New Residuals in Multivariate Bilinear Models

Testing hypotheses, diagnosing models and validating model assumptions

Jemila Seid Hamid
Faculty of Natural Resources and Agricultural Sciences
Department of Biometry and Engineering
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2005

Acta Universitatis Agriculturae Sueciae
2005:83

ISSN 1652-6880
ISBN 91-576-6982-1

Tryck: SLU Service/Repro, Uppsala 2005

Abstract

Seid Hamid, J. 2005. *New Residuals in Multivariate Bilinear Models - Testing Hypotheses, diagnosing models and validating model assumptions.* Doctoral dissertation. ISSN 1652-6880. ISBN 91-576-6982-1

New residuals taking the bilinear structure into account are defined for the Extended Growth Curve Model. It is shown that the ordinary residuals are defined by projecting the observation matrix on the space orthogonal to the one generated by the design matrices which turn out to be the sum of two tensor product spaces. The space on which the ordinary residuals are defined is then decomposed into four orthogonal spaces and new residuals are defined by projecting the observation matrix on the resulting four spaces. The information contained in them is used to check the adequacy of the model and to check if there are extreme observations.

Tests are proposed for the Growth and Extended Growth Curve models which turn out to be functions of appropriate residuals. It is shown that the distributions of the tests under the null hypotheses are independent of the unknown covariance matrix. The distributions are difficult to find, however, two suggestions are made to tackle this problem.

We consider a conditional approach and discuss why it is appropriate in our situation. Moreover, it is shown that the distribution of the conditional test under the null and alternative hypotheses can be written as sums of independent central and non-central chi-square random variables, respectively. However, the exact distributions, which are available as an infinite series, are too complicated to be used in practice and approximations are needed. We use Satterthwaite's approximation to find the critical point.

Under the alternative, an approximation similar to that of Satterthwaite is provided for obtaining an approximate power. However, our approach is different and new ideas are utilized to get the approximation. Numerical examples are given to illustrate the results.

Keywords: ancillary statistics, conditional test, decomposition of linear spaces, estimated likelihood, extended growth curve model, growth curve model, restricted likelihood,
Satterthwaite approximation, tensor product of linear spaces.

Author's address: Jemila Seid Hamid, Department of Biometry and Engineering, SLU, P.O.Box 7032, SE-750 07 Uppsala, Sweden. E-mail: Jemila@bt.slu.se

Contents

Chapter 1

Introduction

1.1 Residuals and statistical diagnosis

Modelling is an important part of understanding different phenomena around us. Statistical models are used in analyzing data, making inference, making predictions and important decisions. One can follow two approaches in constructing a statistical model. The first approach is to assume a predefined model and check if the assumed model describes the situation well. The second approach is to find a model based on data by requiring as few assumptions as possible. In either case, we need to investigate the adequacy of the proposed and fitted models. Moreover, the models constructed using the first approach are usually based on several assumptions that might be unrealistic and need to be treated carefully.

An important part of modelling is diagnosing the flaws in the models as well as checking if the assumptions are true or rather check if the data violates the assumptions. Furthermore, it could also be important to see if there are outliers and/or influential observations. The most common and natural approach for validating models and model assumptions is by looking at the residuals.

In most model fitting problems, whether they are linear or non linear, residuals play an important role in diagnosing the model or the model assumptions. They are also used to detect outliers and/or influential observations in the data. Residuals are the part of the data which is left unexplained by the fitted model. They are defined as the difference between the observed and fitted values:

$$e_i = y_i - \hat{y}_i.$$

In univariate model fitting problems, the resulting residuals are also univariate and it is relatively easy to examine them. In fact, there has been many discussions regarding the residuals in such models and many different types of residuals have been defined and studied, see for example Sen & Srivastava (1990) and Draper & Smith (1998). However, in multivariate models, the residuals are also multivariate and it is relatively difficult to diagnose the model through them. It is not also obvious as to how to use them to detect extreme observations. There has been few studies regarding the residuals in multivariate models although most tests which have been proposed, such as the likelihood ratio test, are in some way functions of the residuals. For example, see Srivastava & Khatri (1979) for some proposed tests.

Ordinary residuals have also been used to assess the adequacy of the fitted models and to check for the presence of outliers in longitudinal analysis (Fitzmaurice et al., 2004). However, under analysis of such data, we believe that ordinary residuals have more structure due to the within and between individual assumptions. The same is true in the Growth Curve Model because of the bilinear nature of the model. For this model, the ordinary residuals which are defined as above consist of two parts. One part gives information about the between individual structure, which we suggest should be used in assessing between individual model assumptions, such as the normality assumption. Moreover, this part of the residuals can be used to check the presence of extreme observations in the data. The second part gives information about the within individual structure which in fact is a part of the residual that tells us if the estimated model fits the data. This part can be shown to be the difference between the observed and estimated means.

Therefore, when dealing with the Growth and Extended Curve models one should be careful when examining the ordinary residuals. For example, the two parts mentioned in the previous paragraph may happen to have opposite signs and could cancel with each other and give an impression that the model fits the data well when it in fact is otherwise. This is also true for other models used to analyze repeated measures and longitudinal data. This explains why there is a need to define other residuals which take the bilinear structure in the model into consideration. This has been done by von Rosen (1995b) where he defined and discussed three residuals. It was indicated that each of the residuals provide valuable information about different aspects of the model.

1.2 Aim and outline of the thesis

Two GMANOVA models, namely the Growth Curve (GC) model and a special case of the Extended Growth Curve (EGC) model are of interest. Inspired from von Rosen's residuals for the GC model, residuals for the EGC model taking the bilinear structure into consideration are defined in Paper I. The information contained in these residuals and that of von Rosen's for the GC model are investigated. If defining and understanding these residuals, it is also important to come up with some ideas concerning how one can make use of these residuals and the information contained in them to validate the statistical models and validate model assumptions.

One can simply use the standard errors as cutoff points and see if the residuals are small enough to decide if the estimated model fits the data. Moreover, one can also see if there are outliers and/or influential observations since they tend to have large residual values. The standard errors for the residuals at different time points can be obtained from the estimated dispersion matrices which are given in Paper I.

The other approach is through hypothesis testing, i.e. by constructing certain statistics for testing different hypotheses regarding the models. There have been some studies regarding hypotheses tests in the GC and ECG models. We refer to the papers by Khatri (1966), Fujikoshi (1974) and Kariya (1978) for discussions about the likelihood ratio test and other tests such as the trace test for the GC model. However, there have not been any studies what so ever connecting the tests in these models with the residuals. Moreover, as to our knowledge, no one has ever tried to study residuals in such bilinear models as well as use them to check the adequacy of the models or check if the assumptions are violated although it is the most convenient and natural way of doing it.

This is the topic of Paper II where we have proposed and studied tests for common hypotheses arising in the GC and EGC models. The tests are constructed using restricted likelihood followed by estimated likelihood approaches. Moreover, we show that the tests constructed are functions of appropriate residuals and this enables us to understand the structure of the tests.

If using the tests proposed in Paper II in practice, one needs to find the critical points. The distribution of most tests defined for MANOVA and GMANOVA models are difficult to deal with. This is also true for our tests. However, two suggestions are made in Paper II to overcome this problem. In Paper III we present and discuss the second approach which is based on conditioning by an ancillary statistics.

The thesis consists of seven chapters. In the second and third chapters the GC and EGC models, respectively, are introduced briefly. In the fourth chapter we give a brief background about conditional tests and we discuss a conditional approach for the GC model. Some technical results developed and used in the papers are presented in the fifth chapter. Short summary and main results about the three papers are given in Chapter 6. Finally, a discussion together with future research is presented in the seventh chapter.

The following three papers which are the basis for the thesis are given in the Appendix. The papers will be refereed to by their Roman numerals.

I. Seid Hamid, J. & von Rosen, D. (2005) Residuals in the Extended Growth Curve Model. (Accepted by Scandinavian Journal of Statistics)

II. Seid Hamid, J. & von Rosen, D. (2005) Hypothesis Testing via Residuals in two GMANOVA models. (Submitted)

III. Seid Hamid, J. & von Rosen, D. (2005) An Approximate Critical Point for a Test in the Growth Curve Model: A Conditional Approach. (Submitted)

Chapter 2

The Growth Curve model

2.1 Introduction

The Growth Curve Model was introduced by Potthoff & Roy (1964) and subsequently studied among others by Rao (1965), although the first paper considering growth curves was presented by Wishart (1938). Discrimination between growth curves was discussed by Burnaby (1966). Since then different aspects of the model has been considered by many authors including Khatri (1966), Gleser & Olkin (1970) and von Rosen (1989). There is a book by Kshirsagar & Smith (1995) about growth curves which are the principal applications of the GC model. Some discussions about the Potthoff & Roy model are also given in the book. Statistical diagnostics for the model is discussed in the book by Pan & Fang (2002). The book also gives an excellent background about the model with illustrations using practical examples. For further discussions about the model, see also Kollo & von Rosen (2005). In this chapter we give a brief introduction about the model and present some results which are used in the papers.

Definition 2.1. Let $\mathbf{X} : p \times n$ and $\mathbf{B} : q \times k$ be the observation and parameter matrices, respectively, and let $\mathbf{A} : p \times q$ and $\mathbf{C} : k \times n$ be the within and between individual design matrices, respectively. Suppose that $q \leq p$ and $\rho(\mathbf{C}) + p \leq n$, where $\rho(\bullet)$ denotes the rank of a matrix. The GC model is given by

$$\mathbf{X} = \mathbf{ABC} + \mathbf{E}, \tag{2.1}$$

where the columns of $\mathbf{E}$ are assumed to be independently p-variate normally distributed with mean zero and an unknown positive definite covariance matrix $\mathbf{\Sigma}$.

The above model is sometimes denoted by MLNM($\mathbf{ABC}$), where MLNM stands for Multivariate Linear Normal Model. We have used similar notations in the papers.

It is important to note here that if $\mathbf{A} = \mathbf{I}$ the GC model reduces to the classical MANOVA model. Note also that the matrix $\mathbf{C}$ in Definition 2.1 is the same design matrix as in univariate and classical multivariate linear models.

The GC model is a generalized multivariate analysis of variance (GMANOVA) model and has many applications and may arise in many different situations. One of its principal applications being in the analysis of growth curves which are applied extensively in biostatistics, medical research and epidemiology. The model also plays important roles in the study of repeated measurements and longitudinal analysis.

Suppose, for example, that we have different groups where repeated observations are taken on a given experimental unit in each group. If these observations can be associated with some continuous variable, such as time and temperature, then they may form a response curve, where the curve for the ith individual can be described as

$$b_{o,i} + b_{1,i}t + ... + b_{p-1,i}t^{p-1} + \epsilon_i.$$

It is assumed that measurements are taken at the same time points and that they are assumed to have the same covariance matrix. Moreover, it is important to note that all the individuals in all the groups are assumed to have polynomial growth curves of the same degree. If the degrees of the polynomials are different, they will be handled by the Extended Growth Curve Model which will be introduced in the next section.

The following examples show how the GC model may arise and illustrate the observation, parameter and design matrices involved in the model. For more examples, you can see Srivastava & Carter (1983), Kshirsagar & Smith (1995) and Pan & Fang (2002).

Example 1 (Potthoff & Roy Dental Data)

Dental measurements on eleven girls and sixteen boys at four different ages (8, 10, 12, 14) were taken. Each measurement is the distance, in millimeters, from the center of pituitary to pteryo-maxillary fissure. Suppose linear growth curves describe the mean growth for both the girls and the boys. Then we may use the Growth Curve model where the observation, parameter and design matrices are given by

$$\mathbf{X}_{4\times 27} = \begin{pmatrix} 21 & 21 & 20.5 & 23.5 & 21.5 & 20 & 21.5 & 23 & 20 & 16.5 \\ 24.5 & 26 & 21.5 & 23 & 20 & 25.5 & 24.5 & 22 & 24 & 23 \\ 27.5 & 23 & 21.5 & 17 & 22.5 & 23 & 22, & & & \\ 20 & 21.5 & 24 & 24.5 & 23 & 21 & 22.5 & 23 & 21 & 19 \\ 25 & 25 & 22.5 & 22.5 & 23.5 & 27.5 & 25.5 & 22 & 21.5 & 20.5 \\ 28 & 23 & 23.5 & 24.5 & 25.5 & 24.5 & 21.5, & & & \\ 21.5 & 24 & 24.5 & 25 & 22.5 & 21 & 23 & 23.5 & 22 & 19 \\ 28 & 29 & 23 & 24 & 22.5 & 26.5 & 27 & 24.5 & 24.5 & 31 \\ 31 & 23.5 & 24 & 26 & 25.5 & 26 & 23.5, & & & \\ 23 & 25.5 & 26 & 26.5 & 23.5 & 22.5 & 25 & 24 & 21.5 & 19.5 \\ 28 & 31 & 26.5 & 27.5 & 26 & 27 & 28.5 & 26.5 & 25.5 & 26 \\ 31.5 & 25 & 28 & 29.5 & 26 & 30 & 25 & & & \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \quad \text{and} \quad \mathbf{C}_{2\times 27} = \begin{pmatrix} \mathbf{1}_{11} & \mathbf{0}_{16} \\ \mathbf{0}_{16} & \mathbf{1}_{16} \end{pmatrix}$$

where the $\mathbf{1}_{11}$ and $\mathbf{0}_{16}$ in the first row indicate that there are 11 1's and 16 0's. Note also how we have presented $\mathbf{X}$, i.e. we have used commas to separate the rows.

