School of Graduate Studies, McMaster University for providing TA support and scholarships and bursaries for my PhD studies.

Special thanks to Hamilton Street Railway (HSR) for their best service in commuting as I travelled the length and breadth of the city of Hamilton. I would like to extend my thanks to McMaster Outdoor Club and Arrive and Thrive group for giving me some of the best memories in McMaster. Special thanks to Cree, Jasmine, Krupesh, Mike, Genn, Jackie, Connor, Megan and Rachel for arranging those trips.

I acknowledge the help I received from School of Graduate studies from their seminars and for the fabulous Thesis Writing Boot Camps. Special cheers to Andrea and the writing consultants for making them happen. I would also like to mention Graduate Study Room where I wrote most of the thesis.

I also express my gratitude to the administrative staff of the Dept. of Mathematics and Statistics and special thanks to Kenneth Moyle for helping me with computing support whenever I needed.

I thank my gurudev Swami Kripamayanandaji and Swami Satyabodhanandaji for inspring me and motivating me. Special thanks to Swami Sarvapriyanandaji and his lectures. I thank Prof. Eileen Schuller, Anand Bhaiya, Finnick, Tao, TJ, Joycelene, Regina, Thuva, Zelalem, Caitlyn, Tian, Antara di for their time and support in various ways. Special thanks to Benu kaku and Kakima for their sumptous meals.

Finally I thank the support from my family, especially my parents for their unconditional love and their prayers, that gave me the freedom and strength to pursue higher studies. I dedicate this thesis to Goddess Durga for making it possible.

List of Abbreviations

AIC	Akaike's Information Criteria
c.d.f.	cumulative distribution function
EGCM	Extended Growth Curve Model
EGMANOVA	Extended Generalized Multivariate Analysis of Variance
EM	Expectation Maximization
REM	Restricted Expectation Maximization
i.i.d.	independent and identically distributed
IQR	Inter Quartile Range
GEO	Gene Expression Omnibus
GCM	Growth Curve Model
GMANOVA	Generalized Multivariate Analysis of Variance
LRT	Likelihood Ratio Test
MANOVA	Multivariate Analysis of Variance
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
MSN	Multivariate Skew Normal
NR	Newton-Raphson method
p.d.f.	probability density function

Contents

List of Figures

Chapter 1

Introduction and Objectives

1.1 Introduction

Linear models are one of the most widely used statistical tools for modelling and analyzing both univariate and multivariate data, of various outcome types, e.g., continuous or discrete. The applications of linear models are widespread across different fields, including, trade and commerce, engineering, physical, biological and social sciences (Kollo and von Rosen, 2005; Fox, 1997; Hastie et al., 2001). Even when the model is not exactly linear, researchers prefer an approximate linear model for its simplicity, interpretability and parsimony (Kollo and von Rosen, 2005; Hastie et al., 2001).

One such multivariate linear model is the Growth Curve Model (GCM). GCM is a Generalized Multivariate Analysis of Variance (GMANOVA) model especially useful in the analysis of longitudinal data, growth curves and other data with repeated measurements associated with a continuous scale (e.g., dose-response curves) (Hamid et al., 2011; Potthoff and Roy, 1964). Unlike the traditional Multivariate Analysis of

Variance (MANOVA) model, the GCM allows a structured mean, where the between and within individual (across time) design matrices are incorporated in a bilinear set up (Kollo and von Rosen, 2005; Hamid and von Rosen, 2006; Hamid et al., 2011; Srivastava and Singull, 2107, 2017a). The model was introduced by Potthoff and Roy (1964) and explored further by Khatri (1966, 73), Rao (1957, 1958, 1959, 1961, 1965, 1966, 1977, 1984, 1987), Grizzle and Allen (1969), Geisser (1970, 1980, 1981), Lee and Geisser (1972, 1975) and von Rosen (1989, 1991, 1995, 1995a).

A GCM consists of p repeated measurements, from several individuals, across time, temperature, pressure gradients, etc. The model is used when standard time series methods are not applicable due to short to moderate time-series (Pan and Fang, 2002). The repeated measurements over time lead to a within-individual model. However, it is also essential to consider a between-individual model as the repeated measurements are recorded on several individuals, often distributed across more than one group. With the inclusion of a between individual design matrix, the model becomes bilinear (Kollo and von Rosen, 2005).

GCMs have extensive applications in the field of economics, social and natural sciences, medical research, epidemiology, psychology, pharmaceutical studies and agriculture (Hamid and Beyene, 2009; Pan and Fang, 2002; Hamid et al., 2011; von Rosen, 1989, 1991; Kollo and von Rosen, 2005; Jana, 2013; Jana et al., 2016; Nzabanita et al., 2012). It is to be noted that the GCM, like any other model involves several assumptions and limitations, which are discussed in the next section. In this thesis, we aim to relax some of the assumptions and address some of the limitations of the model.

The GCM assumes the same degree of polynomial responses to represent across all the groups. That is, it assumes the expected response in each of the groups to be a q-th order polynomial of the continuous predictor variable, say, time. However, in practice, this assumption might often be violated when responses from different groups have different shapes, and hence different degrees of polynomials. Another model is, therefore, required to accommodate the need for this difference across group means.

In this thesis, we consider a natural extension of the GCM, referred to as the Extended Growth Curve Model (EGCM). This model is motivated from the necessity of a more general mean structure, absent in the classical GCM framework (Kollo and von Rosen, 2005). In other words, this extension arises when different degrees of polynomials are required to model the mean for the different groups (Verbyla and Venables, 1988). It was formally introduced by von Rosen in his 1989 paper. However, it was first mentioned in his research report and dissertation (von Rosen 1984, 1985). A similar model under a different name was also introduced by Elswick (1985).

The EGCM possesses all the properties of a GCM. Being an extension, it additionally allows for a more general mean structure to be incorporated in the model. In fact, it can be easily observed that the GCM is a special case of the EGCM, as is the traditional MANOVA model (Kollo and von Rosen, 2005).

The EGCM can also be regarded as a special case of the sum of profiles model introduced by Verbyla and Venables (1988). The EGCM assumes hierarchically related design matrices, while, the sum of profiles model does not possess any such restriction (Hamid, 2001; Hamid and von Rosen, 2006). The EGCM is also a special case of Zellner's multivariate seemingly unrelated regression (SUR) model which finds abundant mention in the econometrics literature (Zellner 1962, 1963; Mendoza et al., 1995; Stanek and Koch, 1985). Banken (1984) considered a canonical form of the model, and Srivastava and Khatri (1979) considered a special case (Banken, 1984; Srivastava and Khatri, 1979; Hamid, 2001; Hamid and von Rosen, 2006). The EGCM is also referred to as the EGMANOVA model in the literature (Elswick, 1985; Cheng and Street, 1997).

1.2 Motivation and Objectives

The traditional GCM involves some assumptions and limitations including distributional assumptions. The overall objective of this thesis is to improve inference for the GCM by allowing some model assumptions to be relaxed and by addressing existing limitations related to high-dimensional applications. Three specific objectives are considered and discussed in detail here.

In the traditional model, the underlying distribution of the error terms is assumed to be multivariate normal, and existing inference is based on this assumption. However, a growth curve data might not be symmetric and hence it becomes important to consider models that allow skewed distributions. To our knowledge, no significant theoretical and computational work has been done regarding inference on non-normal or skewed GCMs. One of the specific objectives of this thesis is, therefore, to explore sensitivity of existing estimators towards violation of the normality assumption and provide estimators under the Multivariate Skew Normal (MSN) distribution, which is a more general class of distributions including the multivariate normal as a special case. This constitutes paper I of this thesis (Jana et al, 2017b).

We also consider the EGCM and relax the normality assumption to allow modelling using skewed distributions, where we assume the distribution of the error terms to be MSN. This is the second specific objective of the thesis, leading to paper II of this thesis (Jana et al., 2017c).

Finally, we consider another key assumption of the GCM, where the number of time points in the model is assumed to be less than the sample size, i.e., $p < n$. In high-dimensional scenarios involving longitudinal data, when the sample size, n, is less than the number of time points, p, it leads to singularity (Jana, 2013; Jana et al., 2017). In high-dimensional scenarios, the unknown population covariance matrix is also poorly estimated, which is also another cause leading to singularity (Ma et al., 2006). To overcome the problem of singularity of the sample covariance matrix, we previously considered its Moore-Penrose generalized inverse, for testing the significance of the mean parameter of a GCM (Jana, 2013; Jana et al., 2017). The estimators for the model parameters were also shown to be associated with increased bias and MSE in near-singularity zone.

However, the method based on the Moore-Penrose generalized inverse suffers from lack of power around near-singularity (when $p \approx n$). The third objective of this thesis is, therefore, to improve inference on GCM under high-dimensionality, where a Bayesian method is considered as an alternative approach to using the Moore-Penrose generalized inverse. We provide estimators for the model parameters and a test for the mean parameter. This work is the basis for Paper III of this thesis (Jana et al., 2017a).

1.3 Scope of the Thesis

This thesis is organized as an article based thesis, where the three objectives are presented as papers in Chapters 3, 4 and 5. In Chapter 3, we consider sensitivity of existing estimators for both the mean and the dispersion parameters, to departures from normality. This work is done under non high-dimensional scenarios, where the number of time points is assumed to be less than the sample size. Since the existing MLEs exhibit lack of robustness, we provided estimators for both parameters when the underlying distribution for the error terms is assumed to follow Multivariate Skew Normal (MSN) distribution. Since closed-form solutions to the likelihood of an MSN GCM is not feasible, the EM algorithm was adopted for this purpose. We also provided an estimator for the additional skewness parameter arising from the MSN distribution. The performances of the estimators have been assessed using empirical bias and MSE under different conditions. We also provide an illustration of the estimation procedure in practical problems, and demonstrate its utility by fitting the model to a real-life growth curve data.

In Chapter 4, we consider the estimation problem for the Extended Growth Curve Model under MSN distribution. The EGCM involves a nested subspace condition, where the column space of the consecutive design matrices are nested in previous one. Hence the EM algorithm fails to provide MLEs for an MSN EGCM. Nevertheless, the EGCM can also be modelled as a linearly restricted GCM. Consequently, the Restricted EM (REM) algorithm can be used to provide estimators for EGCM under MSN errors. We develop extensions of the multivariate Newton-Raphson (NR) method as well as the REM algorithm for estimating matrix parameters. The extended REM algorithm is then used to derive estimators for the mean and variance parameters of the model. The performance of the new estimators were examined under different scenarios using extensive simulations. We would like to highlight that the motivation for this study came from a real-life longitudinal data, which required different polynomial fits across the different groups involved. We illustrate the practical usefulness of the proposed estimation technique by fitting the EGCM to this motivating example.

Inferential techniques for the traditional GCM, in a non high-dimensional setting are widely available in literature. In a previous work, we considered the Moore-Penrose inverse and provided methods for testing the mean parameters of a GCM under a high-dimensional regime ($n < p$) (Jana, 2013; Jana et al., 2017). Although, this test overall exhibited optimal properties, it was associated with poor performance in the singularity zone ($n \approx p$). It was also observed that the estimator for the mean parameter, $\mathbf{B}$, was associated with high bias and MSE near singularity. Chapter 5 of this thesis deals with inference for multivariate normal distribution under the high-dimensional regime, where a Bayesian framework was used to overcome the limitations of the method using Moore-Penrose generalized inverse, around near singularity. A full Bayesian approach was considered to provide an estimator for the unknown population covariance matrix, $\mathbf{\Sigma}$, and a likelihood ratio test (LRT) was constructed using the estimated likelihood, where $\mathbf{\Sigma}$ was replaced by its estimator. This estimated likelihood is maximized to provide an estimator for the mean parameter, $\mathbf{B}$. Both the test statistic and the estimator were examined under different conditions for optimality, using extensive simulations. The full Bayesian approach was also examined for sensitivity to mis-specifications of the hyper-parameters. An illustration using real genetic data, is provided to demonstrate the practical usefulness of the approach. Finally, chapter 6 provides a summary of the thesis with concluding remarks and some discussions around future directions is also provided.

Chapter 2

Background

2.1 The Growth Curve Model

Consider a statistical problem, where we have k groups with n_i $(i = 1, 2, \ldots, k)$ observations in the i-th group. Suppose, p repeated observations were taken from each individual (or experimental unit) in each of the groups. Suppose, further, that the p repeated observations are associated with time or other entities such as doses of treatment, temperature or concentration. The measurements of interest may, among others, be blood pressure, blood glucose level, cholesterol level, calorie intake, body weight, height of a tree, dental measurements, or tensile strength of a material (Chen and Gupta, 2005; Pan and Fang, 2002). Observations across individuals or experimental units are assumed to be independent, while the p measurements within an individual (e.g., across time) are correlated, and can be modeled as a function of a continuous variable such as time. When analyzing such data, we require a model that not only incorporates the inherent correlation across the different time-points, but also the temporal ordering and the relationship between the outcome of interest and time (von Rosen, 1991; Pan and Fang, 2002; Hamid et. al., 2011; Jana, 2013; Jana et al., 2016).

The Growth Curve Model (GCM) is one such model and has been extensively used in several applied areas including economics, medical research and agriculture (Pan and Fang, 2002; von Rosen, 1989; 1991; Kollo and von Rosen, 2005; Hamid and Beyene, 2009; Hamid et al., 2011; Jana et al., 2016). GCMs have been indicated to be especially useful in analyzing longitudinal data with short to moderate time series, where standard time series approaches are not appropriate (Kowalski and Guire, 1974; Bryant and Gillings, 1985; Healy, 1981; Ware, 1985; Woolson and Leeper, 1980; Crowder and Hand, 1990; Pan and Fang, 2002; Hamid et al., 2011). Note that the GCM arises when we have structured mean and hence it can be viewed as the traditional multivariate analysis of variance (MANOVA) model with linear restrictions (Kollo and von Rosen, 2005; Pan and Fang, 2002; von Rosen, 1989, 1991, 1995; Hamid and von Rosen, 2006; Hamid et al., 2011).

A GCM has k groups, with repeated measurements taken from each individual across p time points, being compared with each other. The i-th group consists of n_i $(i = 1, 2, \ldots, k)$ individuals with a total of n $(\sum_{i=1}^{n} n_i = n)$ individuals (or experimental subjects) in the data. Suppose now that the mean outcome measure for the i-th group can be represented by a polynomial of time (t), where the degree of the polynomial is assumed to be $(q-1)$:

$$b_{0i} + b_{1i}t + b_{2i}t^2 + \ldots + b_{q-1,i}t^{q-1},\ i = 1, 2, \ldots, k$$

where the b_i's are coefficients of the polynomial over time. The GCM can be represented in matrix notation as

$$\mathbf{Y} = \mathbf{ZBX} + \mathbf{E}, \tag{2.1}$$

where $\mathbf{Y} : p \times n$ is the observation or response matrix, $\mathbf{B} : q \times k$ is the unknown parameter matrix consisting of all the unknown coefficients of time, sometimes also referred to as the regression coefficient, $\mathbf{Z} : p \times q$ is the within individual design matrix, $\mathbf{X} : k \times n$ is the between individual design matrix, and $\mathbf{E} : p \times n$ is the error matrix. Each column of the error matrix $\mathbf{E}$ independently follows a multivariate normal distribution with mean zero and unknown variance-covariance matrix $\mathbf{\Sigma}$. It is often assumed that $q \leq p$ and $rank(\mathbf{X}) + p \leq n$ (Potthoff and Roy, 1964; Khatri, 1966; Pan and Fang, 2002; Hamid et al., 2011; Jana et al., 2016; von Rosen, 1991, 1994). Time is incorporated through $\mathbf{Z}$ as a continuous variable and the between individual design matrix $\mathbf{X}$ describes which individual belongs to which group, as in ANOVA and MANOVA models.

It is important to note that when $\mathbf{Z} = \mathbf{I}$, the identity matrix, a GCM reduces to a MANOVA model, indicating that the GCM is a natural extension of the MANOVA model with a linearly constrained mean, and for this reason GCMs are considered as a generalized multivariate analysis of variance (GMANOVA) model (Potthoff and Roy, 1964; Khatri, 1966; Hamid et al., 2011; Hamid and Beyene, 2009; Jana et al., 2016; von Rosen, 1995; Kollo and von Rosen, 2005; Woolson and Leeper, 1980). Hamid and Beyene describe a GCM as a model appending a pre-matrix of within individual structure in the MANOVA model (Hamid and Beyene, 2009).

2.1.1 Inference on GCM

Although Wishart (1938) was the first to publish a work on GCMs, Potthoff and Roy (1964) introduced the model formally using a multivariate setup (Wishart, 1938; Potthoff and Roy, 1964). Varied literature on different aspects of the GCM are available, including Rao (1965), Khatri (1966), Chi and Weerahandi (1998), von Rosen (1989, 1995, 1995a), Pan and Fang (1996, 2002), Kollo and von Rosen (2005), Hamid and Beyene (2009) and Hamid et al. (2011). Pan and Fang (2002) and Kshirsagar and Smith (1995) introduced GCMs with several practical examples in their books on GCM. A brief review of GCMs can be found in von Rosen (1991) and Srivastava and von Rosen (1999).

Residual analysis and diagnostics have been discussed more precisely in von Rosen (1995, 1995a), Pan and Fang (1996, 2002), Liski (1991), Hamid (2001) and Hamid and von Rosen (2006). Literature prior to the formal introduction of GCM by Potthoff and Roy (1964) include Box (1950), Rao (1958, 1959, 1961) and Leech and Healy (1959), Healy (1961) and Elston and Grizzle (1962). For more advanced theory on GCM, we refer the readers to Kollo and von Rosen (2005).

Potthoff and Roy (1964) suggested a transformation which would transform a GMANOVA model to a MANOVA model, and consequently the existing tests for MANOVA can be readily used to test the significance of the mean parameter of a GMANOVA model (Potthoff and Roy, 1964). They also provided estimators of the unknown mean and variance parameters, $\mathbf{B}$ and $\boldsymbol{\Sigma}$, using some ad-hoc two-step technique. The choice of the transformation matrix is arbitrary, and it could be any matrix satisfying certain regularity conditions. The optimal choice of the transformation matrix is the unknown covariance matrix, $\boldsymbol{\Sigma}$, and the estimators as well as the test statistic are sensitive to any departure from the optimal choice (Potthoff and Roy, 1964). This approach received several criticisms, because of its choice of matrices required for the transformation (Rao, 1965, 1966).

In 1966, Khatri constructed the likelihood of the GCM, and provided MLEs for the unknown parameters, and a likelihood ratio test (LRT) for testing the mean parameter of a GCM. But the LRT was too complicated, hence, practical applications with the LRT remained limited. Also, Khatri (1966) assumed the design matrices to be of full rank (Pan and Fang, 2002). In 1989, von Rosen provided estimators without any full rank assumption.

Hamid et al. (2011) proposed the trace test for GCM, which was easy to interpret, and hence, more appropriate for practical implementations. An additional advantage of the trace test is that it does not require the full rank assumption for design matrices, used by Khatri (1966). Hamid et al. (2011) used a different estimator for the unknown covariance matrix, $\mathbf{\Sigma}$, for constructing the trace test statistic (Hamid et al., 2011). A few of the inferential techniques mentioned above are discussed in detail now.

2.1.2 Estimation of the Unknown Parameters

Potthoff and Roy (1964) proposed that the Growth Curve Model introduced in Section 2.1 can be transformed to a MANOVA model by post-multiplying the model in (2.1), with the expression $\mathbf{G}^{-1}\mathbf{X}'(\mathbf{X}\mathbf{G}^{-1}\mathbf{X}')^{-1}$, where $\mathbf{G}(n \times n)$ is any arbitrary symmetric positive definite matrix. However, it is also possible to use any nonsingular matrix $\mathbf{G}$, for the transformation, if $\mathbf{X}\mathbf{G}^{-1}\mathbf{X}'$ is of the full rank. In fact, Potthoff and Roy (1964) also mentioned that we can choose a stochastic variation of $\mathbf{G}$, as long as it is stochastically independent of the response matrix, $\mathbf{Y}$ (Potthoff and Roy, 1964) . The transformed model has the form

$$\mathbf{Y}_{trans} = \mathbf{Z}\mathbf{B} + \mathbf{E}_{trans},$$

where $\mathbf{Y}_{trans} = \mathbf{Y}\mathbf{G}^{-1}\mathbf{X}'(\mathbf{X}\mathbf{G}^{-1}\mathbf{X}')^{-1}$. Columns of the transformed error matrix, $\mathbf{E}_{trans}$, follow a multivariate normal distribution with mean $\mathbf{0}(p \times k)$ and unknown positive definite variance covariance matrix, $\mathbf{\Sigma}_{trans}$, where,

$$\mathbf{\Sigma}_{trans} = (\mathbf{X}(\mathbf{G}')^{-1}\mathbf{X}')\mathbf{X}(\mathbf{G}')^{-1}\mathbf{\Sigma}\mathbf{G}^{-1}\mathbf{X}'(\mathbf{X}\mathbf{G}^{-1}\mathbf{X}')^{-1}.$$

The estimator proposed by them for the mean parameter, $\mathbf{B}$, is

$$\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{G}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}.$$

A choice of $\mathbf{G}$ could be the identity matrix or any estimate of the dispersion matrix if available from past data. Potthoff and Roy (1964) also mentioned that the choice of $\mathbf{G}$ affects the variance of the estimator. They demonstrated that the unknown population variance matrix, $\mathbf{\Sigma}$, is the optimal choice for $\mathbf{G}$, since this choice yields the Minimum Variance Unbiased Estimator.

Although any choice of the matrix $\mathbf{G}$ yields unbiased estimates, however, the farther $\mathbf{G}$ is from its optimal choice, the greater is the variance of the estimator, and consequently wider is the confidence interval and the power of the test also decreases. Hence, a transformation almost always leads to loss of information and hence loss of power. Rao (1965, 1966) criticized the estimator $\hat{\mathbf{B}}$ provided by Potthoff and Roy (1964) due to the arbitrariness of the choice of $\mathbf{G}$ (Pan and Fang, 2002; von Rosen, 1991).

Following the arbitrariness and sensitivity of Potthoff and Roy's approach, Khatri (1966) constructed the likelihood of the GCM under matrix variate normal distribution and provided MLEs for its the mean and variance parameters, assuming full rank design matrices (Khatri, 1966; Pan and Fang, 2002):

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}, && (2.2)\\ \hat{\mathbf{\Sigma}} &= \frac{1}{n}(\mathbf{Y}-\mathbf{Z}\hat{\mathbf{B}}\mathbf{X})(\mathbf{Y}-\mathbf{Z}\hat{\mathbf{B}}\mathbf{X})', && (2.3)\end{aligned}$$

where $\mathbf{S} = \mathbf{Y}(\mathbf{I}_p - \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X})\mathbf{Y}'$. For details on the estimation procedure, the readers are referred to Pan and Fang (2002). Pan and Fang (2002) have also shown the MLE of $\mathbf{B}$, provided by Khatri (1966) to be an unbiased estimator and have provided an expression for its dispersion matrix. The expectation and the dispersion matrix for $\hat{\mathbf{\Sigma}}$ are also available (Pan and Fang, 2002; Srivastava and Singull, 2017). Khatri showed that this MLE of $\mathbf{B}$ is a weighted estimator with $\mathbf{S}$ (Khatri, 1966; Srivastava and Singull, 2017, 2017a).

Srivastava and Singull (2017) proposed an unweighted estimator for the mean parameter $\mathbf{B}$ by removing $\mathbf{S}$ to be

$$\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}. \tag{2.4}$$

This unweighted estimator is also unbiased. The authors provided an exact distribution for the unweighted estimator along with its exact moments. They also provided a consistent and unbiased estimator for $\mathbf{\Sigma}$ as

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-k}\mathbf{S}.$$

Srivastava and Singull (2017) provided an exact distribution for $\mathbf{S}$ with its exact moments.

Rao (1959) also considered parameter estimation for GCM under a special case (one group), which turned out to be in agreement with Khatri's estimator and hence the MLE for the special case (Pan and Fang, 2002; von Rosen, 1991). A review of Rao's (1959) and Khatri's (1966) estimation procedures can be found in Grizzle and Allen (1969). Khatri's estimator was also in agreement with Potthoff and Roy's (1964) estimator for $\mathbf{B}$ when $\mathbf{G}$ is chosen to be $\mathbf{S}$ (Pan and Fang, 2002). Both Khatri's (1966) and Potthoff and Roy's (1964) estimators are unbiased (Potthoff and Roy, 1964; Woolson and Leeper, 1980; Pan and Fang, 2002).

Furthermore, alternative approaches for obtaining the MLE have been given by Gleser and Olkin (1966, 1970, 1972) who used a canonical reduction of GCM (Gleser and Olkin, 1966, 1970, 1972; Pan and Fang, 2002; von Rosen, 1991). For some other approaches of deriving the MLE, we refer the readers to Srivastava and Khatri (1979), Elswick (1985), Kabe (1975), Kenward (1986) and Pan and Fang (2002). Chakravorti (1975) proposed an MLE for $\mathbf{B}$ when the columns of the error matrix $\mathbf{E}$ have different dispersion matrices. The assumption of full rank in Khatri's ML estimators was relaxed by the estimators proposed by von Rosen (1989):

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-} + (\mathbf{Z}')^{0}\mathbf{A}_1 + \mathbf{Z}'\mathbf{A}_2(\mathbf{X}^{0})', \\ \hat{\mathbf{\Sigma}} &= \frac{1}{n}(\mathbf{Y}-\mathbf{Z}\hat{\mathbf{B}}\mathbf{X})(\mathbf{Y}-\mathbf{Z}\hat{\mathbf{B}}\mathbf{X})', \qquad (2.5)\end{aligned}$$

where $\mathbf{X}^{0}$ is a matrix of full rank spanning the orthogonal complement of the linear space generated by the columns of $\mathbf{X}$ and $\mathbf{X}^{-}$ is any generalized inverse of $\mathbf{X}$. The matrix $\mathbf{X}^{0}$ or $(\mathbf{Z}')^{0}$ are not unique, and neither are $\mathbf{A}_1$ and $\mathbf{A}_2$, and so $\hat{\mathbf{B}}$ is also not unique. But, the estimator $\hat{\mathbf{\Sigma}}$ and the fitted response $\hat{\mathbf{Y}}$ are always unique:

$$\hat{\mathbf{Y}} = \mathbf{Z}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}.$$

It is also to be noted that under full rank conditions, $\hat{\mathbf{B}}$ is unique and is the same as

Khatri's MLE (von Rosen, 1989, 1995). For further discussion of Khatri's (1966) and von Rosen's (1989) estimators for GCM, we refer the readers to Pan and Fang (2002) and Kollo and von Rosen (2005).

2.1.3 Hypothesis Testing

In this thesis, we consider a hypothesis testing problem where we test the significance of the mean parameter of a GCM under different conditions. Thus, we are interested in testing $H_0 : \mathbf{B} = \mathbf{0}$ against $H_1 : \mathbf{B} \neq \mathbf{0}$. The parameter matrix, $\mathbf{B}$, consists of all the coefficients of time, e.g., slope, quadratic coefficients and also the intercept. In most statistical problems including longitudinal analysis, the intercept is usually not of much interest. A researcher is often just interested in studying the effects of time, by testing the linear, quadratic or other coefficients of time. So testing the intercept terms seem redundant and useless. However, in genetic studies which consist of longitudinal measurements on tens of thousands of genes, simultaneous testing of a huge pool of genes is never recommended as it increases statistical complications. And since a considerable percentage of genes among them are suspected to be noise, a filtering tool is required to filter the noise or housekeeping genes. Traditional filtering techniques developed for univariate data are not appropriate for longitudinal data. Our hypothesis of interest, $H_0 : \mathbf{B} = \mathbf{0}$, serves this purpose of filtering significant genes in a longitudinal genetic study such as a timecourse microarray study (Jana, 2013; Jana et al., 2017).

After filtering, it might still be of interest to identify genes that are differentially expressed over time. To achieve this, one needs to consider testing the general linear hypothesis $H_0 : \mathbf{GBF} = \mathbf{0}$ against $H_1 : \mathbf{GBF} \neq \mathbf{0}$, where $\mathbf{G}$ and $\mathbf{F}$ are matrices consisting of only ones and zeros. These matrices are chosen appropriately depending on which coefficients of time we are interested in testing. Another objective in the analysis of longitudinal genetic data might be to identify genes with differential expressions among groups, e.g., control and treatment groups. This can also be tested using the general linear hypothesis $H_0 : \mathbf{GBF} = \mathbf{0}$, after filtering has been done using the simple hypothesis $H_0 : \mathbf{B} = \mathbf{0}$ (Jana, 2013; Jana et al., 2017).

In this thesis, our hypothesis of interest is the simple hypothesis $H_0 : \mathbf{B} = \mathbf{0}$ against $H_1 : \mathbf{B} \neq \mathbf{0}$, however, the results can be easily extended to the general linear hypothesis. Gene filtering in longitudinal data is an example where the simple hypothesis is relevant, but the practical implementation of the simple hypothesis is not limited solely for filtering purposes. Sometimes, the researcher might also be interested in testing the baseline effects in a GCM which are reflected solely in the intercepts. The simple hypothesis is, in fact, a special case of the general hypothesis (Hamid et al., 2011; Jana, 2013; Jana et al., 2017).

Potthoff and Roy (1964) proposed that once we transform the GCM to a regular MANOVA model, we can use any test developed for MANOVA models, on the transformed data, for example, Roy's largest root test, Pillai's trace test, Wilks' Lambda, etc., to name a few (Potthoff and Roy, 1964). Khatri constructed a likelihood ratio test (LRT) using the MLEs of the parameters provided by him, for testing the general linear hypothesis $H_0 : \mathbf{GBF} = \mathbf{0}$ (Khatri, 1966). His test criterion was based on Wilks' Lambda. He also provided a few other test criteria based on the trace and the maximum characteristic root of the test statistic. Khatri's approach did not involve any transformation but was computationally tedious and practically difficult to interpret and implement as it required iterative computation of two estimators until they converge simultaneously.

Likelihood Ratio Test

Khatri's objective, in his 1966 paper, was to develop a test criterion for testing the mean parameter of a GCM and compare the means between the groups. Once Khatri (1966) constructed the likelihood for a GCM, developing a Likelihood Ratio Test (LRT) was a natural choice. Khatri (1966) proposed an LRT for testing the general linear hypothesis of a GCM as

$$\Lambda = \frac{\mid \mathbf{Q} \mid}{\mid \mathbf{P} + \mathbf{Q} \mid},$$

where $\mathbf{Q} = \mathbf{G}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-1}\mathbf{G}'$, $\mathbf{P} = (\mathbf{G}\hat{\mathbf{B}}\mathbf{F})(\mathbf{F}'\mathbf{R}\mathbf{F})^{-1}(\mathbf{G}\hat{\mathbf{B}}\mathbf{F})'$,
and $\mathbf{R} = (\mathbf{X}\mathbf{X}')^{-1} + (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{Y}'(\mathbf{S}^{-1} - \mathbf{S}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}^{-1})\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}$, and suggested three test procedures for the LRT statistic whose rejection regions are:

1. $tr(\mathbf{P}\mathbf{Q}^{-1}) \geq \lambda_2$,
2. $tr(\mathbf{P} + \mathbf{Q})^{-1}\mathbf{P} \geq \lambda_3$,
3. maximum characteristic root of $\mathbf{P}\mathbf{Q}^{-1} \geq \lambda_1$,

where the critical values λ_1, λ_2 and λ_3 are calculated from the level of significance. This test was difficult to implement and it also required the design matrices to be of full rank.

The Trace Test

Lack of practical usefulness of Khatri's LRT motivated Hamid et al. (2011) to propose a simpler, intuitive and practical test statistic for the mean parameter of the GCM.

A restricted maximum likelihood approach followed by an estimated likelihood was used for constructing the trace test. This new test was called the trace test and as stated earlier it was developed without assuming full rank design matrices. For a MANOVA, this test reduces to the Lawley-Hotelling trace test. Hence, this trace test for a GMANOVA model seems to be a natural extension of the classical trace test for MANOVA. The construction of the trace test was motivated by the residuals of GCM, and the authors showed that the test was actually a function of the model residuals.

The likelihood of the GCM was decomposed into two independent components, and the one consisting the population covariance matrix, $\mathbf{\Sigma}$, was maximized to obtain an estimator for $\mathbf{\Sigma}$: $\frac{1}{n}\mathbf{S}$. However, it is to be noted that $\frac{1}{n}\mathbf{S}$ is not an unbiased estimator for $\mathbf{\Sigma}$, rather $\frac{1}{n-k}\mathbf{S}$ is an unbiased estimator for $\mathbf{\Sigma}$ (Srivastava and Singull, 2017). The authors also established a relationship between Khatri's MLE for $\mathbf{\Sigma}$ and the sample covariance matrix, $\mathbf{S}$:

$$\hat{\mathbf{\Sigma}} = \mathbf{S} + \mathbf{V}\mathbf{V}',$$

where $\mathbf{V} = \mathbf{Y}\mathbf{X}(\mathbf{X}\mathbf{X}')^{-}\mathbf{X} - \mathbf{Z}\hat{\mathbf{B}}\mathbf{X}$, and $\hat{\mathbf{\Sigma}}$ is as defined in (2.5) (Hamid et al., 2011; Pan and Fang, 2002).

Hamid et al. (2011) replaced $\mathbf{\Sigma}$ with the estimator, $\frac{1}{n}\mathbf{S}$, in the original likelihood, to construct the estimated likelihood. An LRT was developed based on the estimated likelihood, and then simplified and expressed as a trace of the product of some matrices, namely, the known design matrices, response matrix and the sample covariance matrix. It is used for testing the mean parameter of the GCM:

$$\phi = tr\{\mathbf{S}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}\mathbf{Y}'\}. \tag{2.6}$$

This test has a relatively simpler form than Khatri's LRT and it does not assume full rank. Another clear advantage of the trace test is that the null distribution of the test statistic is independent of the unknown covariance matrix, and hence is free of any unknown parameters, allowing the null distribution and hence the critical value to be generated empirically. This has very useful practical implications. The authors also provided the exact and approximate distributions of the test statistic under both the null and alternative hypothesis (Hamid et al., 2011).

This test can be used for testing the simple hypothesis $H_0 : \mathbf{B} = \mathbf{0}$. However, Hamid et al. (2011) also provided a test statistic for testing the general linear hypothesis $H_0 : \mathbf{GBF} = \mathbf{0}$.

2.2 The Extended Growth Curve Model

Consider k groups, with n_i $(i = 1, 2, \ldots, k)$ individuals in the i-th group, and let p repeated observations be recorded at p time-points for each of the n $(\sum_{i=1}^{k} n_i = n)$ individuals. We further assume that the mean in each group is a polynomial in time

t, and every group requires a different polynomial fit, as

$$\begin{aligned}
E(Y_i) &= \beta_{01} + \beta_{11}t + \beta_{21}t^2 + \ldots + \beta_{(q-k)1}t^{q-k}, \\
&\qquad\qquad\qquad\qquad \text{for, } i = 1, 2, \ldots n_1, \\
E(Y_i) &= \beta_{02} + \beta_{12}t + \beta_{22}t^2 + \ldots + \beta_{(q-k)2}t^{q-k} + \beta_{(q-k+1)2}t^{q-k+1}, \\
&\qquad\qquad\qquad\qquad \text{for, } i = n_1 + 1, \ldots, n_1 + n_2, \\
& \ldots\ldots \\
E(Y_i) &= \beta_{0k} + \beta_{1k}t + \beta_{2k}t^2 + \ldots + \beta_{(q-1)k}t^{q-1}, \\
&\qquad\qquad\qquad\qquad \text{for, } i = \sum_{i=1}^{k-1} n_i + 1, \ldots, n,
\end{aligned}$$

where Y_i $(i = 1, 2, \ldots, n)$ is the response of the i-th individual and the β-s are the unknown coefficients of time. It is to be noted that in this example, the mean structure is such that the polynomial for the mean of each group is just one order higher than that of the previous group. However, it may not always be so simple. The difference in the order of the polynomials among the different groups could be more general. The example stated here for k groups is solely for illustration. It is clearly evident from this example that an EGCM allows us to model clusters in the data (Kollo and von Rosen, 2005). Hence, the significance of an EGCM is that it linearizes a non-linear model (Verbyla and Venables, 1988).

The assumption of the same profile for all the individuals across different groups, inherent in a GCM, can be relaxed in an EGCM and the EGCM can be represented in matrix format as

$$\mathbf{Y} = \sum_{i=1}^{k} \mathbf{Z}_i \mathbf{B}_i \mathbf{X}_i + \mathbf{E}, \tag{2.7}$$

where $\mathbf{Z}_i(p \times q_i)$ is the within-individual design matrix, $\mathbf{B}_i(q_i \times k)$ is the unknown parameter matrix or the matrix of regression coefficients, $\mathbf{X}_i(k \times n)$ is the between-individual design matrix of the i-th group ($i = 1, 2, \ldots, k$), $\mathbf{Y}(p \times n)$ is the response matrix and $\mathbf{E}(p \times n)$ is the error matrix, consisting of the responses or the error terms, respectively, for all the individual across all the groups (Kollo and von Rosen, 2005). The error terms for different individuals in an EGCM are assumed to be independently distributed with mean zero and equal variance. It is also assumed that $q_i \leq p$ ($i = 1, 2, \ldots, k$), $rank(\mathbf{X}_1) + p \leq n$ and $\mathcal{C}(\mathbf{X}_i') \subset \mathcal{C}(\mathbf{X}_{i-1}')$ ($i = 2, 3, \ldots, k$), where $\mathcal{C}(\cdot)$ denotes the column space of a matrix. Alternatively, this nested subspace assumption can equivalently be made on the within-individual design matrices as well: $\mathcal{C}(\mathbf{Z}_i) \subset \mathcal{C}(\mathbf{Z}_{i-1})$ ($i = 2, 3, \ldots, k$) (Kollo and von Rosen, 2005; Hamid, 2001; Hamid and von Rosen, 2006).

This model also finds mention in the literature under a different name - the multivariate linear model with mean $\sum_{i=1}^{k} \mathbf{Z}_i \mathbf{B}_i \mathbf{X}_i$ which is abbreviated as MLNM($\sum_{i=1}^{k} \mathbf{Z}_i \mathbf{B}_i \mathbf{X}_i$) (Kollo and von Rosen, 2005). It can be easily noticed that for $k = 1$, the EGCM reduces to a GCM. Thus the GCM is a special case of the EGCM.

