AN INVESTIGATION OF CERTAIN METHODS IN THE ANALYSIS OF GROWTH CURVES

Vassilios G. S. Vasdekis

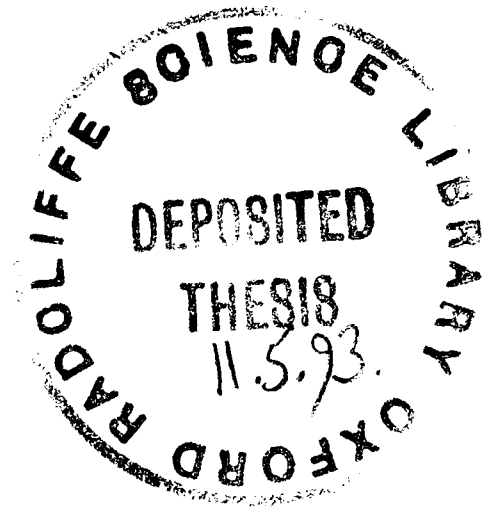
St. Peter's College

A thesis submitted to the Faculty of Mathematical Sciences

for the degree of Doctor of Philosophy of the University of Oxford.

Department of Statistics

University of Oxford



Hilary Term, 1993

AN INVESTIGATION OF CERTAIN METHODS IN THE ANALYSIS OF GROWTH CURVES

Vassilios G. S. Vasdekis
St. Peter's College, Oxford

D.Phil. thesis
Hilary Term, 1993

ABSTRACT

In growth curve studies, measurements are made on individuals over a moderately large period of time and the main problem is set as the best possible evaluation of the curve which is believed to underlie the phenomenon. One important meaning of the term 'best possible evaluation' is the efficient estimation of the coefficients which, postulated by the model, characterize this curve. The important key for doing this, is the 'best' (in a certain way) estimation of the covariance matrix of observations which is supposed here to be the same for all individuals. The purpose of this thesis is to investigate certain methods which have been suggested to be appropriate for this problem both theoretically and empirically as well as to prove certain results concerning the problem of efficiency itself when an estimate of the covariance matrix of observations, irrespective of the method which calculated it, is at hand. More specifically, it is shown that REstricted Maximum Likelihood (REML) gives in general estimates of the regression coefficients whose estimated variance lies nearer to their true value quite independently from the form of the covariance matrix, than Maximum Likelihood (ML) and it is proved that the general estimated scatter of the estimated regression coefficients is greater for REML. When we measure more than one characteristics on time and for a special class of covariance models, the property that the REML estimates of the variance of estimated regression coefficients are larger than the corresponding ML,

which holds for one characteristic, is lost. Asymptotically however, the two methods give identical results.

Another method which is not based on the adoption of a certain parametric form for the covariance matrix is considered. A new method is suggested for its optimal choice and a comparison is made between this and three other already known methods. Empirical results suggest that it retains a very good balance between the variance of the regression coefficients and their real and optimal value for the majority of the covariance models which have been tested as possible population covariance matrices.

Finally, upper and lower bounds of different types of efficiency are obtained by assuming that an estimate of the population covariance matrix has already been calculated and is not distant from the true covariance matrix more than a certain constant.

Acknowledgements

First, I would like to thank my supervisor Dr. F. H. C. Marriott for his advice throughout the course and especially for his valuable and patient guidance during this last year.

I would also like to thank Dr. I. G. Vlachonikolis for supervising me during the first two years of the course and supported me in many ways.

All members of the department were very kind to me and provided me a very pleasant working environment. It is my pleasure to record my thanks especially to Ms. S. R. Hutchinson for her advice in computing matters as well as to Mrs. A. S. Margetts and to Miss E. B. Green.

Without the financial support of the State Scholarships Foundation of Greece (I.K.Y), this research could not have been fulfilled.

I have made some new friends in this town during my 4 and so years of residence. Among them Stefanos, Petros, Christoforos and Arif made feel the meaning and the warmth of being a friend.

My parents supported me morally and materially throughout my life and I give them my deepest respect and gratitude.

Finnaly, I would like to thank Gianna for making me the happiest man in this world last year.

V. G. S. Vasdekis

Oxford, England

January, 1993.

To my parents

To Gianna

Contents

Chapter 1

Introduction

Τά πάντα ρεῖ

Growth curve analysis applies to data consisting of measurements taken on different individuals and over different time points. Many scientific areas deal with the growth of specific natural phenomena (physics, biology, geology, seismology), but the underlying mechanisms are very complex for significant progress to be made. By the term "mechanisms" we mean that although many things have been explained experimentally, no one can predict with absolute certainty the future "action" of a specific phenomenon (for example, we cannot predict earthquakes). If probability and statistics is a human attempt to understand, in a rough way, how nature works, then growth curve analysis can become a powerful tool in the understanding of the effects of evolutionary factors in natural phenomena.

However, there are two distinct points of view in growth curve literature. The first is the biological one, which seeks models with a biological basis and biologically meaningful parameters. Usually these models are deterministic, in the sense that they do not include any random source of variation and they usually arise as a result of assumptions concerning the type of growth, environmental factors etc. One then could write down the differential or difference equations which represent these assumptions and solve them to obtain a growth model. Such a typical differential equation will have the form:

$$dP/dt = \mathrm{f(P)}$$

where f is a function of the population size P. For example, we can consider the growth rate dP/dt to be such that:

$$dP/dt = \frac{kP(\alpha - P)}{\alpha}$$

that is, to be proportional to the product of the present size and the future amount of growth, since α is some limiting growth value. This equation is called the logistic equation or the Verhulst–Pearl equation. Using fraction decomposition and integrating, we obtain:

$$P(t) = \frac{\alpha}{1 + \beta \exp(-kt)}$$

where $\beta = \frac{\alpha}{k}c$ and c is the constant of integration. It is easily seen that $\lim_{t\to\infty} P = \alpha$ which is the limiting size of the population (Examples of fitting the logistic curve can be found in Nelder (1961) or Oliver (1964)).

Another well known curve is the Gompertz curve which arises if we assume that:

$$dP/dt = kP \log(\alpha/P)$$

By integration, we obtain that

$$P = \alpha \exp(-\beta \exp -kt)$$

This curve has been used for population studies and animal growth (Draper and Smith 1981). We note that $\lim_{t\to\infty} P(t) = \alpha$ which is again the limiting growth value.

The critical assumption behind these deterministic (mechanistic) models is that the entire future behaviour of the system is exactly and explicitly determined by the present status of the system i.e. if we know everything about the system at a particular moment we can predict its behaviour at any future instance. In practice these models provided good predictions of the future, but the above assumption is judged to be rather unrealistic.

The second point of view under which a growth curve has been considered, is the statistical one. This deals with a class of growth models, usually polynomial ones which are analysed allowing for the dependence between different measurements. Historically,

one of the most classical papers and one of the first on the subject, was that of John Wishart (1938). In this publication, Wishart introduced a class of models which had great influence on applied and theoretical studies of growth. His approach was actually to use least squares and fit linear and quadratic terms using orthogonal polynomials. Linking this approach with the point of view described previously (the biological one) one could say that Wishart proposed the use of polynomials as a powerful tool of describing the response of short term phenomena, when nothing is known about the parametric form of their true curve.

Later, this approach was extended by Rao (1959) and Elston and Grizzle (1962): Let us denote by $\boldsymbol{y}$ the $p \times 1$ vector of the mean responses of observations of an individual taken on each at p time points. Suppose that the growth curve of each (of n in total) individual is represented by the model:

$$E(\boldsymbol{y}) = A\boldsymbol{\beta} \tag{1.1}$$

where A is a known $p \times m$ matrix and $\boldsymbol{\beta}$ is a $m \times 1$ column vector of unknown parameters. Then making no assumptions about the true variance-covariance matrix of the p time points, Rao (1959) pointed out that the best unbiased estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (A^\top S^{-1} A)^{-1} A^\top S^{-1} \boldsymbol{y}$ where S is the sample $p \times p$ variance-covariance matrix of the p time points. He also tested the adequacy of the model and finally, he obtained confidence limits for the unknown parameters $\boldsymbol{\beta}$.

On the other hand, Elston and Grizzle (1962) formed the equivalent of the mixed model in the univariate analysis of variance by assuming that:

$$y_{ij} = \sum_{k=0}^{m-1} x_{kj}\beta_k + \sum_{k=0}^{m-1} x_{kj}\lambda_{ik} + \epsilon_{ij}$$

where y_{ij} is the response of the ith individual at the jth time point $i = 1, \ldots, n$, $j = 1, \ldots, q$, x_{ij} are the elements of the design matrix with (possibly) orthogonal columns, β_k are some fixed effects $k = 0, \ldots, m-1$ and $\lambda_{ik} = \beta_{ik} - \beta_k$ are the deviations from the ith individual's own growth curve. In matrix notation, this is equivalent to the model:

$$Y = BX + E$$

where Y is a $n \times q$ matrix, B is a $n \times m$ matrix of unknown (stochastic) coefficients whose each element of column $k+1$ has the same mean value $\beta_k \ \ \forall\, k = 0, \ldots, m-1$ (that is, each individual has the same mean vector), X is a $m \times q$ known design matrix with orthogonal columns and E is a $n \times q$ matrix of errors. The assumptions now, concerning the components λ_{ik} , $i = 1, \ldots, n$, $k = 0, \ldots, m-1$ are that for each i and k, λ_{ik} are independently and normally distributed with variance σ_k^2 and that ϵ_{ij} are independently and normally distributed with variance σ_ϵ^2. It follows that y_{ij} and $y_{i'j}$ $(i \neq i')$ are independent and $y_{ij}, y_{ij'}$ have covariance

$$\sum_k \sigma_k^2 x_{kj} x_{kj'} \qquad (j \neq j')$$

The authors obtained estimates only for the fixed effects β_k, $k = 0, \ldots, m-1$, a test for the adequacy of the model and confidence limits for the parameters $\boldsymbol{\beta}$, but as they noted in the last part of their paper, the results obtained by just one example made it impossible to say which of the two methods (Rao's or theirs) should be preferred in general. This was also the first attempt for a comparison between a method which postulates a parametric form for the covariance matrix of observations and a method which does not. This comparison would be more explicit in more recent years.

Potthoff and Roy (1964) noted that these different approaches dealt with specialized aspects of the growth curve problem and in their attempt to furnish new tools for growth curve analysis, they provided a generalization of the multivariate analysis of variance.

The usual MANOVA model is:

$$E(Y) = A\xi \tag{1.2}$$

where Y is a $n \times q$ matrix whose different rows are mutually independently distributed and the q elements follow a $q-$variate distribution with unknown variance covariance matrix Σ, so that:

$$Var(vecY) = \Sigma \otimes I$$

Here, A is a $n \times m$ matrix of known constants with $r(A) = m$ and ξ is a $m \times q$ matrix of unknown parameters. Roy (1957) has shown that the testing of hypotheses of the

form:

$$C\xi V = 0 \tag{1.3}$$

where C is a $s \times m$ matrix, ξ an $m \times q$ matrix and V is a $q \times v$ matrix, is connected directly with two quantities S_h and S_e which for the special case of the one-way multivariate analysis of variance represent the between samples sum of squares and products (SSP) and the within SSP matrices respectively. To test (1.3) we have three proposed test statistics:

- Roy's test which uses the largest characteristic root of $S_h S_e^{-1}$
- A test developed by Hotelling (called the sum-of-the-roots test) based on the trace of $S_h S_e^{-1}$, and
- The Wilks's λ-criterion which uses the statistic $\mid S_e \mid / \mid S_h + S_e \mid$.

Although MANOVA models have been found to be extremely useful in practical work, they have some limitations if, for example, the q different columns represent q time points and one would like to fit growth curves to the average growth of the population over time. Potthoff and Roy (1964) provided a generalization to the MANOVA problem by parameterizing ξ matrix to be of the form $\xi_0 P$ where now ξ_0 is a $m \times q$ matrix of unknown parameters and P a $p \times q$ matrix of known constants. We, thus, obtain (by dropping the subscript 0 for convenience):

$$E(Y_{n\times q}) = A_{n\times m}\xi_{m\times p}P_{p\times q} \tag{1.4}$$

The authors described some of the possible applications of such a model to the context of growth modelling. With respect to estimation, first, they considered a transformation of the initial data, of the form:

$$Y_0 = YG^{-1}P^T(PG^{-1}P^T)^{-1} \tag{1.5}$$

Clearly, if $G = \Sigma$, Y_0 is the best row-regression estimator of the regression of Y on P. In general, G must be a symmetric positive definite (weight) matrix, either non-stochastic

or stochastic but independent of Y, in order to allow normality or conditional normality for the transformed Y_0 matrix. Hence, model (1.4) is reduced to model (1.2) and can be analysed in the same way as an ordinary MANOVA model. The new covariance matrix for each row of the transformed data is easily seen to be:

$$\Sigma_0 = (PG^{-1}P^T)^{-1}PG^{-1}\Sigma G^{-1}P^T(PG^{-1}P^T)^{-1} \tag{1.6}$$

Proposition 1 *Among all estimators of the form* $\boldsymbol{c}^T Y \boldsymbol{a}$, *the best unbiased estimator of* $\boldsymbol{b}^T C\xi V\boldsymbol{d}$ *is:*

$$\boldsymbol{b}^T C_1(A^TA)^{-1}A^TY\Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}V\boldsymbol{d} \tag{1.7}$$

where C_1 $(s \times r)$ *is a submatrix of* C *with* $r(C_1) = r(C) = s$

Proof : One can easily see that necessary and sufficient conditions for $\boldsymbol{c}^T Y\boldsymbol{a}$ to be an unbiased estimator of $\boldsymbol{b}^T C\xi V\boldsymbol{d}$ are:

$$k\boldsymbol{c}^T A = \boldsymbol{b}^T C \text{ and } 1/k(P\boldsymbol{a}) = V\boldsymbol{d} \;\; k \in \boldsymbol{R} \tag{1.8}$$

Without loss of generality, we can assume that $k = 1$. Now, $Var(\boldsymbol{c}^T Y\boldsymbol{a}) = \boldsymbol{c}^T\boldsymbol{a}^T\Sigma\boldsymbol{a}\boldsymbol{c} = \boldsymbol{c}^T\boldsymbol{c}(\boldsymbol{a}^T\Sigma\boldsymbol{a})$ since $\boldsymbol{a}^T\Sigma\boldsymbol{a}$ is a scalar. Thus, if we want to minimize $Var(\boldsymbol{c}^T Y\boldsymbol{a})$, we should minimize $\boldsymbol{c}^T\boldsymbol{c}$ under the restriction $\boldsymbol{c}^T A = \boldsymbol{b}^T C$ and $\boldsymbol{a}^T\Sigma\boldsymbol{a}$ under the restriction $P\boldsymbol{a} = V\boldsymbol{d}$. Applying Lagrange multipliers we obtain

$$\boldsymbol{c} = A(A^TA)^{-1}C_1^T\boldsymbol{b}$$

$$\boldsymbol{a} = \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}V\boldsymbol{d}$$

Hence

$$\boldsymbol{b}^T C_1(A^TA)^{-1}A^TY\Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}V\boldsymbol{d}$$

is the best unbiased estimator of $\boldsymbol{b}^T C\xi V\boldsymbol{d}$.