Example 2

Consider the Potthoff & Roy data. Now suppose that a quadratic growth curve is to be fitted to both the girls and the boys. The observation matrix, $\mathbf{X}$, and the between individual design matrix, $\mathbf{C}$, remain the same. However, the parameter and the within individual design matrices become

$$\mathbf{B} = \begin{pmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 1 & 8 & 64 \\ 1 & 10 & 100 \\ 1 & 12 & 144 \\ 1 & 14 & 196 \end{pmatrix}.$$

Potthoff & Roy (1964) in their paper gave an estimator for the parameter of the model based on a matrix $\mathbf{G}$. The estimator is given by

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{G}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1},$$

where $\mathbf{A}$ and $\mathbf{C}$ were assumed to have full rank. However, the problem with the method utilized by them was that the choice of $\mathbf{G}$ is arbitrary. One possible choice suggested by them was that $\mathbf{G} = \mathbf{I}$, however, they also mentioned that if some information about the dispersion matrix, $\mathbf{\Sigma}$ is available, then $\mathbf{G} = \mathbf{I}$ may not be the best choice. In fact, they indicated that the more $\mathbf{G}$ differs from $\mathbf{\Sigma}$, the worse the power of the tests will be and the wider the confidence intervals will be although they mentioned that the estimators remain unbiased.

Generalized least square estimation and the admissibility of the estimates are discussed in Pan & Fang (2002). They have also shown that the generalized least square estimators are best linear unbiased in the sense of a matrix loss function.

We consider estimators obtained by using the maximum likelihood approach and will be presented shortly. It is possible to show that the maximum likelihood estimator of the mean structure is obtained by projecting the observation matrix on the space generated by the two design matrices. This is also true for the EGC model. For the later model, we have shown that the space is a sum of two tensor product spaces and used this fact to get the space on which the ordinary residuals are defined.

Khatri (1966) provided the maximum likelihood estimator for the parameter matrix $\mathbf{B}$ which is given by,

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1},$$

where $\mathbf{S} = \mathbf{X}(\mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C})\mathbf{X}'$ and it was assumed that $\mathbf{A}$ and $\mathbf{C}$ are of full rank. The maximum likelihood estimator under the general situation, i.e., without assuming full rank conditions, is given by

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-} + (\mathbf{A}')^{o}\mathbf{Z}_1' + \mathbf{A}'\mathbf{Z}_2\mathbf{C}^{o'},$$

where, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are arbitrary matrices, $\mathbf{C}^{o}$ is a matrix of full rank spanning the orthogonal complement of the linear space generated by the columns of $\mathbf{C}$, and $\mathbf{G}^{-}$ denotes an arbitrary generalized inverse in the sense of $\mathbf{G}\mathbf{G}^{-}\mathbf{G} = \mathbf{G}$. For different methods of maximizing the likelihood function, we refer to Srivastava & Khatri (1979), von Rosen (1989) and the book by Kollo & von Rosen

(2005). Due to the arbitrariness of the vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$ it is evident that the maximum likelihood estimator is not unique. However, it is important to note that

$$\mathbf{A}\hat{\mathbf{B}}\mathbf{C} = \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}$$

is always unique. For more details about maximum likelihood based inference in the GC model and restricted likelihood estimation we refer to Pan & Fang (2002).

2.2 Residuals in the Growth Curve model

Consider the classical multivariate linear (MANOVA) model which is given by

$$\mathbf{X} = \mathbf{B}\mathbf{C} + \mathbf{E},$$

where $\mathbf{X} : p \times n$, $\mathbf{B} : p \times k$ and $\mathbf{C} : k \times n$ are the observation, parameter and design matrices, respectively. In univariate and classical MANOVA cases residuals are obtained by projecting $\mathbf{X}$ on the space orthogonal to $\mathcal{C}(\mathbf{C}')$, the column space of $\mathbf{C}'$, which is the space generated by the design matrix, i.e. $\mathbf{R} = \mathbf{X}(\mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C})$, where $\mathbf{R}$ stands for residuals. However, in the GC model the space has a bilinear structure which generates a tensor product $\mathcal{C}(\mathbf{C}') \otimes \mathcal{C}_{\mathbf{S}}(\mathbf{A})$, where the $\mathbf{S}$ in $\mathcal{C}_{\mathbf{S}}$ indicates that the inner product is defined with the help of $\mathbf{S}^{-1}$, i.e. $< \mathbf{x}, \mathbf{y} >= \mathbf{x}'\mathbf{S}^{-1}\mathbf{y}$. If there is no subscript as in $\mathcal{C}(\mathbf{C}')$, the standard inner product is assumed. Therefore, in the GC model ordinary residuals are obtained by projecting the observation matrix on the orthogonal complement to $\mathcal{C}(\mathbf{C}') \otimes \mathcal{C}_{\mathbf{S}}(\mathbf{A})$.

However, it is not obvious to understand and interpret the residuals due to the complexity of the space generated by the design matrices which is the result of the bilinear nature of the model. It is shown by von Rosen (1995b) that the residuals in fact are different from those in univariate and MANOVA models in the sense that they contain both the within and between individual information.

Residuals taking the bilinear structure in the model into account were defined by von Rosen (1995b). This work was the first paper which takes the bilinear structure into consideration, although diagnostic tools for assessing influential observations in the model has been discussed by, for instance, Liski (1991), Pan & Fang (1995, 1996) and von Rosen (1995a). However, we believe that understanding the residuals in the model is an important step in developing diagnostic tools for bilinear models such as the GC and EGC models. Three residuals were defined by von Rosen (1995b) by projecting $\mathbf{X}$ on the resulting three spaces obtained from decomposing the space orthogonal to the one generated by the design matrices. The residuals are given bellow:

$$\begin{aligned}
\mathbf{R}_{g1} &= \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}(\mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}),\\
\mathbf{R}_{g2} &= (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1})\mathbf{X}(\mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}),\\
\mathbf{R}_{g3} &= (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1})\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}.
\end{aligned}$$

Note that we put g in the subscript so that we can distinguish these residuals with those for the EGC model defined in Paper I.

Each of the above residuals contain important information about the model. The sum of the first two residuals, for example, represents the difference between the observations and the mean and therefore can be used to check the between individual variations. The last residual is the difference between the observed and estimated mean and therefore can be used to check within individual variations. In other words, this gives information about how the estimated growth curve fits the data. More about the residuals and their properties can be found in von Rosen (1995b). His results are also extended to the special case of the EGC model in Paper I. However, the method utilized in our paper is more natural and has an advantage that it can be used to extend the results to the more general model. Furthermore, interpretations as to what information is contained in the residuals and how one can use that information to validate the model and model assumptions are clearly presented.

Chapter 3

Extended Growth Curve model

3.1 Introduction

In this chapter we briefly introduce the Extended Growth Curve Model together with some important results which are used in the papers. The model is given in Definition 3.1 below and was introduced by von Rosen (1989) although a canonical form of the model was considered by Banken (1984). A special case of the model was considered by Srivastava & Khatri (1979). The EGC model is a special case of a more general model which in econometrics literature is known as a multivariate seemingly unrelated regression (SUR) model. The model without the nested subspace condition was considered by Verbyla & Venables (1988) under a different name; sum of profiles model. The model is sometimes referred to as MLNM($\sum_{i=1}^{m} \mathbf{A}_i\mathbf{B}_i\mathbf{C}_i$). This notation is also used in the papers.

Definition 3.1. Let $\mathbf{X} : p \times n$, $\mathbf{A}_i : p \times q_i$, $\mathbf{B}_i : q_i \times k_i$ and $\mathbf{C}_i : k_i \times n$. Suppose that $q_i \leq p$, $\rho(\mathbf{C}_1) + p \leq n$ and $C(\mathbf{C}_m') \subseteq C(\mathbf{C}_{m-1}') \subseteq ... \subseteq C(\mathbf{C}_1')$, where $\rho(\bullet)$ and $C(\bullet)$ represent the rank and column space of a matrix, respectively. Then the Extended Growth Curve (EGC) model is given by,

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

where the columns of $\mathbf{E}$ are assumed to be independently distributed as a p-variate normal distribution with mean zero and an unknown dispersion matrix $\mathbf{\Sigma}$.

Verbyla & Venables in their paper provided an algorithm for obtaining maximum likelihood estimators. They also gave some examples to illustrate how the model may arise and mentioned some remarks as to the applications of the model. Explicit forms of the maximum likelihood estimators of the model parameters under the nested subspace condition between the design matrices were given by von Rosen (1989). The estimators are obtained recursively and are given in the following theorem which is due to von Rosen (1989).

Let us first give the following notations which make the formulas in the theorem a bit shorter.

$$\mathbf{Q}_{\mathbf{C}_i} = \mathbf{C}_i'(\mathbf{C}_i\mathbf{C}_i')^-\mathbf{C}_i, \tag{3.1}$$
$$\mathbf{Q}_{\mathbf{A}_i} = (\mathbf{A}_i'\mathbf{P}_i'\mathbf{S}_i^{-1}\mathbf{P}_i\mathbf{A}_i)^-\mathbf{A}_i'\mathbf{P}_i'\mathbf{S}_i^{-1}. \tag{3.2}$$

Theorem 3.2. *Consider the Extended Growth Curve Model given in Definition 3.1. Representations of maximum likelihood estimators of $\boldsymbol{B}_i$ and $\boldsymbol{\Sigma}$ are given by*

$$\hat{\boldsymbol{B}}_r = \boldsymbol{Q}_{\boldsymbol{A}_r}\{\boldsymbol{X} - \sum_{i=r+1}^{m} \boldsymbol{A}_i\hat{\boldsymbol{B}}_i\boldsymbol{C}_i\}\boldsymbol{C}_r'(\boldsymbol{C}_r\boldsymbol{C}_r')^- + (\boldsymbol{A}_r'\boldsymbol{P}_r')^o\boldsymbol{Z}_{r1} + \boldsymbol{A}_r'\boldsymbol{P}_r'\boldsymbol{Z}_{r2}\boldsymbol{C}_r^{o\prime}$$

$$\begin{aligned} n\hat{\boldsymbol{\Sigma}} &= \{\boldsymbol{X} - \sum_{i=r+1}^{m} \boldsymbol{A}_i\hat{\boldsymbol{B}}_i\boldsymbol{C}_i\}\{\boldsymbol{X} - \sum_{i=r+1}^{m} \boldsymbol{A}_i\hat{\boldsymbol{B}}_i\boldsymbol{C}_i\}' \\ &= \boldsymbol{S}_m + \boldsymbol{P}_{m+1}\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{C}_m}\boldsymbol{X}'\boldsymbol{P}_{m+1}', \end{aligned}$$

where $\boldsymbol{Z}_{r1}$ and $\boldsymbol{Z}_{r1}$ are arbitrary matrices, and

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

where $\boldsymbol{Q}_{\boldsymbol{C}_i}$ and $\boldsymbol{Q}_{\boldsymbol{A}_i}$ are as given in (3.1) and (3.2), respectively.