For simplicity, in this thesis, we just consider two groups. However, without loss of generality, the results from this thesis can be well extended to more than two groups. Assuming that the mean for the second group has m more polynomials than the first one, the average responses in the two groups can be modelled as

$$\begin{aligned} E(Y_i) &= \beta_{01} + \beta_{11}t + \beta_{21}t^2 + \ldots + \beta_{(q-m-1)1}t^{q-m-1}, \\ &\qquad \text{for, } i = 1, 2, \ldots n_1, \\ E(Y_i) &= \beta_{02} + \beta_{12}t + \beta_{22}t^2 + \ldots + \beta_{(q-m-1)2}t^{q-m-1} + \ldots + \beta_{(q-1)2}t^{q-1}, \\ &\qquad \text{for, } i = n_1 + 1, \ldots, n. \end{aligned}$$

The design matrices and the parameter matrices for such a model have the form

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{q-m-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{q-m-1} \\ \dots\dots & & & & \\ 1 & t_p & t_p^2 & \dots & t_p^{q-m-1} \end{pmatrix}, \mathbf{Z}_2 = \begin{pmatrix} t_1^{q-m} & \dots & t_1^{q-1} \\ t_2^{q-m} & \dots & t_2^{q-1} \\ \dots\dots & & \\ t_p^{q-m} & \dots & t_p^{q-1} \end{pmatrix},$$

$$\mathbf{X}_1 = \begin{pmatrix} 1_{n1} & 0_{n2} \\ 0_{n1} & 1_{n2} \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} 0_{n1} & 0_{n2} \\ 0_{n1} & 1_{n2} \end{pmatrix},$$

$$\mathbf{B_1} = \begin{pmatrix} \beta_{01} & \beta_{02} \\ \dots & \\ \beta_{(q-m-1)1} & \beta_{(q-m-1)2} \end{pmatrix} = \begin{pmatrix} \beta_1^{(1)}, \beta_2^{(1)} \end{pmatrix}, \text{ and}$$

$$\mathbf{B_2} = \begin{pmatrix} \beta_{(q-m)1} & \beta_{(q-m)2} \\ \dots & \\ \beta_{(q-1)1} & \beta_{(q-1)2} \end{pmatrix} = \begin{pmatrix} \beta_1^{(2)}, \beta_2^{(2)} \end{pmatrix}.$$

where 1_{n1} is a vector of ones of dimension $n1$ and so on for 0_{n1}. Hence, we can re-write the columns of $\mathbf{Y}$ as

$$\begin{aligned} Y_i &= \mathbf{Z}_1\beta_1^{(1)} + E_i \ , \ i = 1, 2, \dots, n_1, \\ Y_i &= \mathbf{Z}_1\beta_2^{(1)} + \mathbf{Z}_2\beta_2^{(2)} + E_i \ , \ i = (n_1 + 1), \dots, n. \end{aligned}$$

It is not possible to express an EGCM as a GCM unless we introduce some restrictions on the parameter matrices (von Rosen, 1989; Kollo and von Rosen, 2005).

In this thesis, since we consider parameter estimation for the special case $k = 2$, so, we introduce and discuss the linear restrictions on the parameter matrices, for $k = 2$, using which we will be able to express an EGCM as a GCM. For further details on the linear restrictions, the readers are referred to the book by Kollo and von Rosen (2005).

The mean of the first group can be written as

$$\begin{aligned} E(y_i) &= \beta_{01} + \beta_{11}t + \beta_{21}t^2 + \ldots + \beta_{(q-m-1)1}t^{q-m-1} + \ldots + \beta_{(q-1)1}t^{q-1}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for, } i = 1, 2, \ldots, n_1, \end{aligned}$$

if we assume $\beta_{(q-m)1} = \beta_{(q-m+1)1} = \ldots\ldots = \beta_{(q-1)1} = 0$, that is, $\beta_1^{(2)} = 0$. Under matrix setup, these m linear restrictions can be expressed as $\mathbf{GBH} = \mathbf{0}$, where $\mathbf{G} = (\mathbf{0}_{m\times(q-m)}, \mathbf{I}_m)$, $\mathbf{H}' = (1, 0)$ and $\mathbf{B}' = (\mathbf{B}_1', \mathbf{B}_2')$. Hence, an EGCM can be expressed as a GCM as

$$\mathbf{Y} = \mathbf{ZBX} + \mathbf{E}, \tag{2.8}$$

subject to the constraint $\mathbf{GBH} = \mathbf{0}$, where $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ and $\mathbf{X} = \mathbf{X}_1$.

The significance of introducing this restriction is that, instead of treating the EGCM as a model with a nested subspace condition, we can frame it as a model with a simple linear restriction, and proceed with the task of parameter estimation.

2.2.1 Inference on EGCM

In this section, we would like to briefly introduce the estimators of a traditional EGCM, that is, an EGCM whose underlying distribution is multivariate normal. von Rosen constructed the likelihood equations for the classical EGCM and provided MLEs for the regression coefficients, $\mathbf{B}_1, \mathbf{B}_2$ and the dispersion matrix, $\mathbf{\Sigma}$ (von Rosen;

1989). For $k = 2$, the MLEs are

$$\begin{aligned}
\hat{\mathbf{B}}_2 &= (\mathbf{Z}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{Z}_2)^-\mathbf{Z}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\mathbf{Y}\mathbf{X}_2'(\mathbf{X}_2\mathbf{X}_2')^- + (\mathbf{Z}_2'\mathbf{T}_1')^0\mathbf{A}_{21} + \mathbf{Z}_2'\mathbf{T}_1'\mathbf{A}_{22}\mathbf{X}_2^{0\prime}, \\
\hat{\mathbf{B}}_1 &= (\mathbf{Z}_1'\mathbf{S}_1^{-1}\mathbf{Z}_1)^-\mathbf{Z}_1'\mathbf{S}_1^{-1}(\mathbf{Y} - \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2)\mathbf{X}_1'(\mathbf{X}_1\mathbf{X}_1')^- + (\mathbf{Z}_1')^0\mathbf{A}_{11} + \mathbf{Z}_1'\mathbf{A}_{12}\mathbf{X}_1^{0\prime}, \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}(\mathbf{Y} - \mathbf{Z}_1\hat{\mathbf{B}}_1\mathbf{X}_1 - \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2)(\mathbf{Y} - \mathbf{Z}_1\hat{\mathbf{B}}_1\mathbf{X}_1 - \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2)', \qquad (2.9)
\end{aligned}$$

where $\mathbf{T}_1 = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{S}_1^{-1}\mathbf{Z}_1)^-\mathbf{Z}_1'\mathbf{S}_1^{-1}$, $\mathbf{S}_1 = \mathbf{Y}\{\mathbf{I} - \mathbf{X}_1'(\mathbf{X}_1\mathbf{X}_1')^-\mathbf{X}_1\}\mathbf{Y}'$, $\mathbf{S}_2 = \mathbf{S}_1 + \mathbf{T}_1\mathbf{Y}\mathbf{X}_1'(\mathbf{X}_1\mathbf{X}_1')^-\mathbf{X}_1\{\mathbf{I} - \mathbf{X}_2'(\mathbf{X}_2\mathbf{X}_2')^-\mathbf{X}_2\}\mathbf{X}_1'(\mathbf{X}_1\mathbf{X}_1')^-\mathbf{X}_1\mathbf{Y}'\mathbf{T}_1'$, and $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ are arbitrary matrices. The matrix $(\mathbf{Z}_1)^0$ is a matrix of full rank spanning the orthogonal complement of the column space of $\mathbf{Z}_1$, and $\mathbf{Z}_1^-$ is any generalized inverse of $\mathbf{Z}_1$. It is worth mentioning that the estimators $\hat{\mathbf{B}}_1$ and $\hat{\mathbf{B}}_2$ are not unique, while $\hat{\boldsymbol{\Sigma}}$ is unique.

It is to be noted that, in this thesis, we have presented inferential techniques for the special case, $k = 2$, hence in this section, we are presenting the estimators of the traditional EGCM, for $k = 2$ only (Kollo and von Rosen, 2005). These estimators for $k = 2$ have also been presented in Nzabanita et al. (2012). However, the results can be extended to more than two groups, without loss of generality.

Under a GCM, the estimators $\hat{\mathbf{B}}_1$ and $\hat{\boldsymbol{\Sigma}}$ in (2.9) reduce to von Rosen's (1989) estimators for $\mathbf{B}$ and $\boldsymbol{\Sigma}$, when the design matrices are not of full rank. Under the full rank assumption, the estimators reduce to Khatri's (1966) MLEs for $\mathbf{B}$ and $\boldsymbol{\Sigma}$.

Fujikoshi and Satoh (1996) provided MLEs for the mean and covariance matrix of the EGCM using canonical form. Nzabanita et al. (2012) provided consistent estimators for the covariance matrix and estimators for the mean parameter of an EGCM when the unknown population covariance matrix has a linear structure. Significance tests for the mean parameters of an EGCM are also available (Srivastava and Singull, 2012; Kollo and von Rosen, 2005; Yokoyama and Fujikoshi, 1993). The tests proposed by Srivastava and Singull (2012) seemed to perform better than Yokoyama and Fujikoshi's (1993) test.

2.3 High-dimensional Extensions

Traditionally, practical problems require small number of unknown parameters or features to be estimated based on relatively large number of experimental units (Johnstone and Titterington, 2009; Hamilton 1970). Thus, traditional theory and practice mostly focused on 'small p large n' problems (Johnstone and Titterington, 2009; Huber, 1981), where p is the number of unknown parameters and n is the number of observations measured.

However, over the last couple of decades, the dramatic advancement in biomedical sciences has revolutionized data collection technologies, allowing researchers to acquire large amount of complex and high-dimensional biomedical data, where large number of variables, p, are measured from relatively small number of study units, n (Johnstone and Titterington, 2009; Kendziorski et al., 2003; Hamid and Beyene, 2009). Such high-dimensional data pose several methodological and computational challenges to statisticians and other researchers (Hamid and Beyene, 2009; Kendziorski et al., 2003).

For instance, in genomic studies, measurements from hundreds of thousands of genes were taken from individuals, often from different groups (e.g., cancer patients and health individuals) leading to high-dimensional data. One of the objectives of such genomic experiments, for example, is identifying genes that are differentially expressed between cancer patients (cases) and healthy individuals (controls), with the objective of understanding the genetic aspect of the disease and for possible targeted treatment or therapy. High-dimensional genetic data involve many pre-processing steps before a formal hypothesis on differential expression is statistically tested. An important step among them is gene filtering, where genes that are not expressed (e.g., house keeping genes or genes that are not relevant to the study) are removed from the data to reduce high-dimensionality and alleviate the number of multiple tests.

The methodological challenges involved in high-dimensional data are often multiplied in studies involving time course genomic data, where repeated measurements over time on hundreds of thousands of genes are taken from individuals across different groups (Efron et al., 2001; Johnstone and Titterington, 2009; Kendziorski et al., 2003; Smyth, 2004; Lönnstedt and Speed, 2002; Tai and Speed, 2006). For a given gene, the number of time points are often larger than the number of individuals, leading to gene specific high-dimensionality, which results in singular gene-specific sample covariance matrix in addition to the global high-dimensionality inherently present in genomic experiments because of the thousands of genes being investigated simultaneously (Hamid and Beyene, 2009; Tai and Speed, 2006, 2009; Ma et al., 2006; Yuan and Kendziorski, 2006).

In recent years, several methods and analysis strategies have been proposed for time course genomic experiments including the moderated Hotelling's T^2 statistic and moderated likelihood ratio test statistic under the MANOVA model (moderated Wilks' lambda), which are now being commonly used in practical applications (Tai and Speed, 2006, 2009). The term moderated is often used to indicate that the tests are based on a non-singular estimate of the sample covariance matrix. These tests by Tai and Speed (2006, 2009) were developed using a Bayesian approach where the authors assumed a prior distribution for the unknown covariance matrix and proposed a moderated estimator for it. They used an empirical Bayes' approach to estimate the hyper-parameters of the prior distribution by pooling data from all the genes.

Srivastava (2007) proposed using the Moore-Penrose generalized inverse of the sample covariance matrix instead of the inverse in a Hotelling's T^2 type test for high-dimensional problems. Although these approaches address the problem of high-dimensionality and account for correlation across time points, they are all based on MANOVA models and hence do not account for the temporal ordering (where time is incorporated as factor or group), do not allow structured mean (time dependency) and the actual time-points also do not enter the model.

GMANOVA models, and hence GCMs allow structured mean, incorporate correlation across time points (within individual) in the model and the actual time-points are entered into the model. The mean response is represented as a continuous function of time. Thus, GMANOVA models being an useful tool for modelling longitudinal data are appropriate for modelling time-course genetic data as gene expression over time is a biologically continuous measurement recorded across several time-points.

In an attempt to analyze high-dimensional time-course microarray data, Hamid and Beyene (2009) proposed a method based on GMANOVA model by modelling gene expression profile over time. They used a gene-specific GMANOVA model and assumed the same prior distribution for the unknown covariance matrix as Tai and Speed (2006). Then they transformed the model into a MANOVA model using Potthoff and Roy's (1964) transformation. The transformation has been discussed in detail in Section 2.1.3. The moderated sample covariance matrix based on the empirical Bayes' approach was used as the unknown transformation matrix, **G**. Although this method incorporates within individual correlation in the model and allow for structured mean, however, it suffers from the same drawbacks as Potthoff and Roy's (1964) approach. The transformation almost always leads to loss of information and hence loss of power. Furthermore, it is extremely difficult to determine the distribution of the transformed test statistic, under the null hypothesis. Consequently, it is impossible to perform a hypothesis test or a significance test using this test statistic and its applications remain limited to gene filtering using gene ranking in timecourse microarray data (Hamid and Beyene, 2009; Jana et al., 2016).

Few approaches to address the issue of high-dimensionality have been developed in Chen and Qin (2010), Chen et al. (2012), Rauf et al. (2013), Srivastava and Kubokawa (2013), etc. For a comprehensive review on tests for the mean parameter in high dimensional conditions, under normality, the readers are referred to Kropf et al. (2009). For an elaborate discussion on high-dimensional problems in MANOVA model, the readers are referred to L*ä*uter et al. (2009). However, all of these approaches are based on MANOVA models and hence it is necessary to develop methods for GMANOVA models which address the high-dimensionality issue.

Recently, Srivastava and Singull (2017) proposed tests for the GMANOVA model based on the unweighted estimator for **B**. The MLE of **B** provided by Khatri (1966) is an weighted estimator which depends on the inverse of the sample covariance matrix, **S**. However, **S** is unstable when $p \approx n$ and under strict high-dimensional conditions ($p >> n$), **S** is singular. Hence, the LRTs proposed by Khatri (1966) are invalid under high-dimensional conditions ($p >> n$ and $p \approx n$).

The unweighted estimator in (2.4), proposed by Srivastava and Singull (2017), does not depend on the sample covariance matrix, **S**. Also the sample covariance matrix is independent of the mean parameter, **B**. The tests based on the unweighted estimator for **B** do not require $\mathbf{S}^{-1}$ and are valid in high-dimensional scenarios. In fact, Srivastava and Singull (2017) compared the performance of their tests with Khatri's (1966) LRT and their tests performed better overall.

Tests for the covariance structure under high-dimensional conditions are also available (Srivastava and Singull, 2017a). These tests were demonstrated to have better performance than the LRT under high-dimensional conditions (Srivastava and Singull, 2017a).

2.4 Evaluation Criteria

2.4.1 Multivariate Bias

Since our parameters of interest, **B** and **Σ**, are matrices, their estimators are also matrices. As a result, we would expect their bias to be matrices as well. Analogous to the univariate definition of bias, we define the bias of a matrix parameter as the expected difference of the estimator and the parameter:

$$Bias(\hat{\mathbf{B}}) = E(\hat{\mathbf{B}} - \mathbf{B}),$$

and it is referred to as the Bias matrix (Jana, 2013; Jana et al., 2017). The definition of bias of estimators **Σ** would follow similarly.

2.4.2 Multivariate Mean Square Error (MSE)

Looking at just the bias of a parameter might be misleading, because even if the bias is small, sometimes the variance might be large. It is, therefore, important to look at the variance of the estimators as well. Alternatively, we can look at the mean square error (MSE), which is a combined measure of bias and variance (Casella and Berger, 2002):

$$MSE = Bias^2 + Variance.$$

In univariate estimation, MSE is defined as

$$E_\theta(T-\theta)^2$$

where T is the estimator of the parameter θ (Casella and Berger, 2002). However, since our parameter of interest is a matrix, hence analogous to the definition of univariate MSE, we define multivariate MSE for our estimator as

$$E[(\hat{\mathbf{B}}-\mathbf{B})'(\hat{\mathbf{B}}-\mathbf{B})].$$

This gives us a matrix which we will refer to as the MSE matrix in the thesis (Jana, 2013; Jana et al., 2016). When we expand this expression for MSE we see that

$$\begin{aligned}
MSE &= E[(\hat{\mathbf{B}}-\mathbf{B})'(\hat{\mathbf{B}}-\mathbf{B})] \\
&= E[(\hat{\mathbf{B}}-E(\hat{\mathbf{B}})+E(\hat{\mathbf{B}})-\mathbf{B})'(\hat{\mathbf{B}}-E(\hat{\mathbf{B}})+E(\hat{\mathbf{B}})-\mathbf{B})] \\
&= E[(\hat{\mathbf{B}}-E(\hat{\mathbf{B}}))'(\hat{\mathbf{B}}-E(\hat{\mathbf{B}}))] - (\hat{\mathbf{B}}-E(\hat{\mathbf{B}}))'(\hat{\mathbf{B}}-E(\hat{\mathbf{B}})) \\
&\quad -(E(\hat{\mathbf{B}})-\mathbf{B})'(\hat{\mathbf{B}}-E(\hat{\mathbf{B}})) + (E(\hat{\mathbf{B}})-\mathbf{B})'(E(\hat{\mathbf{B}})-\mathbf{B})] \qquad (2.10)
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{\Sigma}_{\hat{\mathbf{B}}} - (E(\hat{\mathbf{B}}) - E(\hat{\mathbf{B}}))'(\hat{\mathbf{B}} - E(\hat{\mathbf{B}})) \qquad (2.11)\\
&\quad -(E(\hat{\mathbf{B}}) - \mathbf{B})'(E(\hat{\mathbf{B}}) - E(\hat{\mathbf{B}})) + Bias'Bias\\
MSE &= \boldsymbol{\Sigma}_{\hat{\mathbf{B}}} + Bias'Bias \qquad (2.12)
\end{aligned}
$$

The $\boldsymbol{\Sigma}_{\hat{\mathbf{B}}}$ matrix in (2.12) is the variance-covariance matrix of the estimator $\hat{\mathbf{B}}$ and it is also evident that the definition of matrix MSE conforms to the properties of univariate MSE.

2.4.3 Distance Measure: Euclidean Norm

Note that, in this study, we are providing inferential techniques for the mean and variance-covariance matrices. Since the parameters of interest are matrices, hence the respective empirical Bias and MSE are also matrices, and so are the difference between the values under the null and alternative hypotheses. Consequently, investigating some properties of the power curve, e.g., unbiasedness, monotonicity or assessing the performance of the estimators, e.g., consistency, robustness etc., using empirical Bias or MSE is not obvious. Thus, for ease of comparison and graphic representation, we require a measure to transform these matrices into scalar quantities. Frobenius norm or the Hilbert-Schmidt norm which is more popularly known as Euclidean norm seem to be an appropriate measure to serve our purpose. For further details about the Frobenius norm, we refer the readers to Horn and Johnson (1985).

The Frobenius norm expresses the magnitude of any matrix as a scalar, which can also be interpreted as the distance between that matrix and the **0** matrix, since the norm of a **0** matrix is always 0. We briefly introduce the norm here. For any matrix, $\mathbf{A} = ((a_{ij}))_{m\times n}$, its Euclidean norm is defined as $\| \mathbf{A} \| = \sqrt{\sum_{i,j} a_{ij}^2}$. Being a norm, it satisfies all the requirements of a matrix norm. The higher the magnitude of the matrix, higher is the Euclidean norm (Horn and Johnson, 1985). Thus, greater biases and greater MSEs are reflected in higher Euclidean norm.

As already stated, comparison of Bias matrices between different scenarios, e.g., different sample sizes and different parameter matrices of varying magnitude is not straight forward. Therefore, it was necessary to reduce the bias matrix to a scalar quantity and we used the Euclidean norm for this purpose. Since 0 bias represents the ideal unbiased case, thus a matrix of all zeros could be regarded as the unbiased state for matrix estimation, and its Euclidean norm is 0. Therefore, Euclidean norm of the Bias matrix can also be interpreted as the distance between the Bias matrix and the unbiased matrix.

Similarly, the Euclidean norm of the MSE matrix represents the MSE matrix as a scalar measure and it can also be interpreted as the distance between the MSE matrix and the matrix of 0 MSE. Hence, the Euclidean norm seems to be an apt measure for matrix Bias and MSE.

2.5 Multivariate Distributions used in this Thesis

2.5.1 Matrix Normal Distribution

Multivariate normal is one of the most widely used multivariate distributions in applied statistics. In this section, we introduce Matrix variate normal distribution and its relationship with multivariate normal, so that the reader can easily relate this study with traditional GCM literature. Matrix variate normal distribution arises when random samples are drawn from a multivariate normal population (Gupta and

Nagar, 2000; Roy, 1957; Siotani et al., 1985). A matrix $\mathbf{Y}(p \times n)$ is said to follow a matrix variate normal distribution with mean $\mathbf{M}(p \times n)$ and variance $\mathbf{\Sigma} \otimes \mathbf{\Psi}$, where both $\mathbf{\Sigma}(p \times p)$ and $\mathbf{\Psi}(n \times n)$ are positive definite matrices, if

$$vec(\mathbf{Y}') \sim N_{pn}(vec(\mathbf{M}'), \mathbf{\Sigma} \otimes \mathbf{\Psi}).$$

The matrix $\mathbf{\Sigma}$ is the variance-covariance matrix of the columns of $\mathbf{Y}$, and $\mathbf{\Psi}$ is the variance-covariance matrix of the rows of $\mathbf{Y}$. The pdf of $\mathbf{Y}$ is (Gupta and Nagar, 2000)

$$(2\pi)^{-\frac{np}{2}} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}} \mid \mathbf{\Psi} \mid^{-\frac{p}{2}} e^{-\frac{1}{2}tr(\mathbf{Y}-\mathbf{M})'\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{M})\mathbf{\Psi}^{-1}}.$$

For further discussions and properties of this distribution, the readers are referred to Srivastava and Carter (1983) and Gupta and Nagar (2000).

In a traditional GCM setup, the columns of the error matrix are independent and have a variance-covariance matrix $\mathbf{\Sigma}$ and mean $\mathbf{0}(p \times n)$. Hence, under GCMs, the matrix $\mathbf{\Psi}$ is the identity matrix. Consequently, the matrix $\mathbf{Y}$ follows the matrix variate normal distribution with mean $\mathbf{ZBX}$ and variance $\mathbf{\Sigma} \otimes \mathbf{I}$, and it is written as

$$\mathbf{Y} \sim N_{p\times n}(\mathbf{ZBX}, \mathbf{\Sigma}, \mathbf{I}).$$

Hence, each column of $\mathbf{Y}$ follow a multivariate normal distribution with some mean μ, say, for each group, and variance $\mathbf{\Sigma}$ and this fact has been utilized later for implementation of Bayesian theory.

2.5.2 Multivariate Skew Normal Distribution

Practical problems often require us to model data that are not naturally symmetric. This requires us to consider skewed distributions or a broader class of distributions including both skewed and symmetric ones. Since normal distribution is the most commonly used and most well-known distribution in applied statistics, hence coming up with a skewed class of distributions, which would also include the normal distribution as a special case, would be the most appropriate and convenient for modelling and analyzing real life data.

Azzalini introduced a mathematically tractable extension of the univariate normal distribution in 1985. This new distribution possesses several properties of the normal distribution, yet it is a broader class of distributions including the normal as a special case. When the skewness parameter became 0, the skewed distribution reduces to a regular normal distribution. It provides flexibility in terms of fitting real data in the sense that, it could fit both skewed and symmetric data. This distribution was named the skew normal distribution (Azzalini, 1985, 1986; Harrar and Gupta, 2008; Genton et. al., 2001).

A multivariate extension of this skewed distribution was soon deemed a necessity for modelling multivariate skewed data, and Azzalini and Dalla-Valle proposed the Multivariate Skew Normal (MSN) distribution in 1996. If each component of a vector random variable is dependent and the marginal distribution for each of them is the univariate skew normal distribution, then the vector itself follows the MSN distribution (Azzalini and Dalla-Valle, 1996).

The MSN distribution was naturally a mathematical extension of the multivariate normal distribution. It included a vector parameter to regulate skewness, which when 0, the MSN distribution reduces to a multivariate normal distribution, like its univariate counterpart. Thus the multivariate normal distribution is a special case of the MSN distribution, and the MSN distribution being a broader class of density function allows us to model both multivariate symmetric and skewed data (Gupta and Kollo, 2003; Genton et al., 2001; Azzalini and Dalla-Valle, 1996; Azzalini and Capitanio, 1999; Gupta and Chang, 2003).

In this section, we introduce the MSN distribution and elaborate on its stochastic representation which will be used later in the thesis, for the purpose of estimation. We follow the notations and the parametric form introduced by Azzalini (Azzalini and Capitanio, 2013; Azzalini and Capitanio, 1999). If U is a random variable with density function

$$2\phi_p(U; 0, \bar{\mathbf{\Omega}})\Phi(\alpha' U),$$

where, $\phi_p(\cdot; 0, \bar{\mathbf{\Omega}})$ is the pdf of a multivariate normal distribution with mean 0 and variance $\bar{\mathbf{\Omega}}$ and $\Phi(\cdot)$ is the cdf of a univariate standard normal distribution, then it follows the MSN distribution with parameters $\bar{\mathbf{\Omega}}$ and α, and we write it as

$$U \sim SN_p(\bar{\mathbf{\Omega}}, \alpha),$$

where $\bar{\mathbf{\Omega}}$ is the scale matrix and it is a $p \times p$ correlation matrix, and α, a p-dimensional real vector, is the slant parameter. For simplicity, we will assume $\bar{\mathbf{\Omega}}$ to be of full rank. The shape of the distribution is regulated by the combined effects of $\bar{\mathbf{\Omega}}$ and α. When $\alpha = 0$, the distribution reduces to a multivariate normal distribution. The random variable U is called the 'normalized' multivariate skew normal variate (Azzalini and Capitanio, 2013; Azzalini and Capitanio, 1999).

However, for practical applications, a location and scale parameters are necessary. Azzalini and Capitanio (1999) introduced an affine transformation of U, as

$$Y = \xi + \omega U \ ,$$

so that the pdf of Y is

$$2\phi_p(Y;\xi,\mathbf{\Omega})\Phi(\alpha'\omega^{-1}(Y-\xi)).$$

Thus, Y follows the MSN distribution with parameters ξ, $\mathbf{\Omega}$ and α and it is written as

$$Y \sim SN_p(\xi,\mathbf{\Omega},\alpha),$$

where $\xi \in \mathbb{R}^p$ is the location parameter, $\alpha \in \mathbb{R}^p$ is the slant parameter and $\mathbf{\Omega}(p\times p) > \mathbf{0}$ is the scale matrix such that $\mathbf{\Omega} = \omega\bar{\mathbf{\Omega}}\omega$. The matrix ω is a diagonal matrix whose diagonal elements are the square root of the diagonal elements of $\mathbf{\Omega}$. Thus, $\mathbf{\Omega}$ is a covariance matrix and hence does not possess the limitations of a correlation matrix. The mean and variance of Y are

$$\begin{aligned} E(Y) &= \xi + \sqrt{\frac{2}{\pi}}\omega\delta = \mu, \text{ say, and,} \\ Var(Y) &= \mathbf{\Omega} - \frac{2}{\pi}\omega\delta\delta'\omega = \mathbf{\Sigma}, \text{ say,} \end{aligned}$$

where $\delta = \frac{1}{(1+\alpha'\bar{\mathbf{\Omega}}\alpha)^{\frac{1}{2}}}\bar{\mathbf{\Omega}}\alpha$, and δ is a p-dimensional vector whose every element is within the interval (-1,1). This is the general form of the MSN distribution and is more commonly used in practical applications (Azzalini and Capitanio, 2013; Azzalini and Capitanio, 1999; Balakrishnan and Scarpa, 2012; Genton et al., 2001; Gupta and Kollo, 2003; Gupta and Chang, 2003).

Azzalini and Dalla-Valle (1996) introduced a stochastic representation of the 'normalized' MSN variable. This stochastic representation will enable us to express the MSN variable as a mixture of two distributions which will then be utilized later, for solving estimation problems.

The components of the p-dimensional 'normalized' MSN variable, U can be expressed as a mixture of an univariate standardized half-normal variable, $\mid U_0 \mid$, and a multivariate normal variable, U_1, such as

$$U_j = \delta_j \mid U_0 \mid + \sqrt{1-\delta_j^2} U_{1j} \ , \ j = 1, 2, \ldots, p, \tag{2.13}$$

where U_{1j} is the j-th component of U_1 and δ_j are the components of the p-dimensional vector δ such that δ_j lies within the interval (-1,1), for all $j = 1, 2, \ldots, p$. The variables U_0 and $U_1(p \times 1)$ are independently distributed as

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{p+1} \begin{pmatrix} 0_{p+1}, & \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Psi} \end{pmatrix} \end{pmatrix},$$

where $\mathbf{\Psi}(p \times p)$ is a correlation matrix. The relationships between the original parameters, $\bar{\mathbf{\Omega}}$ and α, of the MSN distribution with its alternate parameters, $\mathbf{\Psi}$ and δ, are as follows:

$$\begin{aligned} \bar{\mathbf{\Omega}} &= \mathbf{\Delta}(\mathbf{\Psi} + \lambda\lambda')\mathbf{\Delta} = \mathbf{\Delta}\mathbf{\Psi}\mathbf{\Delta} + \delta\delta', \\ \alpha &= \frac{\lambda'\mathbf{\Psi}^{-1}\mathbf{\Delta}^{-1}}{(1+\lambda'\mathbf{\Psi}^{-1}\lambda)^{\frac{1}{2}}} = \frac{1}{(1-\delta'\bar{\mathbf{\Omega}}^{-1}\delta)^{\frac{1}{2}}}\bar{\mathbf{\Omega}}^{-1}\delta, \\ \mathbf{\Delta} &= diag(\sqrt{(}1-\delta_1^2), \sqrt{(}1-\delta_2^2), \ldots, \sqrt{(}1-\delta_p^2)), \\ \lambda(\delta) &= (\frac{\delta_1}{\sqrt{(}1-\delta_1^2)}, \frac{\delta_2}{\sqrt{(}1-\delta_2^2)}, \ldots, \frac{\delta_p}{\sqrt{(}1-\delta_p^2)}). \end{aligned} \tag{2.14}$$

Azzalini and Capitanio (1999) showed that there exists a pair $(\bar{\mathbf{\Omega}}, \alpha)$ for every choice of $(\mathbf{\Psi}, \lambda(\delta))$, and the MSN distribution is defined for all such choices, provided the choices conform to the respective parameter spaces (Azzalini and Capitanio, 1999). We remind the readers that the reason for introducing the alternate parametrization of the MSN distribution is to let our readers realize the fact that an MSN distribution can be expressed as a mixture of normal densities which will be utilized later for the purpose of parameter estimation.

2.6 Inferential Techniques used in this Thesis

2.6.1 Bayesian Inference

Bayesian inference is a process of inductive learning using Bayes' rule. Bayes' rule quantifies the relation between a set of rational beliefs more generally referred to as previous knowledge and introduced as a probability model, with information from the data. Bayesian inference is an important tool for statistical inference and imputation such as parameter estimation, hypothesis testing, model estimation, selection and imputation of missing data (Hoff, 2009).

The advantage of using Bayesian inference is that it updates information from the data in the light of some new information. In fact, it uses all the information from the data and also takes into account some previous state of knowledge or ignorance about the parameter, in terms of a probability model, called the prior distribution, and uses them both to calculate the posterior distribution, and consequently makes inferences on the parameters. The guiding principle of Bayesian theory is

$$\text{posterior distribution} \propto \text{ likelihood } \times \text{ prior distribution.}$$

The likelihood function represents information about the unknown parameters from the observed data and the prior distribution represents prior knowledge of the unknown parameters (Box and Tiao, 2011).

The problem regarding choice of estimators and confidence intervals which arise in sampling theory are not encountered in Bayesian theory (Box and Tiao, 2011). Bayes' rule has been mathematically proven to be an optimal method of updating previous knowledge i.e., prior distribution, using information from the data (Cox, 1946, 1961; Savage, 1954, 1972). Also Bayesian methods provide a robust alternative for complicated statistical inferential problems where non-Bayesian methods have failed to provide solutions. Bayesian methods have been shown to perform pretty well for non-Bayesian purposes as well (Hoff, 2009). For further discussion on Bayesian and non-Bayesian approaches, the readers are referred to Little (2006) and Efron (2005).

A crucial step in Bayesian inference is the choice of prior distribution. The performance of the Bayesian estimates rely hugely on the selection of the prior distribution (Samaniego, 2010). Raiffa and Schlaifer (1961) outlines some desirable characteristics of the family of prior distributions. The first characteristic is that the family of prior distributions is analytically tractable, so that it is mathematically convenient to obtain a closed-form of the posterior distribution. It is also desired that the posterior distribution belongs to the same family of distributions as the prior (Raiffa and Schlaifer, 1961; Koch, 2007). This characteristic is called the 'closure' property of the family of prior distributions (Samaniego, 2010). This property ensures that if the prior distribution is analytically tractable, the posterior distribution is also analytically tractable (Koch, 2007; Raiffa and Schlaifer, 1961). Families of distributions which obey the closure property are called conjugate prior families.

It is worth mentioning that conjugacy is not limited to mathematical convenience guaranteeing a closed-form of the posterior distribution, rather, conjugacy also possesses practical importance in the sense that it is readily interpretable (Samaniego, 2010). Thus, a conjugate prior seems like a desirable choice of the prior distribution.

Following the steps described by Raiffa and Schlaifer (1961), a conjugate prior can be determined by first finding the sufficient statistic for a distribution and then treating the sample kernel as the prior kernel (Raiffa and Schlaifer, 1961). Thus, if $k(\cdot \mid \theta)$ (θ being the parameter of interest) is the kernel function on the reduced sample space y_S, which is the sample space of the sufficient statistic, then the kernel of the conjugate prior is $k(y_S \mid \cdot)$, where y_S is to be considered as the parameter on the state space Θ ($\theta \in \Theta$).

2.6.2 Expectation-Maximization (EM) Algorithm

To date, maximum likelihood estimation is a popular choice of estimation technique in most inferential problems in statistics. Since in reality, closed form solutions to likelihood equation are sometimes unobtainable, often statisticians employ numerical methods for solving likelihood equations. The most popular among such numerical methods are the Newton-Raphson method and the Expectation-Maximization (EM) algorithm. In fact, the EM algorithm remains a popular choice over the Newton-Raphson method or other numerical methods for estimation problems (Lange, 2013). In this thesis, we have used EM algorithm for the purpose of estimation, when closed form solution to the likelihood equation does not exist.

Applications of the EM algorithm are not merely limited to statistics, rather, it has several applications across various fields including statistics, engineering, medicine, commerce etc. (Yamaguchi and Watanabe, 2004). In fact, more than a thousand articles were published merely in 1991, which has applications of EM. Of these, only 15% were published in statistics journals (Meng, 1997). Meng and Pedlow also conducted a survey to collect literature on applications of EM and found that there were more than 1700 works available on more than 1000 subjects (Meng and Pedlow, 1992; Yamaguchi and Watanabe, 2004).

The EM algorithm is an iterative computational technique for estimating unknown parameters of a probability distribution. It is most commonly applied to missing data problems. But, it can also be successfully applied to situations where missing data is not directly evident, for example, random effects models, mixtures and convolutions, latent class models, etc. The EM algorithm was first proposed by Demspter, Laird and Rubin in their 1977 seminal paper. But, the idea of the EM algorithm have been in implementation even before the 1977 paper came up (McLachlan and Krishnan, 2007).

The EM algorithm involves two steps as the name suggests - the E-step and the M-step. We start with initial estimates of the parameters. In the E-step, we construct the complete data log-likelihood consisting of the observed data and the missing data. The expectation of the complete data log likelihood conditioned on the observed data is called the Q function. The E-step gets its name from the fact that the Q function is an expectation. In the M-step, the Q function is maximized and an estimate of the parameter is obtained. This updated estimate is substituted into the Q function and the E-step and the M-step are repeated until convergence. Since the EM algorithm guarantees a monotone function, so convergence is guaranteed, although it can be extremely slow at times (Gupta and Chen, 2011).

The beauty and importance of EM lies in its simplicity and stability. The EM algorithm guarantees that the value of the likelihood function is increasing or at least non-decreasing with each iteration. Hence, if each iteration does not improve the likelihood function, at least the algorithm does not make it worse (Yamaguchi and Watanabe, 2004).

Another significant advantage of the EM algorithm is that the final solution is relatively robust to the choice of initial value of the parameter of interest. This property saves considerable time involved in playing around with different choices of the initial value. The EM algorithm also guarantees that if the initial value satisfies the constraints of the parameter and belongs to its domain, then the solution automatically maintains all the constraints and belongs to the domain of the parameter. For example, if we are trying to estimate a variance-covariance matrix, and if a positive definite matrix is used as the initial value, then the positive definiteness of the solution is guaranteed at each step (Yamaguchi and Watanabe, 2004).

The solution of the EM algorithm is either the local maximum or the saddle point of the likelihood function (Boyles, 1983; Wu, 1983). Unlike other iterative estimation procedures, such as the Newton-Raphson method, the EM algorithm does not require calculation of the second partial derivative of the likelihood, which saves considerable computation time (Yamaguchi and Watanabe, 2004).

The EM algorithm is often criticized for its slow rate of convergence. To compare the speed of the EM algorithm with Newton-Raphson method, an estimation problem of multivariate data was considered, and the rate of convergence of the two procedures was compared. According to the results, although the Newton-Raphson method mostly failed to converge, the EM algorithm always converged to some value. In fact, whenever both the methods converged, the EM algorithm had converged a lot faster, even though the Newton-Raphson method required fewer iterations. This is because the time required for each iteration of the EM algorithm was way shorter than the time required for each iteration of the Newton-Raphson method (Yamaguchi and Watanabe, 2004).

However, it is to be noted that no concrete study has been done so far claiming the superiority or inferiority of the EM over Newton-Raphson in terms of speed of convergence. It is worth noting that the speed of convergence of the EM algorithm can be determined using the information matrices of the complete and incomplete data (Yamaguchi and Watanabe, 2004).

2.6.3 Restricted EM (REM) Algorithm

The EM algorithm, being computationally simple and numerically stable, has gained much popularity in providing maximum likelihood estimators. It has a wide range of applications including missing data problems (Dempster, Laird and Rubin, 1977; Little and Rubin, 2014). It can also be successfully applied to complete data problems where a closed-form solution of the maximum likelihood equation does not exist, and if the problem can be re-framed as a missing-data problem by introducing new variables (Kim and Taylor, 1995; Gupta and Chen, 2011; Lachos et al., 2007).

However, in the presence of linear restrictions on the model parameters, the M-step becomes complicated and a closed form solution may not always exist. So it becomes essential to use constrained maximization routines (Kim and Taylor, 1995). In 1995, Kim and Taylor proposed a modification of the EM algorithm, called the Restricted EM (REM) algorithm for solving maximum likelihood equations in the presence of linear restrictions on model parameters. The REM algorithm is an extension of the

EM algorithm when we have linear restrictions in the model.