The main justification of the above model is that estimator (1.7) is used in the usual MANOVA (for hypothesis testing or construction of confidence intervals) when Y is replaced by the transformed Y_0 and $G = \Sigma$. The problem of the choice of the best weight matrix G in (1.5) led many researchers to structure the covariance matrix

of observations by providing reasonable forms and using likelihood based methods for the estimation of the components of the linear model. The most obvious of them, Maximum Likelihood (ML), was a method with many desirable properties, but with some covariance forms and in small samples, it was noted to be somewhat biased and inefficient. Attempts to improve its performance led to the introduction of the Restricted (or Residual) Maximum Likelihood (REML) which in small samples is less biased for a certain class of covariance models, while asymptotically it is equivalent to ML. In both cases, the log-likelihood of the observations or of functions of observations is maximized.

A critical point in Potthoff and Roy's approach is that of the choice of matrix G involved in transformation (1.5). G cannot be equal to Σ as this is not known in applications. Although, the authors discussed other possible solutions, such as $G = I_q$, it is clear that matrix G introduces an element of arbitrariness which we wish to avoid. Of course the simplest case would be when in (1.4) $p = q$. Then if P is of full rank one could have $Y_0 = YP^{-1}$, and no choice of G is required. Rao (1965) commenting on Potthoff and Roy's method suggested an alternative approach which avoids the arbitrariness of G. This is as follows: First we construct a non–singular matrix $H = (H_1, H_2)$ such that the columns of H_1 form a basis of the space generated by the rows of P such that $PH_1 = I_p$ and H_2 satisfies $PH_2 = 0$. With such a choice of H (which is not necessarily unique), (1.4) becomes

$$E(YH_1) = A\xi PH_1 \ , \ E(YH_2) = 0$$

By definition of H_1, the product matrix PH_1 is of rank p (since $r(P) = p$). Finally, we have:

$$E(Y_1) = A\xi, \quad E(Y_2) = 0 \ ,$$

where $Y_1 = YH_1$, $Y_2 = YH_2$. The conditional expectation of Y_1 given Y_2 can be written as:

$$E(Y_1 \mid Y_2) = A\xi + Y_2\rho$$

where ρ is a matrix of unknown parameters. By taking

$$H_1 = G^{-1}P^T(PG^{-1}P^T)^{-1} \text{ and } H_2 = (I_q - H_1P)Z$$

where Z is an arbitrary $q \times (q-p)$ matrix, we reach the transformation suggested by Potthoff and Roy and we can see that the information contained in $Y_2 = Y(I_q - H_1P)Z$ is ignored. Intuitively, we should expect that transformation (1.5) ignores some information since the order of the initial data matrix Y is $n \times q$ and that of Y_0 is $n \times p$, $p \leq q$.

It is evident from the above discussion that the purpose of Rao's approach was to improve all initial estimates of the unknown parameters using the additional information of concomitant variables. In his paper, Rao (1965) illustrated this technique and called it *Covariance Adjustment*. In 1967, Rao, in order to clarify the whole technique, approached the problem from a different point of view. He exploited the fact that when two vectors $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ follow a joint multivariate normal distribution then $\boldsymbol{t}_1 \mid \boldsymbol{t}_2$ follows again multinormal distribution with a certain mean and covariance matrix. His approach is as follows: Let us have two vector statistics of order k and r respectively, such that:

$$E(\boldsymbol{t}_1) = \boldsymbol{\tau} \quad E(\boldsymbol{t}_2) = 0$$

where $\boldsymbol{\tau}$ is a vector of k unknown parameters. The vector $\boldsymbol{t}_1$ is obviously an unbiased estimator of τ, but if we know the covariance matrix of $(\boldsymbol{t}_1, \boldsymbol{t}_2)$ then a better estimate can be obtained. Let us denote by Λ, the covariance matrix of $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, denoted by

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$$

Let us also consider the estimator $\boldsymbol{\tau}^* = \boldsymbol{t}_1 - \Lambda_{12}\Lambda_{22}^{-1}\boldsymbol{t}_2$. This is an unbiased estimator of $\boldsymbol{\tau}$ and is more efficient than $\boldsymbol{t}_1$. This is easily seen since $E(\boldsymbol{t}_1 - \Lambda_{12}\Lambda_{22}^{-1}\boldsymbol{t}_2) = \boldsymbol{\tau}$ and $Var(\boldsymbol{\tau}^*) = Cov(\boldsymbol{t}_1 - \Lambda_{12}\Lambda_{22}^{-1}\boldsymbol{t}_2, \boldsymbol{t}_1 - \Lambda_{12}\Lambda_{22}^{-1}\boldsymbol{t}_2) = Cov(\boldsymbol{t}_1, \boldsymbol{t}_1) - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$. Thus, we see that $Var(\boldsymbol{t}_1) - Var(\boldsymbol{\tau}^*) = \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$ which is obviously a nonnegative definite matrix.

This is essentially the covariance adjustment with the matrix Y_1 in the place of $\boldsymbol{t}_1$ and Y_2 in the place of $\boldsymbol{t}_2$. It is evident that (Y_1, Y_2) follow multivariate normal distribution with some covariance matrix Λ.

Unfortunately, Λ is not known in applications. If we have an estimate of Λ

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$$

then an adjusted estimator is

$$\hat{\boldsymbol{\tau}} = \boldsymbol{t}_1 - U_{12}U_{22}^{-1}\boldsymbol{t}_2$$

The problem with this modification is that although $\hat{\boldsymbol{\tau}}$ remains an unbiased estimator of $\boldsymbol{\tau}$, we cannot say that it is necessarily more efficient than $\boldsymbol{t}_1$ since it is easily seen that:

$$\begin{aligned} Var(\hat{\boldsymbol{\tau}}) &= Var(\boldsymbol{t}_1) + E(U_{12}U_{22}^{-1}Var(\boldsymbol{t}_2)U_{22}^{-1}U_{21}) - E(U_{12}U_{22}^{-1}Cov(\boldsymbol{t}_2, \boldsymbol{t}_1)) \\ &\quad -E(Cov(\boldsymbol{t}_1, \boldsymbol{t}_2)U_{22}^{-1}U_{21})) \\ &= \Lambda_{11} + E(U_{12}U_{22}^{-1}\Lambda_{22}U_{22}^{-1}U_{21}) - E(U_{12}U_{22}^{-1}\Lambda_{21}) - E(\Lambda_{12}U_{22}^{-1}U_{21}) \end{aligned}$$

where all expectations are taken over the variations of U_{ij} $i = 1, 2$, $j = 1, 2$. We are not in a position to say anything about the definiteness of $Var(\boldsymbol{t}_1) - Var(\hat{\boldsymbol{\tau}})$ any more. Thus, covariance adjustment may not be an ideal solution if an estimated dispersion matrix of $Cov(\boldsymbol{t}_1, \boldsymbol{t}_2)$ is used. The problem of the choice between the method suggested by Potthoff and Roy, which naturally led to likelihood based methods, and the one introduced by Rao, was somewhat overcome by Baksalary, Corsten and Kala (1978) who showed that both methods are essentially equivalent since for every covariance adjusted estimator there exists a weight matrix in Potthoff and Roy's approach which gives exactly the same result. As the authors noted, the choice between the two methods can be based on the information one has about the covariance matrix Σ.

After this brief introduction of the subjects which we shall deal with, it is easily seen that the notion underlying this thesis is the notion of efficiency. A note however, should be made here. An estimate of (1.6) is obtained in practice, by substituting Σ with G, since in applications the former is unknown. The resulting expression is seen to be

$$(PG^{-1}P^T)^{-1} \tag{1.9}$$

This expression shows the matrix we should believe as the true covariance matrix of our regression coefficients. The diagonal elements of this matrix represent the estimated variances of the regression coefficients and these variances will be closer to the corresponding real ones as the estimate G of Σ lies nearer to the value which estimate. Thus, we should not only care with how far the real or the estimated variances of the regression coefficients lie from their optimal value but also with how far the estimated variances lie from their corresponding true values. It is obvious that matrix (1.9) will be closer (in some sense) to (1.6) depending on the estimate G of Σ we calculate. However, this is directly dependable on the specific methods we use for this calculation.

In the next chapters this idea will be investigated more thoroughly. In Chapter 2 we shall examine the (briefly described) two likelihood based methods of estimating the coefficients of the mean and the variance components (or just the variance matrix) of the MANOVA model as defined in (1.4). What are their differences with respect to the estimated variance matrix (1.9)? We shall derive a new result and we shall try to make an overall comparison between them with respect to the accuracy with which the estimated variances of the regression coefficients approximate their real and optimal variance. In Chapter 3 we shall generalize the situation to the case where many characteristics are observed at each time point. We shall model the situation in a relatively parsimoneous but adequate way and we shall obtain estimates of the variance matrices under the condition that these matrices are positive definite and of the variances of the estimated coefficients of the mean under both methods (ML and REML). Finally, we shall find the relationship between these estimates. In Chapter 4, we shall deal with covariance adjustment. We shall prove certain results concerning the form of the estimator, and in attempting to retain as much efficiency as we can when using this method, we shall suggest a new method for determining its best choice and we shall compare it with three other methods already known. Finally, in the last chapter, we shall deal more theoretically with the problem of efficiency such methods have. Some new results will be proved which could probably suggest some further work in the area.

Chapter 2

A maximum likelihood approach to repeated measurements

2.1 Introduction

Let us assume that we have measurements of n experimental units on time given in the form $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$. These vectors $\boldsymbol{y}_i$'s will be assumed to be a series of independent multivariate observations following a distribution whose density depends on a parameter vector $\boldsymbol{\theta}$ and having the form

$$f(\boldsymbol{y}_i, \boldsymbol{\theta}_i) \quad i = 1, \ldots, n$$

All individuals will be assumed to have the same parameter vector which characterizes the variance but there are differences in the parameter vector of the mean which we wish to examine. Thus, let us distinguish two different parameter vectors $\boldsymbol{\theta}_{i1}$ and $\boldsymbol{\theta}_{i2}$ which will characterize the mean and the variance respectively. A main characteristic of repeated measurements data is that they consist of a relatively small series of observations in the sense that each $\boldsymbol{y}_i$ $i = 1, \ldots, n$ has a relatively small order. So, the mean can be expressed as a smooth function of a small number of parameters $(\boldsymbol{\theta}_{i1})$. This smooth function is locally well approximated by a linear function of these parameters. This expression for the mean allows the fitting of polynomials as mean trends which for the majority of cases is adequate. In the case also where a certain parametric form for the mean is not available, polynomials serve as very good alternatives. In what follows, density f will be considered to be q-variate normal.

The covariance matrix of observations can be left unrestricted. That entails that θ_{o2} consists of $q(q+1)/2$ parameters which have to be estimated. This number cannot be judged as restrictive for very small series. However, many experimenters have suggested (Diggle, 1988, Pantula and Pollock, 1985) that this may result in overparameterization. It is thus desirable even with q moderately small, to introduce functions with fewer parameters which keep as much information as possible of the real form of the covariance matrix. Two different approaches to this important problem can be distinguished.

a) One has partial knowledge about the real form, or information which can determine it and uses it for constructing a suitable structure.

b) One should carefully examine the data and suggest an appropriate structure. This structure should be general enough to include many possible cases for subsequent hypothesis testing.

In this chapter we shall deal with the first case. There are experiments where physiological changes or other external reasons can affect the variance of the response or the covariance between the observations. In these cases this change can effectively be modelled thus presenting a structure which takes into account this information. One purpose of this chapter is to show that failure to incorporate this change into the covariance matrix will lead to misleading standard errors of the estimated regression coefficients and consequently to invalid inferences. Two likelihood based estimation procedures will be examined, maximum likelihood (ML) and restricted (or residual) maximum likelihood (REML). There is much discussion concerning these two methods which has not ended up with clear results about the superiority of one of them. The two methods are asymptotically equivalent. In small samples and for incomplete data, a large study by Swallow and Monahan (1984) pointed out that REML is less biased than ML while ML gives less mean squared error than REML. In a more recent study McGilchrist (1989) argued again, in a time series context, that REML is less biased with respect to the estimated variance components of the model. In this chapter we shall not deal with the variance components but with the implication of their estimation to the efficiency problem. We shall prove that the general scatter of the estimated

covariance matrix of the regression coefficients is larger when this estimation is calculated using REML (see also in (2.72)). This result also indicates that the corresponding estimated variances (the diagonal elements) should in general be expected to be larger for REML rather than for ML, as a simulation study suggests in the last section of the chapter.

The following definition and lemma will be used throughout this chapter and the whole thesis:

Definition 1 *A GLS estimator weighted by matrix G will be a function of the data matrix Y of the form*

$$(A^TA)^{-1}A^TYGP^T(PGP^T)^{-1} \tag{2.1}$$

Lemma 1 *Let B $(p \times q)$ and B^* $(p \times (p-q))$ be of ranks q and $p-q$ respectively such that $B^TB^* = 0$. If S is a symmetric positive definite (p.d) matrix then*

$$S^{-1} - S^{-1}B(B^TS^{-1}B)^{-1}B^TS^{-1} = B^*(B^{*T}SB^*)^{-1}B^{*T} \tag{2.2}$$

Proof: See Khatri (1966).

2.2 The model

The model which will be examined is the GMANOVA model defined by Potthoff and Roy (1964). It has the following form:

$$Y = A\xi P + E, \tag{2.3}$$

where Y $(n \times q)$ is a matrix of observations from n individuals at q different times, A $(n \times m)$ and P $(p \times q)$ are known design matrices of ranks m and p respectively, while ξ $(m \times p)$ is a matrix of mp unknown parameters. Further, E $(n \times q)$ is a matrix of errors, whose n rows ϵ_i $(i = 1, \ldots, n)$ are independent q-dimensional vectors with

$$\epsilon_i \sim N(0, \Sigma). \tag{2.4}$$

A represents between subjects classifications such as experimental treatment groups, while ξP describes the within subject changes in mean response such as development

or growth over time. In the most usual application different treatments are modelled with polynomials in time of order up to $p-1$ as in the case where there are m groups of individuals with n_j individuals in the $j-th$ group. Each group may be subjected to a different treatment. Observations are made on all individuals in all groups at the same q points in time and have the same covariance matrix Σ. The growth curve associated with the $j-th$ group is $\xi_{j0}+\xi_{j1}t+\ldots+\xi_{j,p-1}t^{p-1}$. Here, $Y\ (n\times q)\ (n=n_1+\ldots+n_m)$ is the data matrix; $A\ (n\times m)$, is the design matrix, with n_1 rows of $(1,0,\ldots,0)$, n_2 rows of $(0,1,0,\ldots,0)$ . . . and n_m rows of $(0,0,\ldots,1)$; $\xi\ (m\times p)$, is the matrix of unknown parameters,with the $j-th$ row being $(\xi_{j0},\xi_{j1},\ldots,\xi_{j,p-1})$; and $P\ (p\times q)$ with its $k-th$ column being $(1,t_k,t_k^2,\ldots,t_k^{p-1})^T$. Potthoff and Roy (1964) discussed other possible applications of model (2.3).

