Linköping Studies in Science and Technology. Dissertations.
No. 1665

Bilinear and Trilinear Regression Models with Structured Covariance Matrices

Joseph Nzabanita

Linköping University
INSTITUTE OF TECHNOLOGY

Department of Mathematics
Linköping University, SE–581 83 Linköping, Sweden

Linköping 2015

Linköping Studies in Science and Technology. Dissertations.
No. 1665

Bilinear and Trilinear Regression Models with Structured Covariance Matrices

Joseph Nzabanita

joseph.nzabanita@liu.se
www.mai.liu.se
Mathematical Statistics
Department of Mathematics
Linköping University
SE–581 83 Linköping
Sweden

ISBN 978-91-7519-070-9 ISSN 0345-7524

Printed by LiU-Tryck, Linköping, Sweden 2015

To my Family and Friends

Abstract

Joseph Nzabanita (2015). Bilinear and Trilinear Regression Models with Structured Covariance Matrices

Doctoral dissertation. ISBN 978-91-7519-070-9 . ISSN 0345-7524.

This thesis focuses on the problem of estimating parameters in bilinear and trilinear regression models in which random errors are normally distributed. In these models the covariance matrix has a Kronecker product structure and some factor matrices may be linearly structured. Most of techniques in statistical modeling rely on the assumption that data were generated from the normal distribution. Whereas real data may not be exactly normal, the normal distributions serve as a useful approximation to the true distribution. The modeling of normally distributed data relies heavily on the estimation of the mean and the covariance matrix. The interest of considering various structures for the covariance matrices in different statistical models is partly driven by the idea that altering the covariance structure of a parametric model alters the variances of the model's estimated mean parameters.

Firstly, we consider the extended growth curve model with a linearly structured covariance matrix. In general there is no problem to estimate the covariance matrix when it is completely unknown. However, problems arise when one has to take into account that there exists a structure generated by a few number of parameters. An estimation procedure that handles linear structured covariance matrices is proposed. The idea is first to estimate the covariance matrix when it may be used to define an inner product in a regression space and thereafter re-estimate it when it should be interpreted as a dispersion matrix. This idea is exploited by decomposing the residual space, the orthogonal complement to the design space, into orthogonal subspaces. Studying residuals obtained from projections of observations on these subspaces yields explicit consistent estimators of the covariance matrix. An explicit consistent estimator of the mean is also proposed.

Secondly, we study a bilinear regression model with matrix normally distributed random errors. For those models, the dispersion matrix follows a Kronecker product structure and it can be used, for example, to model data with spatio-temporal relationships. The aim is to estimate the parameters of the model when, in addition, one covariance matrix is assumed to be linearly structured. On the basis of n independent observations from a matrix normal distribution, estimating equations, a flip-flop relation, are established.

At last, the models based on normally distributed random third order tensors are studied. These models are useful in analyzing 3-dimensional data arrays. The 3-dimensional data arrays may be obtained when, for example, a single response is sampled in a 3-D space or in a 2-D space and time, multiple responses are recorded in a 2-D space or in a 1-D space and time. In some studies the analysis is done using the tensor normal model, where the focus is on the estimation of the variance-covariance matrix which has a Kronecker structure. Little attention is paid to the structure of the mean, however, there is a potential to improve the analysis by assuming a structured mean. We formally introduce a 2-fold growth curve model by assuming a trilinear structure for the mean in the tensor normal model and propose an estimation algorithm for parameters. Also some extensions are discussed.

Populärvetenskaplig sammanfattning

Många statistiska modeller bygger på antagandet om normalfördelad data. Verklig data kanske inte är exakt normalfördelad men det är i många fall en bra approximation. Normalfördelad data kan modelleras enbart genom dess väntevärde och kovariansmatris och det är därför ett problem av stort intresse att skatta dessa parametrar. Ofta kan det också vara intressant eller nödvändigt att anta någon struktur på både väntevärdet och/eller kovariansmatrisen.

Den här avhandlingen fokuserar på problemet att skatta parametrarna i multivariata linjära modeller, speciellt den utökade tillväxtkurvemodellen med en linjär struktur för någon kovariansmatris. I allmänhet är det inget problem att skatta kovariansmatriserna när de är helt okända. Problem uppstår emellertid när man måste ta hänsyn till att det finns en struktur som genereras av ett färre antal parametrar. I många exempel kan maximum-likelihoodskattningar inte erhållas explicit och måste därför beräknas med någon numerisk optimeringsalgoritm. Vi beräknar explicita skattningar som ett bra alternativ till maximum-likelihoodskattningarna. En skattningsprocedur som skattar kovariansmatriser med linjära strukturer föreslås. Tanken är att först skatta en kovariansmatris som används för att definiera en inre produkt, för att sedan skatta den slutliga kovariansmatrisen.

Även tillväxtkurvemodeller med tensornormalfördelning studeras i den här avhandlingen. För dessa modeller är kovariansmatrisen en Kroneckerprodukt och dessa modeller kan användas exempelvis för att modellera data med spatio-temporala förhållande. Syftet är att skatta parametrarna i modellen där möjligen även en av kovariansmatriserna antas följa en linjär struktur.

Acknowledgments

My deepest gratitude goes first to my supervisors Professor Dietrich von Rosen and Dr Martin Singull for their invaluable guidance, encouragement and insightful ideas on scientific matters without which this thesis would not be completed. Thank you Dietrich, thank you Martin. The patience, the enthusiasm and many more humane qualities you constantly show helped me a lot during my studies.

My deep gratitude goes to Bengt Ove Turesson, to Björn Textorius, to Magnus Herberthson, to Fredrik Berntsson, to Göran Forsling, to Tomas Sjödin, to Theresia Carlsson Roth, to Theresa Lagali and to all members of the Department of Mathematics, Linköping University, for their constant help in different things.

I am thankful to the staff of the University of Rwanda, especially Froduald Minani, Alexandre Lyambabaje, Raymond Ndikumana, Alex Karara, Colin Karuhanga and Charles Gakomeye for their invaluable contribution to the smooth running of my studies.

I have also to thank my fellow PhD students at the Department of Mathematics for making life easier during my studies.

Last but not least, I owe my special thanks to my family and friends for moral support and distraction.

My studies have been supported by the UR-Sweden Programme for Research, Higher Education and Institutional Advancement and all involved institutions and people are acknowledged.

Linköping, May 6, 2015
Joseph Nzabanita

Contents

1 Introduction **1**
1.1 Background 1
1.2 Aims 2
1.3 Outline 2
1.3.1 Outline of Part I 2
1.3.2 Outline of Part II 3

I Bilinear and Trilinear Regression Models **5**

2 Multivariate Distributions **7**
2.1 Normal distributions 7
2.2 Wishart distribution 12

3 Regression Models **15**
3.1 General linear regression model 15
3.2 Bilinear regression models. Growth curve models 16
3.3 Trilinear regression model 19
3.4 Estimation in bilinear regression models 21
3.4.1 Maximum likelihood estimators 21
3.4.2 Explicit estimators when the covariance matrix is linearly structured 24
3.5 Estimation in trilinear regression model 28

4 Concluding Remarks **31**
4.1 Summary of contributions 31
4.2 Further research 32

Bibliography **33**

II Papers 37

A Paper A **39**
1 Introduction . . . 42
2 Maximum likelihood estimators . . . 43
3 Estimators in the extended growth curve model with a linearly structured covariance matrix . . . 46
4 Properties of the proposed estimators . . . 51
5 Numerical examples . . . 54
References . . . 58

B Paper B **61**
1 Introduction . . . 64
2 Main idea and space decomposition . . . 65
3 Estimators when the covariance matrix is linearly structured . . . 69
4 Simulation study . . . 75
4.1 Procedures and methods . . . 75
4.2 Results . . . 77
References . . . 77

C Paper C **79**
1 Introduction . . . 82
2 Explicit estimators of linearly structured $\mathbf{\Sigma}$ with unknown parameters and known $\mathbf{\Psi}$ . . . 82
3 Estimators of linearly structured $\mathbf{\Sigma}$ with unknown parameters and unknown $\mathbf{\Psi}$ . . . 84
4 Simulated numerical example . . . 91
References . . . 91

D Paper D **93**
1 Introduction . . . 96
2 The tensor normal distribution with a structured mean . . . 96
3 Maximum likelihood estimation . . . 99
3.1 Estimation of parameters in the 2-fold Growth Curve Model . . . 99
3.2 Estimation of parameters in the 2-fold Growth Curve Model when $\mathbf{\Sigma}$ has a uniform structure . . . 102
3.3 Estimation of $\boldsymbol{B}$, $\mathbf{\Sigma}$, $\mathbf{\Psi}$ and $\mathbf{\Omega}$: General case . . . 103
4 Simulation study . . . 105
4.1 Procedures . . . 105
4.2 Results . . . 106
References . . . 107

1

Introduction

THE goals of statistical sciences are about planning experiments, setting up models to analyze experiments and to study properties of these models. Statistical application is about connecting statistical models to data. Statistical models are essentially for making predictions; they form the bridge between observed data and unobserved (future) outcomes (Kattan and Gönen, 2008). The general statistical paradigm constitutes of the following steps: (i) set up a model, (ii) evaluate the model via simulations or comparisons with data, (iii) if necessary refine the model and restart from step (ii), and (iv) accept and interpret the model. From this paradigm it is clear that the concept of statistical model lies in the heart of Statistics. In this thesis our focus is on linear models, a class of statistical models that play a key role in statistics. If exact inference is not possible then at least a linear approximate approach can often be carried out (Kollo and von Rosen, 2005). In particular, we are concerned with the problem of estimating parameters in multivariate linear normal models with structured mean (bilinear and trilinear regression models) and structured covariance matrices (Kronecker and linear structures).

1.1 Background

Regression analysis includes several statistical techniques for investigating dependencies among variables. It is used essentially when the focus is to understand the relationships between a set of dependent variables and a set of independent variables. The regression analysis appeared in earliest form as the method of least squares in the beginning of the nineteenth-century, where Legendre and Gauss applied the method to the problem of determining orbits of comets and planets about the sun from astronomical observations. The term "regression" was introduced in late nineteenth-century by Francis Galton while he was studying the inheritance problem (Allen, 1997).

Although regression methods are in use since the last two centuries, there are still interesting problems that makes regression analysis to be an area of active research nowa-

days. The newest research directions include regression involving correlated responses such as time series and growth curves, and the focus of this thesis can be traced there.

In regression models often one assumes that the underlying random errors follow a Gaussian distribution. When the set up of the model is matrix or tensor, it becomes natural to model the covariance matrix with a Kronecker product structure. For some other structures for the covariance matrix, there might be theoretical ground to justify a particular choice of the covariance structure (Fitzmaurice et al., 2012). In particular, the linear structures for the covariance matrices emerged naturally in statistical applications and they are in the statistical literature for some years ago. These structures are, for example, uniform structure, also known as intraclass structure, compound symmetry structure, banded matrix, Toeplitz or circular Toeplitz, etc. The uniform structure, a linear covariance structure which consists of equal diagonal elements and equal off-diagonal elements, emerged for the first time in (Wilks, 1946) while dealing with measurements on k psychological tests. An extension of the uniform structure due to Votaw (1948) is the compound symmetry structure, which consists of blocks each having uniform structure. In (Votaw, 1948) one can find examples of psychometric and medical research problems where the compound symmetry covariance structure is applicable. The block compound symmetry covariance structure was discussed by Szatrowski (1982) who applied the model to the analysis of an educational testing problem. Ohlson et al. (2011) proposed a procedure to obtain explicit estimator of a banded covariance matrix. The Toeplitz or circular Toeplitz discussed in (Olkin and Press, 1969) is another generalizations of the intraclass structure. The interest of considering various structures for the covariance matrices in different statistical models is partly driven by the idea that altering the covariance structure of a parametric model alters the variances of the model's estimated mean parameters (Lange and Laird, 1989).

1.2 Aims

The main theme of this thesis is to study the problem of estimation of parameters in the bilinear and trilinear regression models with structured covariance matrices. Specific objectives are (i) to derive explicit estimators of parameters in the extended growth curve model when the covariance matrix is linearly structured, (ii) to propose an algorithm for estimating parameters in the bilinear regression model where the random errors are assumed to be matrix normally distributed with one linearly structured covariance matrix, and (iii) to extend the classical growth curve model by Pothoff and Roy (1964) to a tensor version and to propose an algorithm for estimating model parameters.

1.3 Outline

This thesis consists of two parts and the outline is as follows.