There are two REM algorithms proposed by Kim and Taylor (1995). One that requires convergence with the Newton-Raphson method in its restricted maximization (RM) step and one that is computationally simpler and does not require any Newton-Raphson iteration in the RM step. In this thesis, we employ the simpler second approach to serve our purpose. The steps of the REM algorithm are summarized here. Suppose the parameter of interest is $\theta(p \times 1)$ and its m linear restrictions are represented as $\mathbf{A}\theta = a$, where $\mathbf{A}(m \times p)$ is a known linearly independent matrix with $rank(\mathbf{A}) = m < p$, and a is a known p-dimensional vector. Let $Q(\theta)$ represent the Q function obtained in the E-step and let $(\theta_R^{(0)}, \theta_R^{(1)}, \ldots, \theta_R^{(m)}, \theta_R^{(m+1)}, \ldots)$ be a sequence of solutions obtained using the REM algorithm. The first step in the restricted maximization (RM) step is calculating the observed information matrix, $\mathbf{I}_U$:

$$\mathbf{I}_U = -\frac{\partial^2 Q(\theta \mid \theta_R^{(m)})}{\partial \theta^2} \mid_{\theta=\theta_R^{(m)}} . \tag{2.15}$$

Let θ_U^+ be the unrestricted solution which is in closed form, usually obtained by solving the likelihood using the EM algorithm, by assuming that there are no restrictions on the model parameters. The RM step now involves finding the restricted solution for the $(m+1)$-th iteration, $\theta_R^{(m+1)}$. For that, we first calculate θ_R^+ using the equation

$$\theta_R^+ = \theta_U^+ + \mathbf{I}_U^{-1}\mathbf{A}'(\mathbf{A}\mathbf{I}_U^{-1}\mathbf{A}')^{-1}(a - \mathbf{A}\theta_U^+). \tag{2.16}$$

If $Q(\theta_R^+ \mid \theta_R^{(m)}) > Q(\theta_R^{(m)} \mid \theta_R^{(m)})$, then $\theta_R^{(m+1)} = \theta_R^+$. Otherwise, a step-halving procedure is used and the entire RM step has to be repeated.

The REM algorithm does not require any complicated constrained maximization procedure and is much convenient to use. Like the EM algorithm, the REM algorithm also guarantees a monotonically non-decreasing likelihood function with each iteration. The step-halving procedure ensures that the value of the Q function increases with each iteration. For more elaborate discussion on REM algorithm and its properties, the reader is referred to the original paper by Kim and Taylor (1995).

Chapter 3

Parameter Estimation for the GCM under Multivariate Skew Normal distribution

3.1 Introduction

Traditional literature assumes the error terms in a GCM to be distributed as a multivariate normal random variable. However, practical applications might require a more flexible approach that incorporates a broader class of skewed or non-normal data for GCMs. In situations where one encounters non-normal data, it is a common practice to transform data using a suitable transformation, with the aim of achieving normality or near normality. However, such a transformation, although suitable for ANOVA and MANOVA applications, may not be suitable for GCM applications, since such a transformation ignores the within-individual structure inherent in growth curve data. Moreover, the transformed multivariate data is difficult to interpret (Azzalini and Capitanio, 1999). For several such reasons, a transformation may not be recommended for a GCM when the underlying distribution is skewed or non-normal.

In this study, we extend growth curve models to incorporate skewed data and derive estimators for the unknown parameters. We assume the underlying distribution to be multivariate skew normal (MSN), which is a broader class of distributions and includes multivariate normal as a special case. The MSN distribution is a mathematical extension of the multivariate normal distribution with an additional skewness parameter (Azzalini and Capitanio, 1999; Genton et al., 2001). This parametric class possesses many properties of the normal family; on the other hand it takes the model fitting step further by allowing data from skewed distributions (Azzalini and Dalla-Valle, 1996; Azzalini and Capitanio, 1999; Gupta and Chang, 2003; Harrar and Gupta, 2008; Gupta and Kollo, 2003).

The primary objective of this chapter is to derive estimators for the unknown parameters of the GCM, under MSN errors. Please recall that estimation of unknown parameters of GCM has already been considered extensively in previous literature, including Potthoff and Roy (1964), Khatri (1966) and von Rosen (1989). However, all of them assumed the underlying distribution to be multivariate normal. To our knowledge, robustness of these estimators towards violation of the normality assumption have not been evaluated. Through an extensive simulation study presented in this chapter, we show that the estimators indeed lack robustness, where increased bias and mean squared error (MSE) are associated with increased skewness. Our simulations were restrict to illustrations involving Khatri's MLEs only, since we assumed full-rank design matrices. However, the results can be extended to von Rosen's estimators in

(2.9), without loss of generality, when the full-rank assumption is violated.

For the sensitivity analysis, we generated data from the MSN distribution, while we use the MLEs derived assuming normality. From the results of the sensitivity study presented in Section 3.4, we were able to infer that Khatri's MLEs do not seem to be appropriate in the presence of skewness in the dataset and hence a need arises for new estimators for skewed data.

We relax the assumption of normality by considering a GCM problem when the underlying distribution is MSN. The likelihood for the model was constructed and the estimators for the unknown mean and covariance parameters of GCM as well as an estimator for the skewness parameter of the MSN distribution were derived using EM algorithm. We conducted extensive simulations to demonstrate the performance of the new estimators and illustrated the practical usefulness of our estimators using a real-life data.

3.2 Likelihood of the GCM under the MSN Distribution

Consider the GCM under the MSN setup, where the distribution of the error terms is multivariate skew normal instead of multivariate normal. Let $\mathbf{E} = (E_1, E_2, \ldots, E_n)$, where E_j's are the columns of $\mathbf{E}$. The mean for each column is assumed to be zero and they all have the same variance-covariance matrix, $\mathbf{\Sigma}$. Without loss of generality, we assume each column is independent of each other, and is associated with same scale matrix, $\mathbf{\Omega}$, a covariance matrix and the same slant parameter, α, which a p-dimensional vector. Consequently, for all $j = 1, 2, \ldots, p$,

$$E_j \sim SN_p(\xi, \mathbf{\Omega}, \alpha),$$

where the value of the location parameter ξ can be determined from the equation, $\xi + \sqrt{\frac{2}{\pi}}\omega\delta = 0$ and therefore $\xi = -\sqrt{\frac{2}{\pi}}\omega\delta$. Moreover, from Section 2.5.2, we know that $\delta = \frac{1}{(1+\alpha'\bar{\mathbf{\Omega}}\alpha)^{\frac{1}{2}}}\bar{\mathbf{\Omega}}\alpha$ and ω is a diagonal matrix whose diagonal elements are the square root of the diagonal elements of $\mathbf{\Omega}$ such that $\mathbf{\Omega} = \omega\bar{\mathbf{\Omega}}\omega$, where $\bar{\mathbf{\Omega}}$ is a correlation matrix. It is to be noted that the columns of $\mathbf{E}$ are a location and scale shift of the normalized multivariate skew normal variate. Hence, they can be written as

$$E_j = \xi + \omega U_j, \tag{3.1}$$

where $U_j \sim SN_p(0, \bar{\mathbf{\Omega}}, \alpha)$. Now consider the columns of the response matrix, $\mathbf{Y}$, from the GCM setup, we have

$$\begin{aligned}
Y_j &= \mathbf{Z}b_1 + E_j, \text{ for } j = 1, \ldots, n_1, \\
Y_j &= \mathbf{Z}b_2 + E_j, \text{ for } j = (n_1+1), \ldots, n_2, \\
\ldots & \quad \ldots\ldots \\
Y_j &= \mathbf{Z}b_k + E_j, \text{ for } j = (n_1+n_2+\ldots+n_{k-1}+1), \ldots, n,
\end{aligned}$$

where the b_i's $(i = 1, 2, \ldots k)$ are the column vectors of the parameter matrix $\mathbf{B}$, and Y_j's $(j = 1, 2, \ldots, n)$ are the columns of the response matrix $\mathbf{Y}$. Hence, the columns of $\mathbf{Y}$ are a location shift of the columns of $\mathbf{E}$. The location shifts of Y_j's are determined by the group it belongs to and they are as follows:

$$\xi_1 = \mathbf{Z}b_1 + \xi, \text{ for, } j = 1(1)n_1,$$

$$
\begin{aligned}
\xi_2 &= \mathbf{Z}b_2, +\xi \text{ for, } j = (n_1+1)(1)n_2, \\
\dots & \quad \dots\dots \\
\xi_k &= \mathbf{Z}b_k, +\xi \text{ for, } j = (n_1+n_2+\ldots+n_{k-1}+1)(1)n.
\end{aligned}
$$

The MSN distribution is closed under affine transformations (Azzalini and Capitanio, 2013). Since the columns of $\mathbf{Y}$ are a location shift of the columns of $\mathbf{E}$, they also follow the MSN distribution with the same scale and slant parameter as those of $\mathbf{E}$. Thus, the columns of $\mathbf{Y}$ and $\mathbf{E}$ have the same dispersion matrix but different mean which are the conditions of GCM. The joint likelihood of $Y_1, Y_2, \ldots, Y_n$ will give us the required likelihood of the GCM under the MSN distribution:

$$
\begin{aligned}
L &= \frac{exp(-\frac{1}{2}tr[\mathbf{\Omega}^{-1}\{\sum_{j=1}^{n_1}(Y_j-\xi_1)(Y_j-\xi_1)' + \ldots + \sum_{j=n_1+\ldots+n_{k-1}+1}^{n}(Y_j-\xi_k)(Y_j-\xi_k)'\}])}{2^{-n}(2\pi)^{\frac{np}{2}} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}} \\
&\quad \times \prod_{j=1}^{n_1}\Phi[\alpha'\omega^{-1}(Y_j-\xi_1)] \ldots \prod_{j=n_1+n_2+\ldots+n_{k-1}+1}^{n}\Phi[\alpha'\omega^{-1}(Y_j-\xi_k)] \\
&= \frac{exp(-\frac{1}{2}tr[\mathbf{\Omega}^{-1}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U}\}'])}{2^{-n}(2\pi)^{\frac{np}{2}} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}} \\
&\quad \times\Phi_n[\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}'\omega^{-1}\alpha], \qquad (3.2)
\end{aligned}
$$

where $\mathbf{U} = (\xi, \xi, \ldots, \xi) = -\sqrt{\frac{2}{\pi}}\omega(\delta, \delta, \ldots, \delta)$. The second step is based on the fact that $Y_1, Y_2, \ldots, Y_n$ are independently distributed.

3.3 Estimation Procedure

Consider the likelihood of the GCM under the MSN distribution, as described in (3.2). Direct solution of the likelihood equation is not feasible, as a result, we resort to Expectation-Maximization (EM) algorithm to obtain parameter estimates.

To proceed with the EM algorithm in deriving our estimators, we utilize the stochastic representation of a MSN variable and express the distribution as a mixture of 2 distributions, as presented in Section 2.5.2. Earlier applications of the EM algorithm for obtaining estimators of the parameters of the SN distribution have been mentioned in Azzalini and Capitanio (1999, 2013), Lachos et al. (2007) and Lin et al. (2009). Azzalini has also developed software routines for calculating the parameter estimates of the MSN distribution. In this chapter, we take Azzalini's work one step further by considering the parameter estimation for the GCM under the MSN distribution. Recall that:

$$\begin{aligned} Y_j &= \mathbf{Z}b_i + E_j \\ &= \mathbf{Z}b_i + \xi + \omega U_j,\ j = 1(1)n,\ i = 1(1)k, \end{aligned}$$

where $\xi = -\sqrt{\frac{2}{\pi}}\omega\delta$ and $U_j \sim SN_p(0, \bar{\mathbf{\Omega}}, \alpha)$. Hence, U_j can be decomposed as in (2.13). Thus, we can rewrite Y_j as a mixture of two distributions:

$$Y_j = \mathbf{Z}b_i + \xi + \omega(\delta \mid t \mid + \mathbf{\Delta}\ T_j),\ i = i(1)k,\ j = 1(1)n, \tag{3.3}$$

where t is an univariate standard normal distribution so that $\mid t \mid$ is a standard half normal distribution, T_j is a multivariate normal distribution with mean 0 and variance $\mathbf{\Psi}$, which is a correlation matrix. It is assumed that t and T_j are independent. The relationship between the terms $\bar{\mathbf{\Omega}}$ and ψ have already been explained in (2.14). Since Y_j's ($j = 1(1)n$) represent the columns of $\mathbf{Y}$, (3.3) can be expressed using a matrix setup as

$$\mathbf{Y} = \mathbf{ZBX} + \mathbf{U} + \omega(\delta t_0' + \mathbf{\Delta T})$$

$$= \mathbf{ZBX} + \mathbf{U} + \omega_\delta t_0' + \mathbf{WT},$$

where $\mathbf{U} = -\sqrt{\frac{2}{\pi}}\omega(\delta, \delta, \ldots, \delta)$, $\mathbf{T} = (T_1, T_2, \ldots, T_n)$, the matrices $\mathbf{X}$, $\mathbf{B}$, $\mathbf{Z}$ are the same as defined in (2.1), and $\boldsymbol{\Delta}$ is as introduced in (2.14). Also,

$$t_0 \sim HN_p(0, \mathbf{I_p}),$$

$$\mathbf{T} \sim N_{p\times n}(\mathbf{0}, \boldsymbol{\Psi}, \mathbf{I}),$$

and $\omega_\delta = \omega\delta$, $\mathbf{W} = \omega\boldsymbol{\Delta}$ and HN_p denoting half-normal distribution. Since t_0 and $\mathbf{T}$ are structurally independent of each other, hence

$$\mathbf{Y} \mid t_0 \sim N_{p\times n}(\mathbf{ZBX} + \mathbf{U} + \omega_\delta t_0', \boldsymbol{\Gamma}, \mathbf{I}),$$

where $\boldsymbol{\Gamma} = \mathbf{W}\boldsymbol{\Psi}\mathbf{W}$. $\mathbf{Y}$ is our observed variable and the parameter of interest is $\theta = (\boldsymbol{\Omega}, \mathbf{B}, \delta)$. To frame an EM problem, we require missing data and t_0 serves that purpose. Thus, the complete data is $\mathbf{Y}_c = (\mathbf{Y}', t_0)$ and the complete data log-likelihood is

$$\begin{aligned} l(\theta \mid \mathbf{Y}_c) &= l(\theta \mid (\mathbf{Y}' \mid t_0)) + l(\theta \mid t_0) \\ &= -\frac{1}{2}tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')'\boldsymbol{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0') \\ &\quad -\frac{n}{2}log \mid \boldsymbol{\Gamma} \mid -\frac{1}{2}t_0' t_0. \end{aligned}$$

Now to obtain the Q function, we require the complete data log likelihood conditioned on the observed data $\mathbf{Y}$:

$$\begin{aligned} Q(\theta \mid \hat{\theta}) &= -\frac{1}{2}trE[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')'\boldsymbol{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0') \mid \mathbf{Y}] \\ &\quad -\frac{n}{2}log \mid \boldsymbol{\Gamma} \mid -\frac{1}{2}E(t_0' t_0 \mid \mathbf{Y}) \end{aligned} \tag{3.4}$$

$$\begin{aligned} &= -\frac{n}{2} log \mid \mathbf{\Gamma} \mid -\frac{1}{2} tr[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t}_0')'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t}_0')] \\ &\quad -\frac{1}{2} tr[(\widehat{t_0' t_0} - \hat{t}_0' \hat{t}_0)(\omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta)] - \frac{1}{2} \widehat{t_0' t_0}, \end{aligned}$$

where $\hat{t}_0 = E(t_0 \mid \mathbf{Y})$ and $\widehat{t_0' t_0} = E(t_0' t_0 \mid \mathbf{Y})$. Hence, we need to find the conditional distribution of t_0, given $\mathbf{Y}$:

$$\begin{aligned} f(t_0 \mid \mathbf{Y}) &= \frac{f(\mathbf{Y} \mid t_0) f(t_0)}{f(\mathbf{Y})} \\ &= \frac{e^{-\frac{1}{2} tr[\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0{}')(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0{}')']}}{\mid \Gamma \mid^{\frac{n}{2}} \; e^{-\frac{1}{2} tr[\mathbf{\Omega}^{-1}\{\mathbf{Y} - (\mathbf{ZBX} + \mathbf{U})\}\{\mathbf{Y} - (\mathbf{ZBX} + \mathbf{U})\}']}} \\ &\quad \times \frac{e^{-\frac{1}{2} t_0' t_0} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}}{\Phi_n[\{\mathbf{Y} - (\mathbf{ZBX} + \mathbf{U})\}' \omega^{-1} \alpha]}. \end{aligned}$$

Since $\mathbf{\Omega} = \omega \bar{\mathbf{\Omega}} \omega$, we can rewrite it as

$$\begin{aligned} \mathbf{\Omega} &= \omega(\mathbf{\Delta \Psi \Delta} + \delta \delta')\omega \\ &= \mathbf{\Gamma} + \omega_\delta \omega_\delta'. \end{aligned} \tag{3.5}$$

Using this relationship between $\mathbf{\Omega}$ and $\mathbf{\Gamma}$, we can easily find the relationship between their determinants and inverses as

$$\mid \mathbf{\Omega} \mid = \mid \mathbf{\Gamma} \mid (1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta) = k \mid \mathbf{\Gamma} \mid \text{, say,} \tag{3.6}$$

$$\mathbf{\Omega}^{-1} = \mathbf{\Gamma}^{-1} - (\frac{1}{1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta}) \mathbf{\Gamma}^{-1} \omega_\delta \omega_\delta' \mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^{-1} - (\frac{1}{k}) \mathbf{\Gamma}^{-1} \omega_\delta \omega_\delta' \mathbf{\Gamma}^{-1}, \tag{3.7}$$

where $k = 1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta$. Further, using (3.6) and (3.7) and after a few algebraic simplifications, the conditional distribution of t_0 given $\mathbf{Y}$ turns out to be a truncated multivariate normal. We summarize the steps here:

$$
\begin{aligned}
f(t_0 \mid \mathbf{Y}) &= \frac{e^{-\frac{1}{2}tr[\mathbf{\Gamma}^{-1}(\mathbf{Y}-\mathbf{ZBX}-\mathbf{U}-\omega_\delta {t_0}')(\mathbf{Y}-\mathbf{ZBX}-\mathbf{U}-\omega_\delta {t_0}')']}}{e^{-\frac{1}{2}tr[\mathbf{\Gamma}^{-1}\{\mathbf{I}-\frac{1}{k}\omega_\delta\omega_\delta'\mathbf{\Gamma}^{-1}\}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}']}} \\
&\quad \times \frac{e^{-\frac{1}{2}t_0't_0} \mid \mathbf{\Gamma} \mid^{\frac{n}{2}} \; k^{\frac{n}{2}}}{\mid \Gamma \mid^{\frac{n}{2}} \; \Phi_n[\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}'\omega^{-1}\alpha]} \\
&= \frac{k^{\frac{n}{2}} e^{-\frac{1}{2}tr[(\sqrt{k}t_0-\frac{1}{\sqrt{k}}(\mathbf{Y}-\mathbf{ZBX}-\mathbf{U})'\mathbf{\Gamma}^{-1}\omega_\delta)'(\sqrt{k}t_0-\frac{1}{\sqrt{k}}(\mathbf{Y}-\mathbf{ZBX}-\mathbf{U})'\mathbf{\Gamma}^{-1}\omega_\delta)}}{\Phi_n[\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}'\omega^{-1}\alpha]} \\
&= \frac{\phi_n(t_0 \mid \mu, \mathbf{\Sigma})}{1-\Phi_n[-\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}'\omega^{-1}\alpha]}.
\end{aligned}
$$

Hence, it can be concluded that, $t_0 \mid \mathbf{Y} \sim TN_n(\mu, \mathbf{\Sigma}, \mathbf{A})$, where, $\mu = \frac{1}{k}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U})'\mathbf{\Gamma}^{-1}\omega_\delta$, $\mathbf{\Sigma} = \frac{1}{k}\mathbf{I}$, $\mathbf{A} = \{t_0 : t_0 \geq a\}$ and TN_n denoting truncated normal distribution. The lower truncation point a can be obtained from the relationship $\mathbf{\Sigma}^{-\frac{1}{2}}(a-\mu) = -[\{\mathbf{Y} - (\mathbf{ZBX} + \mathbf{U})\}'\omega^{-1}\alpha]$. Hence, $a = \frac{1}{\sqrt{k}}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U})'\{\frac{1}{\sqrt{k}}\mathbf{\Gamma}^{-1}\omega_\delta - \omega^{-1}\alpha\}$. Since the support of t_0 is positive, the support of its conditional distribution also has to be positive. This requires a to be always positive. But, the expression of a does not guarantee that. Hence, we will assume a to be positive and for practical purposes we will replace all negative values of a with 0.

Now, that we know the distribution of $t_0 \mid \mathbf{Y}$, we can find out $\hat{t_0}$ and $\widehat{t_0't_0}$. Our next step is to find the estimator for θ by differentiating the Q function. Differentiating Q with respect to $\mathbf{\Gamma}$ yields

$$
\begin{aligned}
\frac{\partial Q}{\partial \mathbf{\Gamma}} &= \frac{\partial Q}{\partial \mathbf{\Gamma}^{-1}}\frac{\partial \mathbf{\Gamma}^{-1}}{\partial \mathbf{\Gamma}} \\
&= -\frac{1}{2}[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')' \\
&\quad -(\widehat{t_0't_0} - \hat{t_0}'\hat{t_0})\omega_\delta\omega_\delta']\mathbf{\Gamma}^{-2} - \frac{n}{2}\mathbf{\Gamma}^{-1}.
\end{aligned}
$$

Equating this expression to 0 gives

$$\hat{\mathbf{\Gamma}} = \frac{1}{n}[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')' - (t_0'\hat{t_0} - \hat{t_0}'\hat{t_0})\omega_\delta\omega_\delta']. \tag{3.8}$$

To obtain the estimator for the mean parameter, $\hat{\mathbf{B}}$, we find:

$$\begin{aligned} \frac{\partial Q}{\partial \mathbf{B}} &= -\frac{1}{2}\frac{\partial tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')} \\ &\quad \frac{\partial(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial \mathbf{B}} \\ &= -\frac{1}{2}[2(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')'\mathbf{\Gamma}^{-1}(\mathbf{X} \otimes \mathbf{Z}')] \\ &= -\mathbf{X}[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')']\mathbf{\Gamma}^{-1}\mathbf{Z}. \end{aligned}$$

Consequently, $\frac{\partial Q}{\partial \mathbf{B}} = 0$, yields the estimator for $\mathbf{B}$ as

$$\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{U} - \omega_\delta \hat{t_0}')\mathbf{X}'(\mathbf{XX}')^{-1}. \tag{3.9}$$

Similarly, to obtain $\hat{\delta}$,

$$\begin{aligned} \frac{\partial Q}{\partial \delta} &= -\frac{1}{2}\frac{\partial vec'(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial \delta} \\ &\quad \frac{\partial tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial vec(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')} \\ &\quad -\frac{1}{2}\frac{\partial tr(t_0'\hat{t_0} - \hat{t_0}'\hat{t_0})\omega_\delta'\mathbf{\Gamma}^{-1}\omega_\delta}{\partial \delta} \\ &= -\frac{1}{2}[-\frac{\partial vec'(\mathbf{U} + \omega_\delta \hat{t_0}')}{\partial \delta}2vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')\} \\ &\quad +\frac{\partial(t_0'\hat{t_0} - \hat{t_0}'\hat{t_0})\delta'(\omega\mathbf{\Gamma}^{-1}\omega)\delta}{\partial \delta}] \\ &= -\frac{1}{2}[-\frac{\partial vec'((-\sqrt{\frac{2}{\pi}})\omega\delta 1_n' + \omega\delta\hat{t_0}')}{\partial \delta}2vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')\} \\ &\quad +2(t_0'\hat{t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta] \end{aligned}$$

$$
\begin{aligned}
&= \{(-\sqrt{\frac{2}{\pi}})1_n \otimes \omega + \hat{t_0} \otimes \omega\}' vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')\} \\
&\quad -(\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= \{(-\sqrt{\frac{2}{\pi}})1_n + \hat{t_0}\}' \otimes \omega \; vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')\} \\
&\quad -(\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= vec\{\omega\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0})\} \\
&\quad -(\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= \{\omega\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0})\} - (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta.
\end{aligned}
$$

We now use the equation $\frac{\partial Q}{\partial \delta} = 0$ to derive an expression for $\hat{\delta}$, which after few simplifications turns out to be

$$
\begin{aligned}
&\omega\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX})(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}) \\
&= \omega\mathbf{\Gamma}^{-1}(\mathbf{U} + \omega_\delta \hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}) + (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= \omega\mathbf{\Gamma}^{-1}(-\sqrt{\frac{2}{\pi}}\omega\delta 1_n' + \omega_\delta \hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}) + (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= \omega\mathbf{\Gamma}^{-1}\omega\delta[(-\sqrt{\frac{2}{\pi}}1_n' + \hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}) + (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})] \\
\hat{\delta} &= \frac{1}{[\frac{2n}{\pi} - 2\sqrt{\frac{2}{\pi}}(\sum_{i=1}^{n} \hat{t_{0i}}) + \widehat{t_0' t_0}]}\omega^{-1}(\mathbf{Y} - \mathbf{ZBX})(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}). \qquad (3.10)
\end{aligned}
$$

The variance-covariance matrix, $\mathbf{\Sigma}$ is a function of δ and $\mathbf{\Omega}$, and hence is also a function of δ and $\mathbf{\Gamma}$. As a result, $\hat{\mathbf{\Sigma}}$ can be obtained using the invariance property of MLEs. A closer look at the expression of $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{B}}$ in (3.8) and (3.9), respectively, reveals that in the absence of skewness ($\delta = 0$), i.e., when the distribution is multivariate normal, the estimators reduce to Khatri's MLEs for $\mathbf{\Sigma}$ and $\mathbf{B}$, respectively. This indicates that the EM estimators reduce to Khatri's MLEs when the distribution is multivariate normal.

As in the computation of any EM algorithm, the choice of initial estimates is always crucial for the speed of convergence. Therefore, we need to choose our initial estimates appropriately. Since we are working with full rank design matrices, the MLE of $\mathbf{B}$ proposed by Khatri under normal distribution can be used as the initial estimator for $\mathbf{B}$, $\hat{\mathbf{B}}_{initial}$. The sample covariance matrix $\mathbf{S} = \mathbf{Y}(\mathbf{I} - \mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-}\mathbf{Z})\mathbf{Y}'$ serves as the initial estimator for the scale matrix $\mathbf{\Omega}$. Once we have $\hat{\mathbf{\Omega}}_{initial}$, we can then compute the matrix $\hat{\omega}_{initial}$. The p-dimensional vector with all its elements as 0.1, i.e., $(0.1, 0.1, \ldots, 0.1)$ was chosen as the initial estimator for δ. This choice of $\hat{\delta}_{initial}$ yielded positive definite values of $\hat{\mathbf{\Gamma}}_{initial}$, which was essential for the estimation procedure and a necessary condition for the subsequent iterations. With this choice of $\hat{\delta}_{initial}$, if $\hat{\mathbf{\Gamma}}_{initial}$ happened to be non positive definite, we used an even smaller value for $\hat{\delta}_{initial}$ to ensure the positive definiteness of $\hat{\mathbf{\Gamma}}_{initial}$.

3.4 Simulation Study

In this section, we present extensive simulations conducted to evaluate performance of our estimators and perform comparisons with existing estimators based on normality. A total of 3440 scenarios have been considered, where we generated data with different mean, different levels of correlations and skewness. We also present a simulation study conducted to investigate robustness of Khatri's MLE estimators towards departures from normality, the results of which contributed to the rationale for the objective of this study.

The range of parameter values considered in our simulation are presented in Table 3.1. Without loss of generality, the simulations presented in this chapter involve two groups ($k = 2$), number of time points $p = 4$ and $p = 8$, and a linear growth curve. However, we would like to highlight that these choices were made solely for simplicity of presenting the simulation results by reducing the magnitude of scenarios, since we are already considering a large number of scenarios.

Table 3.1: Range of parameters considered in our simulation

Number of groups, k	2
Growth curve	linear
Number of time-points, p	p=4, 8
Sample size, n	n=10, 15, 20, 25, 30, 35, 40 and $n > p$
Euclidean norm of B	0 to 7, increments of 0.1
Scale matrix, $\boldsymbol{\Omega}$	Wishart matrices with different levels of correlations
Slant vector, α	Any real p-dimensional vector
Euclidean norm of δ	0 to 1.5 increments of 0.25
Correlation structures	Weak, moderate and strong correlations
Distributions	Multivariate Normal and Multivariate Skew Normal

We would like to highlight that the estimators are derived under general conditions and the simulation findings are valid for GCMs involving one group or more, as well as higher degrees of polynomial. Since our focus is on estimation, we used bias and

MSE as performance evaluation criteria. Our parameters of interest ($\mathbf{B}$ and $\mathbf{\Sigma}$) are both matrices and hence their estimators are also matrices. As a result, we used matrix bias and matrix MSE as defined and presented in our earlier work (Jana et. al., 2017). To overcome the challenges matrix comparison bring, we used Euclidean norms of Bias and MSE matrices and reduce the matrices to scalar measures, as already discussed (Jana et al., 2017; Horn and Johnson, 1985).

The design matrices used are of the form

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & p \end{pmatrix} \text{ and } \mathbf{X} = \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix},$$

where n_i = number of sample units in the i^{th} group, $i = 1, 2$, and $n_1 = n_2 = n/2$ if n is even, and $n_1 = (n-1)/2$ and $n_2 = n_1 + 1$ if n is odd.

The scale matrices $\mathbf{\Omega}$ were generated from the Wishart distribution to ensure that they were positive definite and then divided by p to scale them down. Several 2×2 matrices were considered for the regression coefficient $\mathbf{B}$ and several p dimensional vectors were considered for the slant parameter α. The mean and the location for the two groups were determined. Then, the responses for the two groups were generated from the MSN distribution separately with different means (one for each group), but same scale matrix and same slant vector α. The n_1 responses generated for the first group and the n_2 responses generated for the second group were combined together as n columns to form the response matrix $\mathbf{Y}$.

In the empirical assessment of the MLEs obtained using EM algorithm, each data point (matrix $\mathbf{Y}$) was calculated at each iteration using the GCM as a function of the three parameters of interest: $\mathbf{B}$, $\mathbf{\Omega}$ and δ and estimators were calculated using EM algorithm until convergence. This process also required us to calculate $\hat{t_0}$ and $\widehat{t_0' t_0}$.

The first order moment of $TN_n(\hat{\mu}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{A}})$ was used for $\hat{t_0}$ and 1000 observations were simulated from the same distribution using Gibbs sampler to obtain an estimate for $\hat{t_0' t_0}$. On an average, it required 2 to 3 iterations (for each data point) for convergence and the L^2 norm (Euclidean norm) for matrices and vectors were appropriately used as the convergence criteria, as is traditional for convergence of multivariate estimators in EM. We used simultaneous convergence of all three estimators $\hat{\mathbf{B}}$, $\hat{\mathbf{\Omega}}$ and δ as our criterion since the estimation problem was dealt and developed as a simultaneous estimation problem. The estimators were all inter-dependent and each iteration improved all three estimates simultaneously. The bias and MSE matrices were calculated for each data point. This was repeated independently 1000 times and the empirical bias and MSE matrices were calculated by averaging over the 1000 simulations. This yielded the empirical bias and MSE of the estimators for each scenario.

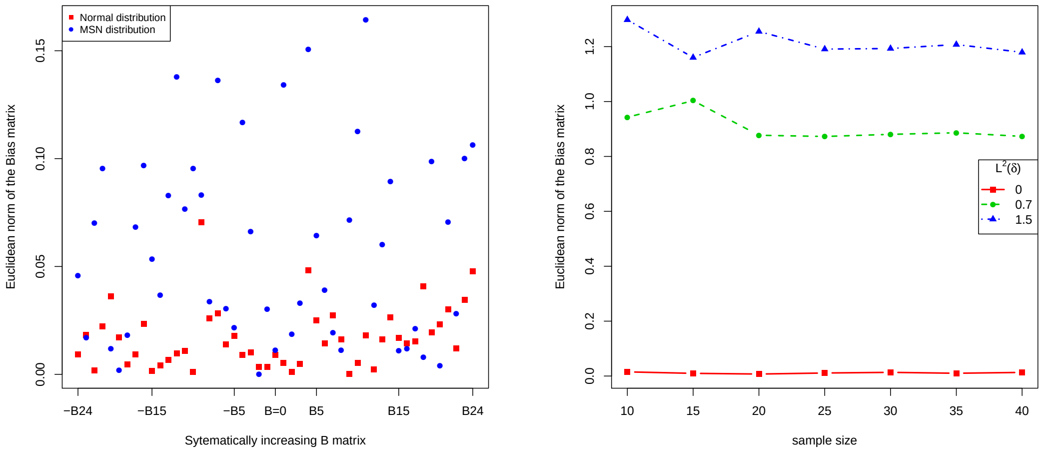
For the sensitivity analysis aimed at evaluating the robustness of Khatri's MLEs, data were generated from both the MSN distribution and the multivariate normal distribution, and Khatri's MLEs (under multivariate normal assumptions) were calculated for the unknown parameters for both cases. We were able to demonstrate how Khatri's MLEs were sensitive to skewness of the distribution.

3.4.1 Simulation Results for the Sensitivity Analysis

Here, data are generated from multivariate normal and skewed normal distributions. However, Khatri's MLEs, derived under multivariate normality are considered with the aim of assessing their robustness to departures from normality. As is evident from the graphs (Figure 3.1), the empirical bias of the MLE for $\mathbf{B}$ is larger for data from the skewed normal distribution compared to bias for data from the normal distribution.

This is true for all sample sizes and correlation strengths considered. It was indicated from previous studies that Bias for MLEs was not affected much by sample size (Jana 2013; Jana et al., 2017). The results here (Figure 3.1(b)) also indicate that sample size does not seem to have any systematic influence on Bias, both under normal and skewed normal distributions, regardless of the magnitude of skewness.

Nevertheless, empirical bias was observed to be uniformly more for data from skewed normal distributions for all sample sizes compared to bias for data from normal distributions. Our results also show that bias, in fact, increases with higher degrees of skewness. Note that skewness parameter $\delta = 0$ corresponds to the normal distribution.



(a) different parameter matrices (b) increasing n and different δ

Figure 3.1: Bias of Khatri's MLE $\hat{\mathbf{B}}$ for fixed variance-covariance matrix, against a) different $\mathbf{B}$, for n=20 with the x-axis representing an ordinal scale of parameter matrices, b) increasing n, fixed $\mathbf{B}$ and different degrees of skewness, δ.

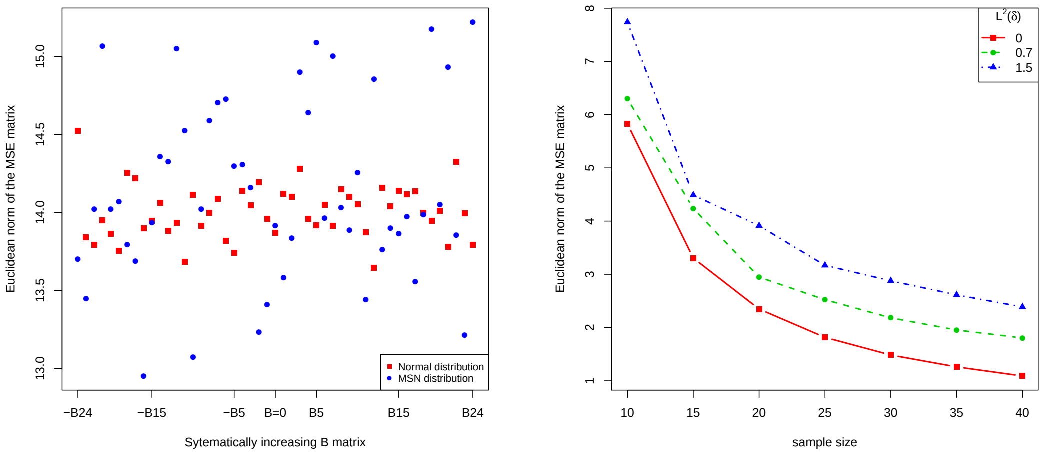
We would like to note that, in both Figures 3.1(a) and 3.1(b), we have presented the Euclidean norm of the Bias matrices on the y-axis and hence it is always positive. However, we also examined the individual elements of the Bias matrix, which are biases associated with each element of the matrix estimators. The results show that bias is indeed scattered around 0, indicating that we can encounter both positive and negative bias values.

Table 3.2: Sensitivity of the Khatri's estimator $\hat{\mathbf{B}}$, in terms of empirical Bias, to MSN errors against increasing Euclidean norm of $\mathbf{B}$, fixed $\mathbf{\Sigma}$.

	n=10		n=20		n=30	
$L^2(\mathbf{B})$	Normal	MSN	Normal	MSN	Normal	MSN
0.1	0.016	0.064	0.013	0.033	0.005	0.010
0.2	0.007	0.057	0.012	0.006	0.010	0.009
0.3	0.008	0.033	0.009	0.015	0.006	0.018
0.4	0.024	0.073	0.011	0.012	0.010	0.030
0.5	0.024	0.021	0.009	0.018	0.011	0.032
0.6	0.006	0.045	0.011	0.022	0.008	0.014
0.7	0.013	0.046	0.009	0.016	0.005	0.017
0.8	0.024	0.059	0.014	0.026	0.014	0.018
0.9	0.014	0.030	0.008	0.007	0.008	0.024
1	0.014	0.005	0.005	0.058	0.002	0.010
1.5	0.014	0.055	0.007	0.046	0.008	0.024
2	0.020	0.026	0.007	0.035	0.002	0.004
3.2	0.011	0.044	0.002	0.027	0.011	0.039
4	0.006	0.046	0.003	0.005	0.004	0.021
5.1	0.012	0.045	0.010	0.048	0.004	0.022
6.3	0.014	0.043	0.006	0.014	0.004	0.028
7.6	0.012	0.030	0.016	0.043	0.009	0.018

The empirical Bias of Khatri's MLEs for selected sample sizes and parameter matrices are presented in Table 3.2. From Table 3.2 and Figure 3.1(a), it can be seen that there is no systematic effect of the parameter matrix on bias, i.e., neither for its magnitude nor its direction. The results, therefore, indicate that bias is associated only with skewness of data.

Our results also show that Khatri's MLEs are associated with larger MSE for skewed data, as demonstrated in Figure 3.2(a), where the MSE for data from normal distribution are more clustered around 0, whereas MSE for data from the skewed distributions are more scattered. This is true for all the correlation structures and sample sizes we considered.



(a) different parameter matrices

(b) increasing n and different δ

Figure 3.2: MSE of Khatri's MLE $\hat{\mathbf{B}}$ for fixed $\boldsymbol{\Sigma}$, against a) different $\mathbf{B}$, for n=20 with the x-axis representing an ordinal scale of the parameter matrices, b) increasing n, fixed $\mathbf{B}$ and different degrees of skewness, δ.

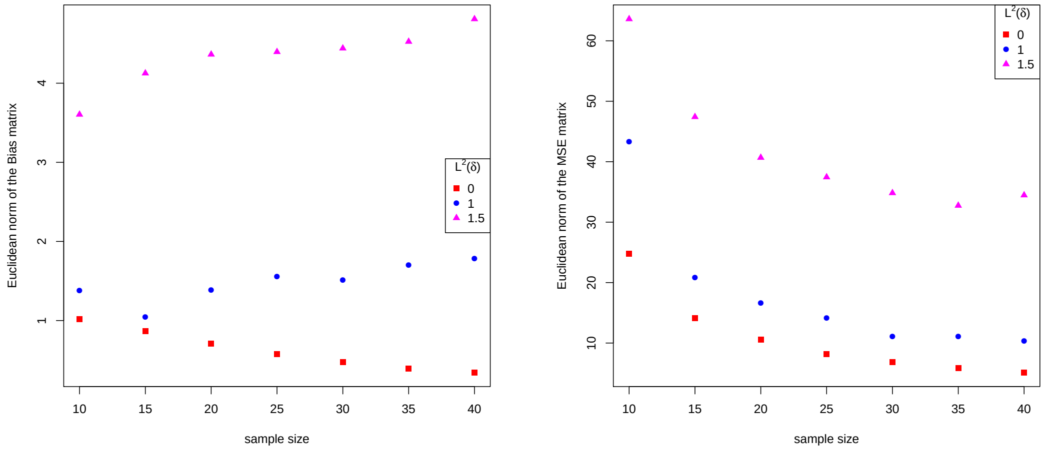
As shown in Figure 3.2(b), MSE of Khatri's MLEs for both normal and skewed normal data are monotonically decreasing with increasing sample size. Moreover, the results (Figure 3.2(b)) also show that MSE is uniformly smaller for data from normal distribution compared to MSE from a skewed distribution. Results also show that MSE increases with increased skewness across all sample sizes. MSE and bias of Khatri's MSE with increased skewness are presented in Table 3.4.