As regards the covariance matrix of observations we shall assume that measurements from different experimental units (individuals) are independent of each other. A modelling for the covariance matrix should take into account the nature of repeated measurements data. It has been suggested that as measurements move away from each other their dependence decreases. This can be plausible in certain cases but in others such as those described in the introduction, such an assumption may be misleading. Whatever the case is, the uniform correlation model suggested by Box (1950), Potthoff and Roy (1964) and others, which assumes the same covariance for all observations may be inappropriate for moderately large series of observations like growth curves. Anderson (1969) considered the case where the covariance matrix has the following linear form:

$$\Sigma(\boldsymbol{\sigma})=\sum_{k=0}^{m}\sigma_i G_i$$

where G_i's are known, linearly independent symmetric matrices and the σ's are unknown scalars such that Σ is a positive definite matrix. However, the fact that G_i's must be known matrices restricts the range of applications for this model and makes it rather inappropriate for such cases as those described in the introduction.

Consider a more general model for the covariance matrix which incorporates a component which will distinguish the response for each individual, a component which

will distinguish the response within individuals and another component which will compensate for the variation which is left unexplained within individuals. Specifically, consider the following model for the errors ϵ_{ij}'s of individual i

$$\varepsilon_{ij} = Z_{ij} + U_i + Q_{ij} \tag{2.5}$$

Z_{ij} are independent and identically distributed (i.i.d.) $N(0, \tau^2)$ variates representing variation within individuals, U_i are independent $N(0, \nu^2)$ variates representing variation in average response between individuals and $\boldsymbol{Q}_i = \{Q_{i1}, \ldots, Q_{iq}\}$ $(i = 1, \ldots, n)$ are independent and identically distributed q-dimensional vectors, comprising within individuals autocorrelated errors. Q_{ij} can also be represented as $Q_{ij} = \boldsymbol{m}_j^T \boldsymbol{W}_i$, where each $\boldsymbol{m}_j$ is a q-dimensional vector of constants and each $\boldsymbol{W}_i = \{W_{i1}, \ldots, W_{iq}\}^T$ is a normal random process with $E(W_{ij}) = 0$ and covariance matrix R (diagonal). Then, each $\boldsymbol{Q}_i$ is itself a normal process with $E(Q_{ij}) = 0$ and covariance matrix MRM^T where M is the matrix consisting of the row vectors $\boldsymbol{m}_j^T$. Finally, Z_{ij}, U_i and W_{ij} (or) Q_{ij} are independent of each other. The idea of incorporating an autocorrelation structure into the covariance matrix is not new. Many authors such as Danford, Hughes and McNee (1960), Patterson and Lowe (1970), Bjornsson (1978), Wilson, Hebel and Sherwin (1981), Pantula and Pollock (1985), and Mansour, Nordheim and Rutledge (1985) have dealt with it among others.

With this specification $\epsilon_i = \{\varepsilon_{i1}, \ldots, \varepsilon_{iq}\}$ is a normal, non-stationary random process with the typical ε_{ij} consisting of three additive components: Z_{ij}, U_i and Q_{ij} $(j = 1, \ldots, q)$ and the covariance matrix of observations has the form

$$\Sigma = \tau^2 I + \nu^2 J + V. \tag{2.6}$$

Note that in some cases it may be necessary to assume that Z_{ij} $(j = 1, \ldots, q)$ have different variances (as if different observations at different times are made by different judges), that is, Z_{ij} are i.i.d. $N(0, \tau_j^2)$ variates. It is not difficult to show that under these assumptions the variance matrix of observations is

$$\Sigma = T^2 + \nu^2 J + V \tag{2.7}$$

where $T^2 = \text{diag}\,(\tau_1^2, \ldots, \tau_q^2)$ and τ_j^2 $(j = 1, \ldots, q)$ are extra unknown parameters. A similar model has been considered by Diggle (1988).

As was remarked in the introduction, in many experiments and for certain reasons, there is a change in the conditions under which the experiment develops. For example, in planting out seedlings, a change in the environment at a time point between two observations may well affect the covariance matrix of observations. These changes are incorporated into the autoregressive scheme by adopting an appropriate model which will be called 'change point' model for obvious reasons.

Consider the following modeling of $\boldsymbol{Q}_i$

$$Q_{ij} = \boldsymbol{m}_j^T \boldsymbol{W}_i = \begin{cases} W_{i1}, & j = 1 \\ \phi_1 Q_{i,j-1} + W_{ij}, & j = 2, \ldots, r \\ \phi_2 Q_{i,j-1} + W_{ij}, & j = r+1, \ldots, q \end{cases} \tag{2.8}$$

where $2 \leq r \leq q-1$ and W_{ij}'s are normal variates with $E(W_{ij}) = 0 \;\forall\; j = 1, \ldots, q$ $Var(W_{ij}) = \sigma_1^2$ $i = 1, \ldots, r$ and $Var(W_{ij}) = \sigma_2^2$ $i = r+1, \ldots, q$. This is the first order $AR(1)$ '1-change point' model with fixed coefficients at a fixed time point. The covariance matrix V of this process will be obtained using a technique also mentioned in Kenward (1981). Let us define new vector variable $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iq})^T$ such that:

$$w_{ij} = \begin{cases} \sigma_1^{-1} Q_{i1}, & j = 1 \\ \sigma_1^{-1}(Q_{ij} - \phi_1 Q_{i,j-1}), & j = 2, \ldots, r \\ \sigma_2^{-1}(Q_{ij} - \phi_2 Q_{i,j-1}), & j = r+1, \ldots, q \end{cases} \tag{2.9}$$

Then $\boldsymbol{w}_i = U\boldsymbol{Q}_i$ where

$$U = \begin{pmatrix} \sigma_1^{-1} & 0 & 0 & \ldots & 0 \\ -\sigma_1^{-1}\phi_1 & \sigma_1^{-1} & 0 & \ldots & 0 \\ 0 & -\sigma_1^{-1}\phi_1 & \sigma_1^{-1} & \ldots & 0 \\ & & \ldots & & \\ 0 & \ldots & 0 & -\sigma_2^{-1}\phi_2 & \sigma_2^{-1} \end{pmatrix} \tag{2.10}$$

Then $Var(\boldsymbol{w}_i) = UVU^T = I_q$ since variables w_{ij} $j = 1, \ldots, q$ are standardized uncorrelated random variables. That gives

$$V = U^{-1}(U^T)^{-1} \text{ and } V^{-1} = U^T U \tag{2.11}$$

Applying elementary matrix operations one can see that

$$U^{-1} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ \sigma_1\phi_1 & \sigma_1 & 0 & 0 & \dots & 0 \\ \sigma_1\phi_1^2 & \sigma_1\phi_1 & \sigma_1 & 0 & \dots & 0 \\ & & \dots & \dots & & \\ \sigma_2\phi_2^{q-r}\phi_1^{r-1} & \sigma_2\phi_2^{q-r}\phi_1^{r-2} & \dots & \dots & \dots & \sigma_2 \end{pmatrix} \tag{2.12}$$

Thus, matrix V can be written as $V = MRM^T$ and the covariance matrix of observations as in (2.5), by setting

$$R = \begin{pmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_1^2 & & & \\ & & & \sigma_2^2 & & \\ & 0 & & & \ddots & \\ & & & & & \sigma_2^2 \end{pmatrix} \tag{2.13}$$

and the rows of M as

$$\boldsymbol{m}_j^T = (\phi_1^{j-1}, \phi_1^{j-2}, \dots, 1, 0, \dots 0), \quad \text{if } j = 1, \dots, r \tag{2.14}$$

$$\text{or } \boldsymbol{m}_j^T = (\phi_2^{j-r}\phi_1^{r-1}, \phi_2^{j-r}\phi_1^{r-2}, \dots, \phi_2^{j-r}\phi_1, \phi_2^{j-r-1}, \phi_2^{j-r-2}, \dots, 1, 0, \dots, 0)$$

$$\text{if } j = r+1, \dots, q. \tag{2.15}$$

This model can easily be extended to an $AR(1)$ 'k-change point' model with fixed coefficients at k fixed time points form, namely

$$Q_{ij} = \boldsymbol{m}_j^T \boldsymbol{W}_i = \phi_c Q_{i,j-1} + W_{ij}, \quad \text{if } j \in J_c;\ c = 1, \dots, k \tag{2.16}$$

where $J_1, \dots, J_k$ form a partition of the set $\{1, \dots, q\}$ and

$$Var(\boldsymbol{W}_i) = \begin{pmatrix} \sigma_1^2 I_{|J_1|} & & & \\ & & 0 & \\ & \ddots & & \\ 0 & & & \\ & & & \sigma_k^2 I_{|J_k|} \end{pmatrix} \tag{2.17}$$

and $|J|$ is the number of elements contained in the set J and $\sum_{i=1}^{k} |J_k| = q$. Using the same technique as before one can show that the corresponding U is

$$U = \begin{pmatrix} \sigma_1^{-1} & 0 & & \dots & & & 0 \\ -\sigma_1^{-1}\phi_1 & -\sigma_1^{-1} & 0 & \dots & & & 0 \\ & & & \dots & & & \\ 0 & \dots & 0 & -\sigma_2^{-1}\phi_2 & -\sigma_2^{-1}0 \dots 0 & & \\ & & & \dots & & & \\ 0 & & & \dots & & 0 \quad -\sigma_k^{-1}\phi_k & -\sigma_k^{-1} \end{pmatrix} \tag{2.18}$$

and U^{-1}

$$U^{-1} = \begin{pmatrix} \sigma_1 & 0 & & \dots & & & 0 \\ \sigma_1\phi_1 & \sigma_1 & 0 & \dots & & & 0 \\ & & & \dots & \dots & & \\ \sigma_1\phi_1^{|J_1|-1} & \sigma_1\phi_1^{|J_1|-2} & \dots & \sigma_1 & 0 \quad \dots & \dots & 0 \\ \sigma_2\phi_2\phi_1^{|J_1|-1} & \sigma_2\phi_2\phi_1^{|J_1|-2} & \dots & \dots & \sigma_2 \quad 0 \quad \dots & & 0 \\ & & & \dots & \dots & & \\ \sigma_k\phi_k^{|J_k|}\dots\phi_2^{|J_2|}\phi_1^{|J_1|-1} & & & & \dots & & \sigma_k \end{pmatrix} \tag{2.19}$$

Matrix V can again be written as $V = MRM^T$ and the covariance matrix of observations as in (2.5) by setting

$$R = \begin{pmatrix} \sigma_1^2 I_{|J_1|} & & \\ & 0 & \\ & \ddots & \\ 0 & & \\ & & \sigma_k^2 I_{|J_k|} \end{pmatrix} \tag{2.20}$$

and the rows of M as

$$\boldsymbol{m}_j^T = (\phi_1^{j-1}, \phi_1^{j-2}, \dots, 1, 0, \dots 0), \quad \text{if } j = 1, \dots, |J_1| \tag{2.21}$$

$$\text{or } \boldsymbol{m}_j^T = (\phi_2^{j-|J_1|}\phi_1^{|J_1|-1}, \phi_2^{j-|J_1|}\phi_1^{|J_1|-2}, \dots, \phi_2^{j-|J_1|}\phi_1, \phi_2^{j-|J_1|-1}, \dots, 1, \dots, 0)$$

$$\text{if } j = |J_1|+1, \dots, |J_1|+|J_2|. \tag{2.22}$$

$$\text{or } \boldsymbol{m}_j^T = (\phi_k^{j-\sum_{i=1}^{k-1}|J_i|}\phi_{k-1}^{|J_{k-1}|}\dots\phi_1^{|J_1|-1}, \dots, 1, 0, \dots, 0)$$

$$\text{if } j = \sum_{i=1}^{k-1}|J_i| + 1, \dots, \sum_{i=1}^{k}|J_i| \tag{2.23}$$

From the above definition, many well known structures arise as special cases.

For example, the first-order autoregressive covariance structure is obtained by setting $\phi_1 = \phi_2 = \phi$ in (2.8) and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, that is, an $AR(1)$ '0–change point'. Thus,

$$R = \sigma^2 \, I \quad \text{and} \quad m_j^T = (\phi^{j-1}, \, \phi^{j-2}, \ldots, \phi, \, 1, 0, \ldots, 0) \tag{2.24}$$

with $\sigma^2 > 0$ and $|\phi| < 1$. In (2.3), Q_i is a first-order autoregressive ($AR(1)$) process with fixed autoregression coefficient ϕ. With this specification $\varepsilon_i = \{\varepsilon_{i1}, \ldots, \varepsilon_{iq}\}$ is a non-stationary process. A further requirement, namely

$$R(1,1) = \text{Var}\,(W_{11}) = \sigma^2/(1-\phi^2)$$

changes ε_i to a stationary process with covariance matrix Σ given by

$$\Sigma = \tau^2 I + \nu^2 J + V, \quad \text{where} \quad V(j,k) = \sigma^2/(1-\phi^2) \begin{cases} 1, & j = k \\ \phi^{|j-k|}, & j \neq k \end{cases} \tag{2.25}$$

By transforming the parameter ϕ one can get the model suggested by Diggle (1988). The model specified by (2.24) provides a useful starting point in many applications of repeated measurements. It has been considered by Geisser (1981), Wilson, Hebel and Sherwin (1981), Azzalini (1984), Pantula and Pollock (1985), Ware (1985), Verbyla and Cullis (1990) and many others.

In the extreme case, the $AR(1)$ '(q-2)–change point' model is obtained by setting

$$Q_{ij} = m_j^T W_i = \begin{cases} W_{i1}, & j = 1 \\ \phi_{j-1} Q_{i,j-1} + W_{ij}, & j = 2, \ldots, q \end{cases} \tag{2.26}$$

that is an $AR(1)$ process with non-stationary autoregression coefficients. With this assumption (2.13) is replaced by

$$R = \begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & 0 & \\ & 0 & \ddots & \\ & & & \sigma_q^2 \end{pmatrix} \tag{2.27}$$

while (2.14), (2.15) are replaced by

$$m_j^T = (\phi_{j-1} \cdots \phi_1, \phi_{j-1} \cdots \phi_2, \ldots, \phi_{j-1}, \, 1, 0, \ldots, 0) \tag{2.28}$$

The latter model is the general form considered by Gabriel (1961, 1962) who called the structure 'ante-dependence structure'. When in (2.6) $\tau^2 = 0$ one reaches the model used by Kenward (1981).

Upon a further simplification one can obtain the 'uniform covariance' structure model by setting $V = 0$ in (2.5), that is

$$\Sigma = \tau^2 I + v^2 J \tag{2.29}$$

or, in (2.3), $Q_{ij} = 0$ $(i = 1, \ldots, n;\ j = 1, \ldots, q)$. The model specified by (2.29) is also known as 'split-plot' or 'compound symmetry' or 'uniform correlation' or '1-random effect' model.

2.3 Parameter estimation

2.3.1 ML estimates

In this section the estimation procedure is described using the maximum likelihood approach. Assuming the validity of the model specified by (2.3) the log-likelihood function $\ell(\xi,\ \Sigma)$ can be written as

$$\ell(\xi,\ \Sigma) = \text{const}\ -\frac{1}{2}\, n \log |\,\Sigma\,| - \frac{1}{2}\, tr(Y - A\xi P)^T (Y - A\xi P)\Sigma^{-1}. \tag{2.30}$$

Here, ξ will be the matrix of unknown coefficients specifying the trend and Σ will be structured as in (2.6) with V being taken as one of the forms discussed above. There are cases where closed formulae exist for $\hat{\xi}$ and $\hat{\Sigma}$, the values which maximize $\ell(\xi,\ \Sigma)$. This is the case where Σ is left unrestricted which requires the estimation of $q(q+1)/2$ unknown parameters. As Khatri (1966, 1973) showed, these estimates are easily computed as follows

$$\hat{\xi} = (A^T A)^{-1} A^T Y S_0^{-1} P^T (P S_0^{-1} P^T)^{-1}, \qquad \text{where} \tag{2.31}$$

$$S_0 = Y^T \left[I - A(A^T A)^{-1} A^T \right] Y, \ \text{and}$$

$$\hat{\Sigma} = \frac{1}{n}\,(Y - A\hat{\xi}P)^T(Y - A\hat{\xi}P). \tag{2.32}$$

The same formula can be derived by using an argument based on the distribution of the sample covariance matrix S_0. This argument was used by Gleser and Olkin (1972) in a similar case. See also Johansen (1982).