1.3.1 Outline of Part I

In Part I the background and relevant results that are needed for an easy reading of this thesis are presented. Part I starts with Chapter 2 which gives a brief review on the multivariate distributions. The main focus is to define the matrix normal distribution, the tensor

normal distribution and the Wishart distribution. The maximum likelihood estimators in the multivariate normal model, the matrix normal model and the tensor normal model, for the unstructured cases, are given. In Chapter 3 the bilinear and trilinear regression models are defined. These models include the growth curve model and the extended growth curve model, which are refereed to as bilinear regression models, and the third order tensor normal model with a structured mean refereed to as the trilinear regression model. Some results on the estimation of parameters in the extended growth curve model are given for unstructured covariance matrix and the procedure to get explicit estimators when the covariance matrix is linearly structured is illustrated. Also, an algorithm for estimating parameters in the trilinear regression model is given. Part I ends with Chapter 4, which gives a summary of contributions and suggestions for further work.

1.3.2 Outline of Part II

Part II consists of four papers. Hereafter a short summary for the papers is presented.

Paper A: Estimation of parameters in the extended growth curve model with a linearly structured covariance matrix

> 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.

In Paper A, the extended growth curve model with two terms and a linearly structured covariance matrix is considered. We propose an estimation procedure that handles linear structured covariance matrices. The idea is first to estimate the covariance matrix when it should be used to define an inner product in a regression space and thereafter re-estimate it when it should be interpreted as a dispersion matrix. This idea is exploited by decomposing the residual space, the orthogonal complement to the design space, into three orthogonal subspaces. Studying residuals obtained from projections of observations on these subspaces yields explicit consistent estimators of the covariance matrix. An explicit consistent estimator of the mean is also proposed and numerical examples are given.

Paper B: Extended GMANOVA model with a linearly structured covariance matrix

> Nzabanita, J., von Rosen, D., and Singull, M. (2015a). Extended GMANOVA model with a linearly structured covariance matrix. *Linköping University Electronic Press, LiTH-MAT-R-2015/07-SE.*

Paper B generalizes results in Paper A to the extended GMANOVA model with an arbitrary number of profiles, say m. We show how to decompose the residual space, the orthogonal complement to the mean space, into $m + 1$ orthogonal subspaces and how to derive explicit consistent estimators of the covariance matrix and an explicit unbiased estimator of the mean.

Paper C: Bilinear regression model with Kronecker and linear structures for the covariance matrix

> Nzabanita, J. (2013). Multivariate linear models with kronecker product and linear structures on the covariance matrices. In *Proceedings, JSM 2013-IMS*, pp. 1582–1588. Alexandria, VA: American Statistical Association.(*A preliminary version*).

This paper deals with models based on normally distributed random matrices. More specifically the model considered is $\boldsymbol{X} \sim N_{p,q}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ with mean $\boldsymbol{M}$, a $p \times q$ matrix, assumed to follow a bilinear structure, i.e., $E[\boldsymbol{X}] = \boldsymbol{M} = \boldsymbol{ABC}$, where $\boldsymbol{A}$ and $\boldsymbol{C}$ are known design matrices, $\boldsymbol{B}$ is unkown parameter matrix, and the dispersion matrix of $\boldsymbol{X}$ has a Kronecker product structure, i.e., $D[\boldsymbol{X}] = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$, where both $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are unknown positive definite matrices. The model may be used for example to model data with spatio-temporal relationships. The aim is to estimate the parameters of the model when, in addition, $\boldsymbol{\Sigma}$ is assumed to be linearly structured. In the paper, on the basis of n independent observations on the random matrix $\boldsymbol{X}$, estimation equations in a flip-flop relation are presented and the consistency of estimators is studied.

Paper D: Maximum likelihood estimation in the tensor normal model with a structured mean

> Nzabanita, J., von Rosen, D., and Singull, M. (2015b). Maximum likelihood estimation in the tensor normal model with a structured mean. *Linköping University Electronic Press, LiTH-MAT-R-2015/08-SE.*

In this paper, we introduce a 2-fold growth curve model by assuming a trilinear structure for the mean in the tensor normal model. More specifically, the model considered may be written as

$$\mathscr{X} = \mathscr{B} \times \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}\} + \mathscr{E},$$

where $\mathscr{X} : p \times q \times r$ is the data tensor, $\mathscr{B} : s \times t \times u$ is the parameter given as a tensor of order three, $\boldsymbol{A} : p \times s$, $\boldsymbol{C} : q \times t$ and $\boldsymbol{D} : r \times u$ are known design matrices, and $\times$ denotes the Tucker product. The random errors follow a tensor normal distribution with mean zero, i.e., $\mathscr{E} \sim N_{p,q,r}(\mathscr{O}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{I}_r)$, and $\mathscr{O}$ is a tensor of zeros. An algorithm for estimating parameters is proposed and some direct generalizations of the model are presented.

Part I

Bilinear and Trilinear Regression Models

2

Multivariate Distributions

THIS chapter focuses on the normal distribution which is very important in statistical analysis. In particular, our interest here is to define the matrix and tensor normal distributions which will play a central role in this thesis. The Wishart distribution will also be looked at for easy reading of papers.

2.1 Normal distributions

The well known univariate normal distribution has been used in statistics for about two hundreds years and the multivariate normal distribution, understood as a distribution of a vector, has been also used for a long time (Kollo and von Rosen, 2005). Due to the complexity of data from various field of applied research, inevitable extensions of the multivariate normal distribution to the matrix normal distribution or even more generalization to multilinear (tensor) normal distribution have been considered. The normal distributions we present here exclude the degenerate cases and thus their density functions exist.

Definition 2.1 (Univariate normal distribution). A random variable x is a univariate normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, denoted as $x \sim N\left(\mu, \sigma^2\right)$ if its density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (-\infty < x < \infty).$$

In particular when $\mu = 0$ and $\sigma = 1$, we get the standard univariate normal distribution, i.e., $u \sim N(0, 1)$. A more general characterization of the univariate normal distribution is

$$x \overset{d}{=} \mu + \sigma u,\ \mu \in \mathbb{R},\ \sigma \geq 0,$$

where $u \sim N(0, 1)$ and the notation "$\overset{d}{=}$" means "has the same distribution as".

Definition 2.2 (Multivariate normal distribution). A random vector $\boldsymbol{x} : p \times 1$ is multivariate normally distributed with mean vector $\boldsymbol{\mu} : p \times 1$ and positive definite covariance matrix $\boldsymbol{\Sigma} : p \times p$ if its density is

$$f(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'\}},$$

where $|\cdot|$ and tr denote the determinant and the trace of a matrix, respectively. We usually use the notation $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The multivariate normal model $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown parameters, is used in the statistical literature for a long time. To find estimators of the parameters, the method of maximum likelihood is often used. Let a random sample of n observation vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ come from the multivariate normal distribution, i.e., $\boldsymbol{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The $\boldsymbol{x}_i$'s constitute a random sample and the likelihood function is given by the product of the densities evaluated at each observation vector

$$\begin{aligned} L(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} f(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\}}. \end{aligned}$$

The maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ resulting from the maximization of this likelihood function, for more details see for example Johnson and Wichern (2007), are respectively

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \frac{1}{n}\boldsymbol{X}\mathbf{1}_n, \\ \widehat{\boldsymbol{\Sigma}} &= \frac{1}{n}\boldsymbol{S}, \end{aligned}$$

where

$$\boldsymbol{S} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})' = \boldsymbol{X}(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\boldsymbol{X}',$$

$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, $\mathbf{1}_n$ is the $n-$dimensional vector of 1's, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

Definition 2.3 (Matrix normal distribution). A random matrix $\boldsymbol{X} : p \times q$ is matrix normally distributed with mean $\boldsymbol{M} : p \times q$ and positive definite covariance matrices $\boldsymbol{\Sigma} : p \times p$ and $\boldsymbol{\Psi} : q \times q$ if its density is

$$f(\boldsymbol{X}) = (2\pi)^{-\frac{pq}{2}} |\boldsymbol{\Sigma}|^{-\frac{q}{2}} |\boldsymbol{\Psi}|^{-\frac{p}{2}} e^{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})'\}}.$$

The model based on the matrix normally distributed is usually denoted as

$$\boldsymbol{X} \sim N_{p,q}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}), \tag{2.1}$$

and it can be shown that $\boldsymbol{X} \sim N_{p,q}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ means the same as

$$\mathrm{vec}\boldsymbol{X} \sim N_{pq}(\mathrm{vec}\boldsymbol{M}, \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}), \tag{2.2}$$

where $\otimes$ denotes the Kronecker product. Since by definition of the dispersion matrix of $\boldsymbol{X}$ is $D[\boldsymbol{X}] = D[\mathrm{vec}\boldsymbol{X}]$, we get $D[\boldsymbol{X}] = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$. For the interpretation we note that $\boldsymbol{\Psi}$ describes the covariances between the columns of $\boldsymbol{X}$. These covariances will be the same for each row of $\boldsymbol{X}$. The other covariance matrix $\boldsymbol{\Sigma}$ describes the covariances between the rows of $\boldsymbol{X}$ which will be the same for each column of $\boldsymbol{X}$. The product $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ takes into account the covariances between columns as well as the covariances between rows. Therefore, $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ indicates that the overall covariance consists of the products of the covariances in $\boldsymbol{\Psi}$ and in $\boldsymbol{\Sigma}$, respectively, i.e., $\mathrm{Cov}[x_{ij}, x_{kl}] = \sigma_{ik}\psi_{jl}$, where $\boldsymbol{X} = (x_{ij})$, $\boldsymbol{\Sigma} = (\sigma_{ik})$ and $\boldsymbol{\Psi} = (\psi_{jl})$.

The following example shows one possibility of how a matrix normal distribution may arise.

Example 2.1

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be an independent sample of n observation vectors from a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let the observation vectors $\boldsymbol{x}_i$ be the columns in a matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$. The distribution of the vectorization of the sample observation matrix $\mathrm{vec}\boldsymbol{X}$ is given by

$$\mathrm{vec}\boldsymbol{X} = (\boldsymbol{x}_1', \ \boldsymbol{x}_2', \ldots, \boldsymbol{x}_n')' \sim N_{pn}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, $\mathbf{1}_n$ is the $n-$dimensional vector of 1s, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. This is written as

$$\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I}_n),$$

where $\boldsymbol{M} = \boldsymbol{\mu}\mathbf{1}_n'$.

The models (2.1) and (2.2) have been considered in the statistical literature. For example Dutilleul (1999), Roy and Khattree (2005) and Lu and Zimmerman (2005) considered the model (2.2), and to obtain MLEs these authors solved iteratively the usual likelihood equations, one obtained by assuming that $\boldsymbol{\Psi}$ is given and the other obtained by assuming that $\boldsymbol{\Sigma}$ is given, by what was called the flip-flop algorithm in Lu and Zimmerman (2005).

Let a random sample of n observation matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be drawn from the matrix normal distribution, i.e., $\boldsymbol{X}_i \sim N_p(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$. The likelihood function is given by the product of the densities evaluated at each observation matrix as it was for the multivariate case. The log-likelihood, ignoring the normalizing factor, is given by

$$\begin{aligned} \ln L(\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) &= -\frac{qn}{2}\ln|\boldsymbol{\Sigma}| - \frac{pn}{2}\ln|\boldsymbol{\Psi}| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \boldsymbol{M})\boldsymbol{\Psi}^{-1}(\boldsymbol{X}_i - \boldsymbol{M})'\}. \end{aligned}$$

The likelihood equations are given by Dutilleul (1999)

$$\begin{aligned}
\widehat{\boldsymbol{M}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i = \overline{\boldsymbol{X}};\\
\widehat{\boldsymbol{\Sigma}} &= \frac{1}{nq}\sum_{i=1}^{n}(\boldsymbol{X}_i - \widehat{\boldsymbol{M}})\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{X}_i - \widehat{\boldsymbol{M}})';\\
\widehat{\boldsymbol{\Psi}} &= \frac{1}{np}\sum_{i=1}^{n}(\boldsymbol{X}_i - \widehat{\boldsymbol{M}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_i - \widehat{\boldsymbol{M}}).
\end{aligned}$$

There is no explicit solutions to these equations and one must rely on an iterative algorithm like the flip-flop algorithm (Dutilleul, 1999). Srivastava et al. (2008) pointed out that the estimators found in this way are not uniquely determined. Srivastava et al. (2008) showed that solving these equations with additional estimability conditions, using the flip-flop algorithm, the estimates in the algorithm converge to the unique maximum likelihood estimators of the parameters.