Table 3.3: Sensitivity of Khatri's estimator $\hat{\mathbf{B}}$, in terms of empirical MSE, to MSN errors against increasing Euclidean norm of $\mathbf{B}$, fixed $\mathbf{\Sigma}$.

	n=10		n=20		n=30	
$L^2(\mathbf{B})$	Normal	MSN	Normal	MSN	Normal	MSN
0.1	0.016	0.064	0.013	0.033	0.005	0.010
0.2	0.007	0.057	0.012	0.006	0.010	0.009
0.3	0.008	0.033	0.009	0.015	0.006	0.018
0.4	0.024	0.073	0.011	0.012	0.010	0.030
0.5	0.024	0.021	0.009	0.018	0.011	0.032
0.6	0.006	0.045	0.011	0.022	0.008	0.014
0.7	0.013	0.046	0.009	0.016	0.005	0.017
0.8	0.024	0.059	0.014	0.026	0.014	0.018
0.9	0.014	0.030	0.008	0.007	0.008	0.024
1	0.014	0.005	0.005	0.058	0.002	0.010
1.5	0.014	0.055	0.007	0.046	0.008	0.024
2	0.020	0.026	0.007	0.035	0.002	0.004
3.2	0.011	0.044	0.002	0.027	0.011	0.039
4	0.006	0.046	0.003	0.005	0.004	0.021
5.1	0.012	0.045	0.010	0.048	0.004	0.022
6.3	0.014	0.043	0.006	0.014	0.004	0.028
7.6	0.012	0.030	0.016	0.043	0.009	0.018

Similar to bias, no pattern of increase or decrease in MSE was observed with increasing or decreasing magnitude of the parameter matrix. The direction of $\mathbf{B}$ also did not seem to affect MSE, indicating that MSE is associated only with sample size and skewness of data. These findings are presented in Figure 3.2(a) and Table 3.3.

Similar findings were obtained for MLE of the variance-covariance matrix $\mathbf{\Sigma}$, where it is evident from Table 3.4 and Figure 3.3 that the empirical bias and MSE of Khatri's MLE are higher for skewed data compared to data from the normal distribution. In fact, empirical bias and MSE are uniformly larger for higher degrees of skewness. This is true for all the scenarios considered - across different sample sizes, different magnitude of the correlation matrices or different values of the mean parameter matrix $\mathbf{B}$.



(a) different parameter matrices (b) increasing n and different δ

Figure 3.3: Sensitivity of Khatri's MLE $\hat{\mathbf{\Sigma}}$, in terms of empirical Bias and MSE, to MSN distribution for fixed $\mathbf{\Sigma}$, increasing sample size, across varying degrees of skewness, a) Euclidean norm of Bias matrix b) Euclidean norm of MSE matrix.

In conclusion, our results demonstrate that the estimators derived under the assumption of multivariate normal errors are sensitive to departures from normality.

Table 3.4: Sensitivity of Khatri's estimators for $\mathbf{B}$ and $\boldsymbol{\Sigma}$, derived under multivariate normality, to MSN distribution, against increasing n, fixed $\mathbf{B}$, fixed $\boldsymbol{\Sigma}$ and different degrees of skewness, δ.

$\widehat{\mathbf{B}}$						
n	Bias			MSE		
	$L^2(\delta) = 0$	$L^2(\delta) = 1$	$L^2(\delta) = 1.5$	$L^2(\delta) = 0$	$L^2(\delta) = 1$	$L^2(\delta) = 1.5$
10	0.015	0.943	1.298	5.827	6.297	7.738
15	0.010	1.003	1.160	3.306	4.238	4.489
20	0.007	0.877	1.255	2.347	2.948	3.913
25	0.011	0.873	1.191	1.816	2.523	3.169
30	0.013	0.880	1.193	1.484	2.183	2.878
35	0.010	0.886	1.208	1.260	1.954	2.614
40	0.013	0.873	1.179	1.090	1.804	2.387
$\widehat{\boldsymbol{\Sigma}}$						
n	Bias			MSE		
	$L^2(\delta) = 0$	$L^2(\delta) = 1$	$L^2(\delta) = 1.5$	$L^2(\delta) = 0$	$L^2(\delta) = 1$	$L^2(\delta) = 1.5$
10	1.021	1.379	3.604	24.798	43.305	63.601
15	0.865	1.043	4.126	14.132	20.799	47.412
20	0.707	1.387	4.364	10.545	16.594	40.673
25	0.579	1.557	4.396	8.195	14.106	37.443
30	0.472	1.514	4.441	6.855	11.118	34.825
35	0.393	1.704	4.525	5.853	11.097	32.742
40	0.344	1.781	4.811	5.133	10.334	34.460

Our results also show that the magnitude of sensitivity depends on the level of skewness in the data. The results also indicate that the estimators are also sensitive under large sample size considerations, where both bias and MSE for MLEs are uniformly larger for skewed data compared to normal data. This lack of robustness of Khatri's MLEs provided the motivation for the current study to derive estimators for the unknown parameters of the GCM under the more general multivariate skewed normal distribution, which includes the multivariate normal distribution as its special case.

3.4.2 Simulation Results for EM based Estimators

Here, we consider the new estimators derived by assuming the error terms in the GCM to come from the multivariate skewed normal distribution. We first establish properties of the developed estimators by considering bias and MSE with respect to increasing sample size as well as increasing skewness and correlation. We present empirical bias and MSE for estimators of $\mathbf{B}$ and $\mathbf{\Sigma}$ in separate graphs.

In Table 3.5, we present empirical bias and MSE, of all parameter estimates against increasing skewness and across increasing correlation, for a fixed sample size and one scenario of regression coefficient, $\mathbf{B}$. It can be easily observed that bias and MSE of $\hat{\mathbf{B}}$ increases with increasing strengths of skewness and increasing correlation. In fact, Khatri's MLE of $\mathbf{B}$ and other estimators of $\mathbf{B}$, such as the BMLE or the MMLE in the high-dimensional cases, were also observed to have better performances with weaker correlation structure in the model (Jana 2013; Jana et al., 2017). This is true for all sample sizes and parameter matrices considered. Note that the special case of $L^2(\delta)$=0 represents Khatri's MLEs.

Similar findings follow for empirical bias and MSE of the estimator of variance-covariance matrix, $\hat{\mathbf{\Sigma}}$ and the skewness vector $\hat{\delta}$. The bias and MSE for both of them increase with increasing magnitude of skewness and increasing correlation (Table 3.5). When we take a look at the lower third of the table, we see that Bias and MSE for δ is not affected by increasing correlation for smaller values of δ. The situation changes for higher values of δ and we see a pattern of increase in empirical Bias and MSE with increasing correlation strengths. It is to be noted, however, that the empirical behavior of the estimator $\hat{\delta}$ is not affected by any increase in sample size.

Table 3.5: Empirical assessment of the EM estimators for $\mathbf{B}$, $\mathbf{\Sigma}$ and δ, against increasing strengths of correlation for $n = 20$, fixed $\mathbf{B}$, and across different degrees of skewness, δ.

$\widehat{\mathbf{B}}$

$L^2(\delta)$	Bias			MSE		
	weak	moderate	strong	weak	moderate	strong
0	0.0162	0.0271	0.0351	0.4368	1.1160	1.8148
0.76	0.0200	0.0376	0.0502	0.1714	0.6716	1.2390
1	0.1130	0.1455	0.1884	1.8920	2.5231	3.6822
1.6	0.2000	0.2170	0.2380	4.0797	5.0939	6.7408

$\widehat{\mathbf{\Sigma}}$

$L^2(\delta)$	Bias			MSE		
	weak	moderate	strong	weak	moderate	strong
0	0.4746	1.0175	1.5470	5.4247	24.4317	56.7073
0.76	0.1935	0.5446	0.9269	1.0061	7.1069	19.8240
1	1.4559	1.9319	2.7409	22.1342	28.1251	44.8315
1.6	2.9576	3.4166	4.1162	46.4971	54.0998	71.9302

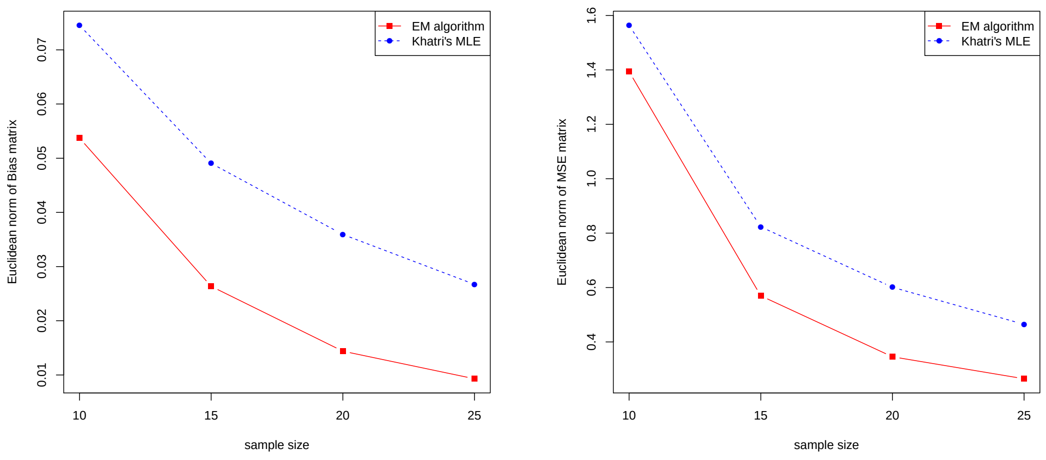
$\widehat{\delta}$

$L^2(\delta)$	Bias			MSE		
	weak	moderate	strong	weak	moderate	strong
0	0.0010	0.0010	0.0010	0.0000	0.0000	0.0000
0.76	0.7573	0.7573	0.7573	0.5735	0.5735	0.5735
1	4.7463	5.7450	6.7439	4.5654	5.5634	6.5614
1.6	8.2487	9.7569	11.2671	8.8583	11.1553	13.4523

Let us now take a look at the comparative assessment of new estimators and

Khatri's MLEs when data are skewed. In Figure 3.4, we presented the bias and MSE for both estimators for the regression coefficients, $\mathbf{B}$. The y-axis represents the Euclidean norm of the bias or the MSE matrices while the x-axis represents sample size. The two curves represent the two methods. Similar scenarios were considered using different values of the parameters $\mathbf{B}$, $\mathbf{\Omega}$ and δ, but we only present one here for the purpose of conciseness.

We can observe from the figures that the EM algorithm yielded estimators with less bias and MSE compared to Khatri's MLE for $\mathbf{B}$, when data are skewed. The relative efficiency of the EM estimators is on an average 50% more than Khatri's MLE for $\mathbf{B}$. We would like to highlight that the two estimators (EM and Khatri's MLE) are identical when data comes from multivariate normal distribution and hence comparative assessment was focused only on skewed data.



(a) Empirical Bias of $\mathbf{B}$ (b) Empirical MSE of $\mathbf{B}$

Figure 3.4: Comparative assessment of EM estimator of $\mathbf{B}$ with Khatri's MLE of $\mathbf{B}$, under an MSN GCM, a) empirical Bias, b) empirical MSE, against increasing n, fixed $\mathbf{B}$, $\mathbf{\Sigma}$ and δ.

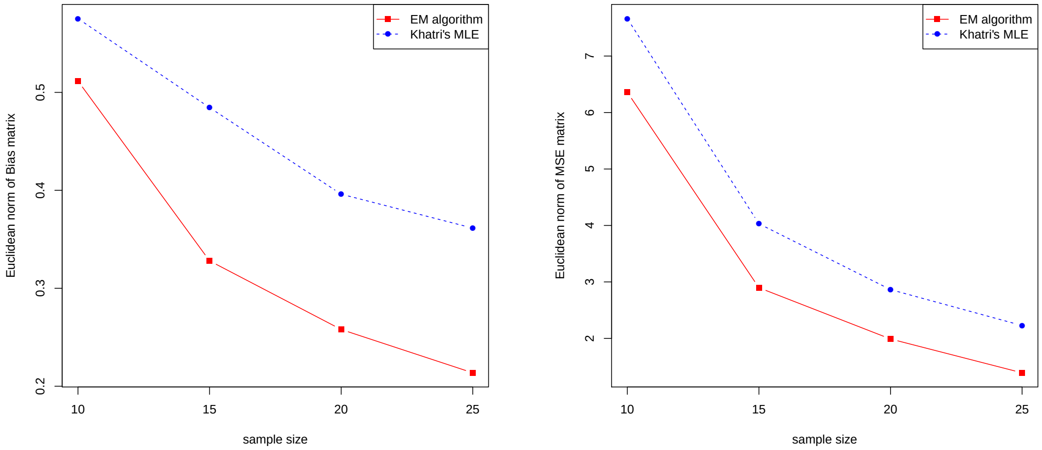
A few more results comparing the estimation of $\mathbf{B}$ under the two approaches are presented in Table 3.6. The results are consistent for other values of the slant parameter, α and different values of the mean parameter, $\mathbf{B}$ whether positive or negative, that is, empirical bias and MSE values were not impacted by a change in $\mathbf{B}$ or α. This demonstrates the robustness of $\hat{\mathbf{B}}$ with respect to the direction or magnitude of the true value of the parameter $\mathbf{B}$.

Table 3.6: Bias and MSE of estimators of $\mathbf{B}$ for a fixed correlation structure

Sample size	Bias		MSE	
n	EM estimators	Khatri's MLE	EM estimators	Khatri's MLE
10	0.0537	0.0787	1.3938	1.5659
15	0.0264	0.0396	0.5695	0.8555
20	0.0144	0.0307	0.5688	0.5857
25	0.0166	0.0267	0.4595	0.4636
30	0.0193	0.0227	0.3610	0.3816
35	0.0169	0.0317	0.3110	0.3144
40	0.0066	0.0258	0.2677	0.2881

Similar assessments follow for the variance-covariance matrix, $\mathbf{\Sigma}$, were we generated data from skewed distributions. The empirical bias and MSE of the estimators for $\mathbf{\Sigma}$ are plotted in Figures 3.5(a) and 3.5(b), respectively. The x-axis represents increasing sample size and the two curves represent two different methods. From the graphs, the MLEs from EM are shown to perform better than Khatri's MLEs in terms of empirical bias and especially MSE. In Figure 3.5(b), we can see that there is a one unit difference in MSE between the two approaches, which seems to be a considerable difference in performance. On an average, there is a remarkable 50% increase in relative efficiency, in terms of bias and MSE, of the EM estimator over

Khatri's MLE. Table 3.7 presents some more comparative results for the estimators of $\boldsymbol{\Sigma}$ under a few more sample sizes.



(a) Empirical Bias of $\boldsymbol{\Sigma}$ (b) Empirical MSE of $\boldsymbol{\Sigma}$

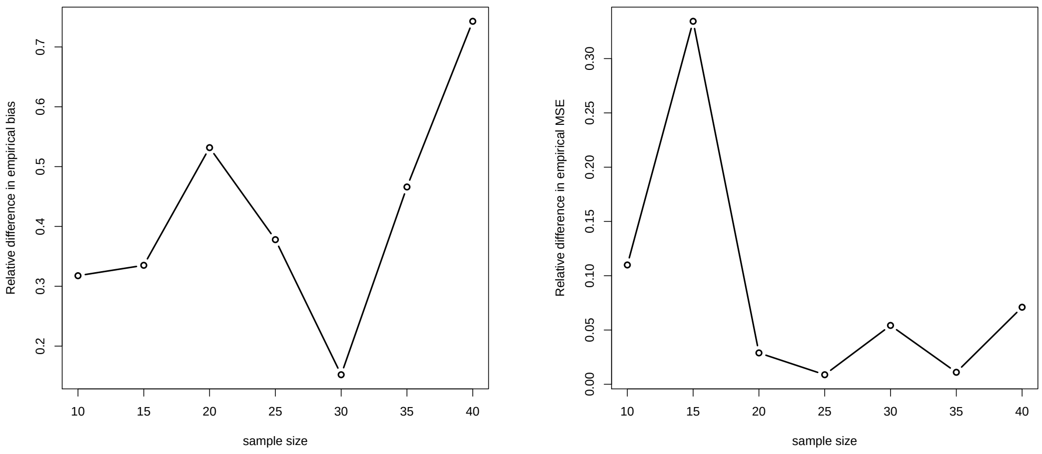
Figure 3.5: Comparative assessment of EM estimator of $\boldsymbol{\Sigma}$ with Khatri's MLE of $\boldsymbol{\Sigma}$, under an MSN GCM, a) empirical Bias b) empirical MSE, against increasing n, fixed $\mathbf{B}$, fixed variance-covariance matrix, $\boldsymbol{\Sigma}$ and fixed skewness vector, δ.

Table 3.7: Bias and MSE of estimators of $\boldsymbol{\Sigma}$ for a fixed correlation structure

Sample size	Bias		MSE	
n	EM estimators	Khatri's MLE	EM estimators	Khatri's MLE
10	0.511 6	0.522 3	6.358 8	7.662 8
15	0.419 3	0.440 2	3.884 6	4.030 5
20	0.341 9	0.422 0	2.827 7	2.874 1
25	0.293 7	0.323 4	2.300 7	2.331 6
30	0.261 1	0.274 8	1.874 3	1.888 3
35	0.193 5	0.226 7	1.631 4	1.637 5
40	0.200 8	0.207 9	1.391 2	1.426 9

It is worth mentioning here, since Khatri's MLEs are a special case of EM based estimators in the absence of skewness, therefore using Khatri's MLEs for parameter estimation under skewed data is equivalent to estimating the skewness level as 0, when it is actually far from 0. In our opinion, it is this underestimation of the skewness parameter that causes Khatri's MLEs to perform worse than EM estimators in the presence of skewness in the data.

Note that from Figures 3.4 and 3.5, it may appear to the reader that the difference in empirical bias and MSE of EM and Khatri's MLEs for $\mathbf{B}$ and $\mathbf{\Sigma}$ might vanish (or becomes negligible) asymptotically. However, we investigated the relative difference between the bias and MSE of the two methods for both $\mathbf{B}$ and $\mathbf{\Sigma}$, and the results indicate that the relative difference is in fact not monotonically decreasing.



(a) Empirical Bias of $\mathbf{B}$

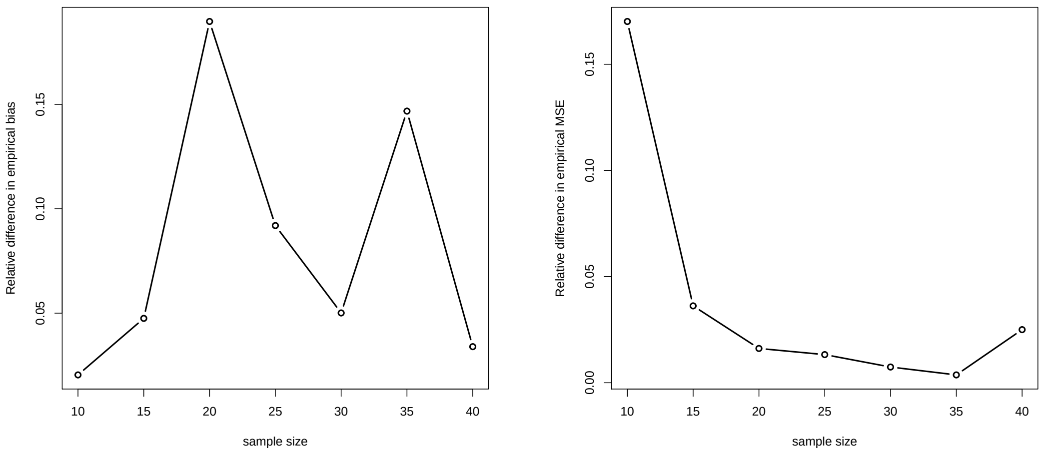
(b) Empirical MSE of $\mathbf{B}$

Figure 3.6: Relative difference in empirical bias and MSE of the two estimators of $\mathbf{B}$ for increasing sample sizes.

To demonstrate this we have plotted the relative difference between the bias and

MSE of the estimators for $\mathbf{B}$ and $\boldsymbol{\Sigma}$ in Figures 3.6 and 3.7. The results show that a positive difference in bias and MLE always exists even for large sample sizes indicating that when data is skewed, the EM based estimators developed under MSN distribution always provides better estimators than MLEs developed under normal assumptions.

As can be seen from Figures 3.6 and 3.7, a gap always exists between Khatri's MLEs and the MLEs obtained using EM algorithm for large sample sizes also. The relative difference and hence the difference also never vanishes asymptotically, rather it remains more or less the same with increasing sample size.



(a) Empirical bias of $\boldsymbol{\Sigma}$ (b) Empirical MSE of $\boldsymbol{\Sigma}$

Figure 3.7: Relative difference in empirical bias and MSE of the two estimators of $\boldsymbol{\Sigma}$ for increasing sample sizes.

3.5 Real Data Analysis

In this section, we illustrate the application of our estimation technique in practical problems involving skewed data. We use longitudinal data from Framingham Heart Study (FHS). FHS, established in 1948, is a notable major epidemiological study that has been undertaken for the purpose of identifying risk factors associated with cardiovascular disease (CVD) (Splansky et al., 2007; Hamid et al., 2009; Roslin et al., 2009).

The objective of the FHS is to understand the development of CVD over a long period of time among a large group of participants, and to identify environmental, demographic and genetic factors associated with increasing risk of CVD. FHS is a prospective study consisting of longitudinal data on variables associated with CVD. Using FHS, several factors including hypertension, diabetes, high blood cholesterol, age, gender have been identified as major risk factors associated with CVD.

We chose TG levels as our response variable, since triglyceride (TG) levels are known to be skewed (Frerichs et al., 1976; Lee et al., 2007). The demographic variable we consider in this illustration is sex, where we consider the GCM for two groups ($k = 2$) - males and females. It is to be noted that the results presented here are solely for illustration purposes and hence should not be considered for clinical interpretation.

Data consists of 708 individuals, among whom 352 were males and 356 were females. Measurements were taken across $p = 4$ time points for all individuals. Figures 3.8 and 3.9 represent histograms for marginal TG levels for both male and females, which indicate presence of skewness in the data. We also performed Mardia's test for multivariate skewness and the results show that there is significant skewness in the data (p-value < 0.0001).

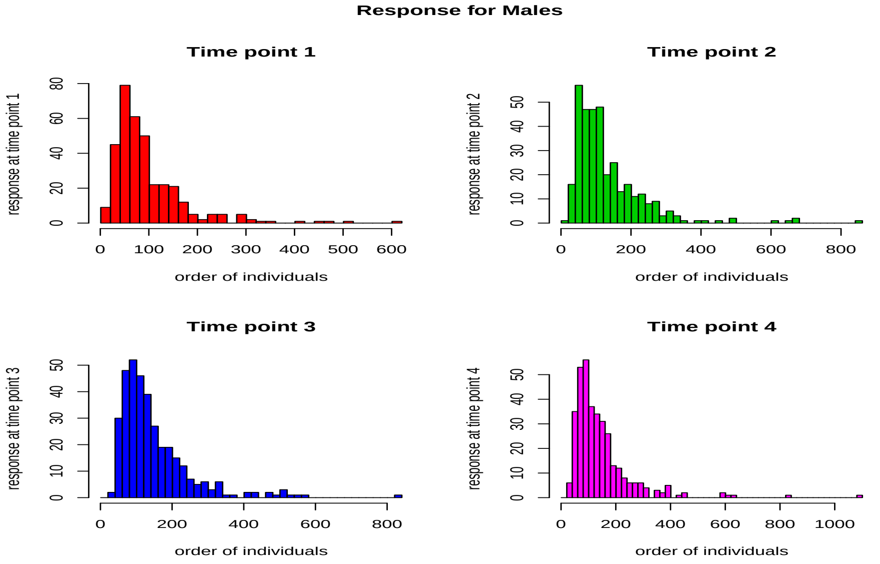


Figure 3.8: Histogram for the TG levels of 352 males from the FHS data for all the 4 time points.

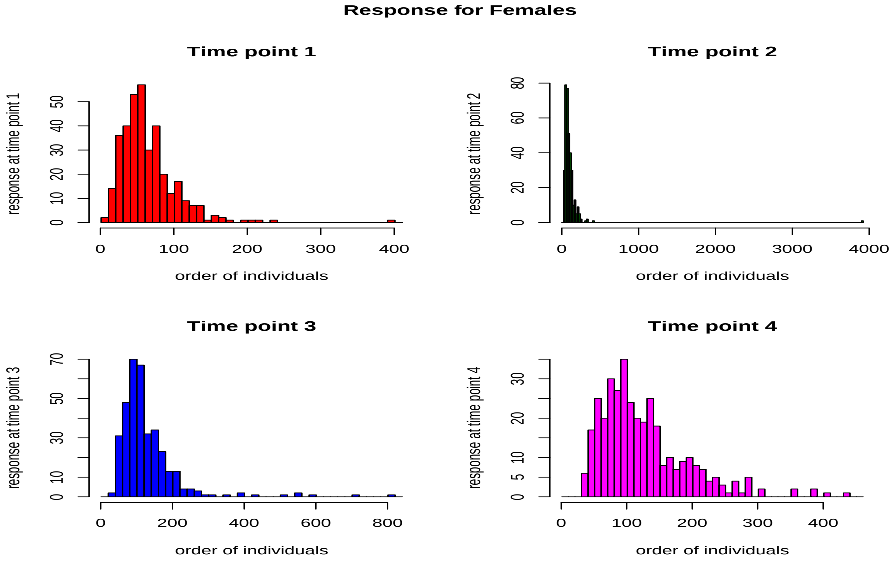


Figure 3.9: Histogram for the TG levels of 356 females from the FHS data for all the 4 time points.

In most practical problems involving longitudinal data, acquiring sample size of 708 is not often feasible. In order to mimic most real world longitudinal studies as well as illustrate small sample size estimation, we randomly selected data with sample sizes of 20, 30, 40 from the 708 observations (with equal allocation for males and females).

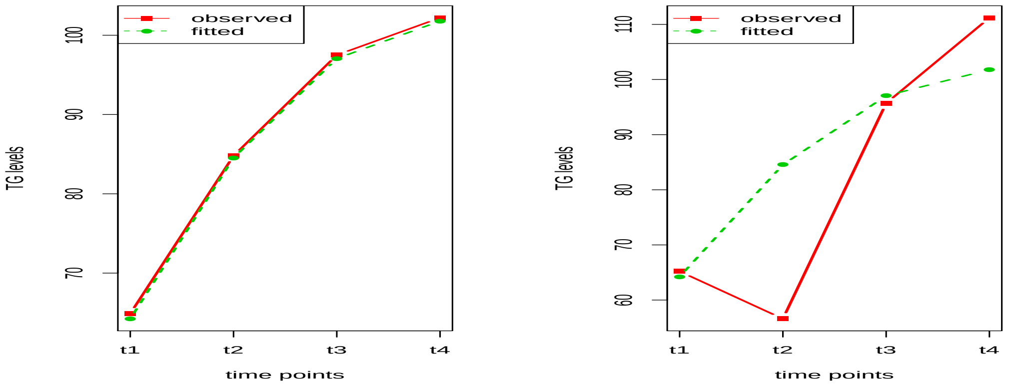
Before choosing a function to represent the relationship between average TG levels and time, we examined profile plots. Using the results from this preliminary investigation of profile plots, we chose a quadratic curve for mean for both males and females. We, therefore, fitted a quadratic GCM and estimated the parameters of interest using the EM algorithm as proposed in this chapter. The estimate for the matrix of regression coefficient, $\mathbf{B}$ for a sample size of 20 is:

$$\hat{\mathbf{B}} = \begin{pmatrix} 51.1671656 & 58.545609 \\ 14.0436242 & -3.264713 \\ -0.9731472 & 1.635385 \end{pmatrix}.$$

In order to investigate goodness of fit, we plotted the estimated mean ($\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\mathbf{B}}\mathbf{X}$) along with the observed mean ($\mathbf{YX}'(\mathbf{XX}')^{-1}\mathbf{X}$). Figure 3.10 indicates that the GCM with quadratic polynomial fits the data well for males, whereas the fit for females indicate that higher degrees are required to model the mean TG levels.

In order to confirm this, a cubic GCM with one group ($k = 1$) was fitted for the females alone and the results indeed show that the third degree polynomial is the best fit.

Note that the GCM does not allow groups with different degrees of polynomials to be fitted in one model (e.g., quadratic for males and cubic for females). Under such circumstances, the extended growth curve model can be used to provide estimators and make comparisons.



(a) Observed vs fitted response for males (b) Observed vs. fitted response for females

Figure 3.10: Observed vs. fitted response for $n = 20$, based on a quadratic fits.

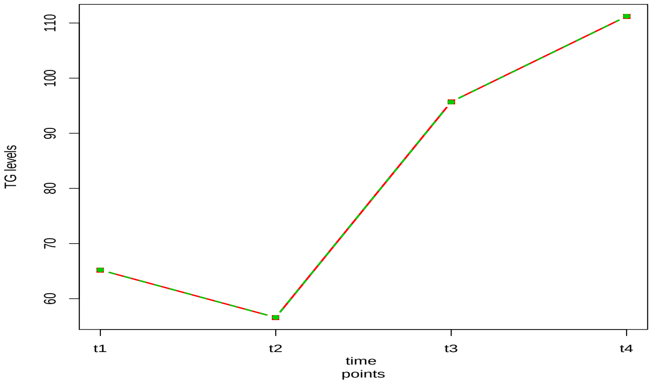


Figure 3.11: Observed vs. fitted response for a sample of size $n = 20$, for females based on a cubic curve.

3.6 Discussion

Motivated by lack of robustness of traditional estimators for GCMs involving skewed data, we proposed estimation techniques for the mean, variance and skewness parameters of GCM when the underlying distribution is MSN. The sensitivity study in Section 3.4.1, demonstrated the lack of robustness of the traditional estimators derived assuming normality i.e., MLEs of the mean and variance parameters provided by Khatri.

In this study, new MLEs were proposed for the unknown parameters of a MSN GCM using EM algorithm, since closed-form for the regular ML estimators were not available. We also examined the performance of these estimators using simulation studies.

Simulation results using skewed data show that bias of our estimates are randomly distributed around zero and decrease with increasing sample size, indicating asymptotic unbiasedness. The results also demonstrate that MSE decreases as sample size increases indicating consistency of the estimators. This is true for all sample sizes, correlations, as well as skewness levels considered.

Comparison with MLEs derived under the normality assumption show that the proposed EM estimators are associated with uniformly smaller bias and MSE for all the scenarios considered - regardless of sample size, correlation, as well as levels of skewness. The difference in bias and MSE between the normal and MSN based estimators remains more or less the same as the sample size increases, indicating that the MSN based estimators are recommended even for large sample size studies. However, the difference between our estimators and normal-based MLEs becomes larger as the skewness level in the data increases. When skewness is zero, the MSN reduces to the multivariate normal distribution and the two estimators are identical. The FHS data have been used for illustrating the practical relevance of the EM estimators proposed in this study.

Chapter 4

Parameter Estimation for the EGCM under Multivariate Skew Normal distribution

4.1 Introduction

The traditional EGCM introduced by von Rosen (1989), and further discussed by Kollo and von Rosen (2005), assumes the underlying distribution of the error terms to be multivariate normal (von Rosen, 1989; Kollo and von Rosen, 2005). In their book, Kollo and von Rosen (2005) discussed estimation techniques for the traditional EGCM, and provided estimators for its unknown parameters (Kollo and von Rosen, 2005). However, the normality assumption may often be violated in practice, as it is common to encounter skewed or non-normal data in different areas of applications (Arellano-Valle and Genton, 2005; Genton, 2004). Estimators developed by assuming normality may not, therefore, be optimal for skewed data. In fact, the previous chapter has established the lack of robustness of estimators of the GCM, developed under normality, to the presence of skewness in the data, indicating that such estimators are not optimal for skewed data. This finding can be generalized to the EGCM without loss of generality. Thus, a need arises to consider the EGCM from a more general perspective, with the aim of allowing a flexibility to model skewed data. Therefore, in this chapter, we consider the EGCM with multivariate skew normal (MSN) errors. To our knowledge, there is no existing literature on inference of skewed EGCMs.

Consider the real data illustration provided in Section 3.5, we see that the quadratic fit of the MSN GCM for males is near perfect (Figure 3.10(a)), although the fit is poor for females (Figure 3.10(b)), leading to large residual errors. The poor fit for females may not be due to the poor performance of the estimators under MSN GCM. Further investigation of data from females indicates that the mean growth for females can be better fit using a third degree polynomial.

The results (Figures 3.10(a) and 3.11), therefore, indicate that we require different polynomial fits for the two groups in this data. However, this is beyond the scope of a GCM, since the model does not allow different within individual matrices. That is, the model assumes equal degree of polynomials for all groups. An EGCM seems to be more appropriate for modelling this data.

In this chapter, we consider the likelihood for a special case of the EGCM (2 groups) and provide MLEs for the model parameters using the Restricted EM (REM) algorithm. Existing REM algorithm is based on the multivariate Newton-Raphson (NR) method, and both the algorithm and the multivariate NR method are defined only for vector parameters. We, therefore, extend the algorithm and the NR method for estimation of matrix parameters and implement them for solving the estimation problem in question. Extensive simulations were conducted to evaluate the performance of the proposed estimators.

4.2 Likelihood of the EGCM under the MSN Distribution

Similar to the GCM, an EGCM under an MSN setup, assumes that the columns of the error matrix, E_i $(i = 1, 2, \ldots, n)$ are independent and follow the MSN distribution with mean 0 and equal variances. Since the columns have the same dispersion matrix, for simplicity, we will assume the same scale matrix $\mathbf{\Omega}(p \times p)$ and same slant parameter, $\alpha(p \times 1)$ for all the n columns. We assume $\mathbf{\Omega}$ to be of full rank. Because the mean is 0, hence the location parameter is $\xi = -\sqrt{\frac{2}{\pi}}\omega\delta$, therefore,

$$E_i \sim SN_p(\xi, \mathbf{\Omega}, \alpha) \ , \ i = 1, 2, \ldots, n.$$

The columns of $\mathbf{Y}$, being a mean shift of the columns of $\mathbf{E}$, are just a location shift of the columns of $\mathbf{E}$, as the dispersion of $\mathbf{E}$ and $\mathbf{Y}$ are equal, according to the conditions of EGCM. Since the MSN distribution is closed under affine transformations, every column of $\mathbf{Y}$ is also an MSN variable with the same scale and slant parameters, $\mathbf{\Omega}$ and α, and differing solely in the location parameter. The location shift depends on the group it belongs:

$$\begin{aligned} Y_i &\sim SN_p(\xi_1, \mathbf{\Omega}, \alpha) \ , \ i = 1, 2, \ldots, n_1, \\ Y_i &\sim SN_p(\xi_2, \mathbf{\Omega}, \alpha) \ , \ i = n_1 + 1, \ldots, n, \end{aligned}$$

where $\xi_1 = \mathbf{Z}_1\beta_1^{(1)} + \xi$, and $\xi_2 = \mathbf{Z}_1\beta_2^{(1)} + \mathbf{Z}_2\beta_2^{(2)} + \xi$. The required likelihood of the problem is the joint likelihood of the columns of $\mathbf{Y}$, which is

$$\begin{aligned} L &= 2^n \frac{exp(-\frac{1}{2}tr[\mathbf{\Omega}^{-1}\{\sum\limits_{i=1}^{n_1}(Y_i - \xi_1)(Y_i - \xi_1)' + \sum\limits_{i=n_1+1}^{n}(Y_i - \xi_2)(Y_i - \xi_2)'\}])}{(2\pi)^{\frac{np}{2}} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}} \\ &\quad \times \prod_{i=1}^{n_1} \Phi[\alpha'\omega^{-1}(Y_i - \xi_1)] \prod_{i=n_1+1}^{n} \Phi[\alpha'\omega^{-1}(Y_i - \xi_2)]. \end{aligned}$$

Since all the columns $Y_1, Y_2, \ldots, Y_n$ are independently distributed, hence the likelihood can be rewritten as

$$\begin{aligned} L &= 2^n \frac{e^{-\frac{1}{2}tr[\mathbf{\Omega}^{-1}\{\mathbf{Y}-(\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1+\mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2+\mathbf{U})\}\{\mathbf{Y}-(\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1+\mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2+\mathbf{U}\}'])}}{(2\pi)^{\frac{np}{2}} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}} \\ &\quad \times \Phi_n[\{\mathbf{Y} - (\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1 + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2 + \mathbf{U})\}'\omega^{-1}\alpha], \end{aligned}$$

where $\mathbf{U} = (\xi, \xi, \ldots, \xi) = -\sqrt{\frac{2}{\pi}}\omega(\delta, \delta, \ldots, \delta)$. This form of the likelihood involves the nested subspace condition on the design matrices of the model: $\mathcal{C}(\mathbf{X}_i') \subset \mathcal{C}(\mathbf{X}_{i-1}')$ $(i = 2, 3, \ldots, k)$. It might be difficult, or even impossible, to maximize this likelihood, with the nested subspace condition, using usual maximum likelihood procedures. So instead, we express the EGCM as a GCM with a simple linear restriction, as we have already shown in (2.8). Then, the likelihood for EGCM reduces to

$$L = 2^n \frac{e^{(-\frac{1}{2}tr[\mathbf{\Omega}^{-1}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U})\}\{\mathbf{Y}-(\mathbf{ZBX}+\mathbf{U}\}'])}\Phi_n[\{\mathbf{Y} - (\mathbf{ZBX} + \mathbf{U})\}'\omega^{-1}\alpha]}{(2\pi)^{\frac{np}{2}} \mid \mathbf{\Omega} \mid^{\frac{n}{2}}}$$

subject to the constraint $\mathbf{GBH} = \mathbf{0}$, where $\mathbf{G}$ and $\mathbf{H}$ have already been defined in Section 2. This will enable us to employ the REM algorithm for the purpose of estimation.

4.3 Estimation

4.3.1 Extension of the REM Algorithm

The REM algorithm, introduced by Kim and Taylor (1995), is meant for deriving ML estimators for vector parameters. The parameters of interest in our case are matrices. Hence, a need arises to extend the REM algorithm, for matrix parameters.

The construction of the REM algorithm depends on the multivariate NR method. Hence, to extend the REM algorithm for matrix parameters, we first require an extension of the NR method for matrix parameters. To our knowledge, there exists no such extension. As a result, we will first develop an extension of the NR method for matrix parameters.

The NR method says that, for a multivariate parameter, the estimate at the $(l+1)$-th iteration is an improvement on the l-th iteration and the improvement is determined by the product of the score function and the inverse of the observed information matrix:

$$\theta^{(l+1)} = \theta^{(l)} + \mathbf{I}_\theta^{-1} S_\theta, \tag{4.1}$$

where S_θ is the score function and $\mathbf{I}_\theta$ is the observed information matrix for θ. The score function is the first order partial derivative of the likelihood function with respect to the parameter θ, and the observed information matrix is the negative of the second order partial derivative of the likelihood function or in other words negative of the first order partial derivative of the score function, if we assume iid observations.