It follows that the maximized log-likelihood is given by

$$\ell_1 \equiv \ell(\hat{\xi}, \hat{\Sigma}; H_1) = -\frac{1}{2}\, n \log \mid S_0 \mid - \frac{1}{2} nq + \text{const.} \tag{2.33}$$

Another interesting case in which the derivation of closed formulae for the maximum likelihood estimates is possible, is the mixed effects model. According to it, ξ is written as (ξ_1, ξ_2) (ξ_1 $(m \times c)$, ξ_2 $(m \times (p-c))$) and P as $\begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$ (P_1 $(c \times q)$ P_2 $((p-c) \times q)$) and ξ_1 is a random variable with a certain mean and variance Γ, with Γ a $c \times c$ covariance matrix. Given ξ_1 the measurements are considered as independent random variables with variance τ^2 and unconditionally the covariance between the measurements arises only from the assumption of the existence of Γ and the inclusion of P_1. The unconditional variance matrix of observations is therefore $\Sigma = P_1^T \Gamma P_1 + \tau^2 I_q$.

Note that the 'Uniform Correlation' model corresponds to $P_1 = (1, \ldots, 1)$ and $\Gamma = \nu^2 I_q$ ($\Sigma = \tau^2 I_q + \nu^2 \mathbf{1}^T \mathbf{1}$). Exhaustive work in this area has shown that the maximum likelihood estimate for ξ coincides with the ordinary least squares estimate (OLS)

$$\hat{\xi} = (A^T A)^{-1} A^T Y P^T (P P^T)^{-1}. \tag{2.34}$$

since $\Sigma^{-1} P^T (P \Sigma^{-1} P^T)^{-1} = P^T (P P^T)^{-1}$ (see also Rao, 1965). Maximum likelihood estimates then, of the two components τ^2 and Γ can be found explicitly as

$$\hat{\tau}^2 = \frac{1}{n(q-c)} tr(Y - \Pi_A Y \Pi_P)(I - \Pi_{P_1})(Y - \Pi_A Y \Pi_P)^T, \tag{2.35}$$

$$\hat{\Gamma} = \frac{1}{n}(P_1 P_1^T)^{-1} P_1 Y^T (I - \Pi_A) Y P_1^T (P_1 P_1^T)^{-1} - \hat{\tau}^2 (P_1 P_1^T)^{-1}, \tag{2.36}$$

where $\Pi_A = A(A^T A)^{-1} A^T$ and $\Pi_P = P^T (P P^T)^{-1} P$ and $\Pi_{P_1} = P_1^T (P_1 P_1^T)^{-1} P_1$, (Lange and Laird, 1989). Apart from these special cases, the log-likelihood function must be

maximized numerically. The following properties concerning differentials will be used:

$$\begin{aligned} dU^T &= (dU)^T \\ dtrU &= trdU \\ d(U+V) &= dU + dV \\ d(UV) &= (dU)V + UdV \\ d|F| &= |F|trF^{-1}dF \\ dF^{-1} &= -F^{-1}dFF^{-1}, \end{aligned} \tag{2.37}$$

(see also Magnus and Neudecker, 1988). Using these results the Fisher scoring method will be described in this context (see also Joreskog, 1970). Denote the vector of the parameters which characterize the covariance matrix as $\boldsymbol{\theta}$. Thus, $\boldsymbol{\theta}$ can be of the form:

$$\boldsymbol{\theta} = (\tau^2, v^2, (vecM)^T, (vecR)^T)^T. \tag{2.38}$$

Let the score vector be

$$U(\widehat{vec(\xi)}, \boldsymbol{\theta}) = \left(\frac{\partial \ell}{\partial(\text{vec } \xi)}, \quad \frac{\partial \ell}{\partial \tau^2}, \quad \frac{\partial \ell}{\partial v^2}, \quad \frac{\partial \ell}{\partial(\text{vec } M)}, \quad \frac{\partial \ell}{\partial(\text{vec } R)} \right)^T, \quad \text{where} \tag{2.39}$$

$vec(B)$ is the operator which stacks the columns of matrix B one under another. Using properties (2.37) it can be seen that the differential of the logarithm of the likelihood (2.30) is

$$dl_M(\xi, \boldsymbol{\theta}) = -\frac{n}{2} tr\Sigma^{-1}d\Sigma + \frac{1}{2} tr\Sigma^{-1}(d\Sigma)\Sigma^{-1}S, \tag{2.40}$$

with

$$S = \frac{1}{n}(Y - A\xi P)^T(Y - A\xi P) \tag{2.41}$$

and M standing for ML. It is not difficult to see from (2.40) that

$$\begin{aligned} \frac{\partial \ell}{\partial(\text{vec } \xi)} &= \text{vec } \left[A^T(Y - A\xi P)\Sigma^{-1}P^T\right] \\ \frac{\partial \ell}{\partial \tau^2} &= -\frac{n}{2} tr(G) \\ \frac{\partial \ell}{\partial v^2} &= -\frac{n}{2} tr\ (JG) \\ \frac{\partial \ell}{\partial(\text{vec } M)} &= -n \text{ vec } (GMR) \\ \frac{\partial \ell}{\partial(\text{vec } R)} &= -\frac{n}{2} \text{ vec } \left[M^T G M\right], \quad \text{where} \end{aligned} \tag{2.42}$$

$$G = \Sigma^{-1} - \Sigma^{-1} S \, \Sigma^{-1}.$$

The last two derivatives are not very operational in applications since matrices M and R do not contain only unknown parameters but are rather functions of a small number of parameters. Consider the case where M and R are modelled as a '1-change-point' $AR(1)$ with fixed coefficients at a fixed time point (c.f. (2.13), (2.14), (2.15)) and with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Then M is a function of two parameters ϕ_1 and ϕ_2 and R of σ^2. Then θ as defined in (2.38) is

$$\theta = (\tau^2, v^2, \sigma^2, \phi_1, \phi_2)^T \tag{2.43}$$

and the corresponding derivatives have the form

$$\begin{aligned} \frac{\partial \ell}{\partial \phi_1} &= -ntr(RM^T G \Omega_1) \\ \frac{\partial \ell}{\partial \phi_2} &= -ntr(RM^T G \Omega_2) \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2} tr M^T G M, \end{aligned} \tag{2.44}$$

where the rows of Ω_1 are given by

$$\omega_j^{1^T} = ((j-1)\phi_1^{j-2}, (j-2)\phi_1^{j-3}, \ldots, 1, 0, \ldots, 0) \quad j = 1, \ldots, r$$

$$\omega_j^{1^T} = ((r-1)\phi_2^{j-r}\phi_1^{r-2}, (r-2)\phi_2^{j-r}\phi_1^{r-3}, \ldots, 1, 0, \ldots, 0) \quad j = r+1, \ldots, q \tag{2.45}$$

and those of Ω_2 by

$$\omega_j^{2^T} = (0, \ldots, 0) \quad j = 1, \ldots, r$$

$$\omega_j^{2^T} = ((j-r)\phi_2^{j-r-1}\phi_1^{r-1}, (j-r)\phi_2^{j-r-1}\phi_1^{r-2}, \ldots, 1, 0, \ldots, 0) \quad j = r+1, \ldots, q. \tag{2.46}$$

Corresponding formulae for the 'k-change-point' model can easily be found.

The information matrix is $J = \left(E(-\frac{\partial^2 l}{\partial(\xi, \theta)^2}) \right)$. It is easily seen that J has a form block partitioned as

$$\begin{pmatrix} J_{\xi,\xi} & J_{\xi,\theta} \\ J_{\theta,\xi} & J_{\theta,\theta} \end{pmatrix}, \tag{2.47}$$

with $J_{\xi,\xi} = E(-\frac{\partial^2 l}{\partial vec(\xi)^2})$, $J_{\xi,\theta} = E(-\frac{\partial^2 l}{\partial vec(\xi)\partial\theta})$ and $J_{\theta,\theta} = E(-\frac{\partial^2 l}{\partial\theta^2})$.

The Fisher scoring algorithm computes new parameter values $\widehat{vec(\xi)}_N,\ \hat{\boldsymbol{\theta}}_N$ from current values of $\widehat{vec(\xi)}_c,\ \hat{\boldsymbol{\theta}}_c$ using

$$(\widehat{vec(\xi)}_N,\ \hat{\boldsymbol{\theta}}_N)^T = (\widehat{vec(\xi)}_c,\ \hat{\boldsymbol{\theta}}_c)^T - J^{-1}U(\widehat{vec(\xi)}_c,\ \hat{\boldsymbol{\theta}}_c). \tag{2.48}$$

Because of the form of the information matrix, (2.47) is simplified to

$$\widehat{vec(\xi)}_N = \widehat{vec(\xi)}_c - J^{-1}_{\xi,\xi}U(\widehat{vec(\xi)}_c,\ \hat{\boldsymbol{\theta}}_c) \text{ and}$$

$$\hat{\boldsymbol{\theta}}_N = \hat{\boldsymbol{\theta}}_c - J^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}U(\widehat{vec(\xi)}_c,\ \hat{\boldsymbol{\theta}}_c), \tag{2.49}$$

since $J_{\xi,\boldsymbol{\theta}} = J_{\boldsymbol{\theta},\xi} = 0$. This is equivalent to using the second equation of (2.49) for the estimation of variance components while $\hat{\xi}$ is given by the well known identity

$$\hat{\xi}_M = (A^TA)^{-1}A^TY\hat{\Sigma}_M^{-1}P^T(P\hat{\Sigma}_M^{-1}P^T)^{-1}, \tag{2.50}$$

where suffix M means that the covariance matrix Σ is estimated by ML. The variance of the resulting estimator conditional on $\hat{\Sigma}_M$ is

$$Var(\hat{\xi}_M) = (A^TA)^{-1} \otimes (P\hat{\Sigma}_M^{-1}P^T)^{-1}P\hat{\Sigma}_M^{-1}\Sigma\hat{\Sigma}_M^{-1}P^T(P\hat{\Sigma}_M^{-1}P^T)^{-1} \tag{2.51}$$

and an estimate of it, using the ML estimated Σ is

$$\widehat{Var(\hat{\xi}_M)} = (A^TA)^{-1} \otimes (P\hat{\Sigma}_M^{-1}P^T)^{-1}. \tag{2.52}$$

2.3.2 REML estimates

Restricted Maximum Likelihood or REML was introduced by Patterson and Thompson (1971, 1974) as an attempt to reduce the bias of the maximum likelihood estimated variance components in the random effects linear model. It can be considered as a variation of the maximum likelihood method since it requires the maximization of that part of the likelihood which is independent of ξ. This involves transformation of the data matrix Y to a new variable whose mean is zero. A vector $\boldsymbol{v}$ which will transform $vec(Y)$, to $\boldsymbol{v}^Tvec(Y)$ such that $E(\boldsymbol{v}^Tvec(Y)) = 0$ will be called an error contrast. It is obvious now from (A11), that model (2.3) is written in vectorized form:

$$vec(Y) = (P^T \otimes A)vec(\xi) + vec(E). \tag{2.53}$$

Since matrix $P^T \otimes A$ has rank mp, we can have $nq - mp$ independent error contrasts in this model. For this reason and by taking into account the form of the vectorized model we can define our contrasts as:

$$(T_1^T \otimes I_n)vec(Y) \quad \text{and} \tag{2.54}$$

$$(T_2^T \otimes T_3)vec(Y). \tag{2.55}$$

T_1 $(q \times (q-p))$, T_2 $(q \times p)$ and T_3 $((n-m) \times n)$ are matrices such that $PT_1 = 0$, $T_3 A = 0$, $T_1^T \Sigma T_2 = 0$ and $T_1^T T_2 = 0$.

Theorem 1 *The linearly transformed variables in (2.54) and (2.55) are the $nq - mp$ required independent error contrasts.*

Proof: Consider the partitioned matrix $\begin{pmatrix} T_1^T \otimes I_n \\ T_2^T \otimes T_3 \end{pmatrix}$ along with the transformation $\begin{pmatrix} T_1^T \otimes I_n \\ T_2^T \otimes T_3 \end{pmatrix} vec(Y)$. Each of the two components of the transformed variable has mean 0 since,

$$E\left((T_1^T \otimes I_n)vec(Y)\right) = (T_1^T P^T \otimes A)vec(\xi) = 0$$

$$E\left((T_2^T \otimes T_3)vec(Y)\right) = (T_2^T P^T \otimes T_3 A)vec(\xi) = 0.$$

Furthermore, the two (linearly independent) components are stochastically independent since

$$(T_1^T \otimes I_n)(\Sigma \otimes I_n)(T_2 \otimes T_3^T) = T_1^T \Sigma T_2 \otimes T_3^T = 0.$$

Since now T_1^T, T_2^T and T_3 are of full row rank, the rank of the transformation matrix is $n(q-p) + (n-m)p = nq - mp$ and the proof is complete.