The model (2.1), where the mean has a bilinear structure was considered by Srivastava et al. (2008). Nzabanita (2013) considered the problem of estimating the parameters in the model (2.1) where the mean has a bilinear structure and, in addition, the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be linearly structured.

A matrix normal model may be thought as a two-array normal model and can be extended to a K-array normal model (also known as tensor normal model). Before we give a formal definition of a K-array normal model, we first introduce few notations and operations on K-arrays to be used later. A K-array or K-way or Kth-order tensor is an element of the tensor product of K vector spaces, each of which has its own coordinate system (Hoff, 2011, Kolda and Bader, 2009, De Lathauwer et al., 2000). For example, a vector $\boldsymbol{x} \in \mathbb{R}^{p_1}$ is a one-array with dimension p_1. A matrix $\boldsymbol{X} \in \mathbb{R}^{p_1\times p_2}$ is a two-array with dimension (p_1, p_2). An array $\mathscr{X} \in \mathbb{R}^{p_1\times p_2\times\cdots\times p_K}$ is a K-array with dimension $(p_1, \ldots, p_K)$ and has elements $\{x_{i_1,\ldots,i_K} : i_k \in \{1, \ldots, p_k\}, k = 1, \ldots, K\}$. A matricization or unfolding or flattening of a tensor is the process of reordering its elements into a matrix. This can be done in several ways. In this thesis we use the so called mode-n matricization and the notation $\boldsymbol{X}_{(n)}$ is used to denote n-mode matrix from the tensor $\mathscr{X}$. For some details about matricization and decomposition of tensors refer to (Hoff, 2011, Kolda and Bader, 2009, De Lathauwer et al., 2000). A vectorization of $\mathscr{X}$ is defined with help of usual vec operator for matrices as $\mathrm{vec}\mathscr{X} = \mathrm{vec}\boldsymbol{X}_{(1)}$.

Definition 2.4 (Tensor normal distribution). A random Kth order tensor $\mathscr{X} \in \mathbb{R}^{p_1\times p_2\times\cdots\times p_K}$ is said to be normally distributed if

$$\mathscr{X} \overset{d}{=} \mathscr{M} + \mathscr{U} \times \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_K\},$$

for some $\mathscr{M} \in \mathbb{R}^{p_1\times p_2\times\cdots\times p_K}$, non-singular matrices $\boldsymbol{\tau}_k \in \mathbb{R}^{p_k\times p_k}$, $k = 1, \ldots, K$ and a random tensor $\mathscr{U} \in \mathbb{R}^{p_1\times p_2\times\cdots\times p_K}$ with i.i.d. standard normal entries.

Here, the symbol "$\times$" denotes the Tucker product (Tucker, 1964, Kolda and Bader, 2009) defined by the identity

$$\mathrm{vec}(\mathscr{U} \times \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_K\}) = (\boldsymbol{\tau}_K \otimes \boldsymbol{\tau}_{K-1} \otimes \ldots \otimes \boldsymbol{\tau}_1)\mathrm{vec}\mathscr{U}.$$

It follows that

$$E[\mathscr{X}] = \mathscr{M},\ D[\mathscr{X}] = D[\mathrm{vec}\mathscr{X}] = \boldsymbol{\tau}_K\boldsymbol{\tau}_K' \otimes \boldsymbol{\tau}_{K-1}\boldsymbol{\tau}_{K-1}' \otimes \ldots \otimes \boldsymbol{\tau}_1\boldsymbol{\tau}_1'.$$

Thus, the tensor normal distribution corresponds to the multivariate normal distribution with separable (Kronecker product structure) covariance matrix. Letting $\boldsymbol{\Sigma}_k = \boldsymbol{\tau}_k\boldsymbol{\tau}_k'$, we write

$$\mathscr{X} \sim N_{p_1,\ldots,p_K}(\mathscr{M}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K), \tag{2.3}$$

which is equivalent to $\mathrm{vec}\mathscr{X} \sim N_{p_1\cdots p_K}(\mathrm{vec}\mathscr{M}, \boldsymbol{\Sigma}_K \otimes \cdots \otimes \boldsymbol{\Sigma}_1)$. The density function is given by

$$f(\boldsymbol{x}) = (2\pi)^{-p/2}\left(\prod_{k=1}^{K}|\boldsymbol{\Sigma}_k|^{-p/(2p_k)}\right)\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:K}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

where $\boldsymbol{\Sigma}_{1:K} = \boldsymbol{\Sigma}_1 \otimes \cdots \otimes \boldsymbol{\Sigma}_K$, $\boldsymbol{x} = \mathrm{vec}\mathscr{X}$, $\boldsymbol{\mu} = \mathrm{vec}\mathscr{M}$ and $p = \prod_{k=1}^{K} p_k$.

The model (2.3) is often used to model variation among entries of the multi-way data, a problem which is of great importance in many research fields. For example Basser and Pajevic (2003) argued on the need to go from the vectorial treatment of some complex data sets to tensor treatment in order to avoid wrong or inefficient conclusions. The Bayesian and the likelihood based approaches are the most used techniques to obtain estimators of unknown parameters in the tensor normal model, see for example (Hoff, 2011, Ohlson et al., 2013). For the third order tensor normal distribution the estimators can be found using the MLE-3D algorithm by Manceur and Dutilleul (2013) or similar algorithms like one proposed by Singull et al. (2012). Let $\mathscr{X}_i$, $i = 1, \ldots, n$, be a random sample from the tensor normal distribution

$$\mathscr{X} \sim N_{p_1,p_2,p_3}(\mathscr{M}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3). \tag{2.4}$$

Then, the maximum likelihood estimator of $\mathscr{M}$ is given by

$$\widehat{\mathscr{M}} = \frac{1}{n}\sum_{i=1}^{n}\mathscr{X}_i = \overline{\mathscr{X}}.$$

The respective maximum likelihood estimators $\widehat{\boldsymbol{\Sigma}}_1$, $\widehat{\boldsymbol{\Sigma}}_2$, $\widehat{\boldsymbol{\Sigma}}_3$ of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ are obtained by solving iteratively the following likelihood equations

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_1 &= \frac{1}{np_2p_3}\sum_{i=1}^{n}(\mathscr{X}_i - \overline{\mathscr{X}})_{(1)}(\widehat{\boldsymbol{\Sigma}}_3 \otimes \widehat{\boldsymbol{\Sigma}}_2)^{-1}((\mathscr{X}_i - \overline{\mathscr{X}})_{(1)})';\\
\widehat{\boldsymbol{\Sigma}}_2 &= \frac{1}{np_1p_3}\sum_{i=1}^{n}(\mathscr{X}_i - \overline{\mathscr{X}})_{(2)}(\widehat{\boldsymbol{\Sigma}}_3 \otimes \widehat{\boldsymbol{\Sigma}}_1)^{-1}((\mathscr{X}_i - \overline{\mathscr{X}})_{(2)})';\\
\widehat{\boldsymbol{\Sigma}}_3 &= \frac{1}{np_1p_2}\sum_{i=1}^{n}(\mathscr{X}_i - \overline{\mathscr{X}})_{(3)}(\widehat{\boldsymbol{\Sigma}}_2 \otimes \widehat{\boldsymbol{\Sigma}}_1)^{-1}((\mathscr{X}_i - \overline{\mathscr{X}})_{(3)})'.
\end{aligned}$$

Most studies on the third order tensor normal model focused on the estimation of parameters with unstructured mean. Nzabanita et al. (2015b) considered a trilinear structure for the mean in model (2.4) and proposed an algorithm for estimating the parameters.

2.2 Wishart distribution

In this section we present the definition and some properties of another important distribution which belongs to the class of matrix distributions, the Wishart distribution. First derived by Wishart (1928), the Wishart distribution is usually regarded as a multivariate analogue of the chi-square distribution. There are many ways to define the Wishart distribution and here we adopt the definition by Kollo and von Rosen (2005).

Definition 2.5 (Wishart distribution). The matrix $\boldsymbol{W} : p \times p$ is said to be Wishart distributed if and only if $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}'$ for some matrix $\boldsymbol{X}$, where $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I})$, and $\boldsymbol{\Sigma}$ is positive definite. If $\boldsymbol{M} = \boldsymbol{0}$, we have a central Wishart distribution which will be denoted $\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}, n)$, and if $\boldsymbol{M} \neq \boldsymbol{0}$, we have a non-central Wishart distribution which will be denoted $W_p(\boldsymbol{\Sigma}, n, \boldsymbol{\Delta})$, where $\boldsymbol{\Delta} = \boldsymbol{M}\boldsymbol{M}'$.

The first parameter $\boldsymbol{\Sigma}$ is usually supposed to be unknown. The second parameter n, which stands for the degrees of freedom is usually considered to be known. The third parameter $\boldsymbol{\Delta}$, which is used in the non-central Wishart distribution, is called the non-centrality parameter.

Some important properties of the Wishart distribution are given in the following theorem.

Theorem 2.1

(i) Let $\boldsymbol{W}_1 \sim W_p(\boldsymbol{\Sigma}, n, \boldsymbol{\Delta}_1)$ be independent of $\boldsymbol{W}_2 \sim W_p(\boldsymbol{\Sigma}, m, \boldsymbol{\Delta}_2)$. Then

$$\boldsymbol{W}_1 + \boldsymbol{W}_2 \sim W_p(\boldsymbol{\Sigma}, n + m, \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2).$$

(ii) Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, where $\mathcal{C}(\boldsymbol{M}') \subseteq \mathcal{C}(\boldsymbol{\Psi})$. Put $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{\Psi}^{-1}\boldsymbol{X}'$. Then

$$\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}, \operatorname{rank}(\boldsymbol{\Psi}), \boldsymbol{\Delta}),$$

where $\boldsymbol{\Delta} = \boldsymbol{M}\boldsymbol{\Psi}^{-1}\boldsymbol{M}'$.

(iii) Let $\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}, n, \boldsymbol{\Delta})$ and $\boldsymbol{A} \in \mathbb{R}^{q \times p}$. Then

$$\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}' \sim W_p(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}', n, \boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{A}').$$

(iv) Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I})$ and $\boldsymbol{Q} : n \times n$ be symmetric. Then $\boldsymbol{X}\boldsymbol{Q}\boldsymbol{X}'$ is Wishart distributed if and only if $\boldsymbol{Q}$ is idempotent.

(v) Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I})$ and $\boldsymbol{Q} : n \times n$ be symmetric and idempotent, so that $\boldsymbol{M}\boldsymbol{Q} = \boldsymbol{0}$. Then $\boldsymbol{X}\boldsymbol{Q}\boldsymbol{X}' \sim W_p(\boldsymbol{\Sigma}, \operatorname{rank}(\boldsymbol{Q}))$.

(vi) Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I})$, $\boldsymbol{Q}_1 : n \times n$ and $\boldsymbol{Q}_2 : n \times n$ be symmetric. Then $\boldsymbol{X}\boldsymbol{Q}_1\boldsymbol{X}'$ and $\boldsymbol{X}\boldsymbol{Q}_2\boldsymbol{X}'$ are independent if and only if $\boldsymbol{Q}_1\boldsymbol{Q}_2 = \boldsymbol{0}$.

The proofs of these results can be found, for example, in Kollo and von Rosen (2005).

Example 2.2

In Section 2.1, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the multivariate normal model $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ were given. These are respectively

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \frac{1}{n}\boldsymbol{X}\mathbf{1}_n, \\ \widehat{\boldsymbol{\Sigma}} &= \frac{1}{n}\boldsymbol{S}, \end{aligned}$$

where

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{Q}\boldsymbol{X}',$$

where $\boldsymbol{Q} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$.

It is easy to show that the matrix $\boldsymbol{Q}$ is idempotent and $\mathrm{rank}(\boldsymbol{Q}) = n - 1$. Thus,

$$\boldsymbol{S} \sim W_p(\boldsymbol{\Sigma}, n-1).$$

Moreover, we note that $\boldsymbol{Q}$ is a projector on the space $\mathcal{C}(\mathbf{1}_n)^{\perp}$, the orthogonal complement to the space $\mathcal{C}(\mathbf{1}_n)$. Hence $\boldsymbol{Q}\mathbf{1}_n = \mathbf{0}$ so that $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{S}$ (or $\widehat{\boldsymbol{\Sigma}}$) are independent.