In our papers we consider the GC model and the EGC model when m=2. However, we believe that the methods applied could be utilized to define residuals in the general case and similar approaches may be used to establish and study different properties of the residuals. Tests based on the residuals for the general model could also be constructed and approximate distributions for the tests could be obtained using the methods presented in Papers I and II. Let us now define the EGC model with m=2 and give some examples to show how the model may arise and illustrate the matrices involved in the model. Following von Rosen (1989), we denote the model by MLNM($\mathbf{A}_1\mathbf{B}_1\mathbf{C}_1 + \mathbf{A}_2\mathbf{B}_2\mathbf{C}_2$). This notation is also used in Papers I-III.

Definition 3.3. Let $\mathbf{X} : p \times n$, $\mathbf{A}_1 : p \times q_1$,$\mathbf{A}_2 : p \times q_2$, $\mathbf{B}_1 : q_1 \times k_1$, $\mathbf{B}_2 : q_2 \times k_2$ $\mathbf{C}_1 : k_1 \times n$ and $\mathbf{C}_2 : k_2 \times n$. Suppose that $q_1, q_2 \leq p$, $\rho(\mathbf{C}_1) + p \leq n$ and $C(\mathbf{C}_2')\subseteq C(\mathbf{C}_1')$, where $\rho(\bullet)$ and $C(\bullet)$ represent the rank and column space of a matrix, respectively. Then the MLNM($\mathbf{A}_1\mathbf{B}_1\mathbf{C}_1 + \mathbf{A}_2\mathbf{B}_2\mathbf{C}_2$) is given by,

$$\mathbf{X} = \mathbf{A}_1\mathbf{B}_1\mathbf{C}_1 + \mathbf{A}_1\mathbf{B}_1\mathbf{C}_1 + \mathbf{E},$$

where the columns of $\mathbf{E}$ are assumed to be independently distributed as a p-variate normal distribution with mean zero and an unknown dispersion matrix $\mathbf{\Sigma}$.

The MLNM($\mathbf{A}_1\mathbf{B}_1\mathbf{C}_1 + \mathbf{A}_2\mathbf{B}_2\mathbf{C}_2$) can be used in the analysis of growth curves when the groups have polynomial growth of different degrees. This model also arises when we have the GC model with a linear restriction on the parameters. The following example illustrates the matrices involved in the model.

Example 3

Consider the Potthoff & Roy data. Suppose now that linear and quadratic growth curves are to be fitted for the girls and boys, respectively. Suppose also that the growth curve for the boys has a linear component. Then, the observation matrix will be the same as the matrix $\mathbf{X}$ given in Example 1, $\mathbf{C}_1 = \mathbf{C}$, $\mathbf{A}_1 = \mathbf{A}$, $\mathbf{B}_1 = \mathbf{B}$, where $\mathbf{C}$, $\mathbf{A}$, $\mathbf{B}$ are the matrices given in Example 1, $\mathbf{C}_2$ is a row vector with 27 elements where the first 11 elements equal zero and the rest equal 1, $\mathbf{B}_2 = b_{22}$ and $\mathbf{A}_2'$= (64 100 144 196).

3.2 Some important results

In this section we give some important results and definitions which are of particular interest. Let us start by giving a theorem due to von Rosen (1989). The special case of the theorem which is presented in Corollary 3.5 is the basis for the results given in Paper I.

Theorem 3.4. *Suppose conditions in Definition 3.1 are satisfied and let $\hat{\boldsymbol{B}}_i$ be the maximum likelihood estimator of $\boldsymbol{B}_i$. Then*

$$\boldsymbol{P}_r \sum_{i=r}^{m} \boldsymbol{A}_i \hat{\boldsymbol{B}}_i \boldsymbol{C}_i = \sum_{i=r}^{m} (\boldsymbol{I} - \boldsymbol{T}_i)\boldsymbol{X}\boldsymbol{C}_i'(\boldsymbol{C}_i\boldsymbol{C}_i')^{-}\boldsymbol{C}_i, \tag{3.3}$$

where $\boldsymbol{P}_r$, $\boldsymbol{T}_i$, $\boldsymbol{S}_i$ and $\boldsymbol{K}_j$ are as given in Theorem 3.2.

Note that if we set r=1, then the expression on the left hand side of (3.3) reduces to $\mathbf{P}_1 \sum_{i=1}^{m} \mathbf{A}_i\hat{\mathbf{B}}_i\mathbf{C}_i = \sum_{i=1}^{m} \mathbf{A}_i\hat{\mathbf{B}}_i\mathbf{C}_i$ which is the estimated mean structure. As mentioned above we consider the special case of the EGC model which is given in Definition 3.3. For this model the above theorem with r=1 reduces to the following result.

Corollary 3.5. *Suppose the conditions in Definition 3.3 are satisfied and let $\hat{\boldsymbol{B}}_1$ and $\hat{\boldsymbol{B}}_2$ be the maximum likelihood estimators of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, respectively. Then, the estimated mean structure for the model is given by*

$$\boldsymbol{A}_1\hat{\boldsymbol{B}}_1\boldsymbol{C}_1 + \boldsymbol{A}_2\hat{\boldsymbol{B}}_2\boldsymbol{C}_2 = (\boldsymbol{I} - \boldsymbol{T}_1)\boldsymbol{X}\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-}\boldsymbol{C}_1 + (\boldsymbol{I} - \boldsymbol{T}_2)\boldsymbol{X}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^{-}\boldsymbol{C}_2,$$

where

$$\begin{aligned}
\boldsymbol{T}_1 &= \boldsymbol{I} - \boldsymbol{A}_1(\boldsymbol{A}_1'\boldsymbol{S}_1^{-1}\boldsymbol{A}_1)^{-}\boldsymbol{A}_1'\boldsymbol{S}_1^{-1},\\
\boldsymbol{T}_2 &= \boldsymbol{I} - \boldsymbol{T}_1\boldsymbol{A}_2(\boldsymbol{A}_2'\boldsymbol{T}_1'\boldsymbol{S}_2^{-1}\boldsymbol{T}_1\boldsymbol{A}_2)^{-}\boldsymbol{A}_2'\boldsymbol{T}_1'\boldsymbol{S}_2^{-1},\\
\boldsymbol{S}_1 &= \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-}\boldsymbol{C}_1)\boldsymbol{X}',\\
\boldsymbol{S}_2 &= \boldsymbol{S}_1^{-1} + \boldsymbol{T}_1\boldsymbol{X}\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-}\boldsymbol{C}_1(\boldsymbol{I} - \boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^{-}\boldsymbol{C}_2)\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-}\boldsymbol{C}_1\boldsymbol{X}'\boldsymbol{T}_1'.
\end{aligned}$$

When considering multivariate models, it is sometimes much easier to deal with the vectorized form of the models. In the first paper we use the *vec*-operator in finding the spaces generated by the design matrices. It is also used when obtaining the covariance matrices between the residuals and dispersion matrices. Next we give the definition of the *vec*-operator.

Definition 3.6. Let $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_q)$ be a $p \times q$ matrix, where $\mathbf{a}_i$, i = 1,2,...,q, is the i-th column vector. The *vec*-operator $vec(\bullet)$ is an operator from $\mathbb{R}^{p\times q}$ to $\mathbb{R}^{pq}$ defined by

$$vec\mathbf{A} = (\mathbf{a}_1', \mathbf{a}_2', ..., \mathbf{a}_q')'. \tag{3.4}$$

In multivariate model fitting problems, the resulting residuals are also multivariate and it is of interest to obtain covariance and dispersion matrices for the residuals. Therefore, it is important to give definitions of covariance and dispersion matrices, which will be presented in the next definition.

Definition 3.7. Let $\mathbf{U}$ and $\mathbf{V}$ be two matrices. The covariance matrix between them is defined by

$$Cov(\mathbf{U}, \mathbf{V}) = E[vec\mathbf{U}vec'\mathbf{V}] - E[vec\mathbf{U}][vec'\mathbf{V}]$$

and the dispersion matrix, $D[\mathbf{U}] = Cov[\mathbf{U},\mathbf{U}]$, where $vec(\bullet)$ is the linear operator given in (3.4) and $vec'(\bullet)$ is the transpose of $vec(\bullet)$.

General cases of the following results which were established and proved by von Rosen (1990) are used in Paper I in obtaining the dispersion matrices for the residuals and the covariance matrices between them.

Theorem 3.8. *Let $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ be as given in Corollary 3.5. Suppose that the conditions in Definition 3.3 are satisfied. Then,*

i) $E[\boldsymbol{I} - \boldsymbol{T}_1'] = \boldsymbol{\Sigma}^{-1}\boldsymbol{L}_1$,

ii) $E[(\boldsymbol{I} - \boldsymbol{T}_1)\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{T}_1'))] = \alpha_1\boldsymbol{L}_1$,

iii) $E[(\boldsymbol{I} - \boldsymbol{T}_2)\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{T}_2')] = \alpha_3\boldsymbol{L}_1 + \alpha_2\boldsymbol{L}_2$,

iv) $E[(\boldsymbol{I} - \boldsymbol{T}_1)\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{T}_2')] = \alpha_3\boldsymbol{L}_1 + \boldsymbol{L}_2$,

where

$$\begin{aligned}
\boldsymbol{L}_1 &= \boldsymbol{A}_1(\boldsymbol{A}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_1)^-\boldsymbol{A}_1' \\
\boldsymbol{L}_2 &= \boldsymbol{\Sigma}\boldsymbol{A}_1^o(\boldsymbol{A}_1^{o'}\boldsymbol{\Sigma}\boldsymbol{A}_1^o)^{-1}\boldsymbol{A}_1^{o'}\boldsymbol{A}_2(\boldsymbol{A}_2'\boldsymbol{A}_1^o(\boldsymbol{A}_1^{o'}\boldsymbol{\Sigma}\boldsymbol{A}_1^o)^{-1}\boldsymbol{A}_1^{o'}\boldsymbol{A}_2)^-\boldsymbol{A}_2'\boldsymbol{A}_1^o \\
&\quad \times (\boldsymbol{A}_1^{o'}\boldsymbol{\Sigma}\boldsymbol{A}_1^o)^{-1}\boldsymbol{A}_1^{o'}\boldsymbol{\Sigma}. \\
\alpha_1 &= \frac{n - \rho(\boldsymbol{C}_1) - 1}{n - \rho(\boldsymbol{C}_1) - p + \rho(\boldsymbol{A}_1) - 1}, \\
\alpha_2 &= \frac{n - \rho(\boldsymbol{C}_2) - 1}{n - \rho(\boldsymbol{C}_2) - p + \rho(\boldsymbol{A}_1 : \boldsymbol{A}_2) - \rho(\boldsymbol{A}_1) - 1}, \\
\alpha_3 &= \frac{(n - \rho(\boldsymbol{C}_1) - 1)(\rho(\boldsymbol{A}_1 : \boldsymbol{A}_2) - \rho(\boldsymbol{A}_1))}{(n - \rho(\boldsymbol{C}_1) - p + \rho(\boldsymbol{A}_1) - 1)(n - \rho(\boldsymbol{C}_2) - p + \rho(\boldsymbol{A}_1 : \boldsymbol{A}_2) - 1)}.
\end{aligned}$$

We close this section by giving another important theorem which appeared in von Rosen (1990). One significant consequence of the theorem is that, since $(\mathbf{G}_{r-1}'\mathbf{W}_r\mathbf{G}_{r-1})^{-1}$ is inverted Wishart distributed, it follows from the theorem that $\mathbf{P}_r'\mathbf{S}_r^{-1}\mathbf{P}_r$ is also inverted Wishart distributed which otherwise is difficult to show. We use this result in Paper I when calculating the covariance between $\mathbf{R}_4$ and the estimated mean structure and obtaining the dispersion matrix of this residual. In Paper II, we use the result to show that the distribution of $\phi_4(\mathbf{X})$ under the null hypothesis is independent of $\boldsymbol{\Sigma}$ as well as to obtain the expected value of the test statistic.

Theorem 3.9. *Let $\boldsymbol{P}_r$ and $\boldsymbol{S}_r$ be as given in Theorem 3.2. Suppose that the conditions in Definition 3.1 are satisfied. Then,*

$$\boldsymbol{P}_r'\boldsymbol{S}_r^{-1}\boldsymbol{P}_r = \boldsymbol{G}_{r-1}(\boldsymbol{G}_{r-1}'\boldsymbol{W}_r\boldsymbol{G}_{r-1})^{-1}\boldsymbol{G}_{r-1}',$$

where,

$$\boldsymbol{G}_{r+1} = \boldsymbol{G}_r(\boldsymbol{G}_r'\boldsymbol{A}_{r+1})^o, \, \boldsymbol{G}_0 = \boldsymbol{I},$$
$$\boldsymbol{W}_r = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}_r'(\boldsymbol{C}_r\boldsymbol{C}_r')^{-}\boldsymbol{C}_r)\boldsymbol{X}'.$$

Chapter 4

Conditional tests

4.1 Introduction

Since the conditional model tends to be simpler than the original unconditional one, the conditional point of view will frequently bring a simplification of the theory (Lehmann, 1986). This is particularly true when considering the Growth Curve model. Apart from a great simplification provided, conditioning, like sufficiency and invariance, leads to a reduction of data. Moreover, in the presence of nuisance parameters, conditional inference has been used to eliminate unwanted parameters.