To extend the NR method for matrix parameters, we require the product of the score function and the inverse of the information matrix. Consequently, we would expect that the $(l+1)$-th iteration has the form

$$\mathbf{B}^{(l+1)} = \mathbf{B}^{(l)} + \mathbf{I}_\mathbf{B}^{-1} S_\mathbf{B}, \tag{4.2}$$

where $\mathbf{B}$ is the parameter of interest. The score function $S_\mathbf{B}$, of matrix parameters is the first order partial derivative of the log likelihood with respect to the parameter $\mathbf{B}$:

$$S_\mathbf{B} = \frac{\partial l}{\partial \mathbf{B}}.$$

The log-likelihood being a scalar and $\mathbf{B}$ being a matrix of dimension $(q \times k)$, the score function is also a matrix of dimension $(q \times k)$. The observed information matrix, like its multivariate counterpart, is the negative of the second order partial derivative of the log-likelihood, or the negative of the partial derivative of the score function:

$$\mathbf{I}_\mathbf{B} = -\frac{\partial^2 l}{\partial \mathbf{B}^2} = -\frac{\partial S_\mathbf{B}}{\partial \mathbf{B}}.$$

Therefore, the dimension of the information matrix is $(qk \times qk)$. Hence, $\mathbf{I}_{\mathbf{B}}^{-1}$ and $S_{\mathbf{B}}$ are non-conformable matrices, rendering equation (4.2) invalid.

There are two possible solutions to this problem, either we can vectorize the matrix parameter and then use the NR method for multivariate vector-valued parameters, or we can modify the information matrix so that it is conformable with the score function. If we use the first approach, the parameter becomes a vector of order qk and so is the dimension of the score function. Hence, we can go ahead with the iterative equation in (4.1):

$$vec(\mathbf{B})^{(l+1)} = vec(\mathbf{B})^{(l)} + \mathbf{I}_{\mathbf{B}}^{-1} vec(S_{\mathbf{B}}). \tag{4.3}$$

Let us now have a look at the structure of the information matrix. It is a derivative of the score function which is a function of the parameter, $\mathbf{B}$. Let us consider the most general form for the score function which would include all possible special cases. Let the score function be represented as

$$S_{\mathbf{B}} = \mathbf{KBN} + \mathbf{H},$$

where $\mathbf{K}(q\times q)$, $\mathbf{N}(k\times k)$ and $\mathbf{H}(q\times k)$ are matrices independent of $\mathbf{B}$. Now if we differentiate the score function with respect to $\mathbf{B}$, we would end up with (Kollo and von Rosen, 2005)

$$\frac{\partial S_{\mathbf{B}}}{\partial \mathbf{B}} = \mathbf{N} \otimes \mathbf{K}'.$$

We can easily find the inverse of the observed information matrix, using the properties of Kronecker product (Zhang and Ding, 2013),

$$\mathbf{I}_{\mathbf{B}}^{-1} = -(\mathbf{N}^{-1} \otimes \mathbf{K}'^{-1}).$$

Referring back to equation (4.3), we see that

$$\begin{aligned} vec(\mathbf{B})^{(l+1)} &= vec(\mathbf{B})^{(l)} - (\mathbf{N}^{-1} \otimes \mathbf{K}'^{-1}) vec(S_{\mathbf{B}}) \\ &= vec(\mathbf{B})^{(l)} - vec(\mathbf{K}'^{-1} S_{\mathbf{B}} \mathbf{N}'^{-1}) \\ &= vec(\mathbf{B}^{(l)} - \mathbf{K}'^{-1} S_{\mathbf{B}} \mathbf{N}'^{-1}). \end{aligned}$$

Converting back to matrices would result in

$$\mathbf{B}^{(l+1)} = \mathbf{B}^{(l)} - \mathbf{K}'^{-1} S_{\mathbf{B}} \mathbf{N}'^{-1},$$

which is our second proposed approach for developing an NR method for matrix parameters. As the second approach, we suggested a modification of the information matrix so that it is conformable with the score function. If the information matrix can be written as a Kronecker product of two matrices, such as

$$\mathbf{I} = \mathbf{I}_1 \otimes \mathbf{I}_2,$$

then the modified NR method for matrix parameters can be written as

$$\mathbf{B}^{(l+1)} = \mathbf{B}^{(l)} + \mathbf{I}_2'^{-1} S_{\mathbf{B}} \mathbf{I}_1'^{-1}. \tag{4.4}$$

It turns out that the matrix $\mathbf{I}_1^{-1}(k \times k)$ contains the within and between group variance terms and the matrix $\mathbf{I}_2^{-1}(q \times q)$ contains the variance of the polynomials.

$$\mathbf{I}_1^{-1} = \begin{pmatrix} Var(\beta_1) & Cov(\beta_1, \beta_2) \\ Cov(\beta_1, \beta_2) & Var(\beta_2) \end{pmatrix},$$

where $\beta_1 = (\beta_1^{(1)}, \beta_1^{(2)})$, are the coefficients in group 1, and, similarly, $\beta_2 = (\beta_2^{(1)}, \beta_2^{(2)})$ are the coefficients in group 2. Thus, $Var(\beta_1)$ and $Var(\beta_2)$ are the variances within the respective groups and the covariance term and $Cov(\beta_1, \beta_2)$ captures the covariance between the two groups. The terms in $\mathbf{I}_2^{-1}$ are the variance and covariances between the polynomials, such as $\beta_{0\cdot} = (\beta_{01}, \beta_{02})$, is the vector of intercepts, and $\beta_{1\cdot} = (\beta_{11}, \beta_{12})$, is the vector of slopes, and so on. Thus, the term $Var(\beta_{0\cdot})$ measures the variance within the intercepts and the term $Cov(\beta_{0\cdot}, \beta_{1\cdot})$ measures the covariance between the slopes and the intercepts, and so on.

$$\mathbf{I}_2^{-1} = \begin{pmatrix} Var(\beta_{0\cdot}) & Cov(\beta_{0\cdot}, \beta_{1\cdot}) & \dots & Cov(\beta_{0\cdot}, \beta_{(q-1)\cdot}) \\ \dots\dots & & & \\ \dots\dots & & & \\ Cov(\beta_{0\cdot}, \beta_{(q-1)\cdot}) & Cov(\beta_{1\cdot}, \beta_{(q-1)\cdot}) & \dots & Var(\beta_{(q-1)\cdot}) \end{pmatrix},$$

These structures of the information matrices indicate the positive definite nature of the matrices and hence ensure that they are invertible. Since the inverses of both the information matrices are symmetric, as is evident from their interpretation, we can rewrite (4.4) as

$$\mathbf{B}^{(l+1)} = \mathbf{B}^{(l)} + \mathbf{I}_2^{-1} S_{\mathbf{B}} \mathbf{I}_1^{-1}, \tag{4.5}$$

which is a possible extension of the NR method for imputing matrices using iterative equations. Now, that there is a possible extension of the NR method for matrix parameters, let us construct an extension of the REM algorithm by Kim and Taylor (1995), for matrix parameters.

Let us consider the estimation problem for EGCM. In any EGCM problem, to obtain an estimator for the parameter matrix $\mathbf{B}$, we need to maximize the log of the likelihood in (4.1) with respect to $\mathbf{B}$. This form of the likelihood involves the constraint $\mathbf{GBH} = \mathbf{0}$. Hence, we will try to maximize $l(\mathbf{B} \mid \mathbf{Y}) = log(L)$ subject to this constraint. A popular and convenient approach for solving such a problem, for both univariate and multivariate problems, is using the Lagrangian optimization procedure (Bertsekas, 1999). So, we will try to optimize the function $l_R(\mathbf{B} \mid \mathbf{Y}, \lambda)$ which is

$$l_R(\mathbf{B} \mid \mathbf{Y}, \lambda) = l_U(\mathbf{B} \mid \mathbf{Y}) + \lambda'(\mathbf{GBH} - \mathbf{0}),$$

where $\lambda(m \times 1)$ is the Lagrangian multiplier and l_U is the unrestricted log-likelihood since it does not consider the constraint. The estimator obtained by maximizing l_U would be denoted as $\mathbf{B}_U$. Now our objective is to determine λ. From the above optimization function, we observe the following relationships

$$\begin{aligned} \mathbf{S}_R &= \mathbf{S}_U + (\lambda'\mathbf{G})'\mathbf{H}', & (4.6) \\ \mathbf{I}_R &= \mathbf{I}_U, & (4.7) \end{aligned}$$

where $\mathbf{S}_U$ and $\mathbf{I}_U$ are the score function and information matrix of the unrestricted solution, $\mathbf{B}_U$, and, $\mathbf{S}_R$ and $\mathbf{I}_R$ are the score function and information matrix of the restricted solution, $\mathbf{B}_R$. We assume $\mathbf{I}_R = \mathbf{I}_{(1)R} \otimes \mathbf{I}_{(2)R}$, $\mathbf{I}_U = \mathbf{I}_{(1)U} \otimes \mathbf{I}_{(2)U}$ and for simplicity we also assume, $\mathbf{I}_{(1)R} = \mathbf{I}_{(1)U}$, $\mathbf{I}_{(2)R} = \mathbf{I}_{(2)U}$ since $\mathbf{I}_R = \mathbf{I}_U$. Hence, the iteration step would be

$$\begin{aligned} \mathbf{B}_R^{(l+1)} &= \mathbf{B}_R^{(l)} + \mathbf{I}_{2(R)}^{-1}\mathbf{S}_R\mathbf{I}_{1(R)}^{-1} \\ &= \mathbf{B}_R^{(l)} + \mathbf{I}_{2(U)}^{-1}(\mathbf{S}_U + \mathbf{G}'\lambda\mathbf{H}')\mathbf{I}_{1(U)}^{-1} \\ &= \mathbf{B}_U^{(l)} + \mathbf{I}_{2(U)}^{-1}\mathbf{G}'\lambda\mathbf{H}'\mathbf{I}_{1(U)}^{-1}. \end{aligned}$$

Our objective now is to determine the value of λ. To obtain that, we use the identity from the constraint

$$\begin{aligned} \mathbf{G}\mathbf{B}_R^{(l+1)}\mathbf{H} &= 0, \\ \mathbf{G}(\mathbf{B}_U^{(l)} + \mathbf{I}_{2(U)}^{-1}\mathbf{G}'\lambda\mathbf{H}'\mathbf{I}_{1(U)}^{-1})\mathbf{H} &= 0, \end{aligned}$$

which yields the solution

$$\lambda = -(\mathbf{G}\mathbf{I}_{2(U)}^{-1}\mathbf{G}')^{-1}(\mathbf{G}\mathbf{B}_U\mathbf{H})(\mathbf{H}'\mathbf{I}_{1(U)}^{-1}\mathbf{H})^{-1}.$$

Hence, the restricted solution for $\mathbf{B}$ is

$$\mathbf{B}_R^{(l+1)} = \mathbf{B}_U^{(l)} - \mathbf{I}_{2(U)}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{I}_{2(U)}^{-1}\mathbf{G}')^{-1}(\mathbf{G}\mathbf{B}_U\mathbf{H})(\mathbf{H}'\mathbf{I}_{1(U)}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{I}_{1(U)}^{-1}. \tag{4.8}$$

In the $(l+1)$-th iteration, if the estimate $\mathbf{B}_R^{(l+1)}$ does not increase the value of the Q function, then we use the step-halving procedure, as suggested by Kim and Taylor (1995).

4.3.2 Estimation procedure

Our objective is to obtain ML estimators for the parameter, $\theta = (\mathbf{B}_1, \mathbf{B}_2, \mathbf{\Omega}, \delta)$. Since closed-form solution of the likelihood is not feasible, we need to use the EM algorithm to obtain the estimators. However, the likelihood in our case involves constraints, so instead we will use the REM algorithm, as stated earlier.

As mentioned in Section 3, in order to use the REM algorithm, we will first try to express the MSN distribution as a mixture of two distributions. Azzalini and Dalla-Valle have provided the stochastic representation of a 'normalized' MSN variable where they have expressed it as a mixture of two distributions (Azzalini and Dalla-Valle, 1996). Since Y_i's $(i = 1, 2, \ldots, n)$ are just a location and scale shift of the 'normalized' variable, hence their stochastic representations are

$$\begin{aligned} Y_i &= \xi_1 + \omega[(\delta t_{0i}) + \mathbf{\Delta} t_i] \ , \ i = 1, 2, \ldots, n_1, \\ Y_i &= \xi_2 + \omega[(\delta t_{0i}) + \mathbf{\Delta} t_i] \ , \ i = n_1 + 1, \ldots, n, \end{aligned}$$

such that in matrix setup they are expressed as

$$\begin{aligned} \mathbf{Y} &= \mathbf{ZBX} + \mathbf{U} + \omega(\delta t_0' + \mathbf{\Delta T}) \\ &= \mathbf{ZBX} + \mathbf{U} + \omega_\delta t_0' + \mathbf{WT} \end{aligned}$$

subject to $\mathbf{GBH} = \mathbf{0}$, where $\omega_\delta = \omega\delta$, $\mathbf{W} = \omega\mathbf{\Delta}$, $t_0' = (t_{01}, \ldots, t_{0n})$, $\mathbf{T} = (t_1, t_2, \ldots, t_n)$, and their stochastic distributions are

$$\begin{aligned} & t_0 \sim HN_p(0, \mathbf{I_p}), \\ & \mathbf{T} \sim N_{p\times n}(\mathbf{0}, \mathbf{\Psi}, \mathbf{I}). \end{aligned}$$

Since t_0 and $\mathbf{T}$ are structurally independent of each other, so

$$(\mathbf{Y} \mid t_0) \sim N_{p\times n}(\mathbf{ZBX} + \mathbf{U} + \omega_\delta t_0', \mathbf{\Gamma}, \mathbf{I}),$$

where $\mathbf{\Gamma} = \mathbf{W\Psi W}$.

Our parameter of interest in this study is $\theta = (\mathbf{B}, \mathbf{\Omega}, \delta)$. Since we wish to employ the REM algorithm to estimate the unknown parameter, θ, we need to re-frame the problem as a missing data problem. The conditional distribution of $\mathbf{Y} \mid t_0$ is a known stochastic distribution, hence the half-normal variable t_0 can be considered as the missing data in our estimation problem. The complete data, which is the observed and the missing data combined together, can be written as $\mathbf{Y}_c = (\mathbf{Y}, t_0)$, and the complete data log-likelihood is

$$\begin{aligned}
l_c(\theta \mid \mathbf{Y}_c) &= l(\theta \mid t_0) + lnf(\mathbf{Y} \mid t_0, \theta) \\
&= -\frac{1}{2} tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')' \mathbf{\Gamma}^{-1} (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0') \\
&\quad -\frac{1}{2} t_0' t_0 - \frac{n}{2} log \mid \mathbf{\Gamma} \mid .
\end{aligned}$$

According to the definition, the Q function for this problem is

$$\begin{aligned}
Q(\theta) &= E[l_c(\theta \mid \mathbf{Y}_c) \mid \mathbf{Y}] \\
&= -\frac{1}{2} trE[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')' \mathbf{\Gamma}^{-1} (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0') \mid \mathbf{Y}] \\
&\quad -\frac{n}{2} log \mid \mathbf{\Gamma} \mid -\frac{1}{2} E(t_0' t_0 \mid \mathbf{Y}) \\
&= -\frac{1}{2} tr[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t}_0')' \mathbf{\Gamma}^{-1} (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t}_0')] \\
&\quad -\frac{n}{2} log \mid \mathbf{\Gamma} \mid -\frac{1}{2} tr[(\widehat{t_0' t_0} - \hat{t}_0' \hat{t}_0)(\omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta)] - \frac{1}{2} \widehat{t_0' t_0},
\end{aligned}$$

where $\hat{t_0} = E[t_0 \mid \mathbf{Y}]$, and $\widehat{t_0' t_0} = E[t_0' t_0 \mid \mathbf{Y}]$. Therefore, we need to determine the conditional distribution of $t_0 \mid \mathbf{Y}$ to calculate the respective moments, and consequently determine the Q function which is the most essential factor in the E-step. We will use the identities (4.11) and (4.12), mentioned below, for ease of computation in determining the conditional distribution $f(t_0 \mid \mathbf{Y})$. To obtain those identities we first need to consider the relationship between $\mathbf{\Omega}$ and $\mathbf{\Gamma}$:

$$\begin{aligned} \mathbf{\Omega} &= \omega(\mathbf{\Delta\Psi\Delta} + \delta\delta')\omega \\ &= \mathbf{\Gamma} + \omega_\delta \omega_\delta'. \end{aligned} \tag{4.9}$$

Further, using (4.10) and properties of determinants and inverses, we can establish the following identities:

$$\mid \mathbf{\Omega} \mid = \mid \mathbf{\Gamma} \mid (1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta) = k \mid \mathbf{\Gamma} \mid \text{ and} \tag{4.10}$$

$$\begin{aligned} \mathbf{\Omega}^{-1} &= \mathbf{\Gamma}^{-1} - (\frac{1}{1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta}) \mathbf{\Gamma}^{-1} \omega_\delta \omega_\delta' \mathbf{\Gamma}^{-1} \\ &= \mathbf{\Gamma}^{-1} - \frac{1}{k} \mathbf{\Gamma}^{-1} \omega_\delta \omega_\delta' \mathbf{\Gamma}^{-1}, \end{aligned} \tag{4.11}$$

where $k = 1 + \omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta$. Now, we proceed to derive the conditional distribution of $t_0 \mid \mathbf{Y}$, which turns out to be truncated normal. The steps have been detailed here:

$$\begin{aligned} f(t_0 \mid \mathbf{Y}) &= \frac{f(\mathbf{Y} \mid t_0) f(t_0)}{f(\mathbf{Y})} \\ &= \frac{\mid \mathbf{\Omega} \mid^{\frac{1}{2}} \ e^{-\frac{1}{2} tr (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')' \mathbf{\Gamma}^{-1} (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta t_0')} e^{-\frac{1}{2} tr (t_0' t_0)}}{\mid \mathbf{\Gamma} \mid^{\frac{1}{2}} \ e^{-\frac{1}{2} tr (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U})' \mathbf{\Omega}^{-1} (\mathbf{Y} - \mathbf{ZBX} - \mathbf{U})} \Phi[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U})' \omega^{-1} \alpha]} \\ &= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2} tr (\mathbf{M} - \omega_\delta t_0')' \mathbf{\Gamma}^{-1} (\mathbf{M} - \omega_\delta t_0')} e^{-\frac{1}{2} tr (t_0' t_0)}}{e^{-\frac{1}{2} tr \mathbf{\Omega}^{-1} \mathbf{M} \mathbf{M}'} \Phi[\mathbf{M}' \omega^{-1} \alpha]} \\ &= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2} tr (\mathbf{\Gamma}^{-1} \mathbf{M}\mathbf{M}' - t_0' (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta) - (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta)' t_0 + \mathbf{\Gamma}^{-1} (\omega_\delta t_0')(\omega_\delta t_0')')} e^{-\frac{1}{2} t_0' t_0}}{e^{-\frac{1}{2} tr \mathbf{\Gamma}^{-1} (\mathbf{I_n} - \omega_\delta k^{-1} \omega_\delta' \mathbf{\Gamma}^{-1}) \mathbf{M}\mathbf{M}'} \Phi[\mathbf{M}' \omega^{-1} \alpha]} \\ &= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2} tr [-t_0' (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta) - (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta)' t_0 + (\omega_\delta' \mathbf{\Gamma}^{-1} \omega_\delta + 1)(t_0 t_0')]}}{e^{\frac{1}{2} tr (\mathbf{\Gamma}^{-1} \omega_\delta k^{-1} \omega_\delta' \mathbf{\Gamma}^{-1} \mathbf{M}\mathbf{M}')} \Phi[\mathbf{M}' \omega^{-1} \alpha]} \\ &= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2} tr [k (t_0 t_0') - (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta)' t_0 - t_0' (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta) + \frac{1}{k} (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta)' (\mathbf{M}' \mathbf{\Gamma}^{-1} \omega_\delta)]}}{\Phi[\mathbf{M}' \omega^{-1} \alpha]} \end{aligned}$$

$$
\begin{aligned}
&= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2}tr(k^{\frac{1}{2}}t_0 - k^{-\frac{1}{2}}(\mathbf{M}'\mathbf{\Gamma}^{-1}\omega_\delta)')'(k^{\frac{1}{2}}t_0 - k^{-\frac{1}{2}}(\mathbf{M}'\mathbf{\Gamma}^{-1}\omega_\delta)')}}{\Phi[\mathbf{M}'\omega^{-1}\alpha]} \\
&= k^{\frac{n}{2}} \frac{e^{-\frac{1}{2}tr(t_0 - k^{-1}(\mathbf{M}'\mathbf{\Gamma}^{-1}\omega_\delta)')'k(t_0 - k^{-1}(\mathbf{M}'\mathbf{\Gamma}^{-1}\omega_\delta)')}}{1 - \Phi[-\mathbf{M}'\omega^{-1}\alpha]},
\end{aligned}
$$

where $\mathbf{M} = \mathbf{Y} - \mathbf{ZBX} - \mathbf{U}$. Hence, $t_0 \mid \mathbf{Y} \sim TN_n(\mu, \mathbf{\Sigma}, \mathbf{A}_T)$, where $\mu = k^{-1}(\mathbf{M}'\mathbf{\Gamma}^{-1}\omega_\delta)'$, $\mathbf{\Sigma} = \frac{1}{k}\mathbf{I}_n$, $\mathbf{A}_T = \{x : x \geq a\}$, $a = \mu + \mathbf{\Sigma}^{\frac{1}{2}}(-\mathbf{M}'\omega^{-1}\alpha)$, and $\hat{t_0}$ is the first order moment of the truncated normal distribution. For computational purposes, we will use the empirical value for $\widehat{t_0' t_0}$. This completes our E-step. Now we proceed to derive the unrestricted estimators for each of the unknown parameters, $\mathbf{\Gamma}$, $\mathbf{B}$ and δ by maximizing the Q function with respect to each of them. The partial derivative of the Q function with respect to $\mathbf{\Gamma}$ is

$$
\begin{aligned}
\frac{\partial Q}{\partial \mathbf{\Gamma}} &= \frac{\partial Q}{\partial \mathbf{\Gamma}^{-1}} \frac{\partial \mathbf{\Gamma}^{-1}}{\partial \mathbf{\Gamma}} \\
&= -\frac{1}{2}[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')' - (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})\omega_\delta \omega_\delta']\mathbf{\Gamma}^{-2} \\
&\quad -\frac{n}{2}\mathbf{\Gamma}^{-1}.
\end{aligned}
$$

Equating $\frac{\partial Q}{\partial \mathbf{\Gamma}}$ to 0 yields the solution $\hat{\mathbf{\Gamma}}$ as

$$
\hat{\mathbf{\Gamma}} = \frac{1}{n}[(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')' - (\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})\omega_\delta \omega_\delta']. \tag{4.12}
$$

Similarly, to obtain $\hat{\delta}$, we first find the partial derivative $\frac{\partial Q}{\partial \delta}$:

$$
\begin{aligned}
\frac{\partial Q}{\partial \delta} &= -\frac{1}{2} \frac{\partial vec'(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial \delta} \\
&\quad \frac{\partial tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')}{\partial vec(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')} \\
&\quad -\frac{1}{2} \frac{\partial tr(\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})\omega_\delta' \mathbf{\Gamma}^{-1}\omega_\delta}{\partial \delta} \\
&= -\frac{1}{2}[-\frac{\partial vec'(\mathbf{U} + \omega_\delta \hat{t_0}')}{\partial \delta} 2vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta \hat{t_0}')\} \\
&\quad + \frac{\partial(\widehat{t_0' t_0} - \hat{t_0}'\hat{t_0})\delta'(\omega \mathbf{\Gamma}^{-1}\omega)\delta}{\partial \delta}]
\end{aligned}
$$

$$\begin{aligned}
&= -\frac{1}{2}[-\frac{\partial vec'((-\sqrt{\frac{2}{\pi}})\omega\delta 1_n' + \omega\delta\hat{t_0}')}{\partial\delta} 2vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')\} \\
&\quad +2(\widehat{t_0't_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta] \\
&= \{(-\sqrt{\frac{2}{\pi}})1_n \otimes \omega + \hat{t_0} \otimes \omega\}' vec\{\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')\} \\
&\quad -(\widehat{t_0't_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta \\
&= \{\omega\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0})\} - (\widehat{t_0't_0} - \hat{t_0}'\hat{t_0})(\omega\mathbf{\Gamma}^{-1}\omega)\delta,
\end{aligned}$$

where the solution to $\hat{\delta}$ could be obtained by equating the partial derivative $\frac{\partial Q}{\partial\delta}$ to 0. After a few simplifications, we obtain the estimator for δ as

$$\hat{\delta} = \frac{1}{[\frac{2n}{\pi} - 2\sqrt{\frac{2}{\pi}}(\sum_{i=1}^{n}\hat{t_{0i}}) + \widehat{t_0't_0}]}\omega^{-1}(\mathbf{Y} - \mathbf{ZBX})(-\sqrt{\frac{2}{\pi}}1_n + \hat{t_0}) \tag{4.13}$$

Now, to obtain the unrestricted solution for $\mathbf{B}$:

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{B}} &= -\frac{1}{2}\frac{\partial(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')}{\partial\mathbf{B}} \\
&\quad \frac{\partial tr(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')}{\partial(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')} \\
&= \mathbf{Z}'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{ZBX} - \mathbf{U} - \omega_\delta\hat{t_0}')\mathbf{X}'.
\end{aligned}$$

Consequently, $\frac{\partial Q}{\partial \mathbf{B}} = 0$, yields, $\hat{\mathbf{B}}_U$ as

$$\hat{\mathbf{B}}_U = (\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mathbf{U} - \omega_\delta\hat{t_0}')\mathbf{X}'(\mathbf{XX}')^{-1}. \tag{4.14}$$

These estimators, $\hat{\mathbf{\Gamma}}$, $\hat{\delta}$ and $\hat{\mathbf{B}}_U$ are the unrestricted estimators. As expected, the unrestricted estimators have the same form as the estimators obtained using the EM algorithm for GCM (Jana et al., 2017b). To obtain the final restricted estimators, we need to use the REM algorithm. However, the linear restriction for our model, $\mathbf{GBH} = \mathbf{0}$, does not involve the parameters $\mathbf{\Gamma}$ and δ. Hence the restricted estimators for $\mathbf{\Gamma}$ and δ remain the same as their unrestricted estimators, (4.12) and (4.13). However, the linear restriction ($\mathbf{GBH} = \mathbf{0}$) involves the mean parameter $\mathbf{B}$ and hence the restricted estimator for $\mathbf{B}$ could be different from its unrestricted estimator and will be obtained by using the REM algorithm.

The variance parameter, $\mathbf{\Sigma}$, is a function of $\mathbf{\Gamma}$ and δ and its estimator can be obtained easily by the invariance property of MLEs. So, it can be easily observed that the restricted estimator for $\mathbf{\Sigma}$ is the same as its unrestricted estimator and hence has the same form as that obtained for the variance parameter of GCM in an earlier study (Jana et al., 2017). This is not surprising, since the variance parameter for the GCM and the EGCM are the same and the two models only differ in the mean parameter.

To obtain the restricted solution $\hat{\mathbf{B}}_R$, we use (4.8). The required observed information matrix is

$$\mathbf{I}_U = (\mathbf{X}\mathbf{X}') \otimes (\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z}) = \mathbf{I}_1 \otimes \mathbf{I}_2, \text{ say.} \tag{4.15}$$

Hence, the final solution for the parameter matrix is

$$\mathbf{B}_R^{(l+1)} = \mathbf{B}_U^{(l)} - (\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z})^{-1}\mathbf{G}'(\mathbf{G}(\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z})^{-1}\mathbf{G}')^{-1}(\mathbf{G}\mathbf{B}_U\mathbf{H})(\mathbf{H}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{H})^{-1}\mathbf{H}'(\mathbf{X}\mathbf{X}')^{-1}. \tag{4.16}$$

Since REM gives us the ML estimators, hence using invariance properties of MLEs, the regression coefficients, $\mathbf{B}_1$ and $\mathbf{B}_2$ can be obtained from the identities

$$\mathbf{B}_1 = (\mathbf{I}_{(q-m)}, \mathbf{0}_{(q-m)\times m})\mathbf{B}, \tag{4.17}$$

$$\mathbf{B}_2 = (\mathbf{0}_{m\times(q-m)}, \mathbf{I}_m)\mathbf{B}. \tag{4.18}$$

4.4 Simulation Study

We conducted extensive simulations, presented in this section, to assess the performance of the proposed estimators under various conditions. Several scenarios were considered with different strengths of correlation, different levels of skewness, varying magnitude of the parameter matrix and different sample sizes. Table 4.1 presents the range of different parameters and settings considered in this study, for the empirical evaluation.

It is to be noted that in this chapter, without loss of generality, we have only presented scenarios with 4 or 8 time points, i.e., $p = 4$ or $p = 8$, for limiting the number scenarios presented, as we are already considering a huge number of scenarios. Without loss of generality, the results can be easily extended to any number of time points as the estimators were derived without any specific assumption on the number of time points. It is to be noted though, as $p < n$ is an inherent assumption in the model, therefore, we feel, that the methods developed in this chapter will be sensitive to departures from this assumption.

Table 4.1: Range of parameters considered in our simulation

Number of groups, k	2
Growth curve	linear, quadratic, cubic
Number of time-points, p	p=4, 8
Sample size, n	n=10, 15, 20, 25, 30, 35, 40 and $n > p$
Scale matrix, $\mathbf{\Omega}$	Wishart matrices with different degrees of freedom
Correlation strengths	Weak, moderate and strong
Slant vector, α	Any real p-dimensional vector
Euclidean norms of B	0 to 3, increment by 0.1
Euclidean norms of δ	0 to 1.56 increment by 0.5
Distribution	Multivariate Skew Normal

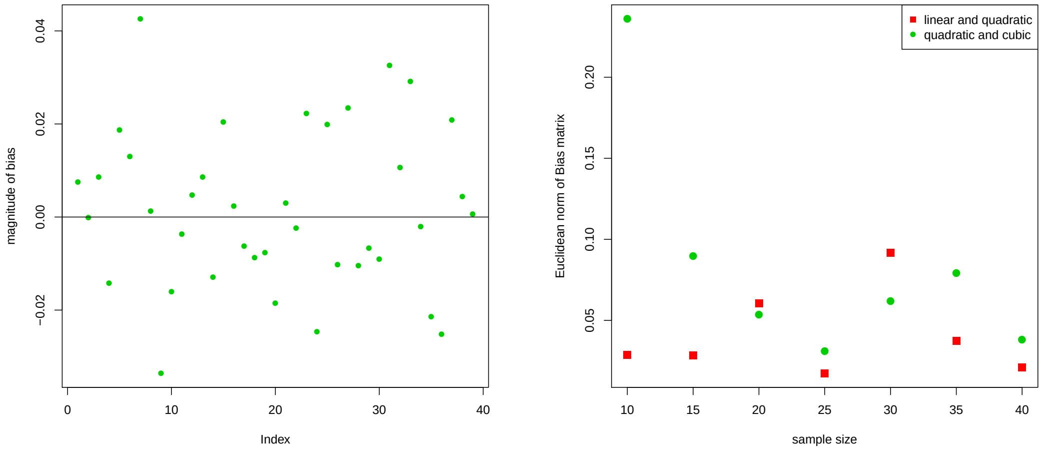
Our primary objective in this section is the empirical assessment of the proposed estimators of the mean, covariance and skewness parameters. As in the previous chapter, we use the Euclidean norm of the empirical bias matrix and MSE matrix as the evaluation criteria for the estimators, for feasibility of comparison of the magnitude of bias and MSE matrices, under the different conditions.

Since we require positive definite scale matrices $\mathbf{\Omega}$, they were generated from the Wishart distribution with varying degrees of freedom, and then divided by p to scale them down. The error terms were generated from the MSN distribution with mean 0 and covariance $\mathbf{\Sigma}$ and the response matrix was calculated using the model in (2.7). This way, the responses and the error terms had the same dispersion but different mean. The first group consisted of n_1 experimental subjects, whereas the second group consisted of n_2 experimental subjects, where $n_1 = n_2 = n/2$ when n is even, and $n_1 = (n-1)/2$ and $n_2 = n_1 + 1$ when n is odd. The results were observed for several parameter matrices, sample sizes, different correlation structures, and varying skewness levels.

A crucial step in any EM algorithm problem is the choice of initial values or starting point. As in the earlier work, we used Khatri's MLE for $\mathbf{B}$, and the sample covariance matrix $\mathbf{S} = \mathbf{Y}(\mathbf{I} - \mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-}\mathbf{Z})\mathbf{Y}'$ as the initial values for $\hat{\mathbf{B}}$ and $\hat{\mathbf{\Omega}}$ respectively (Jana et al., 2017b). We used a very small value as the starting point of δ, $\delta = (0.01, 0.01, \ldots, 0.01)$, to ensure positive definiteness of the matrix $\mathbf{\Gamma}$, which is a necessary condition in the estimation process and for the algorithm to function.

Once each data point $\mathbf{Y}$ is calculated, after generating the error terms, the three estimators $\hat{\mathbf{B}}$, $\hat{\mathbf{\Omega}}$ and $\hat{\delta}$ are computed using the REM algorithm until simultaneous convergence of all three, since the estimators are all inter-dependent. As is traditional for convergence in multivariate estimation using EM, we used the L^2 norm (Euclidean norm) for matrices and vectors as the convergence criteria. Each data point required an average of 6 iterations for convergence. The bias and MSE matrices for all the estimators are calculated for each data point. This was repeated independently 1000 times. The empirical bias and MSE matrices for each scenario are an average of these 1000 simulations. As stated earlier, the L^2 norm of the bias and MSE matrices are calculated for comparison among the different scenarios.

Here we present the results of the empirical analysis. We assume an EGCM where one group has a linear profile and the other possesses a quadratic profile. The results for this model has been represented using red in Figure 4.1(b), and we will refer to this model as a linear-quadratic EGCM. Our objective in Figure 4.1(b) is to study and compare the behaviors of the estimators under different EGCMs. Therefore, we consider another EGCM where one group has a quadratic profile and the other group has a cubic profile. This model will be referred to as quadratic-cubic EGCM and the results corresponding to this model have been represented using green in Figure 4.1(b).



(a) Bias Matrix (b) Euclidean norm of Bias

Figure 4.1: (a) Elementwise bias of $\hat{\mathbf{B}}$ for fixed $\mathbf{\Sigma}$, δ, $n = 20$ and $p = 4$ (b)Euclidean norm of Bias matrix of $\hat{\mathbf{B}}$ for fixed $\mathbf{\Sigma}$ and δ, against increasing n, for a linear-quadratic and quadratic-cubic EGCM .

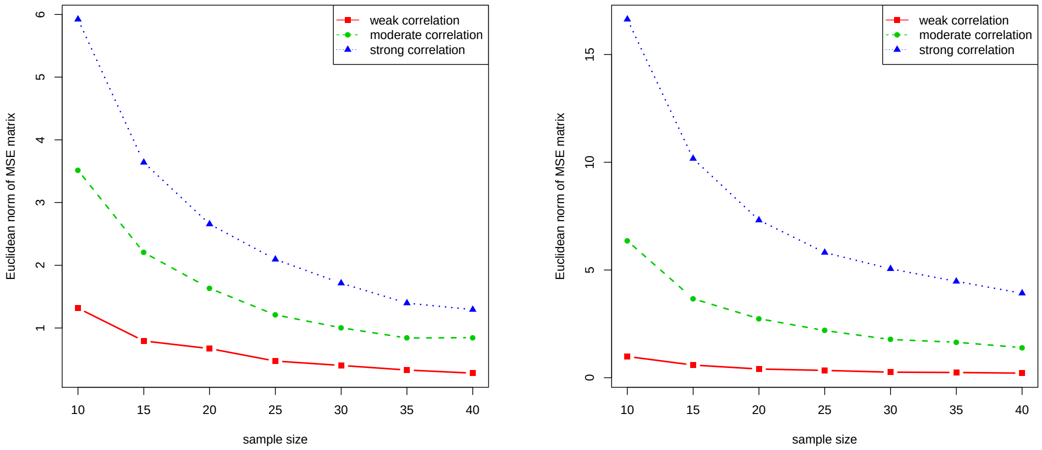
Thus, Figure 4.1(b) compares the two models, namely, the linear-quadratic EGCM and the quadratic-cubic EGCM with respect to bias of $\hat{\mathbf{B}}$. It can be observed from Figure 4.1(b), that empirical bias of $\hat{\mathbf{B}}$ is independent of sample size, n, and the degree of polynomial of the growth curves, q. It might appear from Figure 4.1(b) that bias is always positive. However, it is to be noted that the bias plotted in Figure 4.1(b) is the Euclidean norm of the Bias matrix of the parameter matrix. We provide the elementwise bias in Figure 4.2(b) where the bias of each element of the bias of $\hat{\mathbf{B}}$ is plotted. As can be observed there is both positive and negative bias indicating that actual bias of each element is randomly scattered 0.

Previous studies have also demonstrated that bias of the MLEs of the parameter matrix for GCMs are unaffected by sample size (Jana 2013; Jana et al., 2017, 2017a, 2017b). Contrary to bias, we observed that the MSE of $\hat{\mathbf{B}}$ is remarkably less when the model requires fewer polynomials and the data has a greater sample size. This nature of the bias and MSE of $\hat{\mathbf{B}}$ has also been observed for several other parameter matrices in linear-quadratic EGCMs and quadratic-cubic EGCMs.

It might seem from Figure 4.1(b) that bias is always positive. However, it is to be noted that the bias we have plotted here is the Euclidean norm of the Bias matrix of $\hat{\mathbf{B}}$. Bias of each element of $\hat{\mathbf{B}}$ is however, not always positive, rather it is randomly scattered around 0. This has been demonstrated in Figure 4.1(a), where we plotted the elementwise bias of a few parameter matrices, and we can notice that they are randomly scattered around 0, indicating that we can encounter both positive and negative bias. We would like to mention, that there is no noticeable difference in the performance of $\hat{\mathbf{\Sigma}}$ and $\hat{\delta}$ for the different EGCMs, which is very natural, as these parameters depend solely on the number of time points, p and not on the degree of polynomials, q.

It has already been established in a few studies involving growth curves that the performance of the estimators are affected by the presence of correlation in the data (Jana 2013; Jana et al., 2017, 2017a, 2017b). Similar findings were obtained for the estimators of the

EGCM, in this study as well (Figure 4.2 and Table 4.2).



(a) MSE $\hat{\mathbf{B}}$ (b) MSE $\hat{\boldsymbol{\Sigma}}$

Figure 4.2: MSE of $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$ against increasing sample size, across different correlation strengths for a linear-quadratic EGCM.