The actual form of these $nq - mp$ error contrasts is not of any real interest. As Harville (1974) showed, the likelihood function of these contrasts is always the same apart from a constant which does not affect the maximization procedure. In our case, this form can be found by considering the joint log-likelihood of the two error contrasts which is ('R' stands for REML):

$$l_R = const - \frac{1}{2}\log|T_1^T \Sigma T_1 \otimes I_n| - \frac{1}{2}\log|T_2^T \Sigma T_2 \otimes T_3 T_3^T|$$

$$-\frac{1}{2}\Big((vec(Y))^T(T_1\otimes I_n)\left((T_1^T\Sigma T_1)^{-1}\otimes I_n\right)(T_1^T\otimes I_n)(vec(Y))+$$

$$(vec(Y))^T(T_2\otimes T_3^T)\left((T_2^T\Sigma T_2)^{-1}\otimes(T_3T_3^T)^{-1}\right)(T_2^T\otimes T_3)vec(Y)\Big)$$

$$= const-\frac{1}{2}\log|T_1^T\Sigma T_1|^n|T_2^T\Sigma T_2|^{n-m}|T_3T_3^T|^p-\frac{1}{2}(vec(Y))^T(L_1+L_2)vec(Y), \quad (2.56)$$

where

$$L_1 = T_1(T_1^T\Sigma T_1)^{-1}T_1^T\otimes I_n \quad \text{and}$$

$$L_2 = T_2(T_2^T\Sigma T_2)^{-1}T_2^T\otimes T_3^T(T_3T_3^T)^{-1}T_3.$$

A more simplified form for $|T_1^T\Sigma T_1|^n|T_2^T\Sigma T_2|^{n-m}|T_3T_3^T|^p$ can be found as follows:

Consider the partitioned matrix

$\left(T_1\otimes I_n, T_2\otimes T_3^T, P^T\otimes A\right)$. Since $PT_1 = 0$, $T_1^TT_2 = 0$ and $T_3A = 0$ then the above defined matrix is a square one of order nq and of full rank. For this and from (A17),

$$\left|\begin{pmatrix} T_1\otimes I_n \\ T_2\otimes T_3^T \\ P^T\otimes A\end{pmatrix}^T(\Sigma\otimes I_n)\left(T_1\otimes I_n, T_2\otimes T_3^T, P^T\otimes A\right)\right| \propto |\Sigma\otimes I_n| = |\Sigma|^n. \quad (2.57)$$

However, the left hand side of (2.57) is also equal to

$$\begin{vmatrix} T_1^T\Sigma T_1\otimes I_n & 0 & T_1^T\Sigma P^T\otimes A \\ 0 & T_2^T\Sigma T_2\otimes T_3T_3^T & 0 \\ P\Sigma T_1\otimes A^T & 0 & P\Sigma P^T\otimes A^TA\end{vmatrix}$$

$$= |T_1^T\Sigma T_1|^n\left|\begin{pmatrix} T_2^T\Sigma T_2\otimes T_3T_3^T & 0 \\ 0 & P(\Sigma-\Sigma T_1(T_1^T\Sigma T_1)^{-1}T_1^T\Sigma)P^T\otimes A^TA\end{pmatrix}\right|. \quad (2.58)$$

Recall now that $PT_1 = 0$. Application of Lemma 1 gives

$$\Sigma-\Sigma T_1(T_1^T\Sigma T_1)^{-1}T_1^T\Sigma = P^T(P\Sigma^{-1}P^T)^{-1}P. \quad (2.59)$$

This in turn means that the multiplication of the two determinants in (2.58) is written as:

$$|T_1^T\Sigma T_1|^n|T_2^T\Sigma T_2|^{n-m}|T_3T_3^T|^p|PP^T|^{2m}|P\Sigma^{-1}P^T|^{-m}.$$

But from (2.57), this is proportional to $|\Sigma|^n$. Thus,

$$|T_1^T\Sigma T_1|^n|T_2^T\Sigma T_2|^{n-m}|T_3T_3^T|^p \propto |P\Sigma^{-1}P^T|^m|\Sigma|^n \quad (2.60)$$

Consider now L_1 and L_2. From (2.59) we get that

$$L_1 = (\Sigma^{-1} - \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}) \otimes I_n \tag{2.61}$$

Since now $T_3A = 0$ application of Lemma 1 again gives,

$$T_3^T(T_3T_3^T)^{-1}T_3 = I_n - A(A^TA)^{-1}A^T$$

If we additionally recall that $T_1^T\Sigma T_2 = 0$ then application of the same Lemma with $B = \Sigma T_2$ and $B^* = T_1$ gives

$$T_2(T_2^T\Sigma T_2)^{-1}T_2^T = \Sigma^{-1} - T_1(T_1^T\Sigma T_1)^{-1}T_1^T \Rightarrow$$

$$T_2(T_2^T\Sigma T_2)^{-1}T_2^T = \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}$$

and thus,

$$L_2 = \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1} \otimes (I_n - A(A^TA)^{-1}A^T) \tag{2.62}$$

The above remarks allow us to rewrite log-likelihood (2.56) as:

$$const - \frac{m}{2}\log|P\Sigma^{-1}P^T| - \frac{n}{2}\log|\Sigma|$$

$$-\frac{1}{2}(vec(Y))^T\left((\Sigma^{-1} - \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}) \otimes I_n\right.$$

$$\left.+\Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1} \otimes (I_n - A(A^TA)^{-1}A^T)\right)vec(Y)$$

$$= -\frac{m}{2}\log|P\Sigma^{-1}P^T| - \frac{n}{2}\log|\Sigma|$$

$$-\frac{1}{2}(vec(Y))^T\left(\Sigma^{-1} \otimes I_n - \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1} \otimes \Pi_A\right)vec(Y),$$

where $\Pi_A = A(A^TA)^{-1}A^T$. But easy calculation gives that the last expression is actually (apart from a constant)

$$-\frac{m}{2}\log|P\Sigma^{-1}P^T| - \frac{n}{2}\log|\Sigma| - \frac{1}{2}tr(Y - \Pi_AY\Pi_P)\Sigma^{-1}Y^T,$$

where $\Pi_P = \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P$. However, the expression within trace is also written as:

$$tr(Y - \Pi_AY\Pi_P)\Sigma^{-1}(Y - \Pi_AY\Pi_P)^T,$$

since

$$\begin{aligned} tr(Y-\Pi_A Y\Pi_P)\Sigma^{-1}\Pi_P^T Y^T\Pi_A &= trY\Sigma^{-1}\Pi_P^T Y^T\Pi_A - tr\Pi_A Y\Pi_P\Sigma^{-1}\Pi_P^T Y^T\Pi_A \\ &= trY\Sigma^{-1}\Pi_P^T Y^T\Pi_A - trY\Sigma^{-1}\Pi_P^T\Pi_P^T Y^T\Pi_A \\ &= trY\Sigma^{-1}\Pi_P^T Y^T\Pi_A - trY\Sigma^{-1}\Pi_P^T Y^T\Pi_A \\ &= 0. \end{aligned}$$

Thus the log-likelihood (2.56) is finally written as:

$$const - \frac{m}{2}\log|P\Sigma^{-1}P^T| - \frac{n}{2}\log|\Sigma| - \frac{1}{2}tr(Y-\Pi_A Y\Pi_P)\Sigma^{-1}(Y-\Pi_A Y\Pi_P)^T. \quad (2.63)$$

The form of (2.63) suggests that it is easily related to the likelihood of the ML method. From (2.30) and by using the estimate $\hat{\xi}$ of ξ as in (2.50), we get that this relationship is given by

$$l_R(\theta) = const - \frac{m}{2}\log|P\Sigma^{-1}P^T| + l_M(\hat{\xi},\theta). \quad (2.64)$$

Differentiation of (2.64) gives the relationship between the differentials of the two expressions which are to be maximized. Using properties (2.37) and by dropping the argument $\hat{\xi}$ since this is a function of θ, we get that

$$\begin{aligned} dl_R(\theta) &= \frac{m}{2}tr(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}d\Sigma\Sigma^{-1}p^T - \frac{n}{2}tr\Sigma^{-1}d\Sigma \\ &\quad + \frac{1}{2}tr(Y-\Pi_A Y\Pi_P)\Sigma^{-1}d\Sigma\Sigma^{-1}(Y-\Pi_A Y\Pi_P)^T \\ &= \frac{m}{2}tr\Sigma^{-1}p^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}d\Sigma - \frac{n}{2}tr\Sigma^{-1}d\Sigma \\ &\quad + \frac{1}{2}tr(Y-\Pi_A Y\Pi_P)\Sigma^{-1}d\Sigma\Sigma^{-1}(Y-\Pi_A Y\Pi_P)^T \Rightarrow \end{aligned}$$

$$dl_R(\theta) = \frac{m}{2}tr(Fd\Sigma) + dl_M(\theta), \quad (2.65)$$

where $F = \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}P\Sigma^{-1}$. Using this expression for the differential, it is not difficult to find similar expressions which relate the derivatives of the REML expression with those of ML. These are given by

$$\begin{aligned} \frac{\partial\ell_R}{\partial\tau^2} &= \frac{m}{2}tr(F) + \frac{\partial\ell_M}{\partial\tau^2} \\ \frac{\partial\ell_R}{\partial v^2} &= \frac{m}{2}tr(JF) + \frac{\partial\ell_M}{\partial v^2} \end{aligned}$$

$$\frac{\partial \ell_R}{\partial(\text{vec } M)} = m \text{ vec } (FMR) + \frac{\partial \ell_M}{\partial(\text{vec } M)}$$
$$\frac{\partial \ell_R}{\partial(\text{vec } R)} = \frac{m}{2}\text{vec } \left[M^T F M\right] + \frac{\partial \ell_M}{\partial(\text{vec } R)}. \tag{2.66}$$

Numerical maximization again is needed in most of the cases for the derivation of estimates of variance components. This maximization follows the same lines as in the ML case and can be programmed using relationship (2.65). In a similar context, McGilchrist and Cullis (1991) described an algorithm for estimation purposes. Closed form estimates however, have been shown to exist for the mixed effects model. Lange and Laird (1989) showed that

$$\hat{\tau}^2 = \frac{1}{n(q-c) - m(p-c)} tr(Y - \Pi_A Y \Pi_P)(I - \Pi_{P_1})(Y - \Pi_A Y \Pi_P)^T, \tag{2.67}$$

$$\hat{\Gamma} = \frac{1}{n-m}(P_1 P_1^T)^{-1} P_1 Y^T (I - \Pi_A) Y P_1^T (P_1 P_1^T)^{-1} - \hat{\tau}^2 (P_1 P_1^T)^{-1}, \tag{2.68}$$

A simple comparison of (2.35), (2.36) with (2.67) and (2.68) respectively, shows that the estimates differ by a multiplication constant which makes REML estimates of variance components always larger than the corresponding ML. If we examine now the effect of using these estimates to the expressions which give the real or estimated variance matrices of $\hat{\xi}$ we get that the real covariance matrix given by (2.51) is always the same since $\hat{\xi}_M$ and $\hat{\xi}_R$ coincide with the OLS estimator, while for the estimated covariance matrix given by (2.52) the following corollary is derived:

Corollary 1 *REML estimation gives larger estimated variances than ML for the estimated coefficients $\hat{\xi}$ in the case of the mixed random effects model.*

Proof: Elementary operations can show that

$$(A^T A)^{-1} \otimes (P \hat{\Sigma}_i^{-1} P^T)^{-1} = (A^T A)^{-1} \otimes (PP^T)^{-1} P(\hat{\tau}_i I_q + P_1^T \hat{\Gamma}_i P_1) P^T (PP^T)^{-1}$$

where $i = M$ for ML and $i = R$ for REML. The difference has the form

$$(A^T A)^{-1} \otimes (PP^T)^{-1} P((\hat{\tau}_R - \hat{\tau}_M) I_q + P_1^T (\hat{\Gamma}_R - \hat{\Gamma}_M) P_1) P^T (PP^T)^{-1}$$

which is positive definite.

Given now that the REML estimates of variance are unbiased (Harville, 1977) for the case considered above, we can see that REML gives values which are closer to the common real variance of $\hat{\xi}$ something which gives grounds for a kind of superiority of REML over ML. A question now arises naturally: is this property restricted to the mixed effects models or do wider classes of covariance models share it? This question will be examined in the next section. Before this let us note that the common real variance of $\hat{\xi}$ (whatever the method of estimation) holds only for mixed effects models as Rao (1965) has shown. For other types of structures the variances of $\hat{\xi}_M$ and $\hat{\xi}_R$ are different. Another question of a similar type can be whether we have anything to say about the two estimates $\widehat{Var(\hat{\xi}_M)}$ and $\widehat{Var(\hat{\xi}_R)}$ (which arises from (2.52) with 'R' in the place of 'M').

Let us examine expression (2.64) which can be rewritten as:

$$\begin{aligned} l_R(\boldsymbol{\theta}) &= -\frac{m}{2}\log|P\Sigma^{-1}P^T| + l_M(\boldsymbol{\theta}) \Rightarrow \\ l_R(\boldsymbol{\theta}) &= \frac{m}{2}\log\left|(P\Sigma^{-1}P^T)^{-1}\right| + l_M(\boldsymbol{\theta}) \Rightarrow \\ l_R(\boldsymbol{\theta}) &= \frac{m}{2}\sum_{i=1}^{p}\log\lambda_i(\theta) + l_M(\boldsymbol{\theta}), \end{aligned}$$

where for convention we set $\lambda_i\left((P\Sigma^{-1}P^T)^{-1}\right) \equiv \lambda_i(\boldsymbol{\theta})$, $i = 1, \ldots, p$, with $\boldsymbol{\theta}$ denoting the vector of variance components.

Suppose now that l_M has a maximum at $\hat{\boldsymbol{\theta}}_M$, with maximized value $l_M(\hat{\boldsymbol{\theta}}_M) \equiv \hat{l}_M$. Then, (2.64) becomes

$$l_R(\hat{\boldsymbol{\theta}}_M) = \frac{m}{2}\sum_{i=1}^{p}\log\lambda_i(\hat{\boldsymbol{\theta}}_M) + \hat{l}_M. \tag{2.69}$$

If we also maximize the restricted likelihood, we get as optimum $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_R$, and (2.64)

$$l_R(\hat{\boldsymbol{\theta}}_R) \equiv \hat{l}_R = \frac{m}{2}\sum_{i=1}^{p}\log\lambda_i(\hat{\boldsymbol{\theta}}_R) + l_M(\hat{\boldsymbol{\theta}}_R). \tag{2.70}$$

Subtracting now (2.69) from (2.70) we get

$$\hat{l}_R - l_R(\hat{\boldsymbol{\theta}}_M) = \frac{m}{2}\sum_{i=1}^{p}\log\frac{\lambda_i(\hat{\boldsymbol{\theta}}_R)}{\lambda_i(\hat{\boldsymbol{\theta}}_M)} + l_M(\hat{\boldsymbol{\theta}}_R) - \hat{l}_M.$$

But $\hat{l}_R - l_R(\hat{\theta}_M) > 0$ and $l_M(\hat{\theta}_R) - \hat{l}_M < 0$. Thus, we must necessarily have

$$\sum_{i=1}^{p} \log \frac{\lambda_i(\hat{\theta}_R)}{\lambda_i(\hat{\theta}_M)} > 0. \tag{2.71}$$

Given now that expression (2.71) is also written as $\log \frac{|(P\hat{\Sigma}_R^{-1}P^T)^{-1}|}{|(P\hat{\Sigma}_M^{-1}P^T)^{-1}|} > 0$, and since the log function is positive for arguments greater than 1, we get that

$$|(P\hat{\Sigma}_R^{-1}P^T)^{-1}| > |(P\hat{\Sigma}_M^{-1}P^T)^{-1}|. \tag{2.72}$$

Given now the fact that $(A^TA)^{-1}$ is a diagonal matrix, expression (2.72) means that the general estimated scatter of $\hat{\xi}$ is greater when we use REML than ML. The next Lemma will prove useful for a further examination of (2.71).

Lemma 2 $\forall\ x > 0$, *we have* $\log x \leq x - 1$.

Proof: Let us suppose that $1 < x$. Then, Rolle's theorem gives:

$$\log x = (x-1)\frac{1}{\omega}, \quad 1 \leq \omega \leq x,$$

which in turn gives

$$\log x \leq (x-1), \quad 1 \leq \omega \leq x.$$

If now $x < 1$ then we get

$$-\log x = (1-x)\frac{1}{\omega}, \quad x \leq \omega \leq 1$$

and thus again,

$$\log x \leq (x-1), \quad x \leq \omega \leq 1.$$

Direct application of this Lemma shows that

$$0 < \sum_{i=1}^{p} \log \frac{\lambda_i(\hat{\theta}_R)}{\lambda_i(\hat{\theta}_M)} \leq \sum_{i=1}^{p} \left\{ \frac{\lambda_i(\hat{\theta}_R)}{\lambda_i(\hat{\theta}_M)} - 1 \right\} \Rightarrow$$

$$\sum_{i=1}^{p} \frac{\lambda_i(\hat{\theta}_R)}{\lambda_i(\hat{\theta}_M)} \geq p. \tag{2.73}$$

Result (2.73) as well as (2.72), does not necessarily mean that $(P\hat{\Sigma}_R^{-1}P^T)^{-1} > (P\hat{\Sigma}_M^{-1}P^T)^{-1}$ in the sense that their difference is positive definite. But it certainly means that $(P\hat{\Sigma}_R^{-1}P^T)^{-1}$ can never be less than $(P\hat{\Sigma}_M^{-1}P^T)^{-1}$ in the same above sense. Unfortunately, we have not been able to extract any relationship between the diagonal elements of these matrices which represent the estimated variances of each coefficient. It is however very plausible from (2.72), that these are larger for REML than for ML. This along with the above mentioned questions, will be the subject of the simulation study in the next section.