When the problem involves ancillary statistics conditioning with respect to the ancillary statistic is appropriate since it makes the inference more relevant to the situation at hand. Ancillary statistic is a statistic whose distribution does not depend on the parameter of interest. The term ancillary was first used by Fisher and those statistics are referred to as non-informative since they do not provide any information about the parameters.

In the presence of ancillary statistics say, $\mathbf{Z}$, one can think of the observation $\mathbf{X}$ (with distribution $\mathbf{P}$) as obtained from a two-stage experiment (Lehmann, 1986):

i) Observe the ancillary statistic, $\mathbf{Z}$, with distribution $\mathbf{F}$.

ii) Given $\mathbf{Z}$, observe a quantity $\mathbf{X}$ with distribution $\mathbf{P}(\mathbf{X}|\mathbf{Z})$.

The resulting $\mathbf{X}$ is distributed according to the original distribution $\mathbf{P}$. It was also suggested that the above argument is valid even if the distribution of the ancillary statistic depends on parameters other than the parameter of interest. Such a statistic is usually called S-ancillary or partial ancillary statistic. For more details about ancillary statistics we refer to Fisher (1956) and Basu (1964). There is also an important paper by Basu (1975) about statistical information and likelihood.

The conditioning variables are not always restricted to ancillary statistics. For brief discussion and further references about conditioning with variables other than ancillary statistics and concepts of relevant subsets, we refer to Lehmann (1986).

4.2 The Growth Curve model

Consider the Growth Curve Model given in Definition 2.1. For this model $\mathbf{S} = \mathbf{X}(I - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C})$ is distributed as a Wishart random variable with parameters $\mathbf{\Sigma}$ and $n - \rho(\mathbf{C})$, i.e., $\mathbf{S} \sim W_p(\mathbf{\Sigma}, n - \rho(\mathbf{C}))$, where $\rho(\mathbf{C})$ is the rank of the design matrix $\mathbf{C}$. Its distribution is, therefore, independent of the parameter of interest, $\mathbf{B}$. This shows that $\mathbf{S}$ is S-ancillary for $\mathbf{B}$. As a result, conditioning using $\mathbf{S}$ is appropriate in the sense that it makes the inference more relevant to the situation at hand without losing any information about $\mathbf{B}$, the parameter of interest.

It is important to note that our approach is different from other conditional approaches in which conditioning is usually made at an early stage, and mainly for eliminating nuisance parameters (see Basu, 1977). We are going to use the conditional approach after the test has been constructed using the restricted followed by estimated likelihood approaches. Moreover, in most conditional approaches, the statistic which is used for conditioning is a partial sufficient statistic which gives the advantage that the resulting conditional distribution depends only on the parameter of interest, see Basu (1978) about partial sufficiency. However, we have eliminated the nuisance parameter by using a restricted likelihood approach. The main reason for conditioning in our case, unlike most other cases, is to make the distribution relatively easy to handle so that we can provide a critical point for the test.

Furthermore, in problems of testing Fisher (1956) used ancillary statistics for the determination of the true level of significance. He recommended that, in the presence of ancillary statistics, the level of significance of a test should be computed by referring to the conditional sample space determined by the set of all possible samples for which the value of the ancillary statistics is the one presently observed (Basu, 1964).

Chapter 5

Some technical results used in the papers

In this chapter we give some technical results which are of particular interest. These were necessary to develop in order to derive the results in the papers. Most of the results presented are used in Paper I although some are used in Paper II. Let us first give some notations which help us shorten the expressions for some of the formulas. We use these representations whenever needed.

$$\begin{aligned}
\mathbf{P}_{\mathbf{C}_1} &= \mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1,\\
\mathbf{P}_{\mathbf{C}_2} &= \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2,\\
\mathbf{P}_{\mathbf{A}_1} &= \mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{S}_1^{-1},\\
\mathbf{P}_{\mathbf{T}_1\mathbf{A}_2} &= \mathbf{T}_1\mathbf{A}_2(\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_1^{-1}\mathbf{T}_1\mathbf{A}_2)^-\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_1^{-1},
\end{aligned}$$

where,

$$\begin{aligned}
\mathbf{T}_1 &= \mathbf{I} - \mathbf{P}_{\mathbf{A}_1},\\
\mathbf{S}_1 &= \mathbf{X}(\mathbf{I} - \mathbf{P}_{\mathbf{C}_1})\mathbf{X}'.
\end{aligned}$$

Lemma 5.1. *Let $\boldsymbol{S}_1$ and $\boldsymbol{T}_1$ be given in Corollary 3.5. Then*

$$\boldsymbol{A}_1'\boldsymbol{S}_1^{-1}\boldsymbol{T}_1 = \boldsymbol{0}. \tag{5.1}$$

Proof.

$$\begin{aligned}
\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{T}_1 &= \mathbf{A}_1'\mathbf{S}_1^{-1}(\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{S}_1^{-1})\\
&= \mathbf{A}_1'\mathbf{S}_1^{-1} - \mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{S}_1^{-1}\\
&= \mathbf{A}_1'\mathbf{S}_1^{-1} - \mathbf{A}_1'\mathbf{S}_1^{-1}\\
&= \mathbf{0}.
\end{aligned}$$

□

Theorem 5.2. *Let $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ be as in Corollary 3.5. Then,*

$$\boldsymbol{A}_1'\boldsymbol{S}_1^{-1} = \boldsymbol{A}_1'\boldsymbol{S}_2^{-1}. \tag{5.2}$$

Proof. Recall that $\mathbf{S}_2 = \mathbf{S}_1 + \mathbf{T}_1\mathbf{X}(\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1 - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2)\mathbf{X}'\mathbf{T}_1'$. Consequently, we have

$$\mathbf{A}_1'\mathbf{S}_2^{-1} = \mathbf{A}_1'(\mathbf{S}_1 + \mathbf{T}_1\mathbf{X}(\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1 - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2)\mathbf{X}'\mathbf{T}_1')^{-1}.$$

Now if we let

$$\mathbf{V} = \mathbf{T}_1\mathbf{X}(\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1 - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2),$$

then we have

$$\begin{aligned}\mathbf{A}_1'\mathbf{S}_2^{-1} &= \mathbf{A}_1'(\mathbf{S}_1 + \mathbf{V}\mathbf{V}')^{-1}\\ &= \mathbf{A}_1'(\mathbf{S}_1^{-1} - \mathbf{S}_1^{-1}\mathbf{V}(\mathbf{V}'\mathbf{S}_1^{-1}\mathbf{V} + \mathbf{I})^{-1}\mathbf{V}'\mathbf{S}_1^{-1})\\ &= \mathbf{A}_1'\mathbf{S}_1^{-1} - \mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{V}(\mathbf{V}'\mathbf{S}_1^{-1}\mathbf{V} + \mathbf{I})^{-1}\mathbf{V}'\mathbf{S}_1^{-1}\\ &= \mathbf{A}_1'\mathbf{S}_1^{-1},\end{aligned}$$

where the last statement follows from Lemma 5.1. □

The above result is used in Paper I. Due to the result it is possible to replace $\mathbf{A}_1'\mathbf{S}_2^{-1}$ by $\mathbf{A}_1'\mathbf{S}_1^{-1}$ whenever necessary. This provides a great simplification since $\mathbf{S}_1$ has a much simpler structure than $\mathbf{S}_2$.

The next theorem shows that the matrix $\mathbf{P}$ given in (5.3) below, which is a basis in determining the space generated by the design matrices for the model considered in Paper I, is a projection matrix. The space generated by the columns of $\mathbf{P}$ is also provided.

Theorem 5.3. *Let*

$$\begin{aligned}\boldsymbol{P} = &(\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^-\boldsymbol{C}_1)\otimes(\boldsymbol{A}_1(\boldsymbol{A}_1'\boldsymbol{S}_1^{-1}\boldsymbol{A}_1)^-\boldsymbol{A}_1'\boldsymbol{S}_1^{-1})\\ &+(\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^-\boldsymbol{C}_2)\otimes(\boldsymbol{T}_1\boldsymbol{A}_2(\boldsymbol{A}_2'\boldsymbol{T}_1'\boldsymbol{S}_2^{-1}\boldsymbol{T}_1\boldsymbol{A}_2)^-\boldsymbol{A}_2'\boldsymbol{T}_1'\boldsymbol{S}_2^{-1}),\end{aligned} \tag{5.3}$$

then $\boldsymbol{P}$ is idempotent and

$$\mathcal{C}(\boldsymbol{P}) = \mathcal{C}(\boldsymbol{C}_1')\otimes\mathcal{C}_{\boldsymbol{S}_1}(\boldsymbol{A}_1) + \mathcal{C}(\boldsymbol{C}_2')\otimes\mathcal{C}_{\boldsymbol{S}_2}(\boldsymbol{T}_1\boldsymbol{A}_2). \tag{5.4}$$

Remark: Note that $\otimes$ in (5.4) denotes a tensor product between the spaces. Observe also that we can replace $\mathbf{S}_1$ by $\mathbf{S}_2$.

Proof. Consider (5.3) and let us denote the first and second terms in the sum by $\mathbf{E}$ and $\mathbf{F}$, respectively. Then,

$$\mathbf{PP} = (\mathbf{E}+\mathbf{F})(\mathbf{E}+\mathbf{F}) = \mathbf{EE}+\mathbf{EF}+\mathbf{FE}+\mathbf{FF}.$$

Now consider the two cross product terms, i.e., $\mathbf{EF}$ and $\mathbf{FE}$. Using the fact that $\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{T}_1 = \mathbf{0}$, it follows that $\mathbf{EF} = \mathbf{0}$. Furthermore, we have seen in Theorem 5.2 that $\mathbf{A}_1'\mathbf{S}_1^{-1} = \mathbf{A}_1'\mathbf{S}_2^{-1}$ and hence $\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{T}_1 = \mathbf{A}_1'\mathbf{S}_2^{-1}\mathbf{T}_1 = \mathbf{0}$, which in turn implies that $\mathbf{FE} = \mathbf{0}$. Therefore it remains to show that $\mathbf{E}$ and $\mathbf{F}$ are idempotent. However, $\mathbf{E}$ is an idempotent matrix follows from the fact that both $\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1$ and $\mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{S}_1^{-1}$ are idempotent. Similarly, $\mathbf{F}$ is idempotent since both $\mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2$ and $\mathbf{T}_1\mathbf{A}_2(\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_1^{-1}\mathbf{T}_1\mathbf{A}_2)^-\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_1^{-1}$ are idempotent.