3

Regression Models

THE goal of this chapter is to give definitions and some results on multivariate linear models. It starts with the general linear regression model, which includes well known models like the univariate linear regression model, the analysis of variance model and the analysis of covariance model. The multivariate counterpart includes the multivarate linear regression model, the multivariate analysis of variance and the multivariate analysis of covariance model. Then, the growth curve model and the extended growth curve model, which are refereed to as the bilinear regression models, are presented. At last, we define the trilinear regression model and give an example to illustrate its construction.

3.1 General linear regression model

In the general linear model (GLM) setup, a random set of n correlated observations, an observation vector $\boldsymbol{x}' = (x_1, x_2, \ldots, x_n)$, is related to a vector of k parameters, $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$, through a known nonrandom design matrix $\boldsymbol{C} : k \times n$ plus a random vector of errors $\boldsymbol{e} : n \times 1$, with mean zero, $E(\boldsymbol{e}) = 0$, and covariance matrix $\mathrm{cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}$. Thus, the general linear model (GLM) is represented as

$$\boldsymbol{x}' = \boldsymbol{\beta}'\boldsymbol{C} + \boldsymbol{e}',\ E(\boldsymbol{e}) = 0,\ \mathrm{cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}, \tag{3.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are unkown parameters.

The vector of parameters, $\boldsymbol{\beta}$, can be fixed, random or both (mixed model). In this thesis parameters are assumed to be fixed. The matrices $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ may have different forms and depending on these forms model (3.1) includes well known models like the univariate (linear) regression (UR) model, the analysis of variance (ANOVA) model and the analysis of covariance (ANCOVA) model.

Different techniques to estimate model parameters and hypothesis testing exist. For example, when there is no assumption on the distribution of $\boldsymbol{x}$, one can use the generalized

least squares (GLS) theory and minimum quadratic norm unbiased estimation (MINQUE) theory (Rao and Toutenburg, 1995) to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

In many statistical analysis, one assumes that the vector $\boldsymbol{x}$ has a multivariate normal distribution and model (3.1) becomes

$$\boldsymbol{x} \sim N_n(\boldsymbol{C}'\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{3.2}$$

that is a multivariate normal model with mean vector $\boldsymbol{C}'\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$. In this case the maximum likelihood theory for estimation and hypothesis testing may be used.

Model (3.2) corresponds to n observations on a single dependent response variable. When one has n independent observation vectors, $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})'$, $i = 1, \ldots, n$, on p correlated dependent response variables, model (3.2) generalizes to the model

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{3.3}$$

where $\boldsymbol{X} : p \times n$, $\boldsymbol{B} : p \times k$, $\boldsymbol{C} : k \times n$, $\boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I})$. The matrix $\boldsymbol{C}$ is a known design matrix, and $\boldsymbol{B}$ and the positive definite matrix $\boldsymbol{\Sigma}$ are unknown parameter matrices.

Again, depending on the forms of $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$, model (3.3) includes known models like the multivarate (linear) regression (MR) model, the multivariate analysis of variance (MANOVA) model (Roy, 1957, Anderson, 1958) and the multivariate analysis of covariance (MANCOVA) model.

3.2 Bilinear regression models. Growth curve models

The growth curve analysis is a topic with many important applications within medicine, natural sciences, social sciences, etc. Growth curve analysis has a long history and two classical papers are Box (1950) and Rao (1958). In 1964 the well known paper by Pothoff and Roy (1964) extended the MANOVA model (3.3) to the model which was later termed the growth curve model or the general MANOVA (GMANOVA).

Definition 3.1 (Growth curve model). Let $\boldsymbol{X} : p \times n$, $\boldsymbol{A} : p \times q$, $q \leq p$, $\boldsymbol{B} : q \times k$, $\boldsymbol{C} : k \times n$, $r(\boldsymbol{C}) + p \leq n$, where $r(\,\cdot\,)$ represents the rank of a matrix. The growth curve model is given by

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{3.4}$$

where columns of $\boldsymbol{E}$ are assumed to be independently distributed as a multivariate normal distribution with mean zero and a positive definite dispersion matrix $\boldsymbol{\Sigma}$; i.e., $\boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$.

The matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, often called respectively within-individuals and between-individuals design matrices, are known matrices whereas matrices $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are unknown parameter matrices.

The paper by Pothoff and Roy (1964) is often considered to be the first where the model was presented. The GMANOVA model was introduced to analyze growth in balanced repeated measures data. Several prominent authors wrote follow-up papers, e.g.,

Rao (1965) and Khatri (1966). Notice that the growth curve model is a special case of the matrix normal model where the mean has a bilinear structure. Therefore, we may use the notation

$$\boldsymbol{X} \sim N_{p,n}(\boldsymbol{ABC}, \boldsymbol{\Sigma}, \boldsymbol{I}).$$

Also, it is worth noting that the MANOVA model with restrictions

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{BC} + \boldsymbol{E}, \\ \boldsymbol{GB} &= \boldsymbol{0} \end{aligned} \tag{3.5}$$

is equivalent to the growth curve model. $\boldsymbol{GB} = \boldsymbol{0}$ is equivalent to $\boldsymbol{B} = (\boldsymbol{G}')^o\boldsymbol{\Theta}$, where $(\boldsymbol{G}')^o$ is any matrix spanning the orthogonal complement to the space generated by the columns of $\boldsymbol{G}'$. Plugging $(\boldsymbol{G}')^o\boldsymbol{\Theta}$ in (3.5) gives

$$\boldsymbol{X} = (\boldsymbol{G}')^o\boldsymbol{\Theta C} + \boldsymbol{E},$$

which is identical to the growth curve model (3.4).

Example 3.1: Potthoff & Roy (1964) dental data

Dental measurements on eleven girls and sixteen boys at four different ages ($t_1 = 8$, $t_2 = 10$, $t_3 = 12$, and $t_4 = 14$) were taken. Each measurement is the distance, in millimeters, from the center of pituitary to pteryo-maxillary fissure. These data are presented in Table 3.1 and plotted in Figure 3.1. Suppose linear growth curves describe the mean growth for

Table 3.1: *Dental data*

id	gender	t_1	t_2	t_3	t_4	id	gender	t_1	t_2	t_3	t_4
1	F	21.0	20.0	21.5	23.0	12	M	26.0	25.0	29.0	31.0
2	F	21.0	21.5	24.0	25.5	13	M	21.5	22.5	23.0	26.0
3	F	20.5	24.0	24.5	26.0	14	M	23.0	22.5	24.0	27.0
4	F	23.5	24.5	25.0	26.5	15	M	25.5	27.5	26.5	27.0
5	F	21.5	23.0	22.5	23.5	16	M	20.0	23.5	22.5	26.0
6	F	20.0	21.0	21.0	22.5	17	M	24.5	25.5	27.0	28.5
7	F	21.5	22.5	23.0	25.0	18	M	22.0	22.0	24.5	26.5
8	F	23.0	23.0	23.5	24.0	19	M	24.0	21.5	24.5	25.5
9	F	20.0	21.0	22.0	21.5	20	M	23.0	20.5	31.0	26.0
10	F	16.5	19.0	19.0	19.5	21	M	27.5	28.0	31.0	31.5
11	F	24.5	25.0	28.0	28.0	22	M	23.0	23.0	23.5	25.0
						23	M	21.5	23.5	24.0	28.0
						24	M	17.0	24.5	26.0	29.5
						25	M	22.5	25.5	25.5	26.0
						26	M	23.0	24.5	26.0	30.0
						27	M	22.0	21.5	23.5	25.0

both girls and boy. Then we may use the growth curve model

$$\boldsymbol{X} \sim N_{p,n}(\boldsymbol{ABC}, \boldsymbol{\Sigma}, \boldsymbol{I})$$

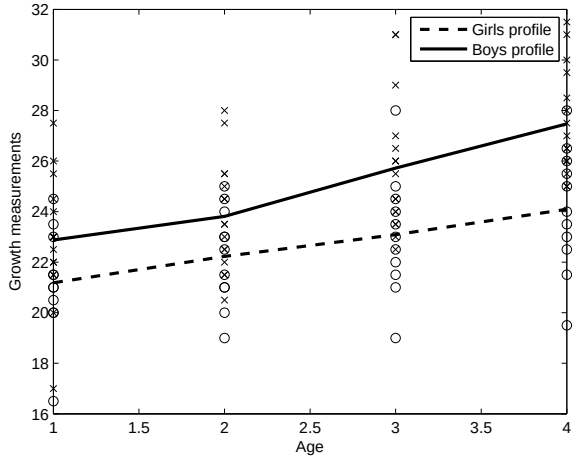

Figure 3.1: *Growth profiles plot (means at each time point joined with straight lines) of Potthoff and Roy (1964) dental data.*

to analysis this data set. In this model, the observation matrix is $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{27})$, in which eleven first columns correspond to measurements on girls and sixteen last columns correspond to measurements on boys. The design matrices are

$$\boldsymbol{A}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}, \quad \boldsymbol{C} = \left(\mathbf{1}'_{11} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \mathbf{1}'_{16} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right),$$

and $\boldsymbol{B}$ is the unknown parameter matrix and $\boldsymbol{\Sigma}$ is the unknown positive definite covariance matrix.

One limitation of the growth curve model is that different individuals should follow the same growth profile. If this does not hold there is a way to extend the model. A natural extension of the growth curve model, introduced by von Rosen (1989), is the following

Definition 3.2 (Extended growth curve model). Let $\boldsymbol{X} : p \times n$, $\boldsymbol{A}_i : p \times q_i$, $\boldsymbol{B}_i : q_i \times k_i$, $\boldsymbol{C}_i : k_i \times n$, $r(\boldsymbol{C}_1) + p \leq n$, $i = 1, 2, \ldots, m$, $\mathcal{C}(\boldsymbol{C}'_i) \subseteq \mathcal{C}(\boldsymbol{C}'_{i-1})$, $i = 2, 3, \ldots, m$, where $r(\,\cdot\,)$ and $\mathcal{C}(\,\cdot\,)$ represent the rank and column space of a matrix respectively. The extended growth curve model is given by

$$\boldsymbol{X} = \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{B}_i \boldsymbol{C}_i + \boldsymbol{E}, \tag{3.6}$$

where columns of $\boldsymbol{E}$ are assumed to be independently distributed as a multivariate normal distribution with mean zero and a positive definite dispersion matrix $\boldsymbol{\Sigma}$; i.e., $\boldsymbol{E} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$.

The matrices $\boldsymbol{A}_i$ and $\boldsymbol{C}_i$, often called design matrices, are known matrices whereas matrices $\boldsymbol{B}_i$ and $\boldsymbol{\Sigma}$ are unknown parameter matrices. As for the growth curve model the notation

$$\boldsymbol{X} \sim N_{p,n}\left(\sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{B}_i\boldsymbol{C}_i, \boldsymbol{\Sigma}, \boldsymbol{I}\right)$$

may be used for the extended growth curve model. The only difference with the growth curve model in Definition 3.1 is the presence of a more general mean structure. When $m = 1$, the model reduces to the growth curve model. The model without subspace conditions was considered before by Verbyla and Venables (1988) under the name of *sum of profiles model*. Also observe that the subspace conditions $\mathcal{C}(\boldsymbol{C}_i') \subseteq \mathcal{C}(\boldsymbol{C}_{i-1}')$, $i = 2, 3, \ldots, m$ may be replaced by $\mathcal{C}(\boldsymbol{A}_i) \subseteq \mathcal{C}(\boldsymbol{A}_{i-1})$, $i = 2, 3, \ldots, m$. This problem was considered for example by Filipiak and von Rosen (2012) for $m = 3$.

Example 3.2

Consider again Potthoff & Roy (1964) classical dental data. From Figure 3.1, it is reasonable to assume that for both girls and boys we have a linear growth component but additionally for the boys there also exists a second order polynomial structure. Then we may use the extended growth curve model with two terms

$$\boldsymbol{X} \sim N_{p,n}(\boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{C}_1 + \boldsymbol{A}_2\boldsymbol{B}_2\boldsymbol{C}_2, \boldsymbol{\Sigma}, \boldsymbol{I}),$$

where

$$\boldsymbol{A}_1' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}, \quad \boldsymbol{C}_1 = \left(\mathbf{1}_{11}' \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \mathbf{1}_{16}' \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$
$$\boldsymbol{A}_2' = \begin{pmatrix} 8^2 & 10^2 & 12^2 & 14^2 \end{pmatrix}, \quad \boldsymbol{C}_2 = (\mathbf{0}_{11}' : \mathbf{1}_{16}'),$$

are design matrices and $\boldsymbol{B}_1 = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$ and $\boldsymbol{B}_2 = (\beta_{32})$ are parameter matrices and $\boldsymbol{\Sigma}$ is the same as in Example 3.1.