2.4 Results of a Monte Carlo study

Data were generated with covariance matrix as specified by (2.5), (2.6) and (2.8), that is allowing for an ($AR(1)$ '1-change point' autoregressive error component with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Parameters τ^2, v^2 and σ^2 were set equal to 2.0, 0.5 and 1.5, respectively while ϕ_1 and ϕ_2 were chosen according to four different settings

- $\phi_1 = 0.6,\ \phi_2 = 0.2$;
- $\phi_1 = 0.9,\ \phi_2 = 0.2$;
- $\phi_1 = 1.5,\ \phi_2 = 1.1$;
- $\phi_1 = 1.9,\ \phi_2 = 0.8$.

For each such case, data were generated for seven time points ($q = 7$) and sample sizes $n = 12, 20, 50$ and eleven time points $q = 11$ and sample sizes $n = 15, 20, 50$. All observations were allocated to one group $m = 1$. The measurements were pseudo-random normal variables with covariance matrix Σ, as specified above with the change point set at the fourth time point, and known mean $A\xi P$ (c.f. (1.1)). For the latter we allowed a a quadratic ($p = 3$) mean trend with $\xi_0 = 20.0$, $\xi_1 = 10.0$ and $\xi_2 = 3.0$.

There were two reasons for which such small sample sizes were considered. First, these are quite common in growth curve studies where the number of individuals, often, is not much greater than the number of the time points at which each individual is

observed. Second, it is of interest to examine the small sample properties of maximum likelihood estimators.

For each setting a total of 50 simulation trials was carried out. For each trial five different covariance models were fitted:

(i) unrestricted covariance matrix;

(ii) uniform covariance model (c.f. (2.12));

Next, three first order autoregressive models:

(iii) with no change point; i.e. AR(1) as in (2.8);

(iv) with one change point at the third time point; i.e. (2.5) with $r = 3$;

(v) with one change point at the fourth time point; i.e $r = 4$ (the true model)

The simulated data were tested for normality using Mardia's measures of skewness and kurtosis. Mardia (1974) provided useful approximations for the distributions of these measures. The results did not show deviations from the normal population hypothesis.

Tables 2.1–2.8 show the Monte Carlo estimates of the ξ_i's $i = 1, 2$ for the four adopted settings of ϕ_1 and ϕ_2 when $q = 7$ (Tables 2.1–2.4) and when $q = 11$ (Tables 2.5–2.8), along with their estimated variances (in parentheses). The row marked with $*$ shows the ratios of the estimated over the corresponding true variances for all 5 models. The theoretical Gauss-Markov (optimal) variance of each coefficient is also shown at the left of the corresponding entry.

The results of these eight tables can be summarized as follows:

1) For all models except the unrestricted one it is clear that REML gives larger variances than ML for all coefficients. It seems therefore that the result of the previous corollary is in general also true for covariance matrices such as those of models (iii) to (v) (since model (ii) is a 1–random effect model). Note also that as the sample size increases, the differences between the REML and ML variances decrease.

2) For the unrestricted model, ML variances are always much larger than REML ones, but this is due to the estimate of variance of $\hat{\xi}$ we used. Grizzle and Allen (1969) showed that an unbiased estimate of variance of $\hat{\xi}$ is given by

$$\frac{n-m-1}{(n-m-c-1)(n-m-c)}(A^TA)^{-1} \otimes (PS^{-1}P^T)^{-1}. \tag{2.74}$$

If we had used instead of constant $\frac{n-m-1}{(n-m-c-1)(n-m-c)}$, the constant $\frac{1}{n}$ then the estimates of variance for ML estimation would be less than those of REML estimation. However, these differences would be negligible and would tend to zero as the sample size would increase. Note that such a problem does not arise for the rest of the covariance structures since the estimate of variance of $\hat{\xi}$ has the form (2.52) always (with subscript 'M' or 'R' corresponding to ML or REML estimation). Such form however, does not necessarily give unbiased estimates of variance as (2.74).

3) Both estimation methods give estimates of variances (for all coefficients) which are closest to the theoretical and the optimal ones, when model (v) (the correct one) is used. In this case they both underestimate the optimal variances for each model (ML more than REML), but the biases are very small and tend to zero as the sample size increases.

4) Model (ii) tends to underestimate the optimum variance when both ϕ_1 and ϕ_2 are less than 1. On the contrary, when at least one of ϕ's is greater than 1, gross overestimation is marked. Model (iii) does not have a consistent trend while in many cases it gives the largest estimated variances among models (ii), (iii), (iv) and (v). Finally, model (iv) underestimates the optimal variance except when $\phi_1 = 1.9$, $\phi_2 = 0.8$. These remarks hold, more or less, for the case when $q = 11$.

5) The ratios of the estimated over the real variances for all models are larger for REML than for ML quite consistently, showing that REML estimates of variance of ξ are closer to their corresponding real ones than those of ML. These ratios increase as the sample size increases and in some cases reach values more than 1, which is its ideal value. When both ϕ's are smaller than 1 and especially when $\phi_1 = 0.6$ and $\phi_2 = 0.2$,

the ratios, for all models, are more or less the same but the situation changes in the rest of the cases. The correct model (model (v)) shows clear stability to values near 1 while the others can be very poor (see Tables 2.3, 2.4, 2.7 and 2.8).

2.5 Conclusions

In growth curve studies it is possible that one or more external changes happen at certain times that may affect the behaviour of the experimental units and thus the variance of, or the correlation between successive measurements. Failure to take account of such drastic changes in modelling the growth curves can lead to inefficient estimation of regression coefficients. With the class of models studied in this chapter, it was demonstrated that when both ϕ_1 and ϕ_2 are less than one, wrong models provide variances of ξ_i's which in general, underestimate the optimal ones, in particular when the series is short. For the cases where ϕ_1 is larger than one, then these models can lead to grossly overestimated optimal variances and subsequently to misleading results. But it is not only the proximity to the optimal, Gauss–Markov, variance which matters. The degree to which the estimated variances of ξ_i's calculated from different models, underestimate or overestimate, their real variances can give a clearer idea to the way we should look at a particular method, if it is, for example, to choose between ML and REML. In this respect, REML is the method to be preferred since in general, it gives ratios and estimated variances of ξ_i's larger than those of ML. In this way the advantages of the one method over the other are two: first, given the fact that the real variances of ξ_i's are systematically underestimated (unless we make a gross error choice about the change point, (see model (iii)) where we frequently have overestimation), REML gives more efficient estimates and since in many cases the estimated variances are much less than the optimal Gauss–Markov variance, we get a less misleading inpression of accuracy.

Table 2.1: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.6$, $\phi_2 = 0.2$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 12$	ML	9.939	10.008	10.016	10.011	10.000
0.2494		(0.3719)	(0.2285)	(0.2455)	(0.2219)	(0.2329)
	*	(0.65)	(0.91)	(0.97)	(0.88)	(0.92)
	REML	9.939	10.008	10.016	10.007	9.997
		(0.1388)	(0.2351)	(0.2574)	(0.2324)	(0.2448)
	*	(0.35)	(0.93)	(1.02)	(0.92)	(0.97)
$n = 12$	ML	3.007	2.998	2.997	2.998	2.999
0.0039		(0.0056)	(0.0034)	(0.0037)	(0.0034)	(0.0036)
	*	(0.63)	(0.88)	(0.94)	(0.86)	(0.92)
	REML	3.007	2.998	2.997	2.999	3.000
		(0.0021)	(0.0035)	(0.0038)	(0.0035)	(0.0038)
	*	(0.34)	(0.90)	(0.99)	(0.91)	(0.97)
$n = 20$	ML	9.945	9.976	9.966	9.971	9.980
0.1497		(0.2017)	(0.1402)	(0.1553)	(0.1394)	(0.1475)
	*	(0.84)	(0.93)	(1.03)	(0.92)	(0.98)
	REML	9.945	9.976	9.966	9.972	9.983
		(0.1230)	(0.1426)	(0.1596)	(0.1440)	(0.1509)
	*	(0.61)	(0.94)	(1.06)	(0.95)	(1.00)
$n = 20$	ML	3.006	3.003	3.003	3.002	3.001
0.0023		(0.0030)	(0.0021)	(0.0023)	(0.0021)	(0.0023)
	*	(0.81)	(0.89)	(1.00)	(0.91)	(0.97)
	REML	3.006	3.003	3.003	3.002	3.001
		(0.0018)	(0.0021)	(0.0024)	(0.0022)	(0.0023)
	*	(0.60)	(0.91)	(1.03)	(0.94)	(1.00)
$n = 50$	ML	9.987	10.001	9.996	9.999	9.996
0.0599		(0.0693)	(0.0565)	(0.0629)	(0.0581)	(0.0605)
	*	(0.97)	(0.93)	(1.05)	(0.97)	(1.01)
	REML	9.987	10.001	9.996	9.999	9.996
		(0.0580)	(0.0568)	(0.0635)	(0.0587)	(0.0611)
	*	(0.86)	(0.94)	(1.06)	(0.97)	(1.02)
$n = 50$	ML	3.002	3.000	3.001	3.001	3.001
0.0009		(0.0011)	(0.0008)	(0.0009)	(0.0009)	(0.0009)
	*	(0.96)	(0.90)	(1.01)	(0.95)	(1.01)
	REML	3.002	3.000	3.001	3.001	3.001
		(0.0009)	(0.0008)	(0.0010)	(0.0009)	(0.0009)
	*	(0.86)	(0.91)	(1.02)	(0.96)	(1.02)

Table 2.2: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.9$, $\phi_2 = 0.2$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 12$	ML	9.903	10.016	10.003	10.001	10.015
0.2636		(0.4305)	(0.2379)	(0.2731)	(0.2334)	(0.2518)
	*	(0.77)	(0.86)	(1.00)	(0.85)	(0.93)
	REML	9.902	10.016	10.004	10.001	10.015
		(0.1594)	(0.2446)	(0.2863)	(0.2428)	(0.2608)
	*	(0.41)	(0.89)	(1.05)	(0.89)	(0.97)
$n = 12$	ML	3.011	2.997	2.998	2.999	2.996
0.0043		(0.0074)	(0.0036)	(0.0041)	(0.0037)	(0.0041)
	*	(0.81)	(0.78)	(0.91)	(0.82)	(0.92)
	REML	3.012	2.997	2.998	2.998	2.996
		(0.0027)	(0.0037)	(0.0043)	(0.0039)	(0.0042)
	*	(0.43)	(0.80)	(0.96)	(0.85)	(0.95)
$n = 20$	ML	10.002	10.001	10.007	9.999	9.995
0.1581		(0.1852)	(0.1462)	(0.1676)	(0.1468)	(0.1528)
	*	(0.75)	(0.88)	(1.02)	(0.90)	(0.95)
	REML	10.002	10.001	10.007	9.997	9.994
		(0.1127)	(0.1487)	(0.1722)	(0.1506)	(0.1551)
	*	(0.55)	(0.90)	(1.05)	(0.92)	(0.97)
$n = 20$	ML	2.998	2.998	2.998	2.998	2.999
0.0026		(0.0030)	(0.0022)	(0.0025)	(0.0023)	(0.0025)
	*	(0.75)	(0.79)	(0.93)	(0.86)	(0.94)
	REML	2.998	2.998	2.998	2.999	2.999
		(0.0018)	(0.0022)	(0.0026)	(0.0024)	(0.0025)
	*	(0.55)	(0.81)	(0.96)	(0.88)	(0.96)
$n = 50$	ML	10.008	10.005	10.010	10.009	10.006
0.0633		(0.0680)	(0.0597)	(0.0688)	(0.0610)	(0.0627)
	*	(0.93)	(0.90)	(1.06)	(0.94)	(0.99)
	REML	10.008	10.005	10.010	10.009	10.006
		(0.0569)	(0.0601)	(0.0696)	(0.0616)	(0.0635)
	*	(0.83)	(0.91)	(1.07)	(0.95)	(1.00)
$n = 50$	ML	2.999	2.999	2.999	2.999	2.999
0.0010		(0.0011)	(0.0009)	(0.0010)	(0.0010)	(0.0010)
	*	(0.94)	(0.81)	(0.97)	(0.91)	(0.98)
	REML	2.999	2.999	2.999	2.999	2.999
		(0.0009)	(0.0009)	(0.0010)	(0.0010)	(0.0010)
	*	(0.83)	(0.82)	(0.98)	(0.91)	(0.99)

Table 2.3: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.5$, $\phi_2 = 1.1$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 12$	ML	10.059	10.078	10.056	10.071	10.025
0.4442		(0.7852)	(0.5834)	(0.3574)	(0.3907)	(0.4222)
	*	(0.71)	(1.16)	(0.75)	(0.76)	(0.93)
	REML	10.061	10.078	10.055	10.074	10.013
		(0.2845)	(0.6000)	(0.3731)	(0.4026)	(0.4295)
	*	(0.37)	(1.19)	(0.79)	(0.78)	(0.95)
$n = 12$	ML	2.988	2.989	2.992	2.990	2.992
0.0041		(0.0070)	(0.0087)	(0.0047)	(0.0040)	(0.0040)
	*	(0.76)	(2.02)	(1.13)	(0.92)	(0.96)
	REML	2.988	2.989	2.992	2.990	2.992
		(0.0026)	(0.0090)	(0.0049)	(0.0041)	(0.0040)
	*	(0.40)	(2.08)	(1.17)	(0.96)	(0.98)
$n = 20$	ML	9.902	9.935	9.926	9.907	9.929
0.2665		(0.3359)	(0.3447)	(0.2190)	(0.2284)	(0.2564)
	*	(0.85)	(1.14)	(0.77)	(0.75)	(0.95)
	REML	9.902	9.935	9.926	9.904	9.930
		(0.2002)	(0.3505)	(0.2248)	(0.2396)	(0.2605)
	*	(0.61)	(1.16)	(0.79)	(0.79)	(0.97)
$n = 20$	ML	3.012	3.006	3.007	3.007	3.006
0.0024		(0.0031)	(0.0051)	(0.0029)	(0.0024)	(0.0024)
	*	(0.84)	(1.99)	(1.15)	(0.94)	(0.98)
	REML	3.012	3.006	3.007	3.007	3.006
		(0.0018)	(0.0052)	(0.0029)	(0.0025)	(0.0024)
	*	(0.61)	(2.02)	(1.17)	(0.98)	(0.99)
$n = 50$	ML	10.006	10.006	10.000	10.000	9.995
0.1066		(0.1153)	(0.1412)	(0.0904)	(0.0948)	(0.1053)
	*	(0.93)	(1.16)	(0.79)	(0.78)	(0.98)
	REML	10.006	10.006	9.999	10.000	9.994
		(0.0958)	(0.1421)	(0.0914)	(0.0946)	(0.1059)
	*	(0.82)	(1.17)	(0.80)	(0.78)	(0.99)
$n = 50$	ML	3.002	3.002	3.002	3.002	3.002
0.0010		(0.0011)	(0.0021)	(0.0012)	(0.0010)	(0.0010)
	*	(0.93)	(2.03)	(1.17)	(0.96)	(0.99)
	REML	3.002	3.002	3.002	3.002	3.003
		(0.0009)	(0.0021)	(0.0012)	(0.0010)	(0.0010)
	*	(0.83)	(2.05)	(1.18)	(0.96)	(1.00)