Now, let us consider the column space of $\mathbf{P}$. Since $\mathcal{C}_{\mathbf{S}_1}(\mathbf{A}_1)$ and $\mathcal{C}_{\mathbf{S}_2}(\mathbf{T}_1\mathbf{A}_2)$ are orthogonal (see Theorem 5.4 below) we have

$$\begin{aligned}\mathcal{C}(\mathbf{P}) = &\,\mathcal{C}((\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1)\otimes(\mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^-\mathbf{A}_1'\mathbf{S}_1^{-1}))\\ &+\mathcal{C}((\mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2)\otimes(\mathbf{T}_1\mathbf{A}_2(\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{A}_2)^-\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1})).\end{aligned}$$

Using the fact that $\mathcal{C}(\mathbf{A}\otimes\mathbf{B}) = \mathcal{C}(\mathbf{A})\otimes\mathcal{C}(\mathbf{B})$ (Takemura, 1983), we get the desired result. □

Theorem 5.4. *Let $\boldsymbol{T}_1$ be as in Corollary 3.5. Then*

$$\mathcal{C}_{\boldsymbol{S}_2}(\boldsymbol{T}_1\boldsymbol{A}_2) \subseteq \mathcal{C}_{\boldsymbol{S}_2}(\boldsymbol{A}_1) \tag{5.5}$$

and

$$\mathcal{C}(\boldsymbol{A}_1) + \mathcal{C}(\boldsymbol{T}_1\boldsymbol{A}_2) = \mathcal{C}(\boldsymbol{A}_1 + \mathcal{C}(\boldsymbol{A}_2). \tag{5.6}$$

Proof. Using the fact that $\mathcal{C}(\mathbf{EF}) \subseteq \mathcal{C}(\mathbf{E})$ for any two matrices $\mathbf{E}$ and $\mathbf{F}$, it follows that

$$\begin{aligned}\mathcal{C}(\mathbf{T}_1\mathbf{A}_2) &\subseteq \mathcal{C}(\mathbf{T}_1)\\ &= \mathcal{C}(\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^{-}\mathbf{A}_1'\mathbf{S}_1^{-1})\\ &= \mathcal{C}(\mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-1}\mathbf{A}_1)^{-}\mathbf{A}_1'\mathbf{S}_1^{-1})^{\perp}\\ &= \mathcal{C}_{\mathbf{S}_1}(\mathbf{A}_1)^{\perp} = \mathcal{C}_{\mathbf{S}_2}(\mathbf{A}_1)^{\perp}.\end{aligned}$$

Now observe that $\mathbf{T}_1$ is a projection matrix. It follows that

$$\begin{aligned}\mathcal{C}(\mathbf{A}_1) + \mathcal{C}(\mathbf{T}_1\mathbf{A}_2) &= \mathcal{C}(\mathbf{A}_1) + \{(\mathcal{C}(\mathbf{A}_2) + \mathcal{N}(\mathbf{T}_1)) \cap \mathcal{C}(\mathbf{T}_1)\}\\ &= \mathcal{C}(\mathbf{A}_1) + \{(\mathcal{C}(\mathbf{A}_2) + \mathcal{C}(\mathbf{T}_1')^{\perp}) \cap \mathcal{C}(\mathbf{T}_1)\}\\ &= \mathcal{C}(\mathbf{A}_1) + \{(\mathcal{C}(\mathbf{A}_2) + \mathcal{C}(\mathbf{A}_1)) \cap \mathcal{C}_{\mathbf{S}_1}(\mathbf{A}_1)^{\perp}\}\\ &= \{\mathcal{C}(\mathbf{A}_1) + \mathcal{C}(\mathbf{A}_2)\} \cap \{\mathcal{C}(\mathbf{A}_1) + \mathcal{C}_{\mathbf{S}_1}(\mathbf{A}_1)^{\perp}\}\\ &= \mathcal{C}(\mathbf{A}_1) + \mathcal{C}(\mathbf{A}_2),\end{aligned}$$

Here we have used the fact that $\mathcal{C}(\mathbf{A}_1) + \mathcal{C}_{\mathbf{S}_1}(\mathbf{A}_1)^{\perp} = \mathbf{V}$ where $\mathbf{V}$ represents the whole space, $\mathcal{N}(\mathbf{P}) = \mathcal{C}(\mathbf{P}')^{\perp}$ and $\mathcal{C}(\mathbf{PA}) = (\mathcal{C}(\mathbf{A}) + \mathcal{N}(\mathbf{P})) \cap \mathcal{C}(\mathbf{P})$ for any matrix $\mathbf{A}$ and any projection matrix $\mathbf{P}$. Note that $\mathcal{N}(\mathbf{P})$ stands for the null space of $\mathbf{P}$. □

The above theorem together with the result in Theorem 5.3 is used in obtaining the spaces on which the new residuals are defined in Paper I. The spaces are provided by a decomposition given in the next theorem. We refer to Kollo & von Rosen (2005) for discussions about decomposition of linear spaces.

Theorem 5.5. *The orthogonal complement of the space given in (5.4) can be decomposed as follows:*

$$\{C(\boldsymbol{C}_1') \otimes C_{\boldsymbol{S}_2}(\boldsymbol{A}_1) + C(\boldsymbol{C}_2') \otimes C_{\boldsymbol{S}_2}(\boldsymbol{T}_1\boldsymbol{A}_2)\}^{\perp} = \mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV}, \tag{5.7}$$

where $\boxplus$ represents the orthogonal sum and

$$\begin{aligned}\mathbf{I} &= C(\boldsymbol{C}_1')^{\perp} \otimes C_{\boldsymbol{S}_2}(\boldsymbol{A}_1),\\ \mathbf{II} &= C(\boldsymbol{C}_1')^{\perp} \otimes C_{\boldsymbol{S}_2}(\boldsymbol{A}_1)^{\perp},\\ \mathbf{III} &= (C(\boldsymbol{C}_1') \cap C(\boldsymbol{C}_2')^{\perp}) \otimes C_{\boldsymbol{S}_2}(\boldsymbol{A}_1)^{\perp},\\ \mathbf{IV} &= C(\boldsymbol{C}_2') \otimes (C_{\boldsymbol{S}_2}(\boldsymbol{A}_1) + C_{\boldsymbol{S}_2}(\boldsymbol{A}_1)).\end{aligned}$$

Proof.

$$\begin{aligned}\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV} &= (\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV}) \cap (\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV})\\ &= (\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV} \boxplus (\mathbf{A} \cap \mathbf{A}^{\perp}))\\ &\quad \cap (\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV} \boxplus (\mathbf{B} \cap \mathbf{B}^{\perp})),\end{aligned}$$

where $\mathbf{A} = C(\mathbf{C}_2') \otimes C_{\mathbf{S}_2}(\mathbf{A}_1)$ and $\mathbf{B} = C(\mathbf{C}_2') \otimes C_{\mathbf{S}_2}(\mathbf{T}_1\mathbf{A}_2)$. Furthermore, note that **I**, **II**, **III** and **IV** are all subsets of **A** and **B**. As a result we get

$$\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV} \boxplus (\mathbf{A} \cap \mathbf{A}^\perp) = (C(\mathbf{C}_1') \otimes C_{\mathbf{S}_2}(\mathbf{A}_1))^\perp$$

and

$$\mathbf{I} \boxplus \mathbf{II} \boxplus \mathbf{III} \boxplus \mathbf{IV} \boxplus (\mathbf{B} \cap \mathbf{B}^\perp) = (C(\mathbf{C}_2') \otimes C_{\mathbf{S}_2}(\mathbf{T}_1\mathbf{A}_2))^\perp.$$

□

Lemma 5.6. *Let $\boldsymbol{G}_1$, $\boldsymbol{G}_2$ and $\boldsymbol{W}_2$ be as in Theorem 3.9 and let*

$$\boldsymbol{G}_1' = \boldsymbol{H}_1(\boldsymbol{I} : \boldsymbol{0})\boldsymbol{\Gamma}^1\boldsymbol{\Sigma}^{-1/2} \tag{5.8}$$

and

$$\boldsymbol{G}_2' = \boldsymbol{H}_2(\boldsymbol{I} : \boldsymbol{0})\boldsymbol{\Gamma}^2(\boldsymbol{I} : \boldsymbol{0})\boldsymbol{\Gamma}^1\boldsymbol{\Sigma}^{-1/2}, \tag{5.9}$$

where $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are non-singular matrices of proper size, and $\boldsymbol{\Gamma}^1$ and $\boldsymbol{\Gamma}^2$ are orthogonal matrices of proper size. Then,

$$\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}\boldsymbol{G}_1(\boldsymbol{G}_1'\boldsymbol{\Sigma}\boldsymbol{G}_1)^{-1}\boldsymbol{G}_1', \tag{5.10}$$

$$\boldsymbol{E}[\boldsymbol{G}_2(\boldsymbol{G}_2'\boldsymbol{W}_2\boldsymbol{G}_2)^{-}\boldsymbol{G}_2'\boldsymbol{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}] = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}\boldsymbol{\Gamma}_1^2 \tag{5.11}$$

and

$$\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^1\boldsymbol{\Gamma}_1^2\boldsymbol{\Gamma}_1^{2'}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{G}_2(\boldsymbol{G}_2'\boldsymbol{\Sigma}\boldsymbol{G}_2)^{-1}\boldsymbol{G}_2'\boldsymbol{\Sigma}, \tag{5.12}$$

where $\boldsymbol{\Gamma}^{1'} = (\boldsymbol{\Gamma}_1^{1'} : \boldsymbol{\Gamma}_2^{1'})$ and $\boldsymbol{\Gamma}^{2'} = (\boldsymbol{\Gamma}_1^{2'} : \boldsymbol{\Gamma}_2^{2'})$.

Proof. First observe that $\mathbf{G}_1 = \mathbf{A}_1^o$ and $\mathbf{G}_2 = \mathbf{G}_1(\mathbf{G}_1'\mathbf{A}_2)^o$, where $\mathbf{A}_1^o$ is any matrix of full rank spanning the orthogonal complement of $\mathcal{C}(\mathbf{A}_1)$ (with respect to the standard inner product). Consider (5.10), by substituting $\mathbf{G}_1$ by $\mathbf{H}_1(\mathbf{I} : \mathbf{0})\boldsymbol{\Gamma}^1\boldsymbol{\Sigma}^{-1/2}$, it follows that

$$\begin{aligned}
\boldsymbol{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\boldsymbol{\Sigma}\mathbf{G}_1)^{-}\mathbf{G}_1' &= \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}(\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'})^{-}\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}.
\end{aligned}$$

Now consider $\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}$. If we replace $\mathbf{G}_2'$ by (5.9), we get

$$\begin{aligned}
&\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'} \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\boldsymbol{\Gamma}_1^2\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2\prime})^{-1}\boldsymbol{\Gamma}_1^2\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'} \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\boldsymbol{\Gamma}_1^2\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'})^{-1}\boldsymbol{\Gamma}_1^2\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}^{2'}\boldsymbol{\Gamma}^2 \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}((\mathbf{I} : \mathbf{0})\mathbf{Z}\mathbf{Z}'(\mathbf{I} : \mathbf{0})')^{-1}(\mathbf{I} : \mathbf{0})\mathbf{Z}\mathbf{Z}'\boldsymbol{\Gamma}^2 \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\mathbf{Z}_1\mathbf{Z}_1')^{-1}(\mathbf{Z}_1\mathbf{Z}_1' : \mathbf{Z}_1\mathbf{Z}_2')\boldsymbol{\Gamma}^2 \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\mathbf{I} : (\mathbf{Z}_1\mathbf{Z}_1')^{-1}\mathbf{Z}_1\mathbf{Z}_2')\boldsymbol{\Gamma}^2,
\end{aligned}$$

where $\mathbf{Z} = \boldsymbol{\Gamma}^2\boldsymbol{\Gamma}_1^1\boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{I} - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-}\mathbf{C}_2)$ and $\mathbf{Z}' = (\mathbf{Z}_1' : \mathbf{Z}_2')$. Now note that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent. Moreover $E[\mathbf{Z}_2] = \mathbf{0}$ which implies that

$E[(\mathbf{Z}_1\mathbf{Z}_1')^{-1}\mathbf{Z}_1\mathbf{Z}_2'] = \mathbf{0}$. This in turn implies that $E[\mathbf{I} : (\mathbf{Z}_1\mathbf{Z}_1')^{-1}\mathbf{Z}_1\mathbf{Z}_2'] = (\mathbf{I} : \mathbf{0})$. Therefore,

$$\begin{aligned} E[\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2'\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}] &= E[\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\mathbf{I} : (\mathbf{Z}_1\mathbf{Z}_1')^{-1}\mathbf{Z}_1\mathbf{Z}_2')\boldsymbol{\Gamma}^2] \\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\mathbf{I} : \mathbf{0})\boldsymbol{\Gamma}^2 \\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\mathbf{I} : \mathbf{0})\boldsymbol{\Gamma}_1^2. \end{aligned}$$

Finally, consider $\boldsymbol{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\boldsymbol{\Sigma}\mathbf{G}_1)^{-}\mathbf{G}_1'\boldsymbol{\Sigma}$. Replacing $\mathbf{G}_2'$ by (5.9) gives

$$\begin{aligned} \boldsymbol{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\boldsymbol{\Sigma}\mathbf{G}_1)^{-}\mathbf{G}_1'\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}(\boldsymbol{\Gamma}_1^{2}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'})^{-1}\boldsymbol{\Gamma}_1^{2}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma} \\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{2'}\boldsymbol{\Gamma}_1^{2}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{1/2}. \end{aligned}$$

□

We close this chapter by presenting the following theorem which is used in obtaining the covariance matrices between the residuals defined in Paper I. This result is also used in Paper II when calculating the expected value of one of the tests.