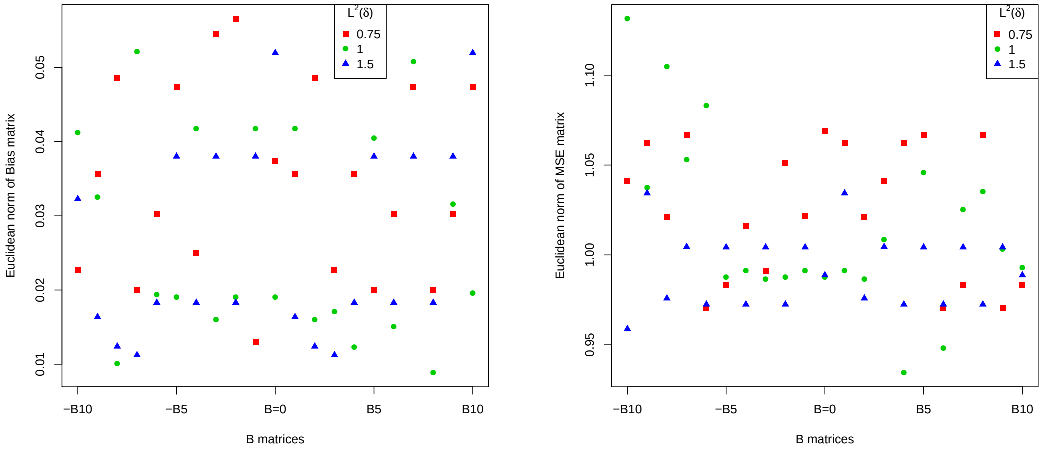
Table 4.2: Bias of $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$ against increasing sample size, across different correlation strengths for a linear-quadratic EGCM

estimators	$\widehat{\mathbf{B}}$			$\widehat{\boldsymbol{\Sigma}}$		
sample size	weak	moderate	strong	weak	moderate	strong
10	0.052 8	0.028 9	0.122 5	0.183 8	0.541 8	1.034 4
15	0.019 0	0.028 3	0.032 8	0.156 9	0.509 9	0.742 0
20	0.028 0	0.060 6	0.098 8	0.127 5	0.384 6	0.667 2
25	0.011 7	0.017 4	0.046 6	0.119 7	0.385 4	0.531 8
30	0.012 9	0.091 6	0.057 6	0.095 3	0.260 2	0.395 8
35	0.010 3	0.037 2	0.033 7	0.076 9	0.260 7	0.273 1
40	0.008 7	0.020 9	0.037 6	0.073 0	0.194 4	0.239 9

In general, bias of $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$ tend to be more, with increasing strengths of correlation

(Table 4.2). But the effect is more pronounced in the results for MSE. In Figure 4.2, we can observe that the MSE of $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$ increase with increasing correlation in the data. These figures also establish the empirical consistency of $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$, as the empirical MSEs decrease with increasing sample size. We would like to highlight that empirical consistency have also been observed and established for estimators of the mean and variance parameters of the GCM as well (Jana 2013; Jana et al., 2017, 2017a, 2017b).

From Figure 4.3, it can be observed that there is no systematic pattern in bias and MSE of $\hat{\mathbf{B}}$ for different magnitude, whether positive or negative, of the parameter matrix, indicating robustness of the estimator $\hat{\mathbf{B}}$ to the true value of the parameter. This reiterates the robustness property of ML estimators of $\mathbf{B}$ established in earlier works involving GCM (Jana, 2013; Jana et al., 2017, 2017a, 2017b). Needless to say, the magnitude or direction (positive or negative) of the parameter matrix has no effect on the estimators $\hat{\boldsymbol{\Sigma}}$ and $\hat{\delta}$.



(a) Bias (b) MSE

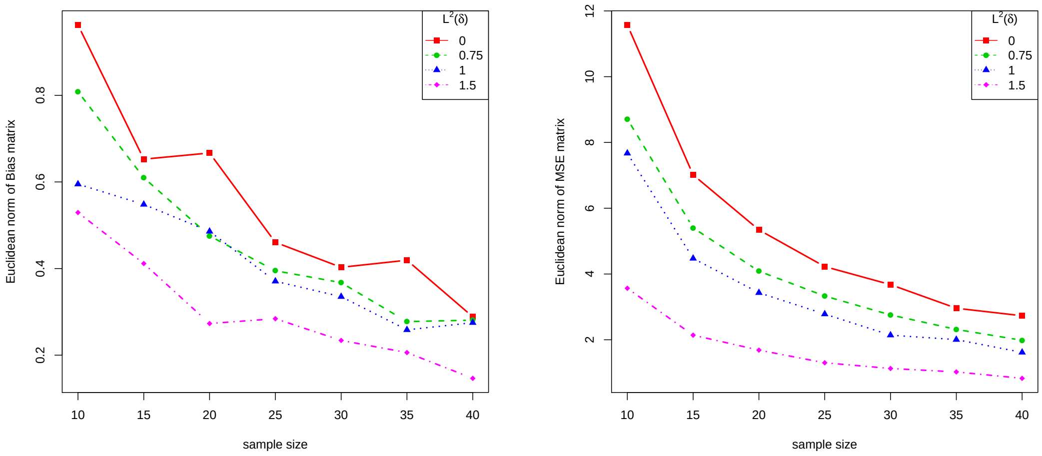
Figure 4.3: Bias and MSE of $\hat{\mathbf{B}}$ against different values of $\mathbf{B}$, across different levels of skewness for n=30.

We also observe from Figure 4.3 and Table 4.3 that the performance of the estimator $\hat{\mathbf{B}}$ does not depend on the magnitude of skewness in the data. There seems to be no systematic influence of skewness on the bias or MSE of $\hat{\mathbf{B}}$. In contrast, however, we see that bias and MSE of $\hat{\boldsymbol{\Sigma}}$ decrease with increasing levels of skewness in the data (Figure 4.4 and Table 4.4). This might be due to $\hat{\delta}$ being underestimated and $\hat{\boldsymbol{\Sigma}}$ being a function of $-\delta$.

Table 4.3: Bias and MSE of $\hat{\mathbf{B}}$ against increasing sample size, across different levels of skewness for a linear-quadratic EGCM

$L^2(\delta)$	Bias						
	$n = 10$	$n = 15$	$n = 20$	$n = 25$	$n = 30$	$n = 35$	$n = 40$
0	0.109 2	0.053 8	0.026 7	0.009 7	0.037 5	0.012 1	0.038 3
0.75	0.020 4	0.052 8	0.010 1	0.009 0	0.022 8	0.017 4	0.016 4
1	0.095 4	0.052 0	0.028 5	0.034 4	0.041 2	0.030 2	0.028 1
1.5	0.094 8	0.008 8	0.003 6	0.019 1	0.032 3	0.008 4	0.045 3
$L^2(\delta)$	MSE						
	$n = 10$	$n = 15$	$n = 20$	$n = 25$	$n = 30$	$n = 35$	$n = 40$
0	3.237 3	1.954 0	1.522 4	1.226 7	1.039 2	0.866 6	0.745 8
0.75	3.147 4	2.023 8	1.565 0	1.087 8	1.041 4	0.846 3	0.727 9
1	3.408 5	1.918 8	1.448 5	1.147 5	1.039 6	0.889 5	0.716 6
1.5	3.592 4	2.077 8	1.478 7	1.166 2	0.958 8	0.824 1	0.758 0

It is worth mentioning here that, although the empirical bias of the variance estimator seem to be always positive from the results, in reality it is not. When we plot the bias of the individual elements of the parameter matrix or the variance-covariance matrix, we notice that they are randomly distributed around 0, indicating that bias can be both positive and negative. However, as already stated, for ease of comparison, since we plot the Euclidean norm of the bias matrices, which is a positive measure, the bias values on the graph seem to be always positive.



(a) Bias (b) MSE

Figure 4.4: Bias and MSE of $\hat{\mathbf{\Sigma}}$ against increasing sample size, across different levels of skewness for a fixed $\mathbf{B}$.

Table 4.4: Bias and MSE of $\hat{\mathbf{B}}$ and $\hat{\mathbf{\Sigma}}$ against different levels of skewness for a quadratic-cubic EGCM, for $n = 20$, across different correlation strengths

$\widehat{\mathbf{B}}$						
$L^2(\delta)$	Bias			MSE		
	weak	moderate	strong	weak	moderate	strong
0	0.097 9	0.109 8	0.181 6	7.970 2	15.787 2	22.493 9
0.75	0.210 6	0.074 3	0.050 9	7.954 0	15.972 3	24.146 8
1	0.045 3	0.068 5	0.080 4	7.800 8	15.625 3	23.365 5
1.5	0.102 1	0.015 9	0.120 3	7.708 3	15.857 2	24.001 4
$\widehat{\mathbf{\Sigma}}$						
$L^2(\delta)$	Bias			MSE		
	weak	moderate	strong	weak	moderate	strong
0	0.447 0	0.778 7	1.384 6	3.347 2	15.680 6	35.946 0
0.75	0.296 6	0.905 2	1.492 6	2.830 9	12.404 8	26.577 8
1	0.264 0	0.726 9	1.205 1	2.344 5	10.276 3	21.570 2
1.5	0.202 3	0.557 3	0.750 8	1.108 0	4.727 8	11.398 4

Similar to the empirical analysis in MSN GCM (Jana et al., 2017b), $\hat{\delta}$ does not exhibit any empirical consistency, although bias and MSE of $\hat{\delta}$ is uniformly less when there is less correlation in the data. These properties have been observed to be true for different values of the parameter matrices. We would like to remark here that the estimate for $\hat{\delta}$ does not depend on its initial value and the estimate remains the same regardless of the initial value specified. However, for the algorithm to function, we need to specify an initial value for $\hat{\delta}$ which is smaller than the true value. Since we never know the true value of the parameter, so we recommend using a very small value as the initial value δ, so that the algorithm functions for any amount of skewness in the data.

In conclusion, the results from this section indicate the robustness of the estimator $\hat{\mathbf{B}}$ with respect to sample size, level of skewness and the true value of the parameter $\mathbf{B}$. We would like to highlight though, that the MSE of $\hat{\mathbf{B}}$ is hugely affected by the value of q, i.e., the degree of the polynomial considered in the study, indicating that the variance of the estimator is less for simpler models. The results also demonstrate that the estimators $\hat{\mathbf{B}}$ and $\hat{\mathbf{\Sigma}}$ are consistent estimators, although $\hat{\delta}$ is not a consistent estimator. However, their performance is affected by presence of correlation in the data. Surprisingly, the estimator $\hat{\mathbf{\Sigma}}$ has better performance with increasing skewness in the data. This could be due to $\hat{\delta}$ being underestimated. Improving the estimator $\hat{\delta}$ could be of interest for further study.

4.5 Real Data Analysis

Our objective in this section is to demonstrate the practical usefulness of the estimators developed in this chapter by fitting a skewed multivariate data. To serve our purpose of illustration, we use the longitudinal data from a major epidemiological study called the Framingham Heart Study (FHS), which has already been introduced and discussed in Section 3.5.

It is to be noted that the longitudinal data from the FHS has been used here as an example, and the results are not intended to be used for any clinical analysis or interpretation. The tools developed in this chapter considers only two groups, hence we consider a predictor variable which consists of only two groups, e.g., gender. However, without loss of generality, the results can be extended to more than two groups.

We use the same illustrative example from Section 3.5. In fact, the lack of fit of an MSN GCM in fitting this data provided motivation for the development of estimation theory for this chapter. The skewness in TG levels has already been demonstrated in Figures 3.8 and 3.9, hence we choose this as our response variable. The TG levels were measured across 4 time points, hence p=4. The sample size for our illustrative example is 20, with equal number of males and females in the sample.

For this data, we assume a quadratic response for males and a cubic response for females, as is evident from their profile plots. Accordingly, we fit a quadratic-cubic EGCM, using the REM algorithm for matrix estimators, developed in this study. The appropriate design matrices for the EGCM are

$$\mathbf{Z}_1' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 5 & 7 \\ 1 & 9 & 25 & 49 \end{pmatrix}, \mathbf{Z}_2' = \begin{pmatrix} 1 & 27 & 125 & 343 \end{pmatrix},$$

$$\mathbf{X}_1 = \begin{pmatrix} 1_{10} & 0_{10} \\ 0_{10} & 1_{10} \end{pmatrix} \text{ and, } \mathbf{X}_2 = 1_{10}.$$

The mean responses for the two groups are

$$E(Y_i) = \beta_{01} + \beta_{11}t + \beta_{21}t^2,\ i = 1, 2, \ldots, 10,$$

$$E(Y_i) = \beta_{02} + \beta_{12}t + \beta_{22}t^2 + \beta_{32}t^3,\ i = 11, 12, \ldots, 20,$$

where β's are the unknown coefficients of the parameter matrices, $\mathbf{B}_1$ and $\mathbf{B}_2$, where

$$\mathbf{B}_1' = \begin{pmatrix} \beta_{01} & \beta_{11} & \beta_{21} \\ \beta_{02} & \beta_{12} & \beta_{22} \end{pmatrix}, \mathbf{B}_2 = (\beta_{32}).$$

The estimate of the parameters matrices are

$$\mathbf{B}_1 = \begin{pmatrix} 51.9163403 & 109.668750 \\ 13.8320510 & -62.314583 \\ -0.9488764 & 19.331250 \end{pmatrix}, \mathbf{B}_2 = -1.485417.$$

In Figure 4.5, we plot the estimated mean from the EGCM and the observed mean, for illustrating the goodness of fit of the predicted response, and we see that the fit is near perfect for both the groups. In fact, the fit for males has improved by fitting an EGCM, although the improvement is negligible.

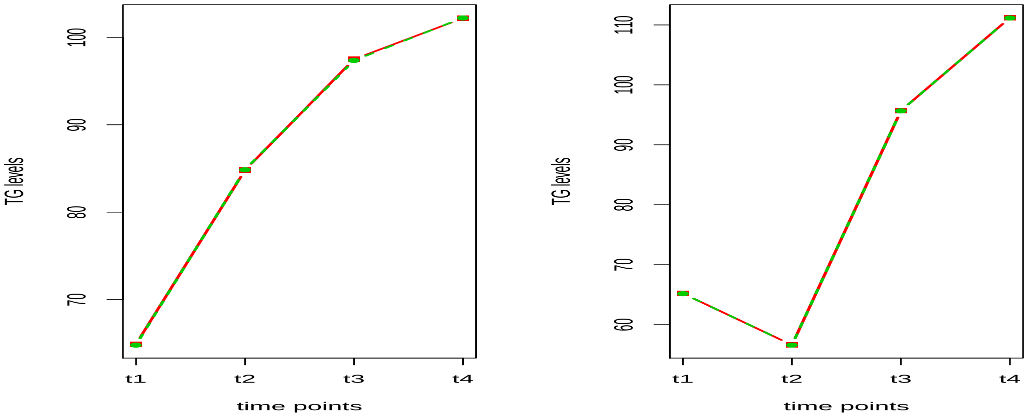


Figure 4.5: Observed vs fitted response for $n = 20$, for males and females based on an EGCM.

We calculated the Akaike's information criterion (AIC) for both the model fits for this example. The AIC assuming a quadratic GCM (Figure 3.10) for this data turned out to be

958.6961. The AIC assuming EGCM (Figure 4.5) is 678.1482. This result is at par with our Figures 3.10 and 4.5, where we can see that the EGCM is a much better fit for the current data. We also wanted to compare the efficiency of the methods developed in this study with the methods developed for the traditional EGCM, for a skewed data.The AIC value for the traditional model, that is the model fitted by using von Rosen's estimators in (3) for the traditional EGCM, which assumes multivariate normality, is 795.0401. Thus, the current fit using EGCM for skewed models seems to be a better fit for skewed data.

4.6 Discussion

This chapter was motivated by the limitations of a GCM in modelling skewed longitudinal data with different profiles across different groups. An EGCM relaxes this limitation and allows us to model longitudinal data with parallel profiles. EGCM is a special case of EGMANOVA models and can be used for the analysis of growth or dose-response when there are clusters in the data.

In this chapter, we considered the estimation problem for an EGCM whose underlying distribution is MSN. Since the EGCM can be expressed as a GCM with linear restrictions, therefore, we employed the REM algorithm, which is an extension of the EM algorithm, in the presence of linear restrictions in the model. The development of the REM algorithm is based on the Newton-Raphson method and both are only applicable to multivariate parameters. An extension of the REM algorithm and the NR method is first developed for matrix parameters and then implemented in this study to propose estimators for the parameter of interest. Extensive simulations were conducted to study the behavior of the estimators developed in this chapter.

The estimators for the mean and variance parameters were empirically demonstrated as consistent estimators. The bias of the parameter matrix is, however, unaffected by increasing or decreasing magnitude and direction of $\mathbf{B}$, sample size and varying skewness level indicating robustness of the estimator. Bias and MSE of both the mean and covariance matrix increases with increasing strengths of correlation, which was also observed in earlier works involving GCMs (Jana, 2013; Jana et al., 2017). The performance of the covariance matrix, in terms of empirical bias and MSE, however improves with increasing skewness in the data. This might be because the covariance matrix is a function of -δ.

The empirical behavior of the estimator for the skewness vector seems to be independent of the sample size indicating that it is not a consistent estimator. It rather seems to be dependent on increasing levels of skewness. However, this is due to the skewness parameter $\hat{\delta}$ being underestimated.

Chapter 5

GCM for high-dimensional data under Matrix Variate Normal

5.1 Introduction

As stated earlier in Chapter 2, a few tests are available for testing the general linear hypothesis in the GCM, among them are Potthoff and Roy's approach based on transformation to MANOVA, Khatri's likelihood ratio test and likelihood based trace test (Potthoff and Roy, 1964; Khatri, 1966; Hamid et al., 2011). However, these tests are not appropriate for high-dimensional data since they all involve the inverse of the sample covariance matrix, which does not exist due to singularity of the sample covariance matrix.

Although few methods are available for modelling high-dimensional longitudinal data, those methods are based on MANOVA models and hence do not allow temporal ordering as well as structured mean in the model. However, in most studies involving longitudinal data, the mean is often structured and can be represented as a continuous function of time (Ma et al., 2006; Hamid and Beyene, 2009; Jana, 2013; Jana et al., 2017). Methods that allow a structured mean and appropriately model the outcome over time, incorporates correlation across time points and temporal ordering in the model, while accounting for high-dimensionality were also considered before (Hamid and Beyene, 2009; Jana, 2013; Jana et al., 2017).

In a previous work, we proposed a test based on the Growth Curve Model (GCM) using the Moore-Penrose generalized inverse to tackle the high-dimensionality problem, examined its performance using simulations and illustrated its usefulness in practical applications for gene-filtering using genetic data from human lung tissues (Jana, 2013; Jana et al., 2017). The proposed test performed well in many situations, and, we have shown that it possesses the desired optimality characteristics including monotonicity, unbiasedness and symmetry. The test has good power in detecting small departures from the null and has a level close to the nominal level. However, we observed that the test has poor performance in the near-singularity area (that is, when $n \approx p$), where power starts to decline monotonically after a specific sample size in the near-singularity area ($p \approx n$). Furthermore, the mean squared error of the estimators of the mean parameters also increase when $p \approx n$. Nevertheless, the results show that the level of the test and bias of the estimators are not affected (Jana, 2013; Jana et al., 2017).

In this chapter, we use a Bayesian approach to overcome the limitations in our previous work and derive a test statistic with improved performance overall and especially in the near singularity area. We performed extensive simulations to examine its performances as well as establish optimality properties for the proposed test. Using publicly available breast cancer data, we illustrated the application of our method in gene-filtering using time course genetic experiments. We annotated the top ranked genes to show their relevance to breast cancer.

5.2 Bayesian Growth Curve Model and the Proposed Test

Consider the likelihood function for the GCM in equation (2.1),

$$L = (2\pi)^{-\frac{1}{2}np} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\mathbf{\Sigma^{-1}}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}},$$

which can be written as

$$L = (2\pi)^{-\frac{1}{2}np} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\mathbf{\Sigma^{-1}}(\mathbf{YX}'(\mathbf{XX}')^{-}\mathbf{X}-\mathbf{ZBX})(\mathbf{YX}'(\mathbf{XX}')^{-}\mathbf{X}-\mathbf{ZBX})'+\mathbf{\Sigma^{-1}S}\}}, \tag{5.1}$$

where $\mathbf{YX}'(\mathbf{XX}')^{-}\mathbf{X} = \bar{\mathbf{Y}}$is the sample mean and $\mathbf{S} = \mathbf{Y}(\mathbf{I}_p - \mathbf{X}'(\mathbf{XX}')^{-}\mathbf{X})\mathbf{Y}'$ (Srivastava, 2002; Hamid et al., 2011; Jana, 2013; Jana et al., 2017). Using the Fisher-Neyman factorization theorem, it can be easily concluded from (5.1) that $(\bar{\mathbf{Y}}, \mathbf{S})$ are jointly sufficient for $(\mathbf{B}, \mathbf{\Sigma})$ since the multivariate normal distribution belongs to the natural exponential family. L is the joint likelihood of the two sufficient statistics given the two parameters $(\mathbf{B}, \mathbf{\Sigma})$. Also, $\mathbf{S}$ is sufficient for $\mathbf{\Sigma}$.

Recall that the columns of $\mathbf{Y}$ independently follow a multivariate normal distribution. For mathematical simplicity (without loss of generality), consider only one group and denote the columns of $\mathbf{Y}$ as $y_1, y_2, \ldots y_n$, thus $y_i \sim$ i.i.d. $N_p(\mu, \mathbf{\Sigma})$ where μ is any column of the matrix of expected mean which is $\mathbf{ZBX}$. Thus, the likelihood of the GCM can be rewritten as (Anderson 1958, Ando and Kauffman, 1965)

$$\begin{aligned} L &= (2\pi)^{-\frac{pn}{2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(y_j-\mu)'\mathbf{\Sigma}^{-1}(y_j-\mu)} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}} \\ &= (2\pi)^{-\frac{pn}{2}} e^{-\frac{1}{2}tr\{n(\bar{y}-\mu)'\mathbf{\Sigma}^{-1}(\bar{y}-\mu)+\mathbf{\Sigma}^{-1}(n-1)\mathbf{S}\}} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}} \end{aligned}$$

where $\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})(y_j - \bar{y})'$. In this chapter, we will use a two step partially Bayesian approach to derive the test statistic, with the objective of extending the trace test for the GCM so that it is applicable for high-dimensional data. We will first consider the second part of the likelihood and estimate $\mathbf{\Sigma}$ using a Bayesian approach. We will then maximize the estimated likelihood under the null hypothesis and the entire parametric space to calculate the targeted test statistic.

Note that $(\bar{y}, \mathbf{S})$ are jointly sufficient for $(\mu, \mathbf{\Sigma})$ and that the kernel of the joint likelihood of $(\bar{y}, \mathbf{S})$ given $(\mu, \mathbf{\Sigma})$ is (Anderson, 1958)

$$e^{-\frac{1}{2}tr\{n(\bar{y}-\mu)'\mathbf{\Sigma}^{-1}(\bar{y}-\mu)+\mathbf{\Sigma}^{-1}(n-1)\mathbf{S}\}} \mid \mathbf{\Sigma} \mid^{-\frac{n}{2}},$$

which is the kernel of a Normal-Wishart distribution. The kernel of the conditional marginal likelihood of $\bar{y}$ is (Anderson, 1958; Ando and Kaufman, 1965)

$$e^{-\frac{1}{2}tr\{n(\bar{y}-\mu)'\mathbf{\Sigma}^{-1}(\bar{y}-\mu)\}} \mid \mathbf{\Sigma} \mid^{-\frac{1}{2}},$$

which is the kernel of a multivariate normal distribution of order p with mean μ and variance-covariance matrix $\mathbf{\Sigma}/n$. The conditional marginal likelihood of $\mathbf{S}$, therefore, is

$$e^{-\frac{1}{2}tr\{(\frac{1}{n-1}\mathbf{\Sigma})^{-1}\mathbf{S}\}} \mid \mathbf{\Sigma} \mid^{-\frac{n-1}{2}},$$

which is the kernel of a Wishart distribution of order p, degrees of freedom $(n-1)$ and scale matrix $\frac{1}{n-1}\mathbf{\Sigma}$. Consequently, the pdf of $(\mathbf{S} \mid \mathbf{\Sigma})$ can be given as

$$\frac{e^{-\frac{1}{2}tr\{(\frac{1}{n-1}\mathbf{\Sigma})^{-1}\mathbf{S}\}} \mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}}}{2^{\frac{(n-1)p}{2}}\Gamma_p(\frac{n-1}{2}) \mid \frac{1}{n-1}\mathbf{\Sigma} \mid^{\frac{n-1}{2}}}.$$

Using Raiffa and Schlaifer's suggested approach, we consider $\mathbf{S}$ to be the parameter in the above kernel and $\mathbf{\Sigma}$ to be the random variable (Raiffa and Schlaifer, 1961). As a result, the kernel becomes the kernel of an Inverse Wishart distribution with $(n-1)$ degrees of freedom (d.f.) and scale matrix $\frac{1}{n-1}\mathbf{S}$. The natural conjugate prior distribution for $\mathbf{\Sigma}$, therefore, belongs to an Inverse Wishart family.

Let us now denote the hyper parameters of the prior distribution as ν and $\mathbf{\Lambda}$, where ν is the d.f. and $(\nu\mathbf{\Lambda})$ is the scale matrix of the Inverse-Wishart distribution, provided $\nu > p-1$ and $(\nu\mathbf{\Lambda})$ is positive definite. To derive the posterior distribution of $\mathbf{\Sigma}$, we require the unconditional distribution of $\mathbf{S}$, which we will denote as $f_{\mathbf{S}}(\mathbf{S})$. The unconditional distribution can be derived by integrating the product of the conditional likelihood of $\mathbf{S}$ given $\mathbf{\Sigma}$ and the prior distribution of $\mathbf{\Sigma}$, with respect to $\mathbf{\Sigma}$, which gives

$$\begin{aligned}
f_{\mathbf{S}}(\mathbf{S}) &= \int_{\mathbf{\Sigma}>\mathbf{0}} f(\mathbf{S} \mid \mathbf{\Sigma}) f(\mathbf{\Sigma}) d\mathbf{\Sigma} \\
&= \int_{\mathbf{\Sigma}>\mathbf{0}} \frac{e^{-\frac{1}{2}tr\{(\frac{1}{n-1}\mathbf{\Sigma})^{-1}\mathbf{S}\}} \mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}}}{2^{\frac{(n-1)p}{2}} \Gamma_p(\frac{n-1}{2}) \mid \frac{1}{n-1}\mathbf{\Sigma} \mid^{\frac{n-1}{2}}} \frac{\mid \mathbf{\Sigma} \mid^{-\frac{\nu+p+1}{2}} \mid \nu\mathbf{\Lambda} \mid^{\frac{\nu}{2}} e^{-\frac{1}{2}tr\{(\nu\mathbf{\Lambda})\mathbf{\Sigma}^{-1}\}}}{2^{\frac{\nu p}{2}}\Gamma_p(\frac{\nu}{2})} d\mathbf{\Sigma} \\
&= \frac{\mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}} \mid \nu\mathbf{\Lambda} \mid^{\frac{\nu}{2}} (n-1)^{\frac{n-1}{2}}}{\Gamma_p(\frac{n-1}{2})\Gamma_p(\frac{\nu}{2})} \int_{\mathbf{\Sigma}>\mathbf{0}} \frac{e^{-\frac{1}{2}[\mathbf{\Sigma}^{-1}\{(n-1)\mathbf{S}+(\nu\mathbf{\Lambda})\}]}}{2^{\frac{p}{2}(\nu+n-1)} \mid \mathbf{\Sigma} \mid^{\frac{n-1+\nu+p+1}{2}}} d\mathbf{\Sigma} \\
&= \frac{\mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}} \mid \frac{\nu\mathbf{\Lambda}}{n-1} \mid^{\frac{\nu}{2}}}{\Gamma_p(\frac{n-1}{2})\Gamma_p(\frac{\nu}{2})} \times \frac{\Gamma_p(\frac{n-1+\nu}{2})}{\mid (n-1)\mathbf{S} + (\nu\mathbf{\Lambda}) \mid^{\frac{n-1+\nu}{2}}} \\
&\quad \times \int_{\mathbf{\Sigma}>\mathbf{0}} \frac{\mid (n-1)\mathbf{S} + (\nu\mathbf{\Lambda}) \mid^{\frac{n-1+\nu}{2}} e^{-\frac{1}{2}[\mathbf{\Sigma}^{-1}\{(n-1)\mathbf{S}+(\nu\mathbf{\Lambda})\}]}}{2^{\frac{p}{2}(\nu+n-1)}\Gamma_p(\frac{n-1+\nu}{2}) \mid \mathbf{\Sigma} \mid^{\frac{n-1+\nu+p+1}{2}}} d\mathbf{\Sigma}.
\end{aligned}$$

Since the integrand is the pdf of an Inverse-Wishart distribution, $f_{\mathbf{S}}(\mathbf{S})$ reduces to

$$\begin{aligned}
f_{\mathbf{S}}(\mathbf{S}) &= \frac{\mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}} \mid \frac{\nu\mathbf{\Lambda}}{n-1} \mid^{\frac{\nu}{2}}}{\Gamma_p(\frac{n-1}{2})\Gamma_p(\frac{\nu}{2})} \times \frac{\Gamma_p(\frac{n-1+\nu}{2})}{\mid (n-1)\mathbf{S} + (\nu\mathbf{\Lambda}) \mid^{\frac{n-1+\nu}{2}}} \\
&= \frac{\mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}} \mid \frac{\nu\mathbf{\Lambda}}{n-1} \mid^{\frac{\nu}{2}}}{\beta_p(\frac{n-1}{2}, \frac{\nu}{2}) \times \mid (n-1)\mathbf{S} + (\nu\mathbf{\Lambda}) \mid^{\frac{n-1+\nu}{2}}},
\end{aligned}$$

where $\beta_p(.)$ is the multivariate beta function as defined by Siegel (1935) and has the following relationship with the multivariate gamma function (Ando and Kaufman, 1965; Tan, 1969; Sivazlian, 1981; Javier and Gupta, 1985; Gupta and Nagar, 2000):

$$\beta_p(a,b) = \frac{\Gamma_p(a)\Gamma_p(b)}{\Gamma_p(a+b)}, \tag{5.2}$$

where $\Gamma_p(a) = \int_{\mathbf{S}>\mathbf{0}} e^{-tr\mathbf{S}} \mid \mathbf{S} \mid^{a-\frac{p+1}{2}} d\mathbf{S}$ (James, 1964). Thus, $f_{\mathbf{S}}(\mathbf{S})$ has the generalized type II beta distribution with parameters $\frac{n-1}{2}, \frac{\nu}{2}$, scale matrix $(\frac{\nu\mathbf{\Lambda}}{n-1})$ and location matrix $\mathbf{0}$. The distribution is denoted by $GB_p^{II}(\frac{n-1}{2}, \frac{\nu}{2}, \frac{\nu\mathbf{\Lambda}}{n-1}, \mathbf{0})$ (Siegel, 1935; Ando and Kaufman, 1965; Javier and Gupta, 1985; Tai and Speed, 2006).

Now we proceed to derive the posterior distribution of $\mathbf{\Sigma}$. Due to the mathematical complexity involved in the direct derivation of the posterior distribution of $\mathbf{\Sigma}$, we will instead indirectly approach the problem by considering the posterior distribution of $\mathbf{\Sigma}^{-1}$ and inverting the solution afterwards. Consider the Jacobian of transformation, J for any $p \times p$ symmetric matrix $\mathbf{W}$ (Kollo and von Rosen, 2005),

$$\mid J(\mathbf{W} \longrightarrow \mathbf{W}^{-1}) \mid = \mid \mathbf{W}^{(p+1)} \mid .$$

Hence the posterior distribution of $\mathbf{\Sigma}^{-1}$ can be written as

$$\begin{aligned} f(\mathbf{\Sigma}^{-1} \mid \mathbf{S}) &= \frac{f(\mathbf{S}, \mathbf{\Sigma}^{-1})}{f(\mathbf{S})} \\ &= \frac{f(\mathbf{S}, \mathbf{\Sigma}) \mid \mathbf{\Sigma} \mid^{(p+1)}}{f(\mathbf{S})} \\ &= \frac{f(\mathbf{S} \mid \mathbf{\Sigma}) f(\mathbf{\Sigma}) \mid \mathbf{\Sigma} \mid^{(p+1)}}{f(\mathbf{S})}. \end{aligned} \tag{5.3}$$

Let us now summarize all the required distributions in the derivation of the posterior distribution. That is the prior distribution, the conditional and unconditional likelihoods, respectively:

$$
\begin{aligned}
f(\boldsymbol{\Sigma}) &= \frac{\mid \boldsymbol{\Sigma} \mid^{-\frac{\nu+p+1}{2}} \mid \nu\boldsymbol{\Lambda} \mid^{\frac{\nu}{2}} \ e^{-\frac{1}{2}tr\{(\nu\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}}}{2^{\frac{\nu p}{2}}\Gamma_p(\frac{\nu}{2})} \ , \\
f(\mathbf{S} \mid \boldsymbol{\Sigma}) &= \frac{\mid \mathbf{S} \mid^{\frac{n-1-p-1}{2}} \ e^{-\frac{1}{2}tr\{(\frac{1}{n-1}\boldsymbol{\Sigma})^{-1}\mathbf{S}\}}}{2^{\frac{(n-1)p}{2}}\Gamma_p(\frac{n-1}{2}) \mid \frac{1}{n-1}\boldsymbol{\Sigma} \mid^{(\frac{n-1}{2})}} \ , \\
f(\mathbf{S}) &= \frac{\mid \mathbf{S} \mid^{\frac{(n-1)-p-1}{2}}}{\beta_p(\frac{n-1}{2}, \frac{\nu}{2}) \mid \frac{\nu\boldsymbol{\Lambda}}{n-1} \mid^{\frac{n-1}{2}} \mid I_p + (\frac{\nu\boldsymbol{\Lambda}}{n-1})^{-1}\mathbf{S} \mid^{\frac{n+\nu-1}{2}}} \ .
\end{aligned}
$$

Consequently, Equation (5.3) can be written as

$$
\begin{aligned}
f(\boldsymbol{\Sigma}^{-1} \mid \mathbf{S}) &= \frac{f(\mathbf{S} \mid \boldsymbol{\Sigma})f(\boldsymbol{\Sigma}) \mid \boldsymbol{\Sigma} \mid^{(p+1)}}{f(\mathbf{S})} \\
&= \frac{\mid \mathbf{S} \mid^{\frac{n-p-2}{2}} \ e^{-\frac{1}{2}tr\{(\frac{1}{n-1}\boldsymbol{\Sigma})^{-1}\mathbf{S}\}}}{2^{\frac{(n-1)p}{2}}\Gamma_p(\frac{n-1}{2}) \mid \frac{1}{n-1}\boldsymbol{\Sigma} \mid^{(\frac{n-1}{2})}} \\
&\quad \times \frac{\mid \boldsymbol{\Sigma} \mid^{-\frac{\nu+p+1}{2}} \mid \nu\boldsymbol{\Lambda} \mid^{\frac{\nu}{2}} \ e^{-\frac{1}{2}tr\{(\nu\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}}}{2^{\frac{\nu p}{2}}\Gamma_p(\frac{\nu}{2})} \times \mid \boldsymbol{\Sigma} \mid^{(p+1)} \\
&\quad \times \frac{\beta_p(\frac{n-1}{2}, \frac{\nu}{2}) \mid \frac{\nu\boldsymbol{\Lambda}}{n-1} \mid^{\frac{n-1}{2}} \mid \mathbf{I_p} + (\frac{\nu\boldsymbol{\Lambda}}{n-1})^{-1}\mathbf{S} \mid^{\frac{n+\nu-1}{2}}}{\mid \mathbf{S} \mid^{\frac{n-p-2}{2}}} \\
&= \mid \boldsymbol{\Sigma} \mid^{-\frac{\nu+p+n-1}{2}+(p+1)} \ e^{-\frac{1}{2}tr[\{(n-1)\mathbf{S}+\nu\boldsymbol{\Lambda}\}\boldsymbol{\Sigma}^{-1}]} \\
&\quad \times \mid \nu\boldsymbol{\Lambda} \mid^{\frac{n+\nu-1}{2}} \mid \mathbf{I_p} + (\frac{\nu\boldsymbol{\Lambda}}{n-1})^{-1}\mathbf{S} \mid^{\frac{n+\nu-1}{2}} \\
&\quad \times \frac{\beta_p(\frac{n-1}{2}, \frac{\nu}{2})}{\Gamma_p(\frac{n-1}{2})\Gamma_p(\frac{\nu}{2})} \times \frac{1}{2^{\frac{p(n+\nu-1)}{2}}} \\
&= \mid \boldsymbol{\Sigma} \mid^{-\frac{(n+\nu-1)-p-1}{2}} \ e^{-\frac{1}{2}tr[\{(n-1)\mathbf{S}+\nu\boldsymbol{\Lambda}\}\boldsymbol{\Sigma}^{-1}]} \mid (\nu\boldsymbol{\Lambda}) + (n-1)\mathbf{S} \mid^{\frac{n+\nu-1}{2}} \\
&\quad \times \frac{1}{\Gamma_p(\frac{n-1}{2} + \frac{\nu}{2})} \times \frac{1}{2^{\frac{p(n+\nu-1)}{2}}} \ .
\end{aligned}
$$

The last step uses the identity mentioned in (5.2). By comparing the form of $f(\boldsymbol{\Sigma}^{-1} \mid \mathbf{S})$ above with the Wishart pdf, we can say that

$$(\boldsymbol{\Sigma}^{-1} \mid \mathbf{S}) \sim Wishart_p(df = n + \nu - 1, scale = \{(n-1)\mathbf{S} + \nu\boldsymbol{\Lambda}\}^{-1}).$$

Using the formula for expectation of a Wishart random variable, which is the product of the scale matrix and its degrees of freedom, the posterior mean is

$$E(\mathbf{\Sigma}^{-1} \mid \mathbf{S}) = \{(n-1)\mathbf{S} + \nu\mathbf{\Lambda}\}^{-1} \times (n+\nu-1) \ .$$

The required estimator for the unknown covariance matrix $\mathbf{\Sigma}$ can be obtained by inverting the posterior mean as

$$\tilde{\mathbf{S}} = [E(\mathbf{\Sigma}^{-1} \mid \mathbf{S})]^{-1} = (1-k)\mathbf{S} + k\mathbf{\Lambda} \ ,$$

where, $k = \frac{\nu}{n-1+\nu}$. Since $\nu > p-1$ and $n > p$, so $k \in (\frac{1}{2}, 1)$. Note that $\tilde{\mathbf{S}}$ is a convex combination of a positive definite matrix, $\mathbf{\Lambda}$ and a positive semi-definite matrix, $\mathbf{S}$ and hence always a positive definite matrix itself, regardless of whether $n < p$ or $n > p$. Note also that the estimator $\tilde{\mathbf{S}}$ looks like a special case of James-Stein estimator (Sch*ä*fer and Strimmer, 2005), where $\mathbf{\Lambda}$ is the target matrix. It is also to be noted that the estimator $\tilde{\mathbf{S}}$ has the same structure as the estimator proposed by Tai and Speed (2006), but that was derived under slightly different conditions.

We would like to highlight again that this derivation has been done assuming one group for simplicity of presenting the derivation and the results. However, this derivation can easily be extended for more than one group by using the vectorial form of the GCM. That is,

$$vec(\mathbf{Y}) = (\mathbf{X}' \otimes \mathbf{Z})vec(\mathbf{B}) + vec(\mathbf{E})$$

which has the same structure as the single group model used in our derivation, as a result the test statistic can be represented as the one presented in equation (5.4).