3.3 Trilinear regression model

The classical growth curve model (3.4) by Pothoff and Roy (1964) comprises two design matrices; one models the within-individuals structure whereas the other one models the between-individuals structure. More specifically, the within-individuals design matrix $\boldsymbol{A}$ contains time regressors and models growth curves, and the between-individuals design matrix $\boldsymbol{C}$ is comprised of group separation indicators. It is suitable to analyze, for example, "one directional" repeated measures data. Nzabanita et al. (2015b) extended the classical growth curve model with an additional within-individuals design matrix which can be used to analyze "two directional" repeated measures data. More specifically, the model considered is the third order tensor normal model

$$\mathscr{X} \sim N_{p,q,r}(\mathscr{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\Omega}),$$

with mean structure of the form

$$\mu_{ijk} = \sum_{\ell=1}^{s}\sum_{m=1}^{t}\sum_{n=1}^{u} b_{\ell mn}a_{i\ell}c_{jm}d_{kn},\ i = 1,\dots,p,\ j = 1,\dots,q,\ k = 1,\dots,r.$$

This mean structure can be written

$$\mathscr{M} = \mathscr{B} \times \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}\},$$

where $\mathscr{B} = (b_{\ell mn}) : s \times t \times u$ is the parameter given as a tensor of order three, $\boldsymbol{A} = (a_{i\ell}) : p \times s$, $\boldsymbol{C} = (c_{jm}) : q \times t$ and $\boldsymbol{D} = (d_{kn}) : r \times u$ are known design matrices, and $\times$ denotes the Tucker product, see Kolda and Bader (2009), and it is defined by the identity

$$\mathrm{vec}(\mathscr{B} \times \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}\}) = (\boldsymbol{D} \otimes \boldsymbol{C} \otimes \boldsymbol{A})\mathrm{vec}\mathscr{B}.$$

The artificial example below illustrates how this kind of model may arise.

Example 3.3

Assume that one has measured pH in r lakes from u regions at q levels of depth and for p time points. The aim is to investigate how pH varies with depth and/or time and how pH differs across regions. Thus, we have spatio-temporal measurements. Data form a random tensor $\mathscr{X} : p \times q \times r$, where $r = r_1 + r_2 + \cdots + r_u$ and r_n is the number of lakes in the n^{th} region. It is assumed that measurements of each lake (a frontal slice in the tensor $\mathscr{X}$) is distributed as a matrix normal distribution with covariance matrices $\boldsymbol{\Sigma} : p \times p$, and $\boldsymbol{\Psi} : q \times q$, and that the measurements of different lakes are independent. If the first r_1 frontal slices of $\mathscr{X}$ represent region one, the next r_2 frontal slices represent region two, and so on, we get the between-individuals design matrix $\boldsymbol{D}' = \mathrm{blockdiag}(\mathbf{1}'_{r_1}, \dots, \mathbf{1}'_{r_u})$. It is also assumed that the expected trend in time is a polynomial of order $s-1$ and that the expected trend in depth is a polynomial of order $t-1$. Thus, we have two within-individuals design matrices

$$\boldsymbol{A} = \begin{pmatrix} 1 & t_1 & \cdots & t_1^{s-1} \\ 1 & t_2 & \cdots & t_2^{s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & \cdots & t_p^{s-1} \end{pmatrix} \text{ and } \boldsymbol{C} = \begin{pmatrix} 1 & d_1 & \cdots & d_1^{t-1} \\ 1 & d_2 & \cdots & d_2^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_q & \cdots & d_q^{t-1} \end{pmatrix}.$$

Hence, the model for the data tensor $\mathscr{X}$ is

$$\mathscr{X} = \mathscr{B} \times \{\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}\} + \mathscr{E}, \tag{3.7}$$

where $\mathscr{E} \sim N_{p,q,r}(\mathscr{O}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{I}_r)$, and $\mathscr{O}$ is a tensor of zeros.

In this thesis the model (3.7) is refereed to as the *2-fold Growth Curve Model* and serves as an example of a trilinear regression model.

3.4 Estimation in bilinear regression models

The problem of estimating parameters in the (extended) growth curve model has been studied by several authors. The book by Kollo and von Rosen (2005) [Chapter 4] contains useful detailed information about uniqueness, estimability conditions, moments and approximative distributions of the maximum likelihood estimators in the model given in Definition 3.2. Recently other authors considered the model with slightly different conditions. For example in Filipiak and von Rosen (2012), the explicit MLEs are presented with the nested subspace conditions on the within design matrices instead. In (Hu, 2010, Hu et al., 2011), the extended growth curve model without nested subspace conditions but with orthogonal design matrices is considered and generalized least-squares estimators and their properties are studied.

3.4.1 Maximum likelihood estimators

To find estimators of parameters, when the covariance matrix $\boldsymbol{\Sigma}$ is not structured, very often the maximum likelihood method is used. The maximum likelihood estimators of parameters in the growth curve model have been studied by many authors, see for instance Srivastava and Khatri (1979) and von Rosen (1989). For the extended growth curve model with nested subspace conditions as in Definition 3.2, von Rosen (1989) derived explicit maximum likelihood estimators (MLEs). The following theorem gives the MLEs of parameters in the extended growth curve model.

Theorem 3.1
Consider the extended growth curve model as in Definition 3.2. Let

$$\begin{aligned}
\boldsymbol{P}_r &= \boldsymbol{T}_{r-1}\boldsymbol{T}_{r-2}\times\cdots\times\boldsymbol{T}_0,\ \boldsymbol{T}_0=\boldsymbol{I},\ r=1,2,\ldots,m+1,\\
\boldsymbol{T}_i &= \boldsymbol{I}-\boldsymbol{P}_i\boldsymbol{A}_i(\boldsymbol{A}_i'\boldsymbol{P}_i'\boldsymbol{S}_i^{-1}\boldsymbol{P}_i\boldsymbol{A}_i)^{-}\boldsymbol{A}_i'\boldsymbol{P}_i'\boldsymbol{S}_i^{-1},\ i=1,2,\ldots,m,\\
\boldsymbol{S}_i &= \sum_{j=1}^{i}\boldsymbol{K}_j,\ i=1,2,\ldots,m,\\
\boldsymbol{K}_j &= \boldsymbol{P}_j\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_{j-1}'}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{C}_j'})\boldsymbol{P}_{\boldsymbol{C}_{j-1}'}\boldsymbol{X}'\boldsymbol{P}_j',\ \boldsymbol{C}_0=\boldsymbol{I},\\
\boldsymbol{P}_{\boldsymbol{C}_j'} &= \boldsymbol{C}_j'(\boldsymbol{C}_j\boldsymbol{C}_j')^{-}\boldsymbol{C}_j.
\end{aligned}$$

Assume that $\boldsymbol{S}_1$ is positive definite.

(i) The representations of maximum likelihood estimators of $\boldsymbol{B}_r$, $r=1,2,\ldots,m$ and $\boldsymbol{\Sigma}$ are

$$\begin{aligned}
\widehat{\boldsymbol{B}}_r &= (\boldsymbol{A}_r'\boldsymbol{P}_r'\boldsymbol{S}_r^{-1}\boldsymbol{P}_r\boldsymbol{A}_r)^{-}\boldsymbol{A}_r'\boldsymbol{P}_r'\boldsymbol{S}_r^{-1}(\boldsymbol{X}-\sum_{i=r+1}^{m}\boldsymbol{A}_i\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i)\boldsymbol{C}_r'(\boldsymbol{C}_r\boldsymbol{C}_r')^{-}\\
&\quad+(\boldsymbol{A}_r'\boldsymbol{P}_r')^{o}\boldsymbol{Z}_{r1}+\boldsymbol{A}_r'\boldsymbol{P}_r'\boldsymbol{Z}_{r2}\boldsymbol{C}_r^{o\prime},\\
n\widehat{\boldsymbol{\Sigma}} &= (\boldsymbol{X}-\sum_{i=1}^{m}\boldsymbol{A}_i\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i)(\boldsymbol{X}-\sum_{i=1}^{m}\boldsymbol{A}_i\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i)'\\
&= \boldsymbol{S}_m+\boldsymbol{P}_{m+1}\boldsymbol{X}\boldsymbol{C}_m'(\boldsymbol{C}_m\boldsymbol{C}_m')^{-}\boldsymbol{C}_m\boldsymbol{X}'\boldsymbol{P}_{m+1},
\end{aligned}$$

where $\boldsymbol{Z}_{r1}$ *and* $\boldsymbol{Z}_{r2}$ *are arbitrary matrices and* $\sum_{i=m+1}^{m} \boldsymbol{A}_i \widehat{\boldsymbol{B}}_i \boldsymbol{C}_i = \boldsymbol{0}$.

(ii) For the estimators $\widehat{\boldsymbol{B}}_i$,

$$\boldsymbol{P}_r \sum_{i=r}^{m} \boldsymbol{A}_i \widehat{\boldsymbol{B}}_i \boldsymbol{C}_i = \sum_{i=r}^{m} (\boldsymbol{I} - \boldsymbol{T}_i) \boldsymbol{X} \boldsymbol{C}_i' (\boldsymbol{C}_i \boldsymbol{C}_i')^{-} \boldsymbol{C}_i.$$

The notation $\boldsymbol{C}^o$ *stands for any matrix of full rank spanning* $\mathcal{C}(\boldsymbol{C})^{\perp}$, *and* $\boldsymbol{G}^{-}$ *denotes an arbitrary generalized inverse in the sense that* $\boldsymbol{G}\boldsymbol{G}^{-}\boldsymbol{G} = \boldsymbol{G}$.

For the proof of this theorem, see for example von Rosen (1989) or Kollo and von Rosen (2005).

A useful results is the corollary of this theorem when $r = 1$, which gives the estimated mean structure.

Corollary 3.1

$$\widehat{E[\boldsymbol{X}]} = \sum\nolimits_{i=1}^{m} \boldsymbol{A}_i \widehat{\boldsymbol{B}}_i \boldsymbol{C}_i = \sum_{i=1}^{m} (\boldsymbol{I} - \boldsymbol{T}_i) \boldsymbol{X} \boldsymbol{C}_i' (\boldsymbol{C}_i \boldsymbol{C}_i')^{-} \boldsymbol{C}_i.$$

Example 3.4

Set $m = 2$ in the extended growth curve model of Definition 3.2. Then, form Theorem 3.1, the maximum likelihood estimators for the parameter matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are given by

$$\begin{aligned}
\widehat{\boldsymbol{B}}_2 &= (\boldsymbol{A}_2' \boldsymbol{P}_2' \boldsymbol{S}_2^{-1} \boldsymbol{P}_2 \boldsymbol{A}_2)^{-} \boldsymbol{A}_2' \boldsymbol{P}_2' \boldsymbol{S}_2^{-1} \boldsymbol{X} \boldsymbol{C}_2' (\boldsymbol{C}_2 \boldsymbol{C}_2')^{-} + (\boldsymbol{A}_2' \boldsymbol{P}_2)^o \boldsymbol{Z}_{21} + \boldsymbol{A}_2' \boldsymbol{Z}_{22} \boldsymbol{C}_2^{o'} \\
\widehat{\boldsymbol{B}}_1 &= (\boldsymbol{A}_1' \boldsymbol{S}_1^{-1} \boldsymbol{A}_1)^{-} \boldsymbol{A}_1' \boldsymbol{S}_1^{-1} (\boldsymbol{X} - \boldsymbol{A}_2 \widehat{\boldsymbol{B}}_2 \boldsymbol{C}_2) \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-} + \boldsymbol{A}_1'^{o} \boldsymbol{Z}_{11} + \boldsymbol{A}_1' \boldsymbol{Z}_{12} \boldsymbol{C}_1^{o'}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{S}_1 &= \boldsymbol{X} \left(\boldsymbol{I} - \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-} \boldsymbol{C}_1 \right) \boldsymbol{X}', \\
\boldsymbol{P}_2 &= \boldsymbol{I} - \boldsymbol{A}_1 (\boldsymbol{A}_1' \boldsymbol{S}_1^{-1} \boldsymbol{A}_1)^{-} \boldsymbol{A}_1' \boldsymbol{S}_1^{-1}, \\
\boldsymbol{S}_2 &= \boldsymbol{S}_1 + \boldsymbol{P}_2 \boldsymbol{X} \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-} \boldsymbol{C}_1 \left(\boldsymbol{I} - \boldsymbol{C}_2' (\boldsymbol{C}_2 \boldsymbol{C}_2')^{-} \boldsymbol{C}_2 \right) \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-} \boldsymbol{C}_1 \boldsymbol{X}' \boldsymbol{P}_2',
\end{aligned}$$

$\boldsymbol{Z}_{kl}$ are arbitrary matrices.