Theorem 5.7. *Let $\boldsymbol{T}_2$ be as in Corollary 3.5 and $\boldsymbol{L}_2$ as in Theorem 3.8. Then,*

$$E[\boldsymbol{I} - \boldsymbol{T}_2'] = \boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2. \tag{5.13}$$

Proof. Recall that

$$\mathbf{I} - \mathbf{T}_2 = \mathbf{T}_1\mathbf{A}_2(\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{A}_2)^{-}\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}.$$

Using Theorem 3.9 and the fact that

$$\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1} = \mathbf{S}\mathbf{A}^{o}(\mathbf{A}^{o'}\mathbf{S}\mathbf{A}^{o})^{-}\mathbf{A}^{o'}, \tag{5.14}$$

we can rewrite $\mathbf{I} - \mathbf{T}_1$ as,

$$\mathbf{S}_1\mathbf{G}_1(\mathbf{G}_1'\mathbf{S}_1\mathbf{G}_1)^{-1}\mathbf{G}_1'\mathbf{A}_2(\mathbf{A}_2'\mathbf{G}_1(\mathbf{G}_1'\mathbf{W}_2\mathbf{G}_1)^{-1}\mathbf{G}_1'\mathbf{A}_2)^{-}\mathbf{A}_2'\mathbf{G}_1(\mathbf{G}_1'\mathbf{W}_2\mathbf{G}_1)^{-1}\mathbf{G}_1'.$$

Once again by using a representation similar to (5.14) and the fact that $\mathbf{G}_1(\mathbf{G}_1'\mathbf{A}_2)^{o} = \mathbf{G}_2$, we get

$$\mathbf{I} - \mathbf{T}_2 = \mathbf{S}_1\mathbf{G}_1(\mathbf{G}_1'\mathbf{S}\mathbf{G}_1)^{-1}\mathbf{G}_1'(\mathbf{I} - \mathbf{W}_2\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2'). \tag{5.15}$$

Now by using the canonical representation of $\mathbf{S}_1\mathbf{G}_1(\mathbf{G}_1'\mathbf{S}\mathbf{G}_1)^{-1}\mathbf{G}_1'$, which is obtained by replacing $\mathbf{G}_1$ by $\mathbf{H}_1(\mathbf{I} : \mathbf{0})\boldsymbol{\Gamma}^1\boldsymbol{\Sigma}^{-1/2}$, and is given by

$$\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_2^{1'}\mathbf{Y}_2\mathbf{Y}_1'(\mathbf{Y}_1\mathbf{Y}_1')^{-1} \tag{5.16}$$

where,
$\mathbf{Y} = \boldsymbol{\Gamma}^1\boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{I} - \mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^{-}\mathbf{C}_1)$, $\mathbf{Y}' = (\mathbf{Y}_1' : \mathbf{Y}_2')$ and $\boldsymbol{\Gamma}^{1'} = (\boldsymbol{\Gamma}_1^{1'} : \boldsymbol{\Gamma}_2^{1'})$, we obtain,

$$\begin{aligned} E[\mathbf{I} - \mathbf{T}_2] &= E[\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_1^{1'}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{W}_2\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2')] \\ &\quad + E[\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}_2^{1'}\mathbf{Y}_2\mathbf{Y}_1'(\mathbf{Y}_1\mathbf{Y}_1')^{-1}\boldsymbol{\Gamma}_1^{1}\boldsymbol{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{W}_2\mathbf{G}_2(\mathbf{G}_2'\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}_2')]. \end{aligned}$$

However, the second term in the sum vanishes since $\mathbf{\Gamma}_1^1\mathbf{\Sigma}^{-1/2}\mathbf{W}_2\mathbf{\Gamma}_1^{1'}\mathbf{\Sigma}^{-1/2}$ is a function of only $\mathbf{Y}_1$ and $\mathbf{X}(\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^-\mathbf{C}_1 - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^-\mathbf{C}_2)$. Moreover, the latter is independent of $\mathbf{Y}$. An application of Lemma 5.6 therefore shows that

$$E[\mathbf{I} - \mathbf{T}_2] = \mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1' - \mathbf{\Sigma}\mathbf{G}_2(\mathbf{G}_2'\mathbf{\Sigma}\mathbf{G}_2)^{-1}\mathbf{G}_2'.$$

Substituting $\mathbf{G}_2$ in the above expression by $\mathbf{G}_1(\mathbf{G}_1'\mathbf{A}_2)^o$ results in

$$\begin{aligned}
&\mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1' - \mathbf{\Sigma}\mathbf{G}_2(\mathbf{G}_2'\mathbf{\Sigma}\mathbf{G}_2)^{-1}\mathbf{G}_2' \\
&= \mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1' - \mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{A}_2)^o((\mathbf{G}_1'\mathbf{A}_2)^{o'}\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{A}_2)^o)^{-1} \\
&\quad \times (\mathbf{G}_1'\mathbf{A}_2)^{o'}\mathbf{G}_1' \\
&= \mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}[\mathbf{I} - (\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)(\mathbf{G}_1'\mathbf{A}_2)^o((\mathbf{G}_1'\mathbf{A}_2)^{o'}(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)(\mathbf{G}_1'\mathbf{A}_2)^o)^- \\
&\quad (\mathbf{G}_1'\mathbf{A}_2)^{o'}]\mathbf{G}_1' \\
&= \mathbf{\Sigma}\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1'\mathbf{A}_2(\mathbf{A}_2'\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1'\mathbf{A}_2)^-\mathbf{A}_2'\mathbf{G}_1(\mathbf{G}_1'\mathbf{\Sigma}\mathbf{G}_1)^{-1}\mathbf{G}_1' \\
&= \mathbf{L}_2'\mathbf{\Sigma}^{-1}.
\end{aligned}$$

To get the last equality, we have used (5.14) and the fact that $\mathbf{G}_1 = \mathbf{A}_1^o$.

Chapter 6

Summary of the papers

6.1 Paper I

In this paper the special case of the Extended Growth Curve model given in Definition 3.3 is considered. New residuals, taking the bilinear structure in the model into account, are defined. The *vec* operator is applied on the estimated model to show that the estimated model is the projection of the observation matrix $\mathbf{X}$ on the space generated by the two design matrices. The space turns out to be the sum of two tensor product spaces. In univariate and multivariate linear models, ordinary residuals are defined by projecting $\mathbf{X}$ on the space orthogonal to the one generated by the design matrices. We have shown that this is also true for the model considered in this paper. The space where the residuals are defined is given by

$$(C(\mathbf{C}_1') \otimes C_{\mathbf{S}_1}(\mathbf{A}_1) + C(\mathbf{C}_2') \otimes C_{\mathbf{S}_2}(\mathbf{T}_1\mathbf{A}_2))^{\perp}.$$

We decomposed the above space into four orthogonal spaces and defined four residuals by projecting $\mathbf{X}$ on the resulting spaces.

The residuals are interpreted and remarks are given regarding what kind of information they provide and how one can use this information to validate the model and model assumptions. The residuals can also provide some information about outliers and/or influential observations. These interpretations also apply to von Rosen's residuals since the model considered in his paper is a special case of the one considered in Paper I.

Properties of the residuals are provided. It is shown that the residuals defined are symmetrically distributed around zero and are uncorrelated with each other. We have also given the dispersion matrices for the residuals and the covariance between them and the estimated model.

The results are illustrated using the Potthoff & Roy (1964) data. The residuals are used to check if the assumed growth curves fit the data well and if there are observations which are extreme in some sense. The calculated standard errors are used as cutoff points.

6.2 Paper II

Here the residuals defined in Paper I and those defined by von Rosen (1995b) are considered. In Paper I, it was suggested that the new residuals taking the bilinear structure into consideration can be used to validate model assumptions as well as detect outliers and/or influential observations. In Paper II, we propose statistics for testing different hypotheses in the Growth and Extended Growth Curve models. The tests are constructed using restricted followed by estimated likelihood approaches. The restricted likelihood approach was introduced by Patterson & Thompson (1974). A brief discussion about the topic is presented by Searle (1992). The method is also applied to the GC model by Pan & Fang (2002).

As expected, the tests turn out to be functions of appropriate residuals in the respective models which enable us to understand and interpret the tests.

For the GC model, we write the likelihood as a product of two terms. We maximize a part of the likelihood which is a function of only $\mathbf{\Sigma}$ to get an estimator for the covariance matrix. Then we replace the covariance matrix by the resulting estimator to get the estimated likelihood which is then maximized under H_o and $H_o \cup H_1$, where H_o and H_1 are the null and alternative hypotheses, respectively.

The EGC model has more structure in the mean since the groups involved could have polynomial growth curves of different degrees. As a result, we have more hypotheses to test and hence more tests to construct. We write the likelihood for the model given in Definition 3.3 as a product of three terms. Depending on the hypothesis to be tested, we maximize a part of the likelihood to get an estimator for the covariance matrix. As before, the estimator replaces $\mathbf{\Sigma}$ to get the estimated likelihood, which is then maximized under H_o and $H_o \cup H_1$.

We have shown that the distributions of all the tests under the null hypotheses are independent of the unknown covariance matrix $\mathbf{\Sigma}$. Moreover, the conditional and unconditional expected values for all the tests are also provided in the paper.

The distributions of the tests proposed in this paper are difficult to obtain, as a result, there is a need to provide some methods to find the critical points. We suggest two alternative approaches to approximate the distributions for the tests under both the null and alternative hypotheses. The first suggestion is to approximate the densities for the tests based on the first two moments. The second one is to use conditional versions of the tests, which is considered in Paper III.

6.3 Paper III

In this paper we consider one of the tests proposed in Paper II, although the approach can be extended to all the tests proposed there. The aim is to consider the conditional version of the test. Apart from a great simplification and reduction provided by conditioning we discuss why conditioning makes the test problem relevant to the situation at hand. Due to the existence of a natural ancillary statistics, we do the reduction and simplification without losing any information about the parameter of interest.

We have shown that under the null hypothesis the test can be written as sums of independent weighted central chi-square random variables. This enables us to use existing results for such a sum. Sum of weighted independent chi-square random variables has been considered by many and the exact distribution has been provided as an infinite series in Kotz et al. (1967), Mathai (1982) and Moschopoulos (1985). However, the distribution is too complicated to be used in practice and hence an approximation is needed. Several approximations have been proposed, see for example Moschopoulos (1985). In this paper we use the well known approximation introduced by Satterthwaite (1941, 1946) and provide an approximate critical point for the test. For more discussions about the approximation we refer to Khuri et al. (1998). The extension of the approximation to linear combination of independent Wishart random variables together with some Monte Carlo results to demonstrate the closeness of the approximation is given in Tan & Gupta (1983).

Under the alternative, it is shown that the test can be written as sums of weighted independent non-central chi-square random variables. Although too complicated to be used in practice, the exact distribution for such a sum has also been provided as an infinite series in Press (1966). Moreover, there exist several algorithms to numerically solve the series, see for example Imhof (1961). In this paper, we provide a new approximation which can be used to get an approximate power for the conditional test. The approximation is similar to that of Satterthwaite's for the weighted sum of independent central chi-square random variables. This kind of approximation, as to our knowledge, has not been done for a weighted sum of non-central chi-square random variables. Moreover, our approach is quite different and new ideas have been utilized to obtain estimators for the parameters.

Both the exact and approximate distributions under the alternative hypothesis depend on the unknown covariance matrix, $\mathbf{\Sigma}$. As a result, in practice one needs to find a reasonable estimator for $\mathbf{\Sigma}$ to get an estimate for the power. In this paper, we have suggested and discussed three alternative estimators. Finally, numerical examples have been provided to illustrate the results.

Chapter 7

Discussion

7.1 General discussion

Statistical models play an important role in understanding different practical problems and they are used in summarizing and describing the underlying structure in the data. Statistical models are also used in making inference, prediction and making important decisions. In many cases, these models relay on several assumptions. Model fitting is not complete without validating the model and model assumptions, and we believe model evaluation is an important part of the model fitting problem.

In univariate linear models residuals play an important role in validating the model, model assumptions and checking if there are extreme observations that do not seem to belong to the data (outliers) and/or that in some sense may alter the model fitting problem (influential observations).