Consider now the estimated likelihood, where we replace $\mathbf{\Sigma}$ in the original likelihood in (5.1) with its estimator $\tilde{\mathbf{S}}$

$$EL = (2\pi)^{-\frac{1}{2}np} \mid \tilde{\mathbf{S}} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}} \ .$$

We follow the likelihood ratio test procedure and maximize the estimated likelihood with respect to the null space ($\mathbf{B} = \mathbf{0}$) and the complete parametric space as

$$EL_{H_0} = (2\pi)^{-\frac{1}{2}np} \mid \tilde{\mathbf{S}} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{YY}')\}} \ ,$$

$$\begin{aligned} \arg\max_{\mathbf{B}} EL &= (2\pi)^{-\frac{1}{2}np} \mid \tilde{\mathbf{S}} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}} \ , \\ &= (2\pi)^{-\frac{1}{2}np} \mid \tilde{\mathbf{S}} \mid^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{Z}\tilde{\mathbf{B}}\mathbf{X})(\mathbf{Y}-\mathbf{Z}\tilde{\mathbf{B}}\mathbf{X})'\}} \ , \end{aligned}$$

where the derivation of $\tilde{\mathbf{B}}$ has been shown in detail in the next section. The test statistic is then constructed by taking the ratio of the above two quantities and is given as

$$\phi = tr\{\tilde{\mathbf{S}}^{-1}\mathbf{Z}(\mathbf{Z}'\tilde{\mathbf{S}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{S}}^{-1}\mathbf{YX}'(\mathbf{XX}')^{-1}\mathbf{XY}'\} \ . \tag{5.4}$$

Although the primary objective of this chapter is hypothesis testing on the mean parameters of the GCM, with particular application to high-dimensional longitudinal data, we also considered estimation of the parameter matrix $\mathbf{B}$ and investigated its properties using simulations.

To obtain an estimator, which we denote by $\tilde{\mathbf{B}}$, we maximize the logarithm of the estimated likelihood EL, using matrix differentiation results from Pan and Fang (2002):

$$log(EL) = -\frac{pn}{2}log(2\pi) - \frac{n}{2}log \mid \tilde{\mathbf{S}} \mid -\frac{1}{2}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\} \ .$$

Taking the derivative of the log of the estimated likelihood with respect to **B** gives

$$\begin{aligned}\frac{\partial log(EL)}{\partial \mathbf{B}} &= \frac{\partial log(EL)}{\partial vec\mathbf{B}} \\ &= -\frac{1}{2}\frac{\partial tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}}{\partial vec(\mathbf{B})} \\ &= -\frac{1}{2}\frac{\partial vec'(\mathbf{Y}-\mathbf{ZBX})}{\partial vec(\mathbf{B})} \times \frac{\partial tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}}{\partial vec(\mathbf{Y}-\mathbf{ZBX})} . \quad (5.5)\end{aligned}$$

Since $vec(\mathbf{Y}-\mathbf{ZBX}) = vec(\mathbf{Y}) - (\mathbf{X}'\otimes\mathbf{Z})vec(\mathbf{B})$,

the first part of the derivative is $\frac{\partial vec'(\mathbf{Y}-\mathbf{ZBX})}{\partial vec(\mathbf{B})} = -(\mathbf{X}\otimes\mathbf{Z}')$.

Now consider the second part of the derivative,

$$\begin{aligned}tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\} &= tr\{(\mathbf{Y}-\mathbf{ZBX})'\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})\} \\ &= vec'(\mathbf{Y}-\mathbf{ZBX})(\mathbf{I}_n\otimes\tilde{\mathbf{S}}^{-1})vec(\mathbf{Y}-\mathbf{ZBX}).\end{aligned}$$

As a result, its derivative is given by

$$\begin{aligned}\frac{\partial tr\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})(\mathbf{Y}-\mathbf{ZBX})'\}}{\partial vec(\mathbf{Y}-\mathbf{ZBX})} &= 2(\mathbf{I}_n\otimes\tilde{\mathbf{S}}^{-1})vec(\mathbf{Y}-\mathbf{ZBX}) \\ &= 2vec\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})\} .\end{aligned}$$

Therefore, the derivative of the estimated likelihood with respect to the parameter matrix **B** in (5.5) can be re-written as

$$\begin{aligned}\frac{\partial log(EL)}{\partial \mathbf{B}} &= (\mathbf{X}\otimes\mathbf{Z}')\ vec\{\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})\} \\ &= vec\{\mathbf{Z}'\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})\mathbf{X}'\}.\end{aligned}$$

Setting $\frac{\partial log(EL)}{\partial \mathbf{B}} = 0$ implies $\mathbf{Z}'\tilde{\mathbf{S}}^{-1}(\mathbf{Y}-\mathbf{ZBX})\mathbf{X}' = 0$, and solving for **B** results in

$$\tilde{\mathbf{B}} = (\mathbf{Z}'\tilde{\mathbf{S}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{S}}^{-1}\mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1} \ . \tag{5.6}$$

Note that this estimator is valid in high-dimensional scenarios since $\tilde{\mathbf{S}}$ is always positive definite irrespective of whether $n < p$ or $n > p$. We would also like to highlight that maximization is done under the assumption of full rank design matrices similar to the likelihood estimators provided by Khatri (1966). Solutions for the general problem (without the full rank assumption) can be derived using von Rosen's (1989) approach. However, the maximum likelihood as well as the estimated likelihood do not have unique solutions under general conditions.

5.3 Simulation Study

In this section, we examine the performance of the proposed test using extensive simulations. Details about the range of parameters used in the simulation are provided in Table 5.1.

Several values for the parameter matrix $\mathbf{B}$ were considered under the alternative to examine the power of the test. Multiple sample sizes, n, and number of time points p were considered, and several covariance structures as well as different strengths of within individual (across time) correlations were evaluated.

For all the scenarios, we calculated the empirical power and empirical level values. The mean parameter of the growth curve model is a matrix and hence the linear hypothesis considered in this chapter mostly involves comparison of several elements of the matrices.

To examine the performance of the estimator also, we use the Euclidean norm of the bias matrix and mean squared error (MSE) matrix from the zero matrix. In all the simulation scenarios considered, we present the results for two groups (i.e., $k = 2$) and for simplicity we assumed a linear growth curve. However, we will show in the results that this is assumed (selected) without loss of generality and the performance shown in the chapter and the optimality criteria established here are valid for any number of groups and any degree of polynomial (dimension of $\mathbf{B}$).

Table 5.1: Range of parameters considered in our simulation study

Number of groups, k	2
Growth curve	linear
Sample size, n	n=10, 15, 20, 25, 30, 35 and $n < p$
number of time-points, p	p=25, 40, 50, 100
Variance-Covariance matrix	Inverse-Wishart matrices
d.f. of the prior distribution of $\boldsymbol{\Sigma}$	$\nu \geq p$
scale matrix of the prior distribution of $\boldsymbol{\Sigma}$	Toeplitz, autocorrelated and unstructured
Euclidean norm of B	0 to 1.8, increment by 0.1
Underlying Distribution	Multivariate Normal

The structure of the design matrices used in the simulations are given by

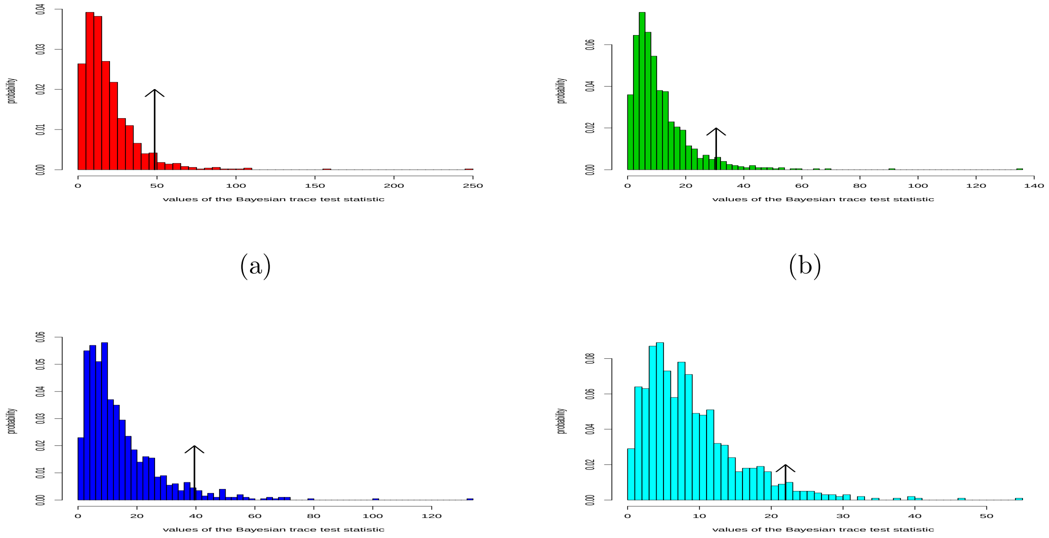
$$\mathbf{Z}' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & p \end{pmatrix} \text{ and } \mathbf{X} = \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix},$$

where $n_1 = n_2 = n/2$ if n is even, and $n_1 = n/2$ and $n_1 = (n-1)/2$, $n_2 = n_1 + 1$ if n is odd. The outcome variable $\mathbf{Y}$ is then generated using the growth curve model provided in equation (2.7), where $\mathbf{\Sigma}$ is generated from the Inverse-Wishart distribution with the specified hyper parameters (ν and $\mathbf{\Lambda}$), for which we have also considered several different values. For each scenario, the empirical distribution for the test statistic was first generated using 50,000 simulations and critical values are calculated using the 95^{th} percentile of the empirical distribution. Another 10,000 independent simulations were then used to calculate empirical level and power of the test statistic under the many scenarios considered.

5.3.1 Simulation Results

We first explored the null distribution of the test statistic presented in equation (5.4) for several scenarios. The null distribution is generated from 10,000 simulations. The histograms for $p = 25$, $\nu = p + 1$ and four different sample sizes are presented in Figure 5.1, where the empirical critical value is indicated using an arrow.

The histograms indicate that for all the scenarios considered the null distribution is positively skewed, similar to the trace test (Hamid et al., 2011). The range of the test statistic values and the corresponding critical value decreases as n increases, where the distribution becomes more skewed to the left also. These results are consistent across different values of p and different sample sizes. We also observed that the range of the test statistic decreases as the hyper parameter ν increases for a fixed p. The same applies for empirical critical values.



(a) (b)

(c) (d)

Figure 5.1: The null distribution of the test for p = 25 and a) n=10, b) n=15, c) n=20 and d) n=25.

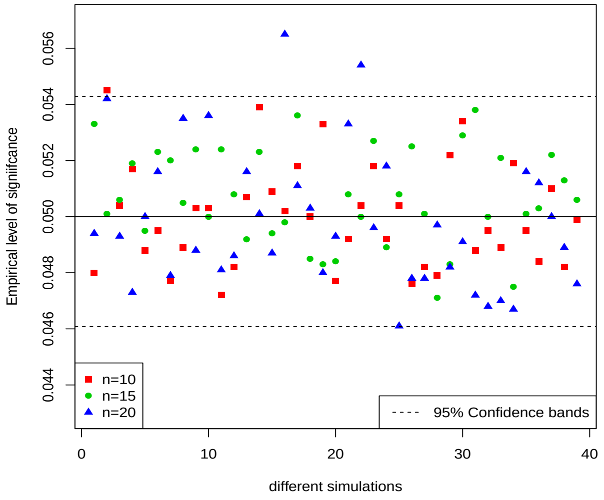


Figure 5.2: Empirical level of the proposed test averaged over 10,000 simulations across n=10, 15, 20, for $p = 25$ and $\nu = p + 1$.

For assessing the level of the proposed test, we set the nominal level of significance to 0.05 and empirical level was calculated after averaging over 10,000 simulations and the results are presented in Figure 5.2. As we can see from the figure, the empirical level values are very close to the nominal level and are randomly distributed within (95% confidence bounds) around the nominal level, irrespective of the sample size.

We considered 38 different values of $\mathbf{B}$ under the alternative hypotheses to investigate the power of the proposed test. In order to examine symmetry, we considered half of the scenarios (19 $\mathbf{B}$ matrices) with positive elements and the remaining 19 with exactly the same values but negative elements, (leading to the same Euclidean norm) but different signs.

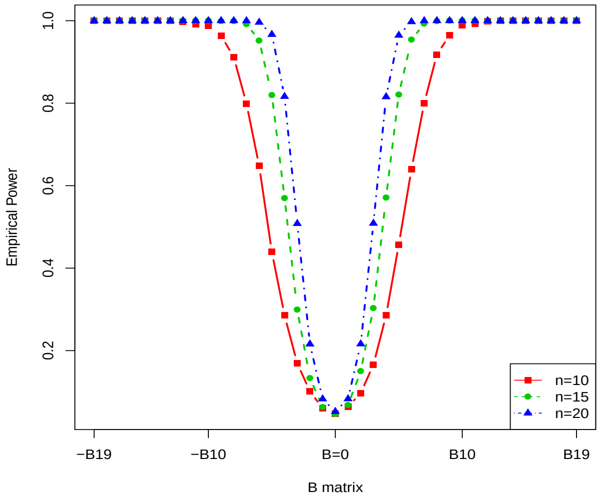


Figure 5.3: Empirical power of the test for $p = 25$ and $\nu = p+1$ for different values of $\mathbf{B}$ across n=10, 15, 20, where Toeplitz structured scale matrix is used for the prior.

The $\mathbf{B}$ matrices are plotted on the x-axis (as ordinal categories) according to their magnitude of deviation from the null matrix, as measured by the Euclidean norm but on opposite sides of the $\mathbf{B} = \mathbf{0}$ matrix, according to the signs of the elements. For clarity, if for instance, a 2×2 matrix $\mathbf{B}_1$ consists of elements (1,1,1,1), the other matrix with the same Euclidean norm will have elements (-1,-1,-1,-1), which is determined by multiplying $\mathbf{B}_1$ by (-1). The results of the simulation are presented in Figure 5.3, where on the x-axis the parameter matrices with only positive elements are on the right side of the null $\mathbf{B}$ and the parameter matrices with only negative elements are on the left side of the null $\mathbf{B}$.

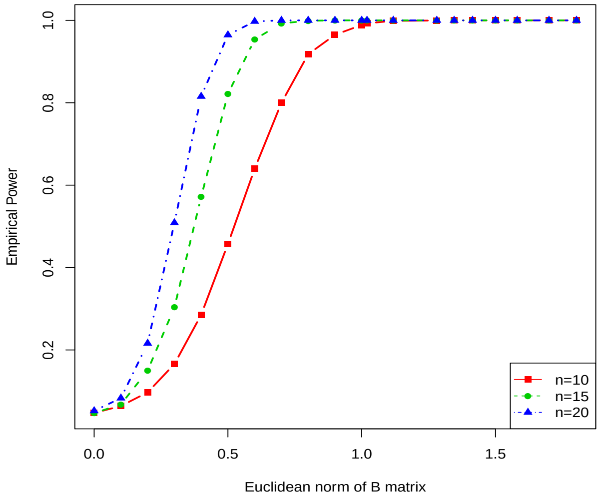


Figure 5.4: Empirical power of the test for $p = 25$ and $\nu = p+1$ for different values of $\mathbf{B}$ across 3 different sample sizes n=10, 15, 20, where each alternative is represented on the x-axis using the Euclidean distance from the null hypothesis.

Figure 5.3 shows that the test statistic is symmetric with respect to the direction of the alternative hypothesis from the null hypothesis, where two alternatives with the same distance from the null hypothesis have identical power of being rejected by the test. So, the sign of the matrix $\mathbf{B}$ does not affect power. The figure also shows that the test is unbiased, because the probability of rejecting an alternative hypothesis is shown to be always greater than the probability of rejecting the null. This is consistent for all the scenarios. Monotonicity with respect to departure from the null is also observed in Figure 5.3. In order to clearly show this, however, we calculated the Euclidean distance for all the alternatives considered and presented the power curve with respect to Euclidean norm in Figure 5.4. The power values are also presented in Table 5.2.

It is evident from the graph in Figure 5.4 that the power of the test statistic is monotonically increasing with respect to increasing distance from the null hypothesis, and is also monotone with respect to sample size. Monotonicity with respect to sample size is displayed in detail in Figure 5.4, where power is presented as a function of sample size. This is true for different covariance structures (prior distribution of $\mathbf{\Sigma}$). In this study, we had considered the Toeplitz structure, autocorrelated structure as well as unstructured covariance structures. For each structure, we also considered different strengths of correlations (e.g., weak, moderate and strong between measurements across time), and we observed increased power for scenarios with weaker correlation compared to scenarios with strong correlation.

It is important to highlight that the proposed test performs well in detecting very small departures from the null hypothesis. For instance, for $p = 25$, departure of 0.7 unit distance was rejected with a minimum of 80% power, with only a sample size of 10 (5 in each group), indicating a very strong power, while the level remains close to the nominal level. Departure of 1 unit distance was rejected with almost a power of 100% with a sample size of only $n = 10$.

Table 5.2: Empirical power of the Bayesian trace test with respect to the Euclidean norm of $\mathbf{B}$ for different values of n and different d.f., ν and $p = 25$, fixed $\mathbf{\Sigma}$

Euclidean norm of $\mathbf{B}$	$\nu = p + 1$			$\nu = p + 5$		
	n=10	n=15	n=20	n=10	n=15	n=20
0	0.0469	0.0467	0.0523	0.0488	0.0512	0.0525
0.1	0.064	0.0679	0.0831	0.0666	0.0815	0.0945
0.2	0.0971	0.15	0.2162	0.1215	0.1892	0.2764
0.3	0.1658	0.3035	0.5087	0.2356	0.4169	0.5808
0.4	0.2854	0.5717	0.8155	0.4153	0.7093	0.8685
0.5	0.4572	0.8216	0.9646	0.6373	0.9094	0.9804
0.6	0.6398	0.9538	0.9975	0.8163	0.9846	0.9984
0.7	0.7999	0.9927	0.9999	0.9333	0.9977	0.9999
0.8	0.9179	0.999	1	0.9803	1	1
0.9	0.9649	1	1	0.9953	1	1
1	0.9888	1	1	0.999	1	1
1.02	0.9928	1	1	0.9998	1	1
1.12	0.9988	1	1	0.9999	1	1
1.28	0.9997	1	1	1	1	1
1.34	1	1	1	1	1	1
1.41	1	1	1	1	1	1
1.5	1	1	1	1	1	1

An additional important feature of the proposed test is its performance in the near singularity zone, where unlike our previously proposed moderated trace test based on the Moore-Penrose generalized inverse, the test proposed in this manuscript continues to have a very strong power near singularity (that is when $n \approx p$).

In fact, power increases monotonically with respect to sample size while for the previously proposed test, it starts to decline in the near-singularity zone as shown in Figure 5.5 and Table 5.3. In order to allow several sample sizes before we reach $n = p$, $p = 40$ is considered in this simulation.

Table 5.3: Empirical power of the Bayesian trace test statistic and the moderated trace test statistic for fixed $\mathbf{B}$ in the non-singularity and near-singularity zone

n	p=40		n	p=100	
	Bayesian	G-inverse		Bayesian	G-inverse
5	0.137	0.067	30	0.19	0.211
10	0.363	0.12	40	0.293	0.354
15	0.673	0.523	50	0.477	0.513
20	0.896	0.831	60	0.611	0.633
25	0.981	0.954	70	0.752	0.664
30	0.997	0.963	80	0.823	0.729
35	0.999	0.887	90	0.88	0.6
39	1	0.491	99	1	0.55

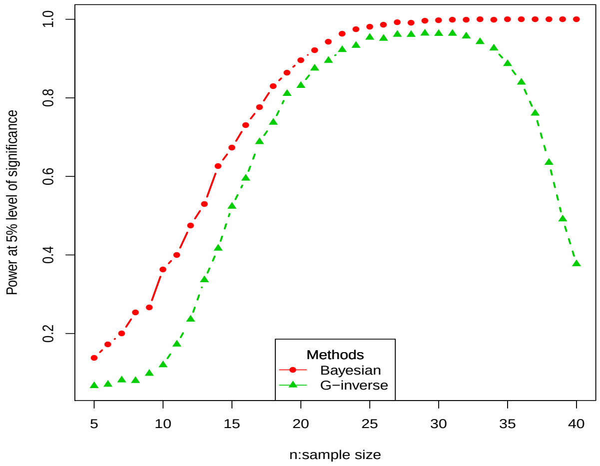


Figure 5.5: Comparison of the Bayesian trace test statistic with the moderated trace test statistic, with respect to empirical power, for $p = 40$ against n for fixed $\mathbf{\Sigma}$, $\mathbf{B}$.

5.3.2 Simulation Results for Estimating B

As discussed in the previous sections, the maximum likelihood estimators (MLE) proposed by Khatri (1966) or von Rosen (1989) in the general case are not defined under high-dimensional scenarios because of singularity of $\mathbf{S}$ due to high-dimensionality. The estimator provided in this chapter is, therefore, an useful alternative. In this section, we investigate bias and MSE for the proposed estimator and compare its performance with an estimator we previously proposed using the Moore-Penrose generalized inverse.

Empirical results from the simulation also indicate that the bias is randomly distributed around zero. This is done by considering element by element bias of the estimator for the parameter matrix.

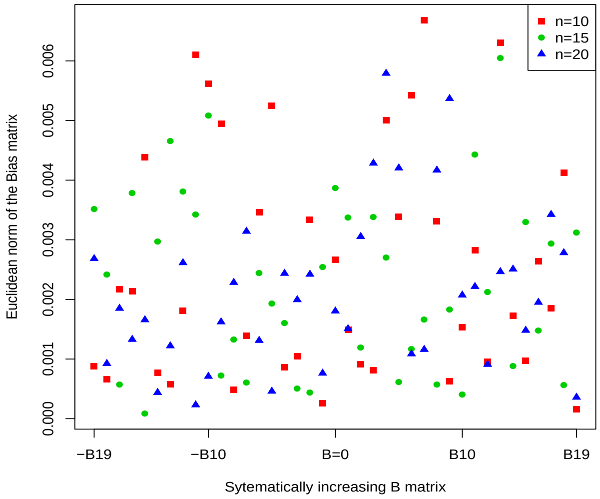


Figure 5.6: Euclidean norm of the bias matrix for $p = 25, \nu = p + 1$ for different scenarios of $\mathbf{B}$, across 3 different sample sizes.

We also calculated the Euclidean distance of the bias matrix from the zero matrix, which is nothing but the Euclidean norm of the bias matrix, and the results are displayed in Figure 5.6 and Table 5.4. The correlation structure considered here is the Toeplitz structure.

Table 5.4: Euclidean norm of the empirical bias matrix of the Bayesian MLE in (5.6) for different values of n and $\nu = p + 5$ for two different p.

Euclidean distance	$p = 25$			$p = 40$		
	n=10	n=15	n=20	n=10	n=20	n=30
0	0.00129	0.000558	0.000849	2.28E-03	0.002118	0.001928
0.1	0.001154	0.002025	0.001022	2.17E-03	5.65E-04	8.66E-05
0.2	0.000739	0.001495	0.001777	0.005275	0.002715	0.000147
0.3	0.0019	0.001372	0.000661	2.68E-03	6.88E-05	8.68E-04
0.4	0.000972	0.002403	0.003745	0.002662	0.001399	0.000266
0.5	0.002325	0.003558	0.001984	0.003337	0.002714	0.000721
0.6	0.003448	0.002122	0.000579	1.31E-03	2.66E-05	2.24E-05
0.7	0.004242	0.002562	0.001331	5.64E-05	2.19E-03	6.26E-04
0.8	2.15E-03	2.95E-05	1.53E-03	0.002565	0.001509	0.000856
0.9	0.001972	0.000436	0.004809	0.001495	0.001437	0.000959
1	0.000823	0.003366	0.002944	0.002485	0.000241	0.00109
1.02	0.001043	0.002432	0.001951	4.17E-04	3.51E-04	8.77E-05
1.11	0.000478	0.001199	0.000401	2.34E-03	0.002921	0.000923
1.28	0.003253	0.000285	0.001432	1.38E-03	0.000546	0.001547
1.34	0.000355	0.000318	0.001701	1.73E-03	0.000551	0.000257
1.41	2.51E-03	1.39E-03	1.80E-05	2.10E-03	0.002095	0.000735
1.5	0.002885	0.00308	0.00039	1.12E-03	0.000748	0.003873

Figure 5.7 shows that the MSE of the proposed estimator decreases monotonically with respect to sample size, indicating consistency of the estimator. For our previously proposed estimator, however, a pattern similar to power was observed, where a

monotone decline in MSE was first observed and when we approach near singularity, it starts to increase. Note also that, away from the near singularity zone, the two estimators have similar MSE values.

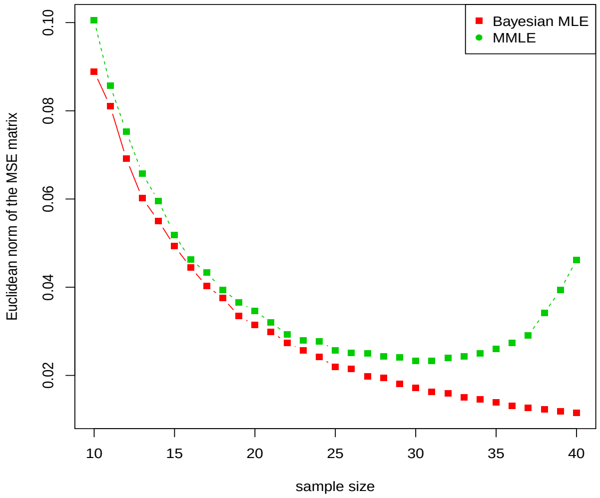


Figure 5.7: Comparison of the Euclidean norms of the empirical MSE matrix for $p = 40$ against increasing sample sizes for the two different estimators.

We have also observed that the empirical MSE decreases with increasing degrees of freedom of the prior distribution, indicating that the performance of the estimator is better for higher degrees of freedom of the prior distribution (Figure 5.8). This result is similar to what we observed in the power comparison, where the proposed test is shown to have higher power for higher degrees of freedom of the prior distribution (Table 5.2).

It is important to note that improved performance seen with increasing d.f. of the prior distribution can also be regarded as a limiting behaviour of the proposed test and estimator. This is because, as the degrees of freedom increases, the prior distribution dominates and the contribution from the data is minimized. We investigated this further and our results suggest that as $\nu \to \infty$, $\tilde{\mathbf{S}} \to \mathbf{\Lambda}$, and so with increasing d.f. of the prior distribution, estimator $\tilde{\mathbf{S}}$ is indeed dominated by the target matrix (which is the scale parameter of the prior distribution) rather than the data. The practical implication of this result is that, if the prior dominates, performance will be sensitive to mis-specification of the parameters of the prior distribution. Nevertheless, this limitation exists for any method relying on informative priors and a careful choice of priors is required to achieve a desired performance.

Below, we present some sensitivity results with respect to the choice of hyper parameters of the prior distribution. Here again, without loss of generality, we considered the Toeplitz and autoregressive structures for the covariance matrix, where we are only required to specify one element of the matrix.

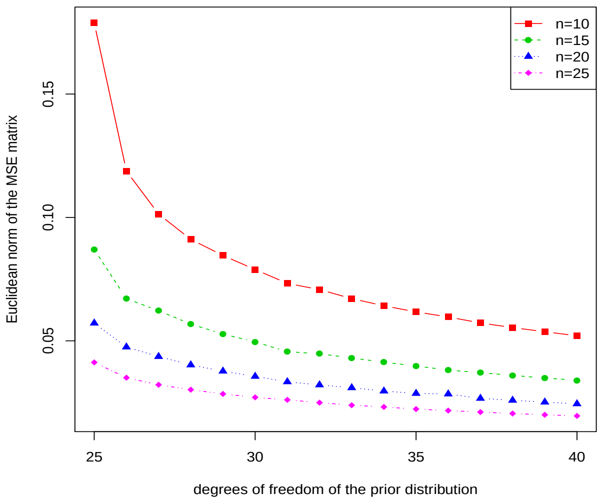


Figure 5.8: Euclidean norm of the empirical MSE matrix for $p = 25$ against increasing d.f. of the prior distribution of $\mathbf{\Sigma}$ for 4 different sample sizes

We first generated data using the specific d.f. and scale matrix, and calculated the critical value using 10,000 simulations from the true distribution. We then mis-specified the first element of the correlation matrix without disrupting the structure. Empirical level and power were then investigated for the mis-specified scale matrix. For comparison purposes, the d.f. for the prior distribution was kept the same. The results are presented in Figures 5.9 and 5.10. The term k represents the degree of misspecification. The confidence bands correspond to those of the true specification.

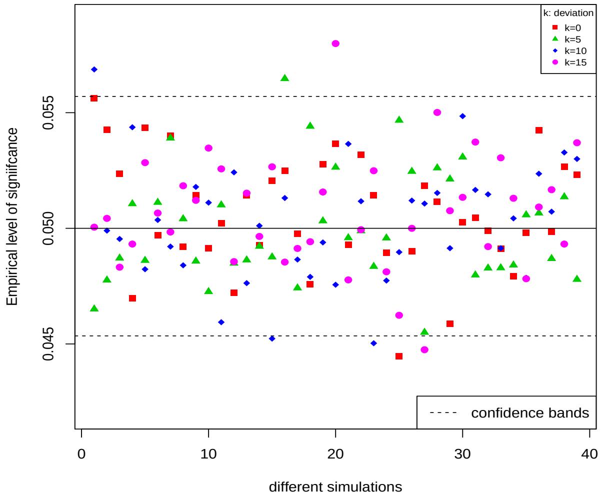


Figure 5.9: Empirical level of the proposed test for $p = 25$, $\nu = p + 1$, $n = 20$ under Toeplitz structure scale matrix across different degrees of mis-specifications.

As can be seen from the figure, the level of the test remains quite close to the nominal level, where on average the empirical level is identical to the nominal level, with random fluctuations within the 95% confidence bands.

The power curve presented in Figure 5.10 also indicates that there is very small sensitivity with respect to mis-specification of the scale matrix of the prior distribution, where larger departure from the true parameter leads to larger decline in power. However, it is important to highlight here that, changing the scale parameter for the prior distribution changes the correlation between observations across time (that is, it changes $\mathbf{\Sigma}$). In our particular simulation formulation, an increase in the scale parameter means an increase in the correlation across time. Therefore, the decline in power with respect to mis-specification that we see in Figure 5.10 can also be attributed to the increase in correlation, and we have discussed earlier in this chapter that a decrease in power (and an increase in MSE) occurs as correlation among the observations increases. We would also like to note that decrease in power with increased correlation is also seen in previously proposed tests (non-high dimensional) in analysis of longitudinal data, including Khatri's LR test (Khatri, 1966) and Hamid et al.'s (2011) trace test.

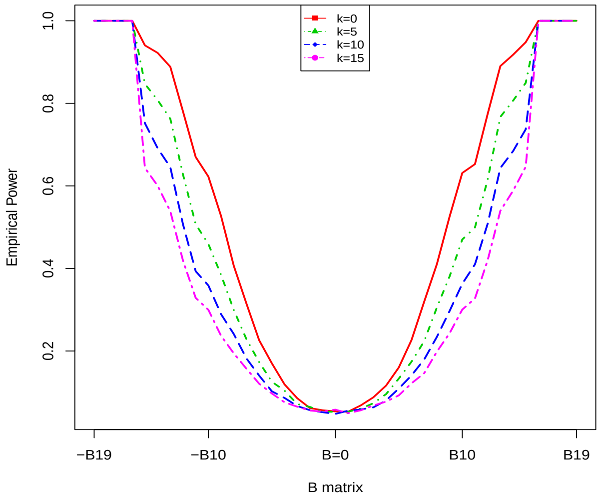


Figure 5.10: Empirical power of the test for $p = 25$, $\nu = p+1$, $n = 20$ under Toeplitz structure scale matrix across different degrees of mis-specifications.

5.4 Real Data Illustration

In most time-course genetic experiments, data from large number of genes (often tens or hundreds of thousands) are collected. However, only a small percentage of the genes are expressed and hence the majority are not relevant to the studies being conducted. Previous findings, in fact, suggest that only about 1%-5% of genes are often relevant. Therefore, gene filtering is an important step in the analysis of genetic data, including time course genetic experiments, where genes that are not expressed (e.g., house keeping genes) are removed from the analysis before a formal statistical analysis for identifying differentially expressed genes is performed. After gene filtering is done, the top ranked genes are often subjected to further analysis (e.g., to identify genes that are differentially expressed over time or that are associated with outcome of interest).

In typical genetic experiments, where no repeated measurements are taken across time, some univariate measures of variability (e.g., inter quartile range (IQR) or variance) between individuals is used as the cut-off value to identify relevant genes and remove house keeping or otherwise noise genes. However, in time course genetic studies such univariate measures of variability are not appropriate.

In this study, we use gene specific Growth Curve Model and use the hypothesis of $H_0 : \mathbf{B}_g = \mathbf{0}$ vs $H_1 : \mathbf{B}_g \neq \mathbf{0}$ for gene filtering, where we rank the genes according to the test statistic proposed in this study and identify noise (irrelevant) genes. We used a publicly available data set obtained from the Gene Expression Omnibus (GEO), under a series named GSE4917. Data consists of genetic measurements from 3 samples of ($n = 3$) human breast cells treated with Dexamethasone for 30 minutes, 2 hours, 4 hours and 24 hours and measurements were taken at these 4 time points ($p = 4$).

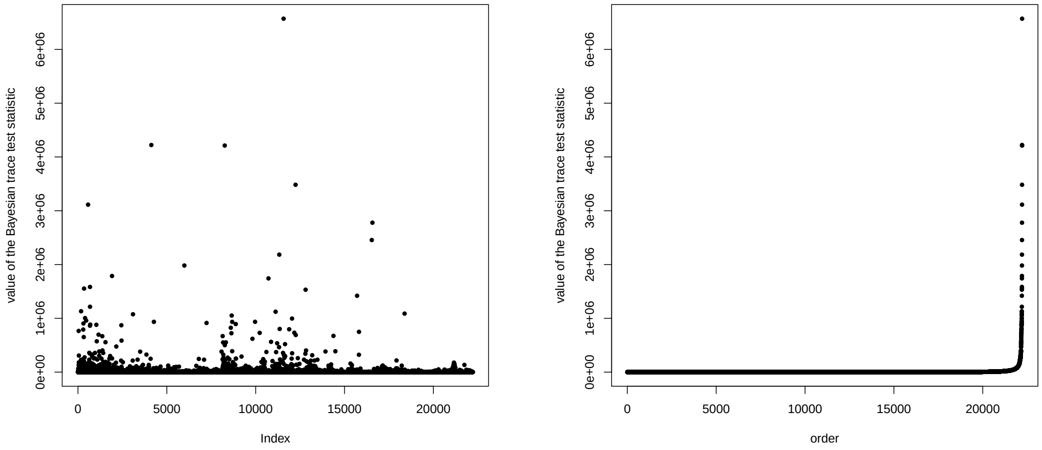
A total of 22,215 probe sets, that represent approximately 16,000 human genes were investigated in the study. In the original study, the breast cells were subjected to 72 hour of growth factor withdrawal before Dexamethasone treatment.

The objective of the original study was to examine change in gene expression after Dexamethasone treatment for glucocorticoid-mediated mammary epithelial cell survival signaling because activation of GR, which is essential for identifying stress response (Wu et al., 2006). Here, we use this data set to illustrate our proposed test for gene filtering, where we fitted gene-specific growth curve models for each of the 22,215 probe sets. That is, for each gene, the following GCM is fitted

$$\mathbf{Y}_g = \mathbf{Z}\mathbf{B}_g\mathbf{X} + \mathbf{E}_g,\ \mathbf{E}_g \sim N_{p\times n}(\mathbf{0}, \boldsymbol{\Sigma}_g, \mathbf{I})\ ,$$

where $\mathbf{Z}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0.5 & 2 & 4 & 24 \end{pmatrix}$ and $\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$.

A linear curve over time is assumed to represent the mean, after a preliminary exploratory analysis was performed and the profile plots for randomly selected genes were investigated. Using data from each gene, the hypothesis of $H_0 : \mathbf{B}_g = \mathbf{0}$ *vs.* $\mathbf{B}_g \neq \mathbf{0}$ was tested using the proposed test provided in (5.4) and genes were ranked according to the values of the test statistic. The hyper-parameter $\boldsymbol{\Lambda}$ was constructed using a Toeplitz structure, after investigating the structure of the sample correlation matrices of a few genes. The first element of the sample correlation matrix was used as an estimate of the first element of the matrix $\boldsymbol{\Lambda}$. We used $\nu = p + 2$ to allow relatively vague (non-informative) prior and to allow the expected value of the inverse Wishart matrices to be defined. A scatter plot of test statistic values for all the genes is presented in Figure 5.11. As we can see from the plots, the test statistic values are very small for majority of the genes and only a small percentage appear to have relatively large test statistic values.



(a) (b)

Figure 5.11: Scatter plot of the test statistic value for all the genes from the real data a) scatter plot b) ordered plot of the test statistic

We annotated the top 10 genes from our analysis using GeneAnnot (`http://genecards.weizmann.ac.il/cgi-bin/geneannot/GA_search.pl`), a tool for microarray gene annotation of Affymetrix probe-sets. The results of the annotation are presented in Table 5.5. The annotation indicates that all of the top 10 ranked genes were previously shown to be associated with breast cancer and other forms of cancer, indicating that our test statistic was able to identify the relevant genes.

Table 5.5: Description of the top 7 ranked genes, ordered according to their ranks

Gene Code	**Description**	**Location**	**Comments**	**References**
SRFP4	secreted frizzled-related protein 4	Chr 7	Expressed in prostate endometrial and invasive breast cancer	Wissman et al. (2003), Abu-Jawdeh et al. (1999)
FKBP15	FK506 binding protein 15, 133kDa	Chr 9	reference gene for breast cancer	Ibusuki et al. (2013)
GRK4	G protein-coupled receptor kinase 4	Chr 4	expressed in human breast cancer	Matsubayashi et al. (2008)
MEF2C	myocyte enhancer factor 2C	Chr 5	expressed in breast cancer cell lines	Ostrander et al. (2007)
SOX10	SRY (sex determining region Y)-box 10	Chr 22	strongly expressed in normal myoepithelial cells present in mammary glands	Vargas et al. (2012), Nonaka et al. (2008)
APOD	apolipoprotein D	Chr 3	Expressed in breast cancer	Lopez-Boado et al. (1994)
RNF44	ring finger protein 44	Chr 5	Upregulated in parous control breast cancer	Balogh et al. (2007)
HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	Chr 10	differentially expressed in cancer cells, regulated in human breast cancer cells	Hai et al. (2008), Wang et al. (2007)
PRDX1	peroxiredoxin 1	Chr 1	overexpressed in breast cancer tissues	Chang et al. (2007), Noh et al. (2001), Karihtala et al. (2003)
PTGER	prostaglandin E receptor 3 (subtype EP3)	Chr 1	expressed in breast cancer cell	Casa et al. (2012)

5.5 Discussion

In this chapter, we considered a Bayesian framework and developed a moderated estimator for the population covariance matrix. We then, constructed an LRT and an estimator for the mean parameter of a GCM. Using extensive simulations, we showed that the test exhibits desirable properties of a good test such as monotonicity, unbiasedness and symmetry. The simulation results also show that the test performs very well, where the level under all scenarios is close to the nominal level.

The results also indicate that the test statistic was able to reject small departures from the null hypothesis with reasonably good power. Unlike a previously proposed test based on the Moore-Penrose generalized inverse (Jana et al., 2017), the test proposed here performs well in the near-singularity area (when $n \approx p$). The estimator for the mean parameter was also shown to be empirically unbiased and consistent, where the empirical bias is negligible and randomly distributed around zero and mean squared error declines monotonically to zero as sample size increases. We also examined sensitivity of its performance to mis-specification of the parameters of the prior distribution, empirically. Using a real data example, we showed that the test can be used in gene ranking and gene filtering in time course genetic experiments.

Finally, we would like to highlight that, although the test is developed for the simple hypothesis of $H_0 : \mathbf{B} = \mathbf{0}$ vs. $H_1 : \mathbf{B} \neq \mathbf{0}$, our approach can be extended to the general linear hypothesis $H_0 : \mathbf{GBF} = \mathbf{0}$ vs. $H_1 : \mathbf{GBF} \neq \mathbf{0}$, allowing several hypotheses on the mean parameter $\mathbf{B}$ to be tested. The practical implication of this generalization is that our proposed test can be used not only for gene filtering as illustrated here, but also to test differential gene expression profiles among different groups (e.g., cancer vs normal), using appropriate values for $\mathbf{G}$ and $\mathbf{F}$.