Table 2.4: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.9$, $\phi_2 = 0.8$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 12$	ML	9.795	9.967	9.949	9.982	9.994
0.6575		(1.0910)	(0.7527)	(0.9851)	(1.0679)	(0.6326)
	*	(0.74)	(0.29)	(0.53)	(0.41)	(0.95)
	REML	9.797	9.967	9.948	9.960	9.994
		(0.4035)	(0.7741)	(1.0300)	(1.0683)	(0.6468)
	*	(0.39)	(0.30)	(0.56)	(0.41)	(0.97)
$n = 12$	ML	3.020	3.003	3.003	3.000	3.000
0.0082		(0.0133)	(0.0112)	(0.0136)	(0.0124)	(0.0079)
	*	(0.73)	(0.37)	(0.62)	(0.42)	(0.95)
	REML	3.020	3.003	3.003	3.003	3.000
		(0.0049)	(0.0116)	(0.0142)	(0.0123)	(0.0080)
	*	(0.39)	(0.39)	(0.65)	(0.41)	(0.97)
$n = 20$	ML	10.009	10.184	10.155	10.170	10.042
0.3945		(0.5331)	(0.4666)	(0.6120)	(0.6539)	(0.3878)
	*	(0.91)	(0.30)	(0.56)	(0.42)	(0.98)
	REML	10.009	10.184	10.155	10.174	10.041
		(0.3200)	(0.4745)	(0.6292)	(0.6519)	(0.3912)
	*	(0.65)	(0.31)	(0.58)	(0.42)	(0.99)
$n = 20$	ML	2.999	2.980	2.983	2.981	2.995
0.0049		(0.0068)	(0.0070)	(0.0085)	(0.0077)	(0.0048)
	*	(0.93)	(0.39)	(0.66)	(0.43)	(0.98)
	REML	2.999	2.980	2.983	2.981	2.995
		(0.0041)	(0.0071)	(0.0087)	(0.0076)	(0.0049)
	*	(0.67)	(0.39)	(0.68)	(0.42)	(0.99)
$n = 50$	ML	9.963	9.996	9.987	10.011	9.995
0.1578		(0.1672)	(0.1878)	(0.2439)	(0.2643)	(0.1569)
	*	(0.90)	(0.31)	(0.56)	(0.42)	(0.99)
	REML	9.963	9.996	9.987	10.005	9.995
		(0.1398)	(0.1891)	(0.2460)	(0.2660)	(0.1572)
	*	(0.81)	(0.31)	(0.57)	(0.42)	(1.00)
$n = 50$	ML	3.002	2.999	2.999	2.997	2.999
0.0020		(0.0021)	(0.0028)	(0.0034)	(0.0031)	(0.0020)
	*	(0.91)	(0.39)	(0.66)	(0.43)	(0.99)
	REML	3.002	2.999	2.999	2.998	2.999
		(0.0017)	(0.0028)	(0.0034)	(0.0031)	(0.0020)
	*	(0.81)	(0.39)	(0.66)	(0.43)	(0.99)

Table 2.5: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.6$, $\phi_2 = 0.2$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 15$	ML	10.085	10.006	10.010	10.007	10.008
0.0466		(0.1169)	(0.0424)	(0.0489)	(0.0463)	(0.0455)
	*	(0.43)	(0.89)	(1.04)	(0.98)	(0.97)
	REML	10.084	10.006	10.010	10.008	10.009
		(0.0185)	(0.0430)	(0.0505)	(0.0474)	(0.0467)
	*	(0.15)	(0.91)	(1.07)	(1.00)	(1.00)
$n = 15$	ML	2.994	3.001	3.000	3.000	3.000
0.0003		(0.0008)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
	*	(0.45)	(0.92)	(1.06)	(0.99)	(0.99)
	REML	2.994	3.001	3.000	3.000	3.000
		(0.0001)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
	*	(0.15)	(0.93)	(1.10)	(1.01)	(1.01)
$n = 20$	ML	10.069	10.024	10.025	10.026	10.026
0.0350		(0.0653)	(0.0311)	(0.0360)	(0.0339)	(0.0333)
	*	(0.67)	(0.87)	(1.02)	(0.96)	(0.95)
	REML	10.069	10.024	10.025	10.027	10.026
		(0.0205)	(0.0314)	(0.0368)	(0.0348)	(0.0340)
	*	(0.34)	(0.88)	(1.04)	(0.98)	(0.97)
$n = 20$	ML	2.995	2.998	2.998	2.998	2.998
0.0002		(0.0004)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
	*	(0.66)	(0.90)	(1.04)	(0.97)	(0.95)
	REML	2.995	2.998	2.998	2.998	2.998
		(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
	*	(0.34)	(0.91)	(1.07)	(0.99)	(0.97)
$n = 50$	ML	9.995	9.992	9.993	9.994	9.993
0.0140		(0.0162)	(0.0123)	(0.0147)	(0.0138)	(0.0136)
	*	(0.82)	(0.87)	(1.04)	(0.98)	(0.97)
	REML	9.995	9.992	9.993	9.994	9.993
		(0.0112)	(0.0124)	(0.0148)	(0.0140)	(0.0137)
	*	(0.66)	(0.87)	(1.05)	(0.99)	(0.98)
$n = 50$	ML	3.000	3.001	3.001	3.001	3.001
0.0001		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	*	(0.80)	(0.89)	(1.07)	(0.98)	(0.97)
	REML	3.000	3.001	3.001	3.001	3.001
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	*	(0.65)	(0.89)	(1.08)	(0.99)	(0.98)

Table 2.6: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.9$, $\phi_2 = 0.2$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 15$	ML	10.073	10.023	10.018	10.022	10.025
0.0450		(0.1182)	(0.0459)	(0.0554)	(0.0490)	(0.0443)
	*	(0.56)	(0.98)	(1.21)	(1.07)	(0.98)
	REML	10.073	10.023	10.018	10.019	10.027
		(0.0188)	(0.0465)	(0.0572)	(0.0500)	(0.0455)
	*	(0.19)	(0.99)	(1.25)	(1.10)	(1.01)
$n = 15$	ML	2.996	2.998	2.999	2.999	2.999
0.0003		(0.0008)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
	*	(0.58)	(1.02)	(1.24)	(1.08)	(0.98)
	REML	2.996	2.998	2.999	2.999	2.998
		(0.0001)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
	*	(0.20)	(1.03)	(1.28)	(1.11)	(1.01)
$n = 20$	ML	9.954	10.000	10.006	10.002	10.002
0.0338		(0.0638)	(0.0324)	(0.0400)	(0.0347)	(0.0331)
	*	(0.58)	(0.92)	(1.17)	(1.01)	(0.98)
	REML	9.955	10.000	10.006	10.003	10.003
		(0.0200)	(0.0327)	(0.0409)	(0.0351)	(0.0335)
	*	(0.30)	(0.93)	(1.20)	(1.03)	(0.99)
$n = 20$	ML	3.004	3.001	3.001	3.001	3.001
0.0002		(0.0004)	(0.0002)	(0.0003)	(0.0002)	(0.0002)
	*	(0.59)	(0.96)	(1.19)	(1.03)	(0.98)
	REML	3.004	3.001	3.000	3.001	3.001
		(0.0001)	(0.0002)	(0.0003)	(0.0002)	(0.0002)
	*	(0.30)	(0.97)	(1.22)	(1.04)	(1.00)
$n = 50$	ML	10.006	10.004	10.002	10.004	10.004
0.0135		(0.0156)	(0.0136)	(0.0172)	(0.0145)	(0.0135)
	*	(0.82)	(0.96)	(1.27)	(1.06)	(1.00)
	REML	10.006	10.004	10.002	10.003	10.004
		(0.0108)	(0.0136)	(0.0174)	(0.0147)	(0.0136)
	*	(0.67)	(0.97)	(1.28)	(1.08)	(1.01)
$n = 50$	ML	2.999	2.999	2.999	2.999	2.999
0.0001		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	*	(0.82)	(1.00)	(1.28)	(1.08)	(1.00)
	REML	2.999	2.999	2.999	2.999	2.999
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	*	(0.66)	(1.00)	(1.30)	(1.10)	(1.01)

Table 2.7: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.5$, $\phi_2 = 1.1$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 15$	ML	9.996	10.014	10.012	10.034	10.011
0.1543		(0.3837)	(0.1682)	(0.1139)	(0.1146)	(0.1421)
	*	(0.45)	(0.92)	(0.65)	(0.65)	(0.90)
	REML	9.996	10.014	10.012	10.040	10.011
		(0.0600)	(0.1704)	(0.1170)	(0.1179)	(0.1434)
	*	(0.15)	(0.93)	(0.67)	(0.67)	(0.91)
$n = 15$	ML	3.002	2.998	2.998	2.998	2.998
0.0005		(0.0013)	(0.0011)	(0.0006)	(0.0005)	(0.0005)
	*	(0.46)	(2.04)	(1.23)	(0.95)	(0.96)
	REML	3.002	2.998	2.998	2.998	2.998
		(0.0002)	(0.0011)	(0.0007)	(0.0005)	(0.0005)
	*	(0.16)	(2.07)	(1.27)	(0.96)	(0.97)
$n = 20$	ML	9.962	9.982	9.999	9.994	10.008
0.1157		(0.2238)	(0.1173)	(0.0871)	(0.0891)	(0.1099)
	*	(0.71)	(0.86)	(0.66)	(0.67)	(0.93)
	REML	9.962	9.982	9.999	9.994	10.009
		(0.0692)	(0.1185)	(0.0888)	(0.0907)	(0.1120)
	*	(0.36)	(0.87)	(0.67)	(0.68)	(0.95)
$n = 20$	ML	2.998	2.995	2.994	2.995	2.994
0.0004		(0.0008)	(0.0008)	(0.0005)	(0.0004)	(0.0004)
	*	(0.72)	(1.90)	(1.23)	(0.93)	(0.99)
	REML	2.998	2.995	2.994	2.995	2.994
		(0.0002)	(0.0008)	(0.0005)	(0.0004)	(0.0004)
	*	(0.37)	(1.91)	(1.26)	(0.95)	(1.01)
$n = 50$	ML	9.971	9.988	9.984	9.991	9.987
0.0463		(0.0543)	(0.0533)	(0.0362)	(0.0363)	(0.0450)
	*	(0.83)	(0.98)	(0.69)	(0.69)	(0.96)
	REML	9.971	9.988	9.984	9.991	9.986
		(0.0373)	(0.0535)	(0.0365)	(0.0368)	(0.0449)
	*	(0.67)	(0.98)	(0.69)	(0.70)	(0.96)
$n = 50$	ML	3.001	3.001	3.001	3.001	3.001
0.0002		(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0001)
	*	(0.86)	(2.15)	(1.29)	(0.98)	(0.98)
	REML	3.001	3.001	3.001	3.001	3.001
		(0.0001)	(0.0004)	(0.0002)	(0.0002)	(0.0001)
	*	(0.69)	(2.16)	(1.30)	(0.99)	(0.98)

Table 2.8: Estimates (averages over 50 simulations) of (ξ_1, ξ_2) for five models of covariance matrix; averages of estimated variances in brackets using ML and REML. Population values: $\xi_1 = 10.0$, $\xi_2 = 3.0$. Variance component setting $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.9$, $\phi_2 = 0.8$

		Unrestricted	Uniform Correlation	Auto-regressive	Autoregressive 1-thr. r=3	Autoregressive 1-thr. r=4
$n = 15$	ML	10.038	10.013	10.019	9.991	9.953
0.1246		(0.2888)	(0.1231)	(0.2726)	(0.1626)	(0.1212)
	*	(0.46)	(0.25)	(0.59)	(0.56)	(0.95)
	REML	10.037	10.013	10.019	9.990	9.951
		(0.0457)	(0.1248)	(0.2801)	(0.1642)	(0.1229)
	*	(0.16)	(0.25)	(0.60)	(0.56)	(0.97)
$n = 15$	ML	2.996	2.998	2.999	3.000	3.004
0.0009		(0.0020)	(0.0008)	(0.0018)	(0.0011)	(0.0008)
	*	(0.47)	(0.22)	(0.59)	(0.52)	(0.96)
	REML	2.996	2.998	2.999	3.001	3.004
		(0.0003)	(0.0008)	(0.0018)	(0.0011)	(0.0008)
	*	(0.16)	(0.22)	(0.60)	(0.52)	(0.97)
$n = 20$	ML	10.039	10.011	10.004	10.014	9.999
0.0935		(0.1591)	(0.0944)	(0.2089)	(0.1219)	(0.0885)
	*	(0.60)	(0.25)	(0.60)	(0.57)	(0.93)
	REML	10.039	10.011	10.004	10.024	9.996
		(0.0497)	(0.0954)	(0.2124)	(0.1194)	(0.0902)
	*	(0.31)	(0.26)	(0.61)	(0.55)	(0.95)
$n = 20$	ML	2.995	2.998	2.998	2.998	2.999
0.0006		(0.0011)	(0.0006)	(0.0014)	(0.0008)	(0.0006)
	*	(0.59)	(0.22)	(0.60)	(0.52)	(0.94)
	REML	2.995	2.998	2.998	2.997	2.999
		(0.0003)	(0.0006)	(0.0014)	(0.0008)	(0.0006)
	*	(0.30)	(0.22)	(0.61)	(0.51)	(0.96)
$n = 50$	ML	9.972	9.954	9.961	9.969	9.978
0.0374		(0.0455)	(0.0401)	(0.0879)	(0.0463)	(0.0367)
	*	(0.84)	(0.27)	(0.64)	(0.56)	(0.98)
	REML	9.972	9.954	9.961	9.966	9.978
		(0.0313)	(0.0403)	(0.0891)	(0.0468)	(0.0370)
	*	(0.68)	(0.27)	(0.65)	(0.56)	(0.99)
$n = 50$	ML	3.003	3.005	3.004	3.003	3.003
0.0003		(0.0003)	(0.0003)	(0.0006)	(0.0003)	(0.0003)
	*	(0.86)	(0.23)	(0.64)	(0.51)	(0.98)
	REML	3.003	3.005	3.004	3.004	3.003
		(0.0002)	(0.0003)	(0.0006)	(0.0003)	(0.0003)
	*	(0.69)	(0.24)	(0.65)	(0.52)	(0.99)

Chapter 3

An extension to multivariate growth curves

3.1 Introduction and structure of the model

In Chapter 2, we presented some models in the covariance structure analysis which could be useful in the context of growth curve models when one characteristic is measured at a time. We also showed that when two similar methods of analysis, ML or REML, are used we expect to get larger estimated variances for the estimated coefficients of the mean when we use REML rather than ML. We also conjectured that this happens quite independently of the particular covariance model in use. In this chapter we shall consider the case of multiple observations which are correlated at each time point. The situation can be called 'multivariate growth curves'. Early references include work by Reinsel (1982, 1984, 1985) with the last two publications devoted to prediction, and Anderson, Anderson and Olkin (1986) who considered the case of ML estimation under the condition that all estimated variance components, which are now matrices, be positive definite. These authors used one random effect, appropriately defined, to explain the variation of the multivariate observations while Reinsel did not consider estimates of variance components to be positive definite necessarily. In this chapter we shall deal exactly with this situation. That is, we shall generalize to the mixed effects model, adopting an arbitrary number of random effects, and we shall also

present ML and REML estimation under the condition that the estimates of variance components matrices will be positive definite. We shall also find the relationship between the ML and REML estimates and in this way, we shall generalize the results of mixed effects models to the multivariate growth curves situation. In the rest of the section, we shall appropriately define the model, in Section 3.2 we shall consider and expand the likelihood of observations, in Section 3.3 we shall derive the sufficient statistics, in Sections 3.4 and 3.5 we shall consider ML and REML estimation respectively and in 3.6 we shall find the relationships between the two estimates. Finally, the chapter will close with a discussion on the results.