Assuming that matrices $\boldsymbol{A}_i$'s, $\boldsymbol{C}_i$'s are of full rank and that $\mathcal{C}(\boldsymbol{A}_1) \cap \mathcal{C}(\boldsymbol{A}_2) = \{0\}$, the unique maximum likelihood estimators are

$$\begin{aligned}
\widehat{\boldsymbol{B}}_2 &= (\boldsymbol{A}_2' \boldsymbol{P}_2' \boldsymbol{S}_2^{-1} \boldsymbol{P}_2 \boldsymbol{A}_2)^{-1} \boldsymbol{A}_2' \boldsymbol{P}_2' \boldsymbol{S}_2^{-1} \boldsymbol{X} \boldsymbol{C}_2' (\boldsymbol{C}_2 \boldsymbol{C}_2')^{-1}, \\
\widehat{\boldsymbol{B}}_1 &= (\boldsymbol{A}_1' \boldsymbol{S}_1^{-1} \boldsymbol{A}_1)^{-1} \boldsymbol{A}_1' \boldsymbol{S}_1^{-1} (\boldsymbol{X} - \boldsymbol{A}_2 \widehat{\boldsymbol{B}}_2 \boldsymbol{C}_2) \boldsymbol{C}_1' (\boldsymbol{C}_1 \boldsymbol{C}_1')^{-1}.
\end{aligned}$$

Obviously, under general settings, the maximum likelihood estimators $\widehat{\boldsymbol{B}}_1$ and $\widehat{\boldsymbol{B}}_2$ are not unique due to the arbitrariness of matrices $\boldsymbol{Z}_{kl}$. However, it is worth noting that the estimated mean

$$\widehat{E[\boldsymbol{X}]} = \boldsymbol{A}_1 \widehat{\boldsymbol{B}}_1 \boldsymbol{C}_1 + \boldsymbol{A}_2 \widehat{\boldsymbol{B}}_2 \boldsymbol{C}_2$$

is always unique and therefore $\widehat{\boldsymbol{\Sigma}}$ given by

$$n\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{X} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)(\boldsymbol{X} - \boldsymbol{A}_1\widehat{\boldsymbol{B}}_1\boldsymbol{C}_1 - \boldsymbol{A}_2\widehat{\boldsymbol{B}}_2\boldsymbol{C}_2)'$$

is also unique.

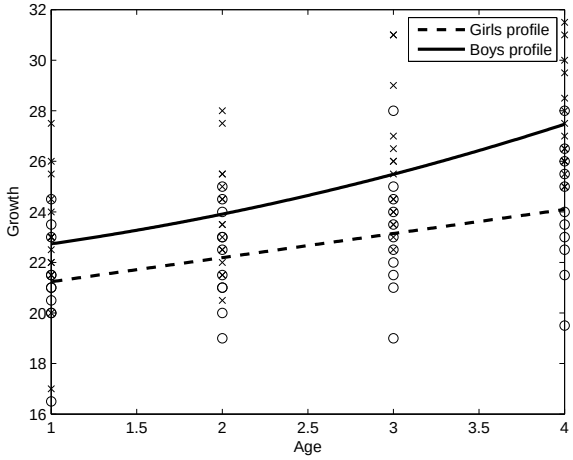

Figure 3.2: *Estimated mean growth curves for Potthoff and Roy (1964) dental data.*

Example 3.5: Example 3.2 continued

Consider again Potthoff & Roy (1964) classical dental data and the model of Example 3.2. Then, the maximum likelihood estimates of parameters are

$$\begin{aligned}
\widehat{\boldsymbol{B}}_1 &= \begin{pmatrix} 20.2836 & 21.9599 \\ 0.9527 & 0.5740 \end{pmatrix}, \widehat{\boldsymbol{B}}_2 = (0.2006), \\
\widehat{\boldsymbol{\Sigma}} &= \begin{pmatrix} 5.0272 & 2.5066 & 3.6410 & 2.5099 \\ 2.5066 & 3.8810 & 2.6961 & 3.0712 \\ 3.6410 & 2.6961 & 6.0104 & 3.8253 \\ 2.5099 & 3.0712 & 3.8253 & 4.6164 \end{pmatrix}.
\end{aligned}$$

The estimated mean growth curves, plotted in Figure 3.2, for girls and boys are respectively

$$\begin{aligned}
\widehat{\mu}_g(t) &= 20.2836 + 0.9527\,t, \\
\widehat{\mu}_b(t) &= 21.9599 + 0.5740\,t + 0.2006\,t^2.
\end{aligned}$$

3.4.2 Explicit estimators when the covariance matrix is linearly structured

A covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ is linearly structured if the only linear structure between the elements is given by $|\sigma_{ij}| = |\sigma_{kl}| \neq 0$ and there exists at least one $(i,j) \neq (k,l)$ so that $|\sigma_{ij}| = |\sigma_{kl}| \neq 0$. Examples of linear structures for the covariance matrix are, e.g., uniform structure, compound symmetry structure, banded structure, Toeplitz structure, etc.

In most of works on the extended growth curve model no particular attention has been made on the structure of the covariance matrix. In fact there are few articles treating the problem of structured covariance matrix although it may be important in the growth curve analysis.

For the classical growth curve model, the most studied structure are the uniform covariance structure and the serial covariance structure, see for example, (Lee and Geisser, 1975, Lee, 1988, Khatri, 1973, Klein and Žežula, 2009, Srivastava and Singull, 2015). The paper by Ohlson and von Rosen (2010) was the first to propose a residual based procedure to obtain explicit estimators for an arbitrary linear structured covariance matrix in the classical growth curve model as an alternative to iterative methods. The idea in Ohlson and von Rosen (2010) was later on applied to the sum of two profiles model by Nzabanita et al. (2012). The results in Nzabanita et al. (2012) have been generalized to the extended GMANOVA model with an arbitrary number of profiles by Nzabanita et al. (2015a). The procedure relies on the decomposition of the residual space into $m+1$ subspaces, see Theorem 3.3, and on the study of residuals obtained from projecting observations onto those subspaces. Hereafter we illustrate how it works.

If $\boldsymbol{\Sigma}$ would have been known, from least squares theory, the best linear unbiased estimator (BLUE) of the mean structure in model (3.6) would be given by

$$\widetilde{E[\boldsymbol{X}]} = \sum_{i=1}^{m} \boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}} \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{C}_i'}, \tag{3.8}$$

where $\boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}} = \widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i (\boldsymbol{A}_i' \widetilde{\boldsymbol{P}}_i' \boldsymbol{\Sigma}_i^{-1} \widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i)^- \boldsymbol{A}_i' \widetilde{\boldsymbol{P}}_i' \boldsymbol{\Sigma}_i^{-1}$, and $\widetilde{\boldsymbol{P}}_i$ is defined as $\boldsymbol{P}_i$ in Theorem 3.1 with $\boldsymbol{S}_i$ replaced with $\boldsymbol{\Sigma}$.

Applying the vec-operator on both sides of (3.8) we get

$$\mathrm{vec}(\widetilde{E[\boldsymbol{X}]}) = \sum_{i=1}^{m} (\boldsymbol{P}_{\boldsymbol{C}_i'} \otimes \boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}}) \mathrm{vec} \boldsymbol{X}.$$

Noting that the matrix $\boldsymbol{P} = \boldsymbol{P}_{\boldsymbol{C}_i'} \otimes \boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}}$ is a projector, see Theorem 3.2, we see that the estimator of the mean structure is based on a projection of observations on the space generated by the design matrices. Naturally, the estimators of the variance parameters are based on a projection of observations on the residual space, that is the orthogonal complement to the design space.

Theorem 3.2
Let $\boldsymbol{P} = \sum_{i=1}^{m} \boldsymbol{P}_{\boldsymbol{C}_i'} \otimes \boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}}$ *and* $\mathcal{V}_i = \mathcal{C}_{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i)$, $i = 1, 2, \ldots, m$. *Then,*

(i) The subspaces $\mathcal{V}_i$*'s are mutually orthogonal and*

$$\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \cdots \oplus \mathcal{V}_i = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1 : \boldsymbol{A}_2 : \cdots : \boldsymbol{A}_i),\ i = 1, 2, \ldots, m;$$

(ii) The matrix $\boldsymbol{P}$ is a projection matrix;

(iii) $\mathcal{C}(\boldsymbol{P}) = \sum_{i=1}^{m} \mathcal{C}(\boldsymbol{C}_i') \otimes \mathcal{V}_i$.

The proof of this theorem can be found in Nzabanita et al. (2015a).

The space $\mathcal{C}(\boldsymbol{P})$ is refereed to as the mean space and it is used to estimate the mean parameters whereas $\mathcal{C}(\boldsymbol{P})^{\perp}$, the orthogonal complement to the mean space, is refereed to as the residual space and it is used to create residuals.

When $\boldsymbol{\Sigma}$ is not known it should be estimated. The general idea is to use the variation in the residuals. For our purposes we decompose the residual space into $m+1$ orthogonal subspaces and Theorem 3.3, proved in Nzabanita et al. (2015a), shows how such a decomposition is made.

Theorem 3.3
Let $\mathcal{C}(\boldsymbol{P})$ and $\mathcal{V}_i$ be given as in Theorem 3.2. Then

$$\mathcal{C}(\boldsymbol{P})^{\perp} = I_1 \boxplus I_2 \boxplus \cdots \boxplus I_{m+1},$$

where

$$\begin{aligned} I_r &= \mathcal{W}_{m-r+2} \otimes (\oplus_{i=1}^{r-1} \mathcal{V}_i)^{\perp},\ r = 1, 2, \ldots, m+1, \\ \mathcal{W}_r &= \mathcal{C}(\boldsymbol{C}'_{m-r+1}) \cap \mathcal{C}(\boldsymbol{C}'_{m-r+2})^{\perp},\ r = 1, \ldots, m+1, \end{aligned}$$

in which by convenience $(\oplus_{i=1}^{0} \mathcal{V}_i)^{\perp} = \emptyset^{\perp} = \mathcal{V}_0 = \mathbb{R}^p$, $\boldsymbol{C}_0 = \boldsymbol{I}$ and $\boldsymbol{C}_{m+1} = \boldsymbol{0}$.

The residuals obtained by projecting data to these subspaces are

$$\boldsymbol{H}_r = (\boldsymbol{I} - \sum_{i=1}^{r-1} \boldsymbol{P}_{\widetilde{\boldsymbol{P}}_i \boldsymbol{A}_i, \boldsymbol{\Sigma}}) \boldsymbol{X} (\boldsymbol{P}_{\boldsymbol{C}'_{r-1}} - \boldsymbol{P}_{\boldsymbol{C}'_r}),$$

$r = 1, 2, 3, \ldots, m+1$, and here we use for convenience $\sum_{i=1}^{0} \boldsymbol{P}_{\boldsymbol{P}_i \boldsymbol{A}_i, \boldsymbol{S}_i} = \boldsymbol{0}$.

For illustrative purposes, let $m = 2$. In this case the BLUE of the mean is

$$\widetilde{E[\boldsymbol{X}]} = \widetilde{\boldsymbol{M}}_1 + \widetilde{\boldsymbol{M}}_2,$$

where

$$\begin{aligned} \widetilde{\boldsymbol{M}}_1 &= \boldsymbol{P}_{\boldsymbol{A}_1, \boldsymbol{\Sigma}} \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{C}'_1}, \\ \widetilde{\boldsymbol{M}}_2 &= \boldsymbol{P}_{\boldsymbol{\mathcal{T}}_1 \boldsymbol{A}_2, \boldsymbol{\Sigma}} \boldsymbol{X} \boldsymbol{P}_{\boldsymbol{C}'_2},\ \boldsymbol{\mathcal{T}}_1 = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_1, \boldsymbol{\Sigma}} = \boldsymbol{T}_1. \end{aligned}$$

From here we see that the estimated mean is obtained by projecting observations on some subspaces. The matrices $\boldsymbol{P}_{\boldsymbol{A}_1, \boldsymbol{\Sigma}}$ and $\boldsymbol{P}_{\boldsymbol{\mathcal{T}}_1 \boldsymbol{A}_2, \boldsymbol{\Sigma}}$ are projectors onto the subspaces $\mathcal{V}_1 = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1)$ and $\mathcal{V}_2 = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1 : \boldsymbol{A}_2) \cap \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1)^{\perp}$, respectively. Figure 3.3 shows the whole space decomposed into mean and residual subspaces.