Chapter 6

Concluding Remarks

The growth curve model (GCM) is a generalized multivariate analysis of variance (GMANOVA) model useful for analysis of repeated outcomes, with means associated with continuous variables such as time, treatment, doses or temperature. Unlike ANOVA and MANOVA models, the mean is structured across time, temperature, dosages etc. The model is constructed in a general bilinear setup, whereas ANOVA and MANOVA are included as special cases.

Maximum likelihood estimates (MLEs) for the model parameters under the assumption of multivariate normal distribution are currently available for both the GCM and the extended GCM (EGCM). The inferential tools developed for the traditional GCM and EGCM might not be optimal when certain assumptions are invalid. In this thesis, we showed sensitivity of existing estimators to departures from the normality assumption. Motivated by that, we relaxed the assumptions and considered inferential for the GCM and the EGCM, under a more general distribution, the multivariate skewed normal (MSN) distribution, which also includes multivariate normal as a special case .

We adopted the EM algorithm and the REM algorithm to derive estimators for the model parameters. To allow REM algorithm for matrix estimation, we extended the multivariate (vector) Newton-Raphson (NR) method and the REM algorithm. Existing tests for the GCM strictly assumes a non high-dimensional regime, that is, it assumes that sample size is greater than the number of time points ($n > p$), leading to singularity of the sample covariance matrix, under a high-dimensional regime. As a result, it was necessary to consider, inference on GCM under high-dimensionality. We used a Bayesian framework to develop an alternate estimator for the sample covariance matrix, which is always positive definite, regardless of the regime, and constructed an estimator and a test for testing the significance of the mean parameter of the GCM under high-dimensionality. Empirical power was used as an evaluation criterion for examining the performance of the proposed test whereas empirical bias and MSE were used as evaluation criteria for studying the behaviour of the developed estimators.

Since the empirical bias and MSE of the parameter matrix are matrices, making comparisons among the different scenarios becomes difficult, unless we reduce the matrix measures to some scalar quantity. We used a matrix norm for this purpose. The Euclidean norm or in other words, the Frobenius norm is a popular matrix norm and is easy to interpret. As a result, we used this norm to reduce our bias and MSE matrices to scalar quantities to compare across different scenarios. The Euclidean norm also measures the distance between the bias or MSE matrix from 0 bias or 0 MSE respectively, and seems to be an apt measure of empirical bias and MSE.

Similar to bias and MSE, evaluating performances using power too becomes relatively complicated. Establishing monotonicity with respect to departure from the null hypothesis also requires some measure of distance from the null. We used a

Euclidean distance (norm) to measure departure from the null hypothesis of $\mathbf{B} = \mathbf{0}$. Since the norm of the null matrix ($\mathbf{B} = \mathbf{0}$) is always 0, the norm of a non-null $\mathbf{B}$ matrix quantifies the deviations of the alternative hypotheses from the null hypothesis. Thus, it can also be interpreted as the distances between the alternative hypotheses from the null hypothesis.

6.1 Summary of Research

In Chapter 3 (Jana et al., 2017b), we investigated the robustness of existing normal based estimators and provide alternatives for analyzing skewed data using the GCM. We first conducted extensive simulations to investigate the robustness of Khatri's MLEs for the parameters of the GCM derived under multivariate normal distribution. The results from our simulation show that Khatri's MLEs are sensitive to departures from normality, where bias and MSE are shown to increase with increasing levels of skewness.

To overcome this lack of robustness, we relaxed the distributional assumption for the error terms by considering the MSN distribution, which is a more general class of density function that includes the multivariate normal distribution as a special case. We derived the likelihood of the GCM under MSN and provided estimators for the model parameters using the EM algorithm. We conducted extensive simulations to evaluate the performance of the estimators and compared bias and MSE with Khatri's MLEs derived under the normality assumption. The comparative assessment of the proposed estimators for the mean parameter and the variance-covariance matrix demonstrated better empirical performance of the EM estimators than Khatri's MLE, for an MSN GCM.

We illustrated our method using real longitudinal data and showed that the estimated model fits the data well, where the difference between the estimated and observed mean is quite negligible. Nevertheless, the limitation of this method is that it is not applicable, when there are clusters in the data, that is, when different group means are represented by different shapes. This is beyond the scope of GCM. Therefore, further work using the EGCM under the MSN distribution was required to derive estimators useful for such scenarios (Jana et al, 2017c).

This provided motivation for the next chapter, where we considered the estimation problem for an EGCM, where the underlying distribution is MSN. An EGCM allows us to model longitudinal data with clusters, where means from each group require different degree of polynomials. Estimation techniques for the parameters of an EGCM already exist (Kollo and von Rosen, 2005).

The EGCM is a general form of GMANOVA models and hence can be expressed as a GCM with linear restrictions. In this thesis, we considered this linearly restricted form of the model and constructed the required likelihood for an MSN EGCM. Since a closed form solution of the likelihood is not possible to derive, the EM algorithm becomes a natural choice for solving the estimation problem. However, the EM algorithm is not applicable for solving likelihoods for models with linear restrictions. The Restricted EM (REM) algorithm, based on multivariate Newton-Raphson (NR) method and the Lagrangian optimization, serves this purpose. The REM algorithm and the multivariate NR method only apply to vector parameters. In our case, however, we are dealing with matrix parameters. We, therefore, extended the NR method and the REM algorithm for estimating matrix parameters, and then employed the extended REM algorithm to solve the estimation problem.

Extensive literature search was done to explore Newton-Raphson method for matrix imputation. To our knowledge, there was no existing methods for estimating matrices using the NR approach. Developing extensions of the multivariate NR method for matrix imputation, which is a novel contribution in numerical analysis, is one of the significant contributions of this thesis, in addition to inferences on the GMANOVA model under non-normality and high-dimensionality. There were several methodological challenges that were encountered while dealing with the matrix imputation problem, especially in determining the associated information matrix. The information matrix of a p dimensional vector only involves the variance and covariance terms of the p elements. In a matrix there are two sets of variances and covariances. One set of variances and covariances correspond to the rows and the other set to the columns, leading to two variance-covariance matrices for the observations in a matrix. Two variance-covariance matrices in turn lead to two information matrices. Consequently, information of a matrix estimator can be decomposed into two individual information matrices or we can also construct a grand information matrix which is nothing but the Kronecker product of the two information matrices. Identifying the correct and complete information matrix for a matrix estimator was the biggest challenge in developing an extension of the NR method and hence the extension of the REM algorithm. We feel both extensions are important contributions to numerical analysis and statistics.

The performances of the proposed estimators were assessed through extensive simulations. The results show that our estimators possess optimal properties. The goodness of fit of the fitted mean structure illustrated using real life skewed longitudinal data, demonstrated that the model and the proposed methods are useful for real life situations.

Finally, we consider a Bayesian framework to relax the assumption of the model, with the aim of extending its applications in high-dimensional data, where the number of parameters is often larger than the sample size (Jana et al., 2017a). In an earlier work (Jana et al., 2017) to overcome the challenges of high-dimensionality, we developed a test based on the Moore-Penrose generalized inverse, to test the significance of the mean parameter. The test overall performed well, however, it was observed that the performance of the test was bad near-singularity.

Our objective in Chapter 5 (Jana et al., 2017a) was to improve the performance of this test and provide a general setup useful for inference in the GCM in high-dimensional conditions, including the near-singularity zone. In doing so, we developed an estimator for the unknown population covariance matrix, constructed an estimated likelihood using this estimator and proposed an LRT and an estimator for the mean parameter. The estimator for the mean parameter and the test perform well under high-dimensional conditions in both the singularity ($n << p$) and near-singularity ($n \approx p$) zone.

We would like to highlight that, in Chapter 5, although we presented results involving $p = 25$ and $p = 40$, we conducted simulations using involving other p values. The results of the simulations indicate that the findings in this paper are generalizable to all values of p, as long as p is less than n. The performances indeed are functions of the ratio n/p. In this simulation we used n/p ratios ranging from 0.1 to 1. Nevertheless, GCMs are used in the analysis of longitudinal data with short to moderate time series and we rarely come across studies with more than 40 time points.

6.2 Future Directions

Further work is required to provide confidence intervals for the estimators for both the growth curve and extended growth curve models under the MSN distribution. The estimation procedures developed in this thesis for the GCM and EGCM, under MSN errors were developed for non-high dimensional ($n > p$) problems and might not be appropriate for high-dimensional ($n < p$) scenarios. Extending our work to include high-dimensional scenarios might be of interest for further study. This will be especially useful for genetic experiments with skewed outcomes.

It is also to be noted that, there are no existing inference techniques on traditional EGCM under high-dimensionality, even under multivariate normality. Therefore, this is also a possible direction for further study.

Another possible direction of future research is choice of distance matrices when evaluating performance of matrix estimators. In this thesis, we only used the Euclidean norm as the distance measure. However, we feel, several other distance measures also need to be considered and compared for the power analysis.

An extension of the Newton-Raphson (NR) method was required in order to extend the REM algorithm to matrix parameters. Nevertheless, the performance of the extended NR approach was not individually evaluated. Some of the limited performance of our estimators might, therefore, be attributed to that. Moreover, studying the behaviour of the matrix NR method under different conditions might be of interest in the field of numerical analysis.

Another limitation of our work is that the EM algorithm based estimators required more computation time than that of Khatri's. Hence, it might be interesting to consider variants of EM algorithm to decrease the computation time, but, they will not maintain the simplicity of the EM algorithm.

We would also like to highlight that the skewness parameters in this thesis were underestimated. Exploring further estimation techniques, like the profile likelihood method is, therefore, required to improve estimation of this parameter. Profiling with respect to the skewness parameter might help us obtain optimal estimators for the mean, skewness and variance parameters. This in turn might also improve the performance of the covariance estimator which is dependent on the skewness parameter. Azzalini and Capitanio (1999) also mentioned about encountering problems in parameter estimation in multivariate skew normal using the direct parameters e.g., the location, scale and slant parameter. However, they mentioned that using the centered parameters such as, the mean and variance parameters and a skewness index, considerably eased the estimation problem. They also suggested using the method of moments estimator as the initial estimates for the EM algorithm estimating the centered parameters.

Similar to any Bayesian method, a limitation of our proposed test for high-dimensional scenarios requires specification of a prior distribution, and a choice of prior can be challenging in practical applications. Here we considered a conjugate prior for mathematical simplicity. We have illustrated in Chapter 5 that the behaviour of the Bayesian test is sensitive to mis-specifications of the hyper-parameters. The results indicate that performance is slightly sensitive to departures from the true prior distribution, but, the performance of the proposed test is still reasonably good even under mis-specification. One approach to overcome this challenge could be by use of non-informative (or vague prior), which allows the data to dominate the results. Alternatively, prior knowledge on the biological condition can also be used to develop informative prior. Further study, therefore, is required to understand the impact of different priors on the performance of our proposed test.

Finally, we would like to highlight that a GCM is used to model only balanced and complete data. Its properties are, therefore, not investigated extensively under unbalanced data and in the presence of missing data. Mixed models are commonly used for analyzing longitudinal data in situations where data are unbalanced and incomplete. However, there are a few studies providing solutions for missing data in GCMs, which was done by considering a generalization of the model (Kleinbaum, 1973; Leeper and Woolson, 1982; Leeper and Chang, 1987). The advantages of using GCM over mixed models is the GCM has exact solutions and is more powerful and precise in practical applications involving small sample size, which is often the case in analysis of studies involving longitudinal data.. Mixed models, however, rely more on large sample theory (Corbeil and Searle, 1976; Stram and Lee, 1994; Hartley and Rao, 1967). Exploring comparative performance of the GCM with mixed model especially in the presence of missing data is a possible direction for future research.

Bibliography

[1] Abu-Jawdeh, G., Comella, N., Tomita, Y., Brown, L. F., et al. (1999). Differential expression of frpHE: a novel human stromal protein of the secreted frizzled gene family, during the endometrial cycle and malignancy. *Laboratory investigation; a Journal of technical methods and Pathology*, 79(4), 439-447.

[2] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. New York, John Wiley and Sons, Inc.

[3] Ando, A. and Kaufman, G.M. (1965). Bayesian Analysis of the Independent Multinormal Process. Neither Mean Nor Precision Known. *Journal of the American Statistical Association*, 60(309), 347-358.

[4] Arellano-Valle, R. B., and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, 96(1), 93-116.

[5] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171-178.

[6] Azzalini, A., and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4), 715-726.

[7] Azzalini, A., and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 579-602.

[8] Azzalini, A., and Capitanio, A. (2013). *The Skew-Normal and Related Families.* Cambridge University Press, England.

[9] Balakrishnan, N., and Scarpa, B. (2012). Multivariate measures of skewness for the skew-normal distribution. *Journal of Multivariate Analysis*, 104(1), 73-87.

[10] Balogh, G.A., et al. (2007). The breast of parous women without cancer has a different genomic profile compared to those with cancer. *International Journal of Oncology*, 31(5), 1165-1175.

[11] Banken, L. (1984), Eine verallgemeinerung des GMANOVA-Modells. Dissertation, University of Trier, Trier, Germany.

[12] Bertsekas, D.P. (1999). *Nonlinear programming.* Athena scientific, Belmont, Massachusetts.

[13] Box, G. E. (1950). Problems in the analysis of growth and wear curves. *Biometrics*, 6(4), 362-389.

[14] Box, G. E., and Tiao, G. C. (2011). *Bayesian inference in statistical analysis.* John Wiley & Sons, New York.

[15] Boyles, R. A. (1983). *On the convergence of the EM algorithm.* Journal of the Royal Statistical Society. Series B (Methodological), 47-50.

[16] Bryant, E., and Gillings, D. (1985). Statistical analysis of longitudinal repeated measures designs. In P.K. Sen (Ed.), *BIOSTATISTICS: Statistics in Biomedical, Public Health and Environmental Sciences* (pp. 251-282). Elsevier, New York.

[17] Casa, A. J., Potter, A. S., Malik, S., et al. (2012). Estrogen and insulin-like growth factor-I (IGF-I) independently down-regulate critical repressors of breast cancer growth. *Breast cancer research and treatment*, 132(1), 61-73.

[18] Casella, G., and Berger, R. L. (2002). *Statistical inference* (Vol. 2). Pacific Grove, CA: Duxbury.

[19] Chakravorti, S. R. (1975). Maximum-likelihood estimation for the growth curve model with unique dispersion matrices. *Communications in Statistics-Theory and Methods*, 4(1), 69-85.

[20] Chang, X., Li, D., Hou, Y., et al. (2007). Identification of the functional role of peroxiredoxin 6 in the progression of breast cancer. *Breast Cancer Research*, 9(6), R76.

[21] Chen, J. T., and Gupta, A. K. (2005). Matrix variate skew normal distributions. *Statistics*, 39(3), 247-253.

[22] Chen, S. X., and Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2), 808-835.

[23] Chen, L. S., Paul, D., Prentice, R. L., and Wang, P. (2011). A regularized Hotellings T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association*, 106(496), 1345-1360.

[24] Cheng, Y. W., and Street, D. J. (1997). The CGMANOVA model. *Communications in Statistics-Theory and Methods*, 26(5), 1083-1098.

[25] Chi, E. M., and Weerahandi, S. (1998). Comparing treatments under growth curve models: exact tests using generalized p-values. *Journal of Statistical Planning and Inference*, 71(1), 179-189.

[26] Corbeil, R. R., and Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18(1), 31-38.

[27] Crowder, M. J., and Hand, D. J. (1990). *Analysis of Repeated Measures.* CRC Press, Boca Raton, Florida.

[28] Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1), 1-13.

[29] Cox, R. T. (1961). *Algebra of Probable Inference.* JHU Press, Baltimore, USA.

[30] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 39(1), 1-38.

[31] Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96(456), 1151-1160.

[32] Efron, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, 100(469), 1-5.

[33] Elston, R. C., and Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics*, 18(2), 148-159.

[34] Elswick, R. K. (1985). The missing data problem as applied to the extended version of the gmanova model. Dissertation, Virginia Commonwealth University, Richmond, Virginia.

[35] Frerichs, R. R., Srinivasan, S. R., Webber, L. S., and Berenson, G. R. (1976). Serum cholesterol and triglyceride levels in 3,446 children from a biracial community: the Bogalusa Heart Study. *Circulation*, 54(2), 302-309.

[36] Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Sage Publications Inc., Thousand Oaks, CA, US.

[37] Fujikoshi, Y., and Satoh, K. (1996). Estimation and model selection in an extended growth curve model. *Hiroshima Mathematics Journal*, 26, 635-647.

[38] Geisser, S. (1970). Bayesian analysis of growth curves. *Sankhya: The Indian Journal of Statistics, Series A*, 32(1), 53-64.

[39] Geisser, S. (1980). 3 Growth curve analysis. *Handbook of Statistics*, 1, 89-115.

[40] Geisser, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. *Biometrika*, 68(1), 243-250.

[41] Genton, M. G. (Ed.). (2004). Skew-elliptical distributions and their applications: a journey beyond normality. CRC Press, Boca Raton.

[42] Genton, M. G., He, L., and Liu, X. (2001). Moments of skew-normal random

vectors and their quadratic forms. *Statistics and Probability Letters*, 51(4), 319-325.

[43] Gleser, L. J., and Olkin, I. (1966). A k-sample regression model with covariance. In *Multivariate Analysis* (pp. 59-72). Academic Press New York.

[44] Gleser, L. J., and Olkin, I. (1970). Linear models in multivariate analysis. *Essays in Probability and Statistics*, 39, 267-292.

[45] Gleser, L.J., and Olkin, I. (1972). Estimation for a regression model with an unknown covariance matrix. In *Proceedings of the Sixth Berkeley Symposium of Mathematical Statistics*, Prob. 1 , Ed. L.M. Le Cam, J. Neyman and E. Scott, (pp. 541-568). Berkeley and Los Angeles, Califomien: University of California Press.

[46] Grizzle, J. E., and Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics*, 25(2), 357-381.

[47] Gupta, A. K., and Chang, F. C. (2003). Multivariate skew-symmetric distributions. *Applied Mathematics Letters*, 16(5), 643-646.

[48] Gupta, M. R., and Chen, Y. (2011). *Theory and Use of the EM Algorithm.* Now Publishers Inc., Hanover, MA.

[49] Gupta, A. K., and Kollo, T. (2003). Density expansions based on the multivariate skew normal distribution. *Sankhya: The Indian Journal of Statistics*, 65(4), 821-835.

[50] Gupta, A. K., and Nagar, D. K. (2000). Matrix-variate beta distribution. *International Journal of Mathematics and Mathematical Sciences*, 24(7), 449-459.

[51] Hai, Y., Cao, W., Liu, G., et al. (2008). A G-tract element in apoptotic agents-induced alternative splicing. *Nucleic acids Research*, 36(10), 3320-3331.

[52] Hamid, J. (2001). Residuals in the extended growth curve model. U.U.D.M. Report 2001:29, ISSN 1101-3591, Department of Mathematics, Uppsala university, Uppsala, Sweden.

[53] Hamid, J. S., and Beyene, J. (2009). A multivariate growth curve model for ranking genes in replicated time course microarray data. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 1-26.

[54] Hamid, J. S., Beyene, J., and von Rosen, D. (2011). A novel trace test for the mean parameters in a multivariate growth curve model. *Journal of Multivariate Analysis*, 102(2), 238-251.

[55] Hamid, J. S., and Rosen, D. V. (2006). Residuals in the extended growth curve model. *Scandinavian Journal of Statistics*, 33(1), 121-138.

[56] Hamid, J. S., Roslin, N. M., Paterson, A. D., and Beyene, J. (2009). Using a latent growth curve model for an integrative assessment of the effects of genetic and environmental factors on multiple phenotypes. *BMC Proceedings*, 3(7), 1.

[57] Hamilton, W. C. (1970). The revolution in crystallography. *Science*, 169(3941), 133-141.

[58] Harrar, S. W., and Gupta, A. K. (2008). On matrix variate skew-normal distributions. *Statistics*, 42(2), 179-194.

[59] Hartley, H. O., and Rao, J. N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54(1-2), 93-108.

[60] Hastie, T., Friedman, J., and Tibshirani, R. (2001). Model assessment and selection. In *The elements of statistical learning* (pp. 193-224). Springer, New York.

[61] Healy, M. J. R. (1981). Some problems of repeated measurements. In Bithell JF, Coppi R (Eds.), *Perspectives in Medical Statistics* (pp. 155-172). Academic Press: London.

[62] Hoff, P. D. (2009). *A first course in Bayesian statistical methods.* Springer Science & Business Media, New York.

[63] Horn, R. A., and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press, New York.

[64] Huber, P.J. (1981). *Robust statistics.* Wiley, New York.

[65] Ibusuki, M., Fu, P., Yamamoto, S., Fujiwara, S., et al. (2013). Establishment of a standardized gene-expression analysis system using formalin-fixed, paraffin-embedded, breast cancer specimens. *Breast Cancer*, 20(2), 159-166.

[66] Javier, W.R. and Gupta, A.K. (1985). On Generalized Matric Variate Beta Distributions. *Statistics*, 16(4), 549-558.

[67] James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, 35(2), 475-501.

[68] Jana, S. (2013). *The Growth Curve Model for high dimensional data and its application in Genomics.* Retreived from Mac sphere (3552261).

[69] Jana, S., Balakrishnan, N., von Rosen, D., and Hamid, J. S. (2017). High dimensional extension of the growth curve model and its application in genetics. *Statistical Methods and Applications*, 26(2), 273-292.

[70] Jana, S., Balakrishnan, N., and Hamid, J. S. (2017a). Bayesian Growth Curve Model useful for high-dimensional longitudinal data. *Journal of Applied Statistics*. Submitted.

[71] Jana, S., Balakrishnan, N., von Rosen, D., and Hamid, J. S. (2017b). Estimation of the parameters of the Growth Curve Model under Multivariate Skew Normal distribution. *Sankhyā : The Indian Journal of Statistics Series B*. Submitted.

[72] Jana, S., Balakrishnan, N., von Rosen, D., and Hamid, J. S. (2017c). Estimation of the parameters of the Extended Growth Curve Model under Multivariate Skew Normal distribution. *Journal of Multivariate Statistics*. Submitted .

[73] Jonhstone, I. M. and Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transaction of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367, 4237-4253.

[74] Kabe, D. G. (1975). Some results for the GMANOVA model. *Communications in Statistics-Theory and Methods*, 4(9), 813-820.

[75] Karihtala, P., Mäntyniemi, A., Kang, S. W., Kinnula, V. L., and Soini, Y. (2003). Peroxiredoxins in breast carcinoma. *Clinical Cancer Research*, 9(9), 3418-3424.

[76] Kendziorski, C. M., Newton, M. A., Lan, H., and Gould, M. N. (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine*, 22(24), 3899-3914.

[77] Kenward, M. G. (1986). The distribution of a generalized least squares estimator with covariance adjustment. *Journal of Multivariate Analysis*, 20(2), 244-250.

[78] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, 18(1), 75-86.

[79] Khatri, C. G. (1973). Testing some covariance structures under a growth curve model. *Journal of Multivariate Analysis*, 3(1), 102-116.

[80] Kleinbaum, D. G. (1973). A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, 3(1), 117-124.

[81] Kim, D. K., and Taylor, J. M. (1995). The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. *Journal of the American Statistical Association*, 90(430), 708-716.

[82] Koch, K. R. (2007). *Introduction to Bayesian Statistics*. Springer, Berlin, Heidelberg.

[83] Kollo, T., and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, The Netherlands.

[84] Kowalski, C. J., and Guire, K. E. (1974). Longitudinal data analysis. *Growth*, 38(2), 131.

[85] Kropf, S., D. Kose, J. Läuter and D. von Rosen (2009). Comparison of exact parametric tests for high-dimensional data. *Computational Statistics and Data Analysis*, 53, 776-787.

[86] Kshirsagar, A., and Smith, W. (1995). *Growth Curves.* CRC Press, Boca Raton, Florida.

[87] Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B., and Montenegro, L. C. (2007). Likelihood-based inference for multivariate skew-normal regression models. *Communications in Statistics-Theory and Methods*, 36(9), 1769-1786.

[88] Lange, K. (2013). *Optimization*, Springer, New York.

[89] Läuter, J., Horn, F., Rosolowski, M., and Glimm, E. (2009). Highdimensional data analysis: Selection of variables, data compression and graphicsApplication to gene expression. *Biometrical journal*, 51(2), 235-251.

[90] Lee, D. S., Evans, J. C., Robins, S. J., et al. (2007). Gamma Glutamyl Transferase and Metabolic Syndrome, Cardiovascular Disease, and Mortality Risk The Framingham Heart Study. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(1), 127-133.

[91] Lee, J. C., and Geisser, S. (1972). Growth curve prediction. *Sankhya: The Indian Journal of Statistics*, Series A, 34(4), 393-412.

[92] Lee, J. C. S., and Geisser, S. (1975). Applications of growth curve prediction. *Sankhya: The Indian Journal of Statistics*, Series A, 37(2), 239-256.

[93] Leech, F. B., and Healy, M. J. R. (1959). The analysis of experiments on growth rate. *Biometrics*, 15(1), 98-106.

[94] Leeper, J. D., and Chang, L. H. (1987). Hypothesis testing for two-group incomplete-data growth curve problems. *Communications in Statistics-Theory and Methods*, 16(12), 3715-3728.

[95] Leeper, J. D., and Woolson, R. F. (1982). Testing hypotheses for the growth curve model when the data are incomplete. *Journal of Statistical Computation and Simulation*, 15(2-3), 97-107.

[96] Lin, T. I., Ho, H. J., and Chen, C. L. (2009). Analysis of multivariate skew normal models with incomplete data. *Journal of Multivariate Analysis*, 100(10), 2337-2351.

[97] Liski, E. P. (1991). Detecting influential measurements in a growth curves model. *Biometrics*, 47(2), 659-668.

[98] Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *The American Statistician*, 60(3), 213-223.

[99] Little, R. J., and Rubin, D. B. (2014). *Statistical analysis with missing data.* John Wiley & Sons, Inc., Hoboken, New Jersey.

[100] Lönnstedt, I. and Speed, T. (2002). Replicated Microarray data. *Statistica Sinica*, 12, 31-46.

[101] Lopez-Boado, Y.S., Jorge T., and Lŏpez-OtLǐn, C. (1994). Apolipoprotein D gene induction by retinoic acid is concomitant with growth arrest and cell differentiation in human breast cancer cells. *Journal of Biological Chemistry*, 269(43), 26871-26878.

[102] Ma, P., Castillo-Davis, C. I., Zhong, W., and Liu, J. S. (2006). A data-driven clustering method for time course gene expression data. *Nucleic Acids Research*, 34(4), 1261-1269.

[103] Matsubayashi, J., Takanashi, M., Oikawa, K., Xu, M., et al. (2008). Expression of G protein-coupled receptor kinase 4 is associated with breast cancer tumourigenesis. *The Journal of Pathology*, 216(3), 317-327.

[104] McLachlan, G., and Krishnan, T. (2007). *The EM Algorithm and Extensions.* John Wiley & Sons, Hoboken, New Jersey.

[105] Mendoza-Blanco, J.R., Tu, X.M. and Gleser, L.J. (1995). On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions, Technical Report, Department of Statistics, University of Pittsburgh.

[106] Meng, X. L. (1997). The EM algorithm and medical studies: A historical link. *Statistical Methods in Medical Research*, 6(1), 3-23.

[107] Meng, X. L., and Pedlow, S. (1992). EM: a bibliographic review with missing articles. In *Proceedings of the Statistical Computing Section, American Statistical Association*, 24-27.

[108] Noh, D. Y., Ahn, S. J., Lee, R. A., et al. (2000). Overexpression of peroxiredoxin in human breast cancer. *Anticancer Research*, 21(3B), 2085-2090.

[109] Nonaka, D., Chiriboga, L., Rubin, B.P. (2008). Sox10: a pan-schwannian and melanocytic marker. *American Journal of Surgical Pathology*, 32(9), 1291-1298.

[110] Nzabanita, J., Singull, M., and von Rosen, D. (2012). Estimation of parameters in the extended growth curve model with a linearly structured covariance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 16(1), 13-32.

[111] Ostrander, J.H., Daniel, A. R., Lofgren, K., Kleer, C. G., et al. (2007). Breast tumor kinase (protein tyrosine kinase 6) regulates heregulin-induced activation of ERK5 and p38 MAP kinases in breast cancer cells. *Cancer Research*, 67(9), 4199-4209.

[112] Pan, J. X., and Fang, K. T. (1996). Influential observation in the growth curve model with unstructured covariance matrix. *Computational Statistics and Data Analysis*, 22(1), 71-87.

[113] Pan, J. X., and Fang, K. T. (2002). *Growth Curve Models and Statistical Diagnostics, Springer*. New York.

[114] Potthoff, R. F., and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51(3/4), 313-326.

[115] Raiffa, H., and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Havard University & MIT Press, Boston.

[116] Rao, C. R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankhya: The Indian Journal of Statistics* , 18(1/2), 139-148.

[117] Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1), 1-17.

[118] Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, 46(1/2), 49-58.

[119] Rao, C. R. (1961). Some observations on multivariate statistical methods in

anthropological research. *Bulletin of the International Statistical Institute*, 37(4), 99-109.

[120] Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52(3/4), 447-458.

[121] Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. *Multivariate Analysis*, 1, 87-103.

[122] Rao, C. R. (1977). Prediction of future observations with special reference to linear models. *Multivariate Analysis*, 4, 193-208.

[123] Rao, C. R. (1984). Prediction of future observations in polynomial growth curve models. *In Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions*, 512-520.

[124] Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science*, 2(4), 434-447.

[125] Rauf A. M., von Rosen, D., and Singull, M. (2012). A note on mean testing for high dimensional multivariate data under nonnormality. *Statistica Neerlandica*, 67(1), 81-99.

[126] von Rosen, D. (1984). Maximum likelihood estimates in Multivariate Linear Normal Models with special references to the Growth Curve Model. Research report 135, Department of Mathematics and Statistics, University of Stockholm, Stockholm, Sweden.

[127] von Rosen, D. (1985). Multivariate linear normal models with special references to the growth curve model. Stockholm: Von Rosen. `http://library.wur.nl/WebQuery/clc/235153`

[128] von Rosen, D. (1989). Maximum likelihood estimators in multivariate linear normal models. *Journal of Multivariate Analysis*, 31(2), 187-200.

[129] von Rosen, D. (1991). The growth curve model: a review. *Communications in Statistics-Theory and Methods*, 20(9), 2791-2822.

[130] von Rosen, D. (1995). Residuals in the growth curve model. *Annals of the Institute of Statistical Mathematics*, 47(1), 129-136.

[131] von Rosen, D. (1995a). Influential observations in multivariate linear models. *Scandinavian Journal of Statistics*, 22(2), 207-222.

[132] Roslin, N. M., Hamid, J. S., Paterson, A. D., and Beyene, J. (2009). Genome-wide association analysis of cardiovascular-related quantitative traits in the Framingham Heart Study. *BMC Proceedings*, 3(7), 1.

[133] Roy, S. N., and Roy, S. M. (1957). *Some aspects of multivariate analysis*. Wiley, New York.

[134] Samaniego, F. J. (2010). *A comparison of the Bayesian and frequentist approaches to estimation*. Springer, New York.

[135] Savage, L.J. (1954). *The Foundations of statistics*. John Wiley & Sons Inc., New York.

[136] Savage, L. J. (1972). *The Foundations of Statistics* (revised edition). Dover Publications, New York.

[137] Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), Article 32.

[138] Siegel, C.L. (1935). *Ü*ber die analytische Theorie der quadratischen Formen. *Annals of Mathematics*, 36, 527-606.

[139] Siotani, T., Fujikoshi, Y., and Hayakawa, T. (1985). *Modern multivariate statistical analysis, a graduate course and handbook*. American Sciences Press, New York.

[140] Sivazlian, B.D. (1981). On a Multivariate Extension of the Gamma and Beta Distributions. *SIAM Journal on Applied Mathematics*, 41(2), 205-209.

[141] Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1).

[142] Splansky, G. L., Corey, D., Yang, Q., et al. (2007). The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *American Journal of Epidemiology*, 165(11), 1328-1335.

[143] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*. John Wiley and Sons Inc., New York.

[144] Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 37(1), 53-86.

[145] Srivastava, M. S. and Carter, E.M. (1983). *An Introduction to Applied Multivariate Statistics.* Elseiver North Holland Inc., New York.

[146] Srivastava, M. S. and Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, 115, 204-216.

[147] Srivastava, M. S., and von Rosen, D. (1999). *Growth curve models. Statistics Textbooks and Monographs.* Marcel Dekker Inc., New York.

[148] Srivastava, M. S., and von Rosen, D. (1999). Growth curve models. In S. Ghosh (Ed.), *Multivariate Analysis, Design of Experiments, and Survey Sampling, Statistics: Textbooks and Monographs* 159, 547-578. Dekker, New York.

[149] Srivastava, M. S., and Khatri, C. G. (1979). *An Introduction to Multivariate Analysis.* Elseiver North Holland Inc., New York.

[150] Srivastava, M. S., and Singull, M. (2012). Profile analysis with random-effects covariance structure. *Journal of the Japan Statistical Society*, 42(2), 145-164.

[151] Srivastava, M. S., and Singull, M. (2017). Test for the mean matrix in a Growth Curve model for high dimensions. *Communications in Statistics-Theory and Methods*, 46(13), 6668-6683.

[152] Srivastava, M. S., and Singull, M. (2017a) Testing sphericity and intraclass covariance structures under a growth curve model in high dimension. *Communications in Statistics-Simulation and Computation* just-accepted.

[153] Stanek III, E. J., and Koch, G. G. (1985). The equivalence of parameter estimates from growth curve models and seemingly unrelated regression models. *The American Statistician*, 39(2), 149-152.

[154] Stram, D. O., and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50(4), 1171-1177.

[155] Tai, Y.C. and Speed, T.P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *The Annals of Statistics*, 34(5), 2387-2412.

[156] Tai, Y.C. and Speed, T.P. (2009). On gene ranking using replicated microarray time course data. *Biometrics*, 65(1), 40-51.

[157] Tan, W. Y. (1969). Note on the multivariate and the generalized multivariate beta distributions. *Journal of the American Statistical Association*, 64(325), pp 230-241.

[158] Vargas, A. C., Reed, A. E. M., Waddell, N., Lane, A., et al. (2012). Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast cancer research and treatment*, 135(1), 153-165.

[159] Verbyla, A. P., and Venables, W. N. (1988). An extension of the growth curve model. *Biometrika*, 75(1), 129-138.

[160] Wang, Y., He, Q. Y., Chen, H., and Chiu, J. F. (2007). Synergistic effects of retinoic acid and tamoxifen on human breast cancer cells: proteomic characterization. *Experimental cell Research*, 313(2), 357-368.

[161] Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, 39(2), 95-101.

[162] Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika*, 30(1/2), 16-28.

[163] Wissmann, C., Wild, P. J., Kaiser, S., Roepcke, S., et al. (2003). WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. *The Journal of Pathology*, 201(2), 204-212.

[164] Woolson, R. F., and Leeper, J. D. (1980). Growth curve analysis of complete and incomplete longitudinal data. *Communications in Statistics-Theory and Methods*, 9(14), 1491-1513.

[165] Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1) 95-103.

[166] Wu, W., Zou, M., Brickley, D. R., Pew, T., and Conzen, S. D. (2006). Glucocorticoid Receptor Activation Signals through Forkhead Transcription Factor 3a in Breast Cancer Cells. *Molecular Endocrinology*, 20(10), 23042314.

[167] Yamaguchi, K., and Watanabe, M. (2004). *Statistical Models and the EM Algorithm.* The EM Algorithm and Related Statistical Models, New York, Basel, 21-36.

[168] Yokoyama, T., and Fujikoshi, Y. (1993). A parallel profile model with random-effects covariance structure. *Journal of the Japan Statistical Society, Japanese Issue*, 23(1), 83-89.

[169] Yuan, M. and Kendziorksi, C. (2006). Hidden Markov Models for Microarray Time Course Data in Multiple Biological Conditions. *Journal of the American Statistical Association*, 101(476), 1323-1332.

[170] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298), 348-368.

[171] Zellner, A. (1963). Estimators for seemingly unrelated regression equations: Some exact finite sample results. *Journal of the American Statistical Association*, 58(304), 977-992.

[172] Zhang, H., and Ding, F. (2013). On the Kronecker products and their applications. *Journal of Applied Mathematics*, 2013, Article ID 296185, doi:10.1155/2013/296185

Appendix A

Your Appendix

A.1 List of B matrices used in the simulation studies

$$\begin{pmatrix} -5 & -2 \\ -5 & -2 \end{pmatrix}, \begin{pmatrix} -4 & -2 \\ -4 & -2 \end{pmatrix}, \begin{pmatrix} -3 & -2 \\ -3 & -2 \end{pmatrix}, \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ -1 & -2 \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ -1 & -1 \end{pmatrix}$$

$$, \begin{pmatrix} -1 & -1.5 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -1 & -0.8 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -1 & -0.5 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -0.8 & -0.5 \end{pmatrix}$$

$$, \begin{pmatrix} -1 & -1 \\ -0.5 & -0.5 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -0.5 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -0.9 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -0.8 & -1 \\ 0 & 0 \end{pmatrix}$$

$$, \begin{pmatrix} -0.5 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -0.2 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.9 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.8 \\ 0 & 0 \end{pmatrix}$$

$$, \begin{pmatrix} 0 & -0.7 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.6 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.5 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.4 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.3 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & -0.2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -0.1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.3 \\ 0 & 0 \end{pmatrix}$$

$$, \begin{pmatrix} 0 & 0.4 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.5 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.6 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.7 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.8 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.9 \\ 0 & 0 \end{pmatrix}$$

$$, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0.2 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0.8 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0.9 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 \\ 0 & 0.5 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0.5 & 0.5 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0.5 & 0.8 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0.5 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1.5 \\ 1 & 1 \end{pmatrix}$$

$$, \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 4 & 2 \end{pmatrix}, \begin{pmatrix} 5 & 2 \\ 5 & 2 \end{pmatrix}.$$

A.2 The variance-covariance matrices used in the simulation studies

Weak correlation

$$\mathbf{\Sigma} = \begin{pmatrix} 1.2705443 & 0.2766856 & 0.1590603 & 0.0740178 \\ 0.2766856 & 1.0273088 & 0.8048891 & 0.5296352 \\ 0.1590603 & 0.8048891 & 1.5091252 & 0.9823369 \\ 0.0740178 & 0.5296352 & 0.9823369 & 1.5663106 \end{pmatrix}$$

Moderate correlation

$$\mathbf{\Sigma} = \begin{pmatrix} 2.7231989 & 0.9051908 & 0.3394739 & -0.1633491 \\ 0.9051908 & 2.3579832 & 1.6924352 & 0.9617295 \\ 0.3394739 & 1.6924352 & 3.1554558 & 1.9769148 \\ -0.1633491 & 0.9617295 & 1.9769148 & 3.2830360 \end{pmatrix}$$

Strong correlation

$$\mathbf{\Sigma} = \begin{pmatrix} 4.1828919 & 1.589901 & 0.5328818 & -0.4378997 \\ 1.5899015 & 3.724326 & 2.5703977 & 1.3435312 \\ 0.5328818 & 2.570398 & 4.7582759 & 2.9381095 \\ -0.4378997 & 1.343531 & 2.9381095 & 4.9360192 \end{pmatrix}$$