In ordinary univariate linear models, residuals have been studied extensively and there are many graphical and formal approaches based on the residuals to validate the model and model assumptions. Moreover, many other residuals such as studentized and standardized residuals have been defined and studied.

In analysis of longitudinal and repeated measurement data, residuals can also be used to assess the adequacy of the model and they can indicate the presence of outliers (Fitzmaurice et al., 2004). However, residuals obtained in analyzing such data are based on results which are usually obtained numerically using iteration and often relay on asymptotic results. Moreover, they are correlated with each other, do not have constant variance and are not normally distributed. Furthermore, variance-covariance matrices for the residuals are different from that of the error terms. This shows that it is not reasonable to use these residuals, for example, in checking the normality assumption or the homogeneity of variances. Moreover, we believe that the residuals contain information about the within and between individual assumptions. As a result we need to investigate these components separately.

One suggestion, made by Fitzmaurice et al. (2004) to tackle the problem pointed out in the previous paragraph, is to transform the residuals. It is mentioned in the book that there are many ways to transform them and one particular transformation, which is presented below, is discussed.
Let $\mathbf{r}_i$ be a vector of residuals for each individual, i.e.,

$$\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i\hat{\beta}.$$

Then a transformation is given by, (for details see Fitzmaurice et al., 2004)

$$\mathbf{r}_i^* = \mathbf{L}_i^{-1}\mathbf{r}_i = \mathbf{L}_i^{-1}(\mathbf{Y}_i - \hat{\mathbf{Y}}_i),$$

where $Cov(\mathbf{r}_i) \approx Cov(\mathbf{e}_i) = \mathbf{\Sigma}_i$ and $\hat{\mathbf{\Sigma}}_i = \mathbf{L}_i\mathbf{L}_i'$ is a Cholesky decomposition of $\hat{\mathbf{\Sigma}}_i$ (an estimate of $\mathbf{\Sigma}_i$) and $\mathbf{L}_i$ is a lower triangular matrix. It is indicated that the transformed residuals are uncorrelated and have a unit variance. However, it is important to note that $\mathbf{\Sigma}_i$ is not a covariance matrix for the residuals and that $\mathbf{L}_i$ are very complicated functions of $\mathbf{Y}$ and hence are random which obviously tells us that they can not be treated as constants. Moreover, it is possible to show that $\hat{\mathbf{\Sigma}}_i$ is not independent of $\mathbf{r}_i$ which makes things even more complicated.

That is one of the reasons why one should study ordinary residuals in repeated measurement and longitudinal analysis more carefully. These have a close relation with the GC and EGC models since they both have between and within individual assumptions. Moreover, the most common applications of the GC and EGC models are in the analysis of repeated measurement and longitudinal data. Due to the bilinear structure in the models, the residuals have different components and hence one needs to define new residuals.

The new residuals defined in von Rosen (1995b) for the GC model and those defined in Paper II for the EGC model contain information about the model and model assumptions. For example, consider the following residual for the GC model,

$$\mathbf{R}_{g1} + \mathbf{R}_{g2} = \mathbf{X}(\mathbf{I} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}), \tag{7.1}$$

where $\mathbf{R}_{g1}$ and $\mathbf{R}_{g2}$ are the residuals given in Chapter 2.

Note that (7.1) is the difference between the observations and the corresponding means, moreover, it is distributed as a multivariate normal random variable. These residuals can, for example, be used to investigate the normality assumption and can also indicate if there are outliers in the data. It can be performed by using existing graphical or formal tests available for multivariate normal distributions. It is also possible to show that (7.1) is uncorrelated with the predicted values. As a result one can plot these residuals against the predicted values to check for homogeneity of variances.

The equivalent expression for the EGC model given in Definition 3.3 is

$$\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{X}(\mathbf{I} - \mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^{-}\mathbf{C}_1), \tag{7.2}$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are the residuals given in Paper I. The expression given in (7.2) has the same interpretation, property and application as (7.1).

Now consider the following for the GC model,

$$\mathbf{R}_{g3} = (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1})\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}, \tag{7.3}$$

which can be rewritten as

$$\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C} - \mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-}\mathbf{C}.$$

The above expression is the difference between the observed and the estimated means. We can use these residuals to check if the estimated linear curve fits the observed mean well.
The equivalent expression for the EGC model is

$$\mathbf{R}_3 = (\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-}\mathbf{A}_1)^{-}\mathbf{A}_1'\mathbf{S}_1^{-1})\mathbf{X}(\mathbf{C}_1'(\mathbf{C}_1\mathbf{C}_1')^{-}\mathbf{C}_1 - \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-}\mathbf{C}_2).$$

This expression tells us how close the observed mean and estimated mean are for individuals with linear mean structure (see Paper I). For individuals with quadratic mean structure the following expression can be used to check if the estimated quadratic curve fits the data

$$\mathbf{R}_4 = \{\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1'\mathbf{S}_1^{-}\mathbf{A}_1)^{-}\mathbf{A}_1'\mathbf{S}_1^{-1} - \mathbf{T}_1\mathbf{A}_2(\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{A}_2)^{-}\mathbf{A}_2'\mathbf{T}_1'\mathbf{S}_2^{-1}\} \\ \times \mathbf{X}\mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-}\mathbf{C}_2.$$

The above mentioned interpretations of our residuals are some of the advantages, to say the least, which can not be obtained by considering ordinary residuals, i.e. this is not possible without a decomposition. Indeed, each of the residuals defined in Paper I bear important information about the bilinear structure in the model and we may use these information to validate model assumptions and detect outliers. Moreover, unlike most methods provided for analyzing longitudinal and repeated data, our approach does not relay on numerical solutions or asymptotic results.

We would like to mention that the groups involved in Paper I do not have to follow linear and quadratic growth curves. The only assumption required is the nested subspace condition mentioned in Definition 3.3. Moreover, it is important to note that the method applied in defining the residuals in Paper I is different from that of von Rosen (1995). Our approach is better in the sense that

- We obtained the space generated by the design matrices in a natural way. As a result, we do not have a problem in choosing the right inner product when we have more than one.
- We have shown that, like univariate and ordinary MANOVA models, residuals are obtained by projecting the observation matrix on the space orthogonal to the one generated by the design matrices.
- We also have the advantage that our methods can be utilized in defining residuals in the more general model given in Definition 3.1.

As mentioned before, we could use our residuals for checking model assumptions such as the assumption of normality and to check for extreme observations. It was also mentioned that we could check how well the estimated curves fit the data. We may do this in two ways. One way is to look at appropriate residuals and see how close they are to zero. In Paper I, we used the standard errors to see how small the values are. In Paper II, however, we proposed and discussed a different approach, via testing appropriate hypotheses.

Inspired from the bilinear nature of the model, we looked upon the problem as a two stage problem, i.e.,

- We estimated the covariance matrix $\mathbf{\Sigma}$ by maximizing an appropriate part of the likelihood.
- We replaced the covariance matrix by its estimator and defined the tests by taking the ratio of the likelihood under H_o and $H_o \cup H_1$.

It is important to note that, in the absence of information about $\mathbf{B}$, it is possible to show that the part of the likelihood used in estimating $\mathbf{\Sigma}$ gives all the information contained in the data. We refer to Sprott (1975) for details about marginal and conditional sufficiency. Here are some of the main results from Paper II:

- A reasonable and natural approach was used to obtain the tests.
- As expected, the tests turned out to be functions of appropriate residuals. As a result it is easy to understand, interpret and study them.
- The tests are easy to calculate and hence are easy to use in practice.
- The distributions of the test statistics under the null hypotheses are independent of any unknown parameters.

- The distributions are difficult to obtain but due to the existence of a natural ancillary statistics it is reasonable to consider the conditional version of the tests.

As mentioned above, the distribution of the tests are difficult to obtain. Two suggestions are given in Paper II to overcome this problem. The first one is to approximate the density using the first two moments. The conditional and unconditional expectations for the tests are given. It is also possible to get the second moments for the tests. However, due to the existence of an ancillary statistic the problem can be treated in the following way as a two stage experiment, i.e.,

- Observe the ancillary statistic $\mathbf{S}$.
- Calculate the critical point conditioned on $\mathbf{S}$.

In paper III, we considered one of the tests defined in Paper II and the above mentioned conditional approach was used to find the critical point for a given level of significance. The conditional distributions, both under the null and alternative hypotheses, are much simpler to deal with. Moreover, conditioning with the ancillary statistic makes the problem more relevant to the situation at hand without losing any information about the parameter of interest. Beside the above two benefits we obtained from utilizing the conditional approach, we also have the following additional advantages which allow as to use existing results for such sums.

- Under the null hypothesis, the conditional test can be written as sums of central chi-square random variables.
- Under the alternative hypothesis, the conditional test can be written as sums of non-central chi-square random variables.

The distribution under the alternative hypothesis depends on the unknown covariance matrix $\mathbf{\Sigma}$. In practice, one, therefore, needs to estimate $\mathbf{\Sigma}$ to get an estimator for the power of the test which may be used as a measure of performance. Three alternative estimators are suggested in Paper III.

Numerical examples were given in Paper I and Paper III to illustrate the results. In Paper I, the Potthoff & Roy (1964) data was considered. It was assumed that the girls and boys follow linear and quadratic mean structures, respectively. The standard errors were used to see if the residuals are reasonably small. The residuals obtained indicate that the assumed growth curves fit the data well. However, only small improvements were obtained due to the quadratic term (for the boys) and there is a need to investigate the significance of this term. This can, for example, be done using $\phi_5(\mathbf{X})$ given in Paper II. Some extreme observations were also detected.

We could also use $\mathbf{R}_1 + \mathbf{R}_2$, given in (7.2) to check the assumption of normality. We may first check if residuals at different ages are univariate normal. There are many ways, both graphical and formal, to check for a univariate normality. We shall only provide a brief discussion about how we may assess the multivariate normality assumption for the Potthoff & Roy data through $\mathbf{R}_1 + \mathbf{R}_2$. We can use any of the available methods for checking multivariate normality. Here we use Small's graphical method, see Sirvastava (2002). Such graphical approaches also help us to see if there are outliers in the data. The idea behind Small's graphical approach is to reduce the multivariate data to a univariate one.

Suppose $\mathbf{x}_1$, $\mathbf{x}_2$,...,$\mathbf{x}_n$ are independently distributed as $N_p(\mu, \mathbf{\Sigma})$. Then the statistic

$$c_i = n(n-1)^{-2}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}), i = 1, 2, ..., n,$$

where

$$\bar{\mathbf{x}} = n^{-1}\sum_{j=1}^{1}\mathbf{x}_j,\ (n-1)\mathbf{S} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

has a beta distribution with parameters $\alpha = \frac{1}{2}p$ and $\beta = \frac{1}{2}(n - p - 1)$ (see Srivastava, 2002). It was also mentioned that asymptotically, c_i may be considered as independently distributed. The following probability plot shows the plot of the $c_i's$ obtained for $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_2$.

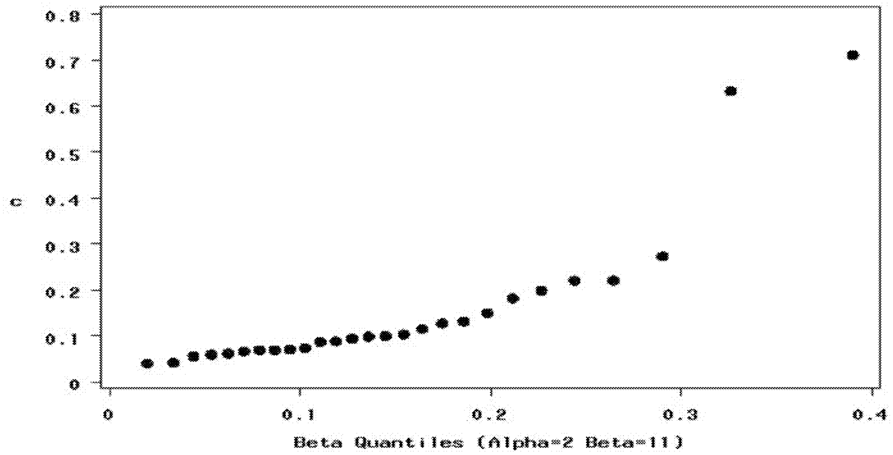

Figure 7.1. The plot of c_i's obtained using Small's graphical method for the Potthoff & Roy data gainst Beta(2,11) quantiles.