In practice $\boldsymbol{\Sigma}$ is not known and should be estimated. A natural way to get an estimator of $\boldsymbol{\Sigma}$ is to use the sum of squared residuals. If $\boldsymbol{\Sigma}$ is not structured we estimate the residuals, $\boldsymbol{H}_i$, $i = 1, 2, 3$, in Figure 3.3 with

$$\boldsymbol{R}_r = (\boldsymbol{I} - \sum_{i=1}^{r-1} \boldsymbol{P}_{\boldsymbol{P}_i \boldsymbol{A}_i, \boldsymbol{S}_i}) \boldsymbol{X} (\boldsymbol{P}_{\boldsymbol{C}'_{r-1}} - \boldsymbol{P}_{\boldsymbol{C}'_r}),\ r = 1, 2, 3,$$

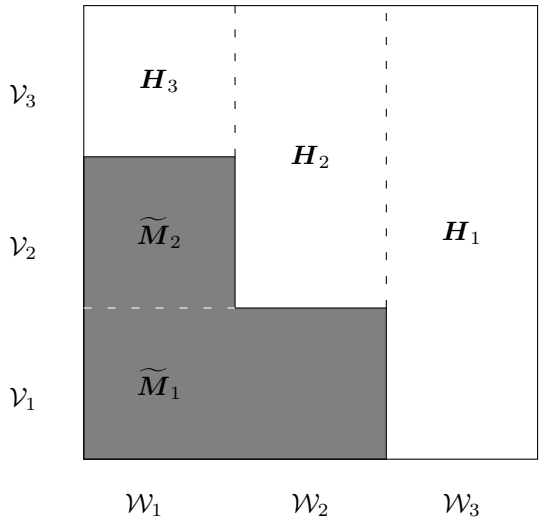

Figure 3.3: *Decomposition of the whole space according to the within and between individuals design matrices illustrating the mean and residual spaces:* $\mathcal{V}_1 = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1)$, $\mathcal{V}_2 = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1 : \boldsymbol{A}_2) \cap \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1)^{\perp}$, $\mathcal{V}_3 = \mathcal{C}_{\boldsymbol{\Sigma}}(\boldsymbol{A}_1 : \boldsymbol{A}_2)^{\perp}$, $\mathcal{W}_1 = \mathcal{C}(\boldsymbol{C}_2')$, $\mathcal{W}_2 = \mathcal{C}(\boldsymbol{C}_1') \cap \mathcal{C}(\boldsymbol{C}_2')^{\perp}$, $\mathcal{W}_3 = \mathcal{C}(\boldsymbol{C}_1')^{\perp}$.

where $\boldsymbol{P}_i$ and $\boldsymbol{S}_i$ are given in Theorem 3.1. Thus a natural estimator of $\boldsymbol{\Sigma}$ is obtained from the sum of squared residuals, i.e.,

$$n\widehat{\boldsymbol{\Sigma}} = \boldsymbol{R}_1\boldsymbol{R}_1' + \boldsymbol{R}_2\boldsymbol{R}_2' + \boldsymbol{R}_3\boldsymbol{R}_3', \tag{3.9}$$

which is the maximum likelihood estimator.

When $\boldsymbol{\Sigma}$ is linearly structured, the estimator in (3.9) does not exhibit the same structure although it may be desirable. One objective of this thesis was to propose an procedure that produces an estimator of the covariance matrix with a desired linear structure. The idea is first to estimate the covariance matrix from $\boldsymbol{Q}_1 = \boldsymbol{H}_1\boldsymbol{H}_1'$ and use it to define the inner product in the spaces $\mathcal{V}_1$, and then estimate $\boldsymbol{M}_1$ and $\boldsymbol{H}_2$ by projecting observations on $\mathcal{C}(\boldsymbol{C}_1') \otimes \mathcal{V}_1$ and $\left(\mathcal{C}(\boldsymbol{C}_1') \cap \mathcal{C}(\boldsymbol{C}_2')^{\perp}\right) \otimes \mathcal{V}_1^{\perp}$ respectively.

By $\mathrm{vec}\boldsymbol{\Sigma}(K)$ we mean the patterned vectorization of the linearly structured matrix $\boldsymbol{\Sigma}$, that is the columnwise vectorization of $\boldsymbol{\Sigma}$ where all 0's and repeated elements (by modulus) have been disregarded. Then there exists a transformation matrix $\boldsymbol{T}$ such that

$$\mathrm{vec}\boldsymbol{\Sigma}(K) = \boldsymbol{T}\mathrm{vec}\boldsymbol{\Sigma} \text{ or } \mathrm{vec}\boldsymbol{\Sigma} = \boldsymbol{T}^{+}\mathrm{vec}\boldsymbol{\Sigma}(K),$$

where $\boldsymbol{T}^{+}$ denotes the Moore-Penrose generalized inverse of $\boldsymbol{T}$.

From results in Nzabanita et al. (2012), the first estimator of $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}}_1 = \arg\min_{\boldsymbol{\Sigma}} \mathrm{tr}\left\{(\boldsymbol{Q}_1 - (n - r_1)\boldsymbol{\Sigma})(\,)'\right\}$$

given by

$$\mathrm{vec}\widehat{\boldsymbol{\Sigma}}_1 = \boldsymbol{T}^{+}\left((\boldsymbol{T}^{+})'\widehat{\boldsymbol{\Upsilon}}_1'\widehat{\boldsymbol{\Upsilon}}_1\boldsymbol{T}^{+}\right)^{-}(\boldsymbol{T}^{+})'\widehat{\boldsymbol{\Upsilon}}_1'\mathrm{vec}\boldsymbol{Q}_1,$$

where $\widehat{\Upsilon}_1 = (n - r_1)\boldsymbol{I}$ and the notation $(\boldsymbol{Y})(\,)'$ stands for $(\boldsymbol{Y})(\boldsymbol{Y})'$.

Assuming that $\widehat{\boldsymbol{\Sigma}}_1$ is positive definite (which always holds for large n), and using it to define the inner product in the space $\mathcal{V}_1$, the estimator of $\boldsymbol{M}_1$ and $\boldsymbol{H}_2$ are given by $\widehat{\boldsymbol{M}}_1 = \boldsymbol{P}_{\boldsymbol{A}_1,\widehat{\boldsymbol{\Sigma}}_1^{(s)}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_1'}$ and $\widehat{\boldsymbol{H}}_2 = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_1,\widehat{\boldsymbol{\Sigma}}_1^{(s)}})\boldsymbol{X}(\boldsymbol{P}_{\boldsymbol{C}_1'} - \boldsymbol{P}_{\boldsymbol{C}_2'})$, respectively.

A second estimator of $\boldsymbol{\Sigma}$ is obtained using the sum of $\boldsymbol{Q}_1$ and $\widehat{\boldsymbol{H}}_2\widehat{\boldsymbol{H}}_2'$ in a similar way and is given by

$$\mathrm{vec}\widehat{\boldsymbol{\Sigma}}_2 = \boldsymbol{T}^+\left((\boldsymbol{T}^+)'\widehat{\boldsymbol{\Upsilon}}_2'\widehat{\boldsymbol{\Upsilon}}_2\boldsymbol{T}^+\right)^-(\boldsymbol{T}^+)'\widehat{\boldsymbol{\Upsilon}}_2'\mathrm{vec}\widehat{\boldsymbol{Q}}_2,$$

where $\widehat{\boldsymbol{Q}}_2 = \boldsymbol{Q}_1 + \widehat{\boldsymbol{H}}_2\widehat{\boldsymbol{H}}_2'$, $\widehat{\Upsilon}_2 = (n-r_1)\boldsymbol{I} + (r_1 - r_2)\widehat{\boldsymbol{\mathcal{T}}}_1 \otimes \widehat{\boldsymbol{\mathcal{T}}}_1$ and $\widehat{\boldsymbol{\mathcal{T}}}_1 = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_1,\widehat{\boldsymbol{\Sigma}}_1}$.

Assume that $\widehat{\boldsymbol{\Sigma}}_2$ is positive definite and use it to define the inner product in $\mathcal{V}_2$, the estimators of $\boldsymbol{M}_2$ and $\boldsymbol{H}_3$ are given by

$$\begin{aligned}
\widehat{\boldsymbol{M}}_2 &= \boldsymbol{P}_{\widehat{\boldsymbol{\mathcal{T}}}_1\boldsymbol{A}_2,\widehat{\boldsymbol{\Sigma}}_2}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}_2'},\\
\widehat{\boldsymbol{H}}_3 &= \widehat{\boldsymbol{\mathcal{T}}}_2\boldsymbol{X}(\boldsymbol{P}_{\boldsymbol{C}_2'} - \boldsymbol{P}_{\boldsymbol{C}_3'}),\\
\widehat{\boldsymbol{\mathcal{T}}}_2 &= \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_1,\widehat{\boldsymbol{\Sigma}}_1} - \boldsymbol{P}_{\widehat{\boldsymbol{\mathcal{T}}}_1\boldsymbol{A}_2,\widehat{\boldsymbol{\Sigma}}_2}.
\end{aligned}$$

At last, a third estimator of $\boldsymbol{\Sigma}$, is obtained using the sum $\widehat{\boldsymbol{Q}}_3 = \widehat{\boldsymbol{Q}}_2 + \widehat{\boldsymbol{H}}_3\widehat{\boldsymbol{H}}_3'$ and is given by

$$\mathrm{vec}\widehat{\boldsymbol{\Sigma}}_3 = \boldsymbol{T}^+\left((\boldsymbol{T}^+)'\widehat{\boldsymbol{\Upsilon}}_3'\widehat{\boldsymbol{\Upsilon}}_3\boldsymbol{T}^+\right)^-(\boldsymbol{T}^+)'\widehat{\boldsymbol{\Upsilon}}_3'\mathrm{vec}\widehat{\boldsymbol{Q}}_3, \tag{3.10}$$

where $\widehat{\Upsilon}_3 = (n - r_1)\boldsymbol{I} + (r_1 - r_2)\widehat{\boldsymbol{\mathcal{T}}}_1 \otimes \widehat{\boldsymbol{\mathcal{T}}}_1 + r_2\widehat{\boldsymbol{\mathcal{T}}}_2 \otimes \widehat{\boldsymbol{\mathcal{T}}}_2$.

The estimators $\widehat{\boldsymbol{\Sigma}}_1$, $\widehat{\boldsymbol{\Sigma}}_2$ and $\widehat{\boldsymbol{\Sigma}}_3$ are all consistent for $\boldsymbol{\Sigma}$, however, as $\widehat{\boldsymbol{\Sigma}}_3$ uses all information contained in all residuals, it can arguably be interpreted as a dispersion matrix. The unbiased estimator of the mean structured is given by $\widehat{E[\boldsymbol{X}]} = \widehat{\boldsymbol{M}}_1 + \widehat{\boldsymbol{M}}_2$.

Example 3.6: Example 3.5 continued

Consider again Potthoff & Roy (1964) classical dental data and the model of Example 3.5. Assume that the covariance matrix has a Toeplitz structure, i.e.,

$$\boldsymbol{\Sigma}^{(s)} = \begin{pmatrix} \sigma & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & \sigma & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & \sigma & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & \sigma \end{pmatrix}.$$

Then, the estimate of the structured covariance matrices given by (3.10) is

$$\widehat{\boldsymbol{\Sigma}}_3 = \begin{pmatrix} 5.2128 & 3.2953 & 3.6017 & 2.7146 \\ 3.2953 & 5.2128 & 3.2953 & 3.6017 \\ 3.6017 & 3.2953 & 5.2128 & 3.2953 \\ 2.7146 & 3.6017 & 3.2953 & 5.2128 \end{pmatrix}.$$

For comparison, the MLE computed with Proc Mixed in SAS® (SAS Institute Inc., 2008) is

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(s)} = \begin{pmatrix} 4.9368 & 3.0747 & 3.4559 & 2.2916 \\ 3.0747 & 4.9368 & 3.0747 & 3.4559 \\ 3.4559 & 3.0747 & 4.9368 & 3.0747 \\ 2.2916 & 3.4559 & 3.0747 & 4.9368 \end{pmatrix}.$$

The procedures illustrated in this section were used to build up a flip-flop algorithm that can handle the linear structured $\boldsymbol{\Sigma}$ in the bilinear regression model $\boldsymbol{X} \sim N_{p,q}(\boldsymbol{ABC}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, see (Nzabanita, 2013).