The above plot shows that the error terms can be considered as normal except for the two outliers. As mentioned earlier, we could plot the residuals given in (7.2) against the predicted values or against age to check if the variances at different ages can be regarded as constant. Below is the plot of the residuals given in (7.2) against age for the Potthoff & Roy data.

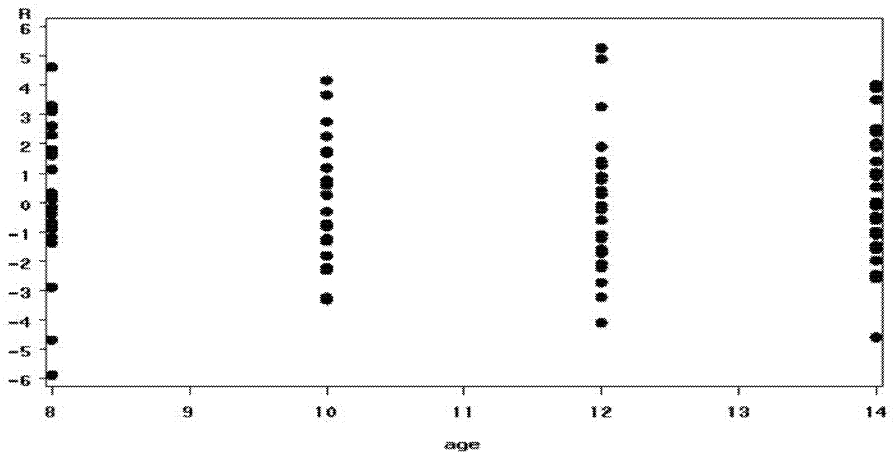

Figure 7.2. The plot of $\mathbf{R}$ against age for the Potthoff & Roy data.

Only a small difference in the variation of residuals at different ages is indicated in the above plot. That is, there is no enough evidence that shows the variances are different at different ages. However, we suggest that other investigations should be made before assuming equal variances.

The same data was considered in Paper III. This time the Growth Curve model were fitted and the growth curves for girls and boys are assumed to be linear. The adequacy of the assumed model was investigated through the hypothesis that $\mathbf{B} = \mathbf{0}$. A conditional approach was used to calculate the critical point and an approximate distribution was provided using Satterthwaite approach. The data gives strong evidence towards rejecting the hypothesis. That indicates that the mean structures for the boys and girls can be regarded as a polynomial of at least first degree. However, results from Paper I show that the growth curves can be regarded as linear although there is a need to check if a quadratic term is necessary for the growth curve of the boys. Moreover, some nice properties that need to be investigated further are observed and will be given in the next section.

7.2 Future research

We believe that our residuals give a new approach in validating bilinear models and that there is much left to be done in the future. However, it is our hope that our approaches will lay the ground for further studies towards developing diagnostic tools for validating such models.

Moreover, we have shown that residuals in bilinear models such as the GC and EGC models are defined by projecting the observation matrix on the space orthogonal to the space generated by the design matrices. It is our hope that this concept and the way we looked upon the space generated by the design matrices and the space on which the residuals are defined can be used to understand residuals in more complicated models, such as linear mixed models. We believe that the concept of decomposing the spaces to understand the residuals could be applied in such models. It is important to note here that linear mixed models are most commonly used in analyzing repeated measures and longitudinal data, and they are especially important when we have observations taken at different time points and when we have unbalanced data.

As mentioned in the previous section, the method used in Paper I can be used to define residuals in the more general model given in Definition 3.1. Moreover, the approaches utilized in Paper III to find approximate distributions for the test in the GC model can be applied to find approximate distributions for the rest of the tests defined in Paper II.

Furthermore, the following nice properties that need to be investigated further was observed from the numerical examples used in Paper III.

- The power of the test is strictly "monotone" in $\mathbf{B}$.
- The test is unbiased.
- The estimated power obtained using $(1/n)\mathbf{S}$ as an estimator of $\mathbf{\Sigma}$ underestimates the power.

We have also used random data from a multivariate normal distribution to see the performance of the test in making the right decision, i.e, the test should not reject the hypothesis that $\mathbf{B} = \mathbf{0}$. As expected, there was not enough evidence to reject the hypothesis. However, there is a need to do more simulation studies to support our findings as well as to discover other properties for the test statistic.

Finally, we would like to note that the hypotheses considered in Paper II and III can be formulated in a more general form to include many possibilities, i.e.,

$$\begin{aligned} H_o &: \mathbf{FBG} = \mathbf{0}, \\ H_1 &: \mathbf{FBG} \neq \mathbf{0}, \end{aligned}$$

where $\mathbf{F}$ and $\mathbf{G}$ are any two matrices. This kind of general formulation can for example be used if one is interested in comparing two growth curves which could be done by choosing suitable elements for the matrices $\mathbf{F}$ and $\mathbf{G}$.

Bibliography

[1] Banken, L. (1984). *Eine Verallgemeinerung des GMANOVA Modells.* Dissertation, University of Trier, Trier, Germany.

[2] Basu, D. (1964). Recovery of ancillary information. *Sankhyā Ser. A* **26**, 3-16.

[3] Basu, D. (1975). Statistical information and likelihood. *Sankhyā Ser. A* **160**, 268-282.

[4] Basu, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* **72**, 355-366.

[5] Basu, D. (1978). On partial sufficiency: A review. *J. Statist. Plann. Inference* **2**, 1-13.

[6] Burnaby, T.P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics* **22**, 96-110.

[7] Draper, N.R. & Smith, H. (1998). *Applied Regression Analysis.* Wiley, New York.

[8] Fisher (1956). *Statistical Methods and Scientific Inference*, Oliver & Boyed, London, UK.

[9] Fizmaurice, M.G., Laird, M.N. & Ware, H.J. (2004). *Applied Longitudinal Analysis.* Wiley, New York.

[10] Fujikoshi, Y. (1974). Asymptotic expansions of the non-null distributions of three statistics in GMANOVA. *Ann. Inst. Statist. Math.* **26**, 289-297.

[11] Gleser, L.J. & Olkin, I. (1970). Linear models in multivariate analysis, *Essays in Probability and Statistics*, 267-292, The University of North Carolina Press, Chapel Hill, North Carolina.

[12] Imhof, J.P. (1961), Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

[13] Kariya, T. (1978). The General MANOVA problem. *Ann. Statist.* **6**, 1, 200-214.

[14] Khatri, C.G. (1966). A note on a MANOVA model applied to the problems in growth curve. *Ann. Inst. Statist. Math.* **18**, 75-86.

[15] Khuri, A.I., Mathew, T. & Sinha, B.K. (1998). *Statistical Tests for Mixed Linear Models.* Wiley, New York.

[16] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Springer, New York.

[17] Kotz, S., Johnson, N.L. & Boyd, D.W. (1967). Series representations of distributions of quadratic forms in normal variables I. central case, *Ann. Math. Statist.* **38**, 823-837.

[18] Kshirsagar, A.M. & Smith, W.B. (1995). *Growth Curves*, Marcel Dekker, New York.

[19] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, Springer, New York.

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

[21] Mathai, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Ann. Inst. Statist. Math.* **34**, 3, A, 591-597.

[22] Moschopoulos, P.G. (1985). The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* **37**, A, 541-544.

[23] Pan, J.X. & Fang, K.T. (1995). Multiple outlier detection in growth curve model with unstructured covariance matrix. *Ann. Inst. Statist. Math.* **47**, 137-153.

[24] Pan, J.X. & Fang, K.T. (1996). Influential observations in the growth curve model with unstructured covariance matrix. *Comput. Statist. Data Anal.* **22**, 71-87.

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

[26] Patterson, H.D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545-554.

[27] Potthoff, R.F. & Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curves. *Biometrika* **51**, 313-326.

[28] Press, S.J. (1966) Linear combination of non-central chi-square random variates. *Ann. Math. Statist.* **37**, 480-487.

[29] 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.

[30] von Rosen, D. (1989). Maximum likelihood estimators in multivariate linear normal models. *J. Multivariate Anal.* **31**, 187-200.

[31] von Rosen. D. (1990). Moments of a multivariate linear model with an application to the growth curve model. *J. Multivariate Anal.* **35**, 243-259.

[32] von Rosen, D. (1995a). Influential observations in multivariate linear models. *Scand. J. Statist.* **22**, 207-222.

[33] von Rosen, D. (1995b). Residuals in the growth curve model. *Ann. Inst. Statist. Math.* **47**, 129-136.

[34] Satterthwaite, F.E. (1941). Synthesis of variance. *Psychometrika* **6**, 309-316.

[35] Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics* **2**, 110-114.

[36] Searle, S.R. et al. (1992). *Variance Components.* Wiley, New York.

[37] Seid Hamid, J. & von Rosen (2005a). Residuals in the extended growth curve model. (Accepted by Scand. J. Statist.)

[38] Seid Hamid, J. & von Rosen (2005b). Hypothesis testing via residuals in two GMANOVA models. Report 2005:4, Centre of Biostochastics, SLU, Sweden.

[39] Seid Hamid, J. & von Rosen (2005c). An approximate critical point for a test in the growth curve model: A conditional approach. Report 2005:5, Centre of Biostochastics, SLU, Sweden.

[40] Sen, A. & Srivastava, M.S. (1990). *Regression Analysis, Theory and Applications*. Springer, New York.

[41] Sprott, D.A. (1975). Marginal and conditional sufficiency. *Biometrika* **62**, 3, 599-605.

[42] Srivastava, M.S. & Carter, E.M. (1983). *An Introduction to Applied Multivariate Statistics*. North Holland, New York.

[43] Srivastava, M.S. & Khatri, C.G. (1979). *An Introduction to Multivariate Statistics*. North Holland, New York.

[44] Srivastava, M.S. (2002). *Methods of Multivariate Statistics*. Wiley, New York.

[45] Takemura, A. (1983). Tensor analysis of ANOVA decomposition. *J. Amer. Statist. Assoc.* **78**, 894-900.

[46] Tan, W.Y. & Gupta, R. P. (1983). On approximating a linear combination of central Wishart matrices with possitive coefficients. *Comm. Statist. Theory Methods.* **12**, 22, 2589-2600.

[47] Verbyla, A.P. & Venables, W.V. (1988). An extension of the growth curve model. *Biometrika* **75**, 129-138.

[48] Wishart, J. (1938). Growth-rate determination in nutrition studies with the bacon pig, and their analysis. *Biometrika* **30**, 16-28.

Acknowledgements

First of all, I would like to thank everyone who directly or indirectly contributed to this thesis in any way.

In particular, I wish to express my sincere and heartfelt gratitude to my supervisor, Professor Dietrich von Rosen, for his unfailing guidance, continues support and encouragements through out my study. Thank you for introducing me to the Growth Curve model and multivariate linear and bilinear models in general. You were also there when I needed a friend to talk too. Thanks, Dietrich, for everything. Some people are born to make life easier for others and you are definitely one of them for me.

I would like to thank all my colleagues at the department of Biometry and Engineering. Special thanks to PhD students at the Biometry group (Razaw, Kristi, Zhanna, Johannes and Karin) for the nice and friendly environment. The department would not be the same with out you. I would also like to thank my former colleagues at the department of Mathematics, Uppsala University. Special thanks to Dr. Dag Johnson and Professor Allan Gut.

Thank you Professor Sinha for the fruitful discussions we had when you were here as a guest. Many thanks to Professor Kollo, my licentiate opponent, for your valuable comments and suggestions.

Thank you my three angels (Rima, Hala and Hiba) for being there for me and the girls when ever we needed you. I love you all and the girls adore you. Many thanks to Maedot for your help on the most difficult day of my life. Thanks Nazu, my dear, for being my friend and for being always there for me.

I would also like to thank all my friends in particular Meaza, Abigeal, Tikle, Wendim, Mussie and Yordanous. Thank you for your support and encouragements.

This thesis was partly financed by SIDA through ISP. I am grateful to both for their support. I would also like to thank people at ISP (Staffan, Roland, Pravina, Mona and Åsa). I wish to express my gratitude to Biostochastism for providing me with a part time job and for financing the last year of my study. Thank you Dietrich for finding me the fund.

My deepest and warmest thanks goes to my family especially Abdu. Thanks for everything especially for letting me bring the children with me. I understand how life must have been hard for you without them. May God be with you.

Last but not the least I would like to thank my precious girls, Yasmine and Iman, for making every day of mine brighter. I am grateful to the Almighty, God, for blessing me with them and making it possible for me to see this day.