3.5 Estimation in trilinear regression model

The model (3.7) can be written in matrix form using three different modes as

$$\begin{aligned} \boldsymbol{X}_{(1)} &\sim N_{p,qr}(\boldsymbol{AB}_{(1)}(\boldsymbol{D} \otimes \boldsymbol{C})', \boldsymbol{\Sigma}, \boldsymbol{I}_r \otimes \boldsymbol{\Psi}), && (3.11) \\ \boldsymbol{X}_{(2)} &\sim N_{q,pr}(\boldsymbol{CB}_{(2)}(\boldsymbol{D} \otimes \boldsymbol{A})', \boldsymbol{\Psi}, \boldsymbol{I}_r \otimes \boldsymbol{\Sigma}), && (3.12) \\ \boldsymbol{X}_{(3)} &\sim N_{r,pq}(\boldsymbol{DB}_{(3)}(\boldsymbol{C} \otimes \boldsymbol{A})', \boldsymbol{I_r}, \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}). && (3.13) \end{aligned}$$

The maximum likelihood approach can be used to find estimators for $\mathscr{B}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$. However, to find explicit estimators is not possible. Instead, we can establish estimating equations that can be solved iteratively using, for example, the flip-flop algorithm. Observe that the parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are defined up to a positive multiplicative constant because, for example, $\boldsymbol{I}_r \otimes c\boldsymbol{\Psi} \otimes c^{-1}\boldsymbol{\Sigma} = \boldsymbol{I}_r \otimes \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ with $c > 0$. This issue has been discussed by some authors, among others (Dutilleul, 1999, Manceur and Dutilleul, 2013, Srivastava et al., 2008, Singull et al., 2012).

To find estimating equations for parameters we first fix $\boldsymbol{\Psi}$ in (3.11) and find estimating equations for $\boldsymbol{\Sigma}$ and $\boldsymbol{B}_{(1)}$. Secondly, we fix $\boldsymbol{\Sigma}$ in (3.12) and find estimating equations for $\boldsymbol{\Psi}$. This procedure is justified by the fact that the models in (3.11)-(3.13) give the same likelihood function (Nzabanita et al., 2015b). From results in Nzabanita et al. (2015b), the estimating equations for parameters in model (3.7) are given by

$$\begin{aligned} \widehat{\boldsymbol{B}}_{(1)} &= (\boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}_1^{-1}\boldsymbol{X}_{(1)}(\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1} \otimes \widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{C}(\boldsymbol{C}'\widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{C})^{-1}), && (3.14) \\ \boldsymbol{S}_1 &= \boldsymbol{X}_{(1)}(\boldsymbol{I}_r \otimes \widehat{\boldsymbol{\Psi}}^{-1} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-}\boldsymbol{D}' \otimes \widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{C}(\boldsymbol{C}'\widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\widehat{\boldsymbol{\Psi}}^{-1})\boldsymbol{X}_{(1)}', \\ qr\widehat{\boldsymbol{\Sigma}} &= (\boldsymbol{X}_{(1)} - \boldsymbol{A}\widehat{\boldsymbol{B}}_{(1)}(\boldsymbol{D} \otimes \boldsymbol{C})')(\boldsymbol{I}_r \otimes \widehat{\boldsymbol{\Psi}}^{-1})(\boldsymbol{X}_{(1)} - \boldsymbol{A}\widehat{\boldsymbol{B}}_{(1)}(\boldsymbol{D} \otimes \boldsymbol{C})')', && (3.15) \\ pr\widehat{\boldsymbol{\Psi}} &= (\boldsymbol{X}_{(2)} - \boldsymbol{C}\widehat{\boldsymbol{B}}_{(2)}(\boldsymbol{D} \otimes \boldsymbol{A})')(\boldsymbol{I}_r \otimes \widehat{\boldsymbol{\Sigma}}^{-1})(\boldsymbol{X}_{(2)} - \boldsymbol{C}\widehat{\boldsymbol{B}}_{(2)}(\boldsymbol{D} \otimes \boldsymbol{A})')', && (3.16) \end{aligned}$$

where $\boldsymbol{S}_1$ is assumed to be positive definite and $\widehat{\boldsymbol{B}}_{(2)}$ is obtained from $\widehat{\boldsymbol{B}}_{(1)}$ by a proper rearrangement of elements.

These estimating equations are nested and cannot be solved explicitly. To obtain maximum likelihood estimators of $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$ and $\mathscr{B}$ the following iterative algorithm is proposed.

Algorithm 3.1. *1. Choose initial solution $\widehat{\boldsymbol{\Psi}} = \widehat{\boldsymbol{\Psi}}_0$;*

2. *Compute* $\widehat{\boldsymbol{B}}_{(1)}$ *using (3.14);*
3. *Compute* $\widehat{\boldsymbol{\Sigma}}$ *using (3.15);*
4. *Compute* $\widehat{\boldsymbol{\Psi}}$ *using (3.16);*
5. *Repeat steps 2–4 until the convergence criterion is met.*

Usually, there is no prior information that may guide to choose the initial solution. Very often the identity matrix would be enough to get the solutions. The convergence criterion may be based on the rate of change in $\widehat{\boldsymbol{\Psi}} \otimes \widehat{\boldsymbol{\Sigma}}$ and not separately on $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Sigma}}$.

4

Concluding Remarks

THIS chapter is reserved to the summary of contributions of this thesis and suggestions for further research.

4.1 Summary of contributions

In this thesis, the problem of estimating parameters in bilinear and trilinear regression models has been considered. The main theme has been to propose algorithms for estimating unkown parameters when the covariance matrices are structured. The main contributions of the thesis are as follows:

- In Paper A, we studied the extended growth curve model with two terms and a linearly structured covariance matrix. A simple procedure based on the decomposition of the residual space into three orthogonal subspaces and the study of the residuals obtained from projections of observations on these subspaces yielded explicit and consistent estimators of the covariance matrix. An explicit unbiased estimator of the mean was also proposed.

- In Paper B, the extended generalized multivariate analysis of variance with a linearly structured covariance matrix was considered. We showed how to decompose the residual space, the orthogonal complement to the mean space, into $m + 1$ orthogonal subspaces and how to derive explicit consistent estimators of the covariance matrix from the sum of squared residuals obtained by projecting observations on those subspaces. Also an explicit unbiased estimator of the mean was derived. Paper B generalizes results of Paper A.

- In Paper C, the bilinear regression models based on normally distributed random matrix was studied. For these models, the dispersion matrix has the so called Kronecker product structure and they can be used for example to model data with

spatio-temporal relationships. The aim was to estimate the parameters of the model when, in addition, one covariance matrix is assumed to be linearly structured. We proposed a flip-flop like algorithm for estimating parameters and showed that the resulting estimators are consistent.

- In Paper D, the classical growth curve model was extended to a tensor version by assuming a trilinear structure for the mean in the tensor normal model. An algorithm for estimating model parameters was proposed.

4.2 Further research

- The algorithm proposed in Paper A & B yields estimators with good properties like unbiasedness and/or consistency. However, to be more useful their other properties (e.g., their distributions) have to be studied. Also, more rigorous studies on the positive definiteness of the estimates for the covariance matrix is of interest.
- The techniques used in Paper C to show the consistency of estimators can be used to prove the consistency of other estimators based on the flip-flop algorithm.
- Paper D proposed an likelihood-based algorithm for estimating parameters in the 2-fold growth curve model. However, there is no warranty whether it produces global solution nor the solution does not depend on the initial guess. These issues merit a deep study. Validation and other model diagnostic techniques can be developed. The model can be extended in various way, e.g., assuming different profiles (in one or both of two growth directions) among groups. The problem of testing on the mean parameters or the covariance matrix structure is of great interest.
- Finally, application of procedures developed in this thesis to concrete real data sets and a comparison with the existing ones may be useful to show their merits.

Bibliography

Allen, M. P. (1997). The origins and uses of regression analysis. In *Understanding Regression Analysis*, pages 1–5. Springer.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Basser, P. J. and Pajevic, S. (2003). A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. *Medical Imaging, IEEE Transactions on*, 22(7):785–794.

Box, G. E. P. (1950). Problems in the analysis of growth and wear curves. *Biometrics*, 6:362–389.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278.

Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, 64(2):105–123.

Filipiak, K. and von Rosen, D. (2012). On MLEs in an extended multivariate linear growth curve model. *Metrika*, 75(8):1069–1092.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied Longitudinal Analysis*, volume 998. John Wiley & Sons, New York.

Hoff, P. D. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Hu, J. (2010). Properties of the explicit estimators in the extended growth curve model. *Statistics*, 44(5):477–492.

Hu, J., Yan, G., and You, J. (2011). Estimation for an additive growth curve model with orthogonal design matrices. *Bernoulli*, 17(4):1400–1419.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education International, New York.

Kattan, W. M. and Gönen, M. (2008). The prediction philosophy in statistics. *Urologic Oncology: Seminars and Original Investigations*, 26(3):316–319.

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.

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

Klein, D. and Žežula, I. (2009). The maximum likelihood estimators in the growth curve model with serial covariance structure. *Journal of Statistical Planning and Inference*, 139(9):3270–3276.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht.

Lange, N. and Laird, N. M. (1989). The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association*, 84(405):241–247.

Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association*, 83(402):432–440.

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

Lu, N. and Zimmerman, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics and Probability Letters*, 73(4):449–457.

Manceur, A. M. and Dutilleul, P. (2013). Maximum likelihood estimation for the tensor normal distribution: Algorithm, minimum sample size, and empirical bias and dispersion. *Journal of Computational and Applied Mathematics*, 239:37–49.

Nzabanita, J. (2013). Multivariate linear models with kronecker product and linear structures on the covariance matrices. In *Proceedings, JSM 2013-IMS*, pages 1582–1588. Alexandria, VA: American Statistical Association.

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.

Nzabanita, J., von Rosen, D., and Singull, M. (2015a). Extended GMANOVA model with a linearly structured covariance matrix. *Linköping University Electronic Press, LiTH-MAT-R-2015/07-SE.*

Nzabanita, J., von Rosen, D., and Singull, M. (2015b). Maximum likelihood estimation in the tensor normal model with a structured mean. *Linköping University Electronic Press, LiTH-MAT-R-2015/08-SE.*

Ohlson, M., Ahmad, M. R., and von Rosen, D. (2013). The multilinear normal distribution: Introduction and some basic properties. *Journal of Multivariate Analysis*, 113:37–47.

Ohlson, M., Andrushchenko, Z., and von Rosen, D. (2011). Explicit estimators under m-dependence for a multivariate normal distribution. *Annals of the Institute of Statistical Mathematics*, 63(1):29–42.

Ohlson, M. and von Rosen, D. (2010). Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices. *Journal of Multivariate Analysis*, 101(5):1284–1295.

Olkin, I. and Press, S. G. (1969). Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, 40(4):1358–1373.

Pothoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51:313–326.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52:447–458.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models.* Springer-Verlang, New York.

Roy, A. and Khattree, R. (2005). On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Statistical Methodology*, 2(4):297–306.

Roy, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.

SAS Institute Inc. (2008). *SAS/STAT® 9.2 User's Guide: The MIXED Procedure (Book Excerpt).* Cary, NC: SAS Institute Inc.

Singull, M., Ahmad, M. R., and von Rosen, D. (2012). More on the Kronecker structured covariance matrix. *Communications in Statistics-Theory and Methods*, 41(13-14):2512–2523.

Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York, USA.

Srivastava, M. S. and Singull, M. (2015). Testing some covariance structures under a growth curve model in high dimension. *Linköping University Electronic Press, LiTH-MAT-R–2015/03*.

Srivastava, M. S., von Rosen, T., and von Rosen, D. (2008). Models with a Kronecker product covariance structure: estimation and testing. *Mathematical Methods of Statistics*, 17(4):357–370.

Szatrowski, T. H. (1982). Testing and estimation in the block compound symmetry problem. *Journal of Educational Statistics*, 7(1):3–18.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. *Contributions to mathematical psychology*, pages 109–127.

Verbyla, A. P. and Venables, W. N. (1988). An extension of the growth curve model. *Biometrika*, 75(1):129–138.

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

Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 19(4):447–473.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17(3):257–281.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20(A):32–52.

Part II

Papers

Papers