Suppose that we have m treatment groups with n_j individuals in each group. Each individual has a complete record of measurements on r characteristics at each of q time-points. The experimental design is such that measurements for different individuals are independent. These measurements are denoted by y_{ijkl} where $i = 1, \ldots, n_j$, $j = 1, \ldots, m$, $k = 1, \ldots, q$ and $l = 1, \ldots, r$. Let us denote by

$$\boldsymbol{y}_{ijk} = (y_{ijk1}, \ldots, y_{ijkr})^T, \tag{3.1}$$

the $r \times 1$ vector which consists of the measurements of the r characteristics on the ith individual in the jth treatment group at the kth time-point. This vector can be modelled by

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_{ijk} + \boldsymbol{u}_{ijk}, \tag{3.2}$$

where

$$\boldsymbol{\mu}_{ijk} = (\mu_{ijk1}, \ldots, \mu_{ijkr})^T, \tag{3.3}$$

is a $r \times 1$ vector of constants representing the means and

$$\boldsymbol{u}_{ijk} = (u_{ijk1} \ldots, u_{ijkr})^T, \tag{3.4}$$

is a $r \times 1$ vector of errors with $E(\boldsymbol{u}_{ijk}) = 0$ and $E(\boldsymbol{u}_{ijk}\boldsymbol{u}_{ijk}^T) = \Sigma_1$

Next, we define the $r \times q$ matrices $\forall \quad i = 1, \ldots, n_j$, $j = 1, \ldots, m$

$$Y_{ij} = (\boldsymbol{y}_{ij1}, \ldots, \boldsymbol{y}_{ijq}) \tag{3.5}$$

consisting of the measurements of the r characteristics at the q time-points of the ith individual in the jth treatment group. Similarly, we define the $r \times q$ matrices of means $\forall \quad i = 1, \ldots, n_j,\ j = 1, \ldots, m$

$$M_{ij} = (\boldsymbol{\mu}_{ij1}, \ldots, \boldsymbol{\mu}_{ijq}) \tag{3.6}$$

and the $r \times q$ matrices of errors

$$U_{ij} = (\boldsymbol{u}_{ij1}, \ldots, \boldsymbol{u}_{ijq}) \tag{3.7}$$

$\forall \quad i = 1, \ldots, n_j,\ j = 1, \ldots, m$, with

$$E(U_{ij}) = 0 \quad \text{and} \quad Var(vec(U_{ij})) = I_q \otimes \Sigma_1 \tag{3.8}$$

where $\otimes$ represents the Kronecker product as defined in the Appendix. From (3.2) we can see that the corresponding model for matrices (3.5), (3.6) and (3.7) is

$$Y_{ij} = M_{ij} + U_{ij} \tag{3.9}$$

However, in many cases we wish to give a structure to M_{ij} which will simplify the model. To that respect, we shall assume that each component of the vector $\boldsymbol{\mu}_{ijk}$ will be explained by a polynomial of degree $p-1$ and that the polynomial parameters will be different for each component. These parameters will characterize the mean. The relationship can be expressed as

$$M_{ij} = \Xi_{ij} P \tag{3.10}$$

where Ξ_{ij} is a $r \times p$ matrix of unknown parameters and P is a $p \times q$ matrix of known (orthogonal) polynomial coefficients and of full row rank $p \leq q$. Matrix Ξ_{ij} could further be simplified by assuming that it depends linearly on a matrix Q $(r \times s)$ (assumed to be known) which is of full column rank $s \leq r$ and is common for all individuals and a matrix ξ_{0j} $(s \times p)$ of unknown parameters which will be different for each treatment group. Consequently, M_{ij} would be written $M_{ij} = Q\xi_{0j}$ and (3.9) as

$$Y_{ij} = Q\xi_{0j}P + U_{ij}.$$

However, this parameterization would complicate the model a lot, something which we shall avoid. We believe that parameterization (3.10) provides an adequate model for the problem and from now on we shall examine only the model

$$Y_{ij} = \Xi_{ij} P + U_{ij} \tag{3.11}$$

Matrix U_{ij} will not be considered as the only source of variation in the model. Assume that Ξ_{ij} is divided into two parts, the first depending on the individual in a specific treatment group and on its treatment group and the other on the treatment group at which this individual belongs. Such a structure can be expressed as

$$\Xi_{ij} = (\Xi^1_{ij}, \Xi^2_j) \tag{3.12}$$

where Ξ^1_{ij} is a $r \times t$ and Ξ^2_j a $r \times (p-t)$ matrix. Consider the corresponding expansion for P as

$$P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \tag{3.13}$$

where P_1 is a $t \times q$ and P_2 a $(p-t) \times q$ matrix. We assume that Ξ^1_{ij} is a random variable with

$$E(\Xi^1_{ij}) = \Xi^0_j \quad \text{and} \quad Var(vec(\Xi^1_{ij})) = I_t \otimes \Sigma_2 \tag{3.14}$$

where Σ_2 is a covariance matrix common for every treatment group. Such a structure will be called a t-random effects one. The assumption of randomness for Ξ^1_{ij} makes us conveniently write

$$\Xi^1_{ij} = \Xi^0_j + V_{ij} \tag{3.15}$$

with V_{ij} a $r \times t$ matrix of errors whose each column has mean 0 and variance Σ_2 and columns are independent each other. Thus,

$$Var(vec(V_{ij})) = I_t \otimes \Sigma_2$$

Under these conditions (3.11) becomes

$$\begin{aligned} Y_{ij} &= \Xi^1_{ij} P_1 + \Xi^2_j P_2 + U_{ij} \Rightarrow \\ Y_{ij} &= (\Xi^0_j + V_{ij}) P_1 + \Xi^2_j P_2 + U_{ij} \Rightarrow \end{aligned}$$

$$
\begin{aligned}
Y_{ij} &= \Xi_j^0 P_1 + \Xi_j^2 P_2 + V_{ij} P_1 + U_{ij} \Rightarrow \\
Y_{ij} &= \bar{\Xi}_j P + V_{ij} P_1 + U_{ij} \qquad (3.16)
\end{aligned}
$$

where $\bar{\Xi}_j = (\Xi_j^0, \Xi_j^2)$. By vectorizing matrix Y_{ij} we get that

$$
vec(Y_{ij}) = (P^T \otimes I_r) vec(\bar{\Xi}_j) + (P_1^T \otimes I_r) vec(V_{ij}) + vec(U_{ij}) \qquad (3.17)
$$

Let us now define matrix Y as

$$
Y = \begin{pmatrix} (vec(Y_{11}))^T \\ (vec(Y_{21}))^T \\ \vdots \\ (vec(Y_{n_m m}))^T \end{pmatrix} \qquad (3.18)
$$

which is a $n \times qr$ matrix consisting of all the observed data and $n = \sum_{i=1}^{m} n_i$. Similar definitions for ξ^* $(m \times pr)$, V $(n \times rt)$ and U $(n \times qr)$ are

$$
\xi^* = \begin{pmatrix} (vec(\bar{\Xi}_1))^T \\ (vec(\bar{\Xi}_2))^T \\ \vdots \\ (vec(\bar{\Xi}_m))^T \end{pmatrix} \qquad (3.19)
$$

$$
V = \begin{pmatrix} (vec(V_{11}))^T \\ (vec(V_{21}))^T \\ \vdots \\ (vec(V_{n_m m}))^T \end{pmatrix} \qquad (3.20)
$$

$$
U = \begin{pmatrix} (vec(U_{11}))^T \\ (vec(U_{21}))^T \\ \vdots \\ (vec(U_{n_m m}))^T \end{pmatrix}. \qquad (3.21)
$$

Then, model (3.17) becomes in matrix form

$$
Y = A\xi^*(P \otimes I_r) + V(P_1 \otimes I_r) + U \qquad (3.22)
$$

where A is a design matrix which assigns the appropriate model equation to a specific individual into its treatment group and V and U are error matrices with

$$
E(V) = 0 \quad \text{and} \quad Var(vec(V)) = (I_t \otimes \Sigma_2) \otimes I_n
$$

and

$$
E(U) = 0 \quad \text{and} \quad Var(vec(U)) = (I_q \otimes \Sigma_1) \otimes I_n
$$

Note that in the case where each treatment group contains the same number of individuals, $n_0 = n/m$, then

$$A = I_m \otimes \mathbf{1}_{n_0}$$

where $\mathbf{1}_{n_0} = (1, \ldots, 1)^T$ is a $n_0 \times 1$ vector.

Under the above assumptions the variance matrix of observations becomes

$$Var(vec(Y)) = (P_1^T \otimes I_r \otimes I_n)(I_t \otimes \Sigma_2 \otimes I_n)(P_1 \otimes I_r \otimes I_n) + (I_q \otimes \Sigma_1 \otimes I_n) \Rightarrow$$

$$Var(vec(Y)) = (I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2) \otimes I_n \tag{3.23}$$

The model considered in (3.22) is a more general one than that used by Reinsel (1982). This is easily seen since Reinsel considered only one random effect to explain the variation within individuals by setting $P_1 = (1, 1, \ldots, 1)$ a $1 \times q$ vector.

3.2 Expansion of likelihood

Using the results of the previous section we can now form the log-likelihood function of the observations. This will be

$$c + \log |(I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2) \otimes I_n|^{-1/2}$$

$$-tr[Y - A\xi^*(P \otimes I_r)](I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2)^{-1}[Y - A\xi^*(P \otimes I_r)]^T$$

$$= c - \frac{n}{2} \log |I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2|$$

$$- tr[Y - A\xi^*(P \otimes I_r)](I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2)^{-1}[Y - A\xi^*(P \otimes I_r)]^T \tag{3.24}$$

where c is a constant independent of parameters of interest. Consider the following definition:

Definition 2 *For a $(m \times n)$ matrix B with full row rank its projection $\Pi_B^{\Sigma^{-1}}$ $(n \times n)$ weighted by Σ^{-1} will be defined as*

$$\Pi_B^{\Sigma^{-1}} = \Sigma^{-1} B^T (B \Sigma^{-1} B^T)^{-1} B$$

If the matrix has full column rank then

$$\Pi_B^{\Sigma^{-1}} = B(B^T\Sigma^{-1}B)^{-1}B^T\Sigma^{-1}$$

There are many nice properties of the projections which will not be mentioned here unless they help in the derivation of the results. However, and in keeping the notation as simple as possible we shall use the following symbols for the projections which will be of extensive use in this chapter:

$$\Pi_{P\otimes I_r}^{\Omega^{-1}} \equiv \Pi_{P\otimes I_r} = \Omega^{-1}(P^T \otimes I_r)\left((P \otimes I_r)\Omega^{-1}(P^T \otimes I_r)\right)^{-1}(P \otimes I_r) \tag{3.25}$$

$$\Pi_{P_1\otimes I_r}^{\Omega^{-1}} \equiv \Pi_{P_1\otimes I_r} = \Omega^{-1}(P_1^T \otimes I_r)\left((P_1 \otimes I_r)\Omega^{-1}(P_1^T \otimes I_r)\right)^{-1}(P_1 \otimes I_r) \tag{3.26}$$

and

$$\Pi_A^{I_n} \equiv \Pi_A = A(A^TA)^{-1}A^T \tag{3.27}$$

where

$$\Omega = I_q \otimes \Sigma_1 + P_1^TP_1 \otimes \Sigma_2 \tag{3.28}$$

A nice property of $\Pi_{P\otimes I_r}$ and $\Pi_{P_1\otimes I_r}$ is given in the following theorem:

Theorem 2 *For $\Pi_{P\otimes I_r}$ as defined in (3.25) we get that*

$$(P_1 \otimes I_r)\Pi_{P\otimes I_r} = P_1 \otimes I_r \tag{3.29}$$

$$(P_2 \otimes I_r)\Pi_{P\otimes I_r} = P_2 \otimes I_r \tag{3.30}$$

Proof: From the definition of the projection, we must have

$$(P \otimes I_r)\Pi_{P\otimes I_r} = P \otimes I_r$$

But it is known (see in the Appendix property (A10)) that

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \otimes I_r = \begin{pmatrix} P_1 \otimes I_r \\ P_2 \otimes I_r \end{pmatrix}$$

Thus,

$$(P \otimes I_r)\Pi_{P\otimes I_r} = \left(\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \otimes I_r\right)\Pi_{P\otimes I_r} =$$

$$\begin{pmatrix} (P_1 \otimes I_r)\Pi_{P\otimes I_r} \\ (P_2 \otimes I_r)\Pi_{P\otimes I_r} \end{pmatrix} = P \otimes I_r = \begin{pmatrix} P_1 \otimes I_r \\ P_2 \otimes I_r \end{pmatrix}$$

and the proof is complete.

The technique which will be followed into the derivation of sufficient statistics is to transform the vector of observations $vec(Y)$ into five uncorrelated random variables whose joint likelihood will be the likelihood of observations. Consider the following five transformations of $vec(Y)$.

1. $(\Pi^T_{P_1\otimes I_r} \otimes \Pi_A)vec(Y) = Y_1$
2. $(\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))vec(Y) = Y_2$
3. $((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)vec(Y) = Y_3$
4. $((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))vec(Y) = Y_4$
5. $((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)vec(Y) = Y_5$

These random variables are all of size $qr \times 1$. Their sum has the form

$$[(\Pi^T_{P_1\otimes I_r} \otimes \Pi_A) + (\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A)) + ((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)$$

$$+((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A)) + ((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)]vec(Y) =$$

$$[(\Pi^T_{P_1\otimes I_r} \otimes I_n) + ((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes I_n) + ((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)]vec(Y) = vec(Y)$$

The above equation shows that these variables split the log-likelihood of observations into five parts. These are also uncorrelated as the following theorem certifies:

Theorem 3 *The vectors $Y_1, \ldots, Y_5$ are uncorrelated random variables, normally distributed, with means and variances given by*

1.

$$E(Y_1) = [\Pi^T_{P_1 \otimes I_r}(P^T \otimes I_r) \otimes A]vec(\xi^*)$$

$$Var(Y_1) = \Pi^T_{P_1 \otimes I_r} \Omega \Pi_{P_1 \otimes I_r} \otimes \Pi_A$$

2.

$$E(Y_2) = 0$$

$$Var(Y_2) = \Pi^T_{P_1 \otimes I_r} \Omega \Pi_{P_1 \otimes I_r} \otimes (I_n - \Pi_A)$$

3.

$$E(Y_3) = [(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T (P^T \otimes I_r) \otimes A]vec(\xi^*)$$

$$Var(Y_3) = (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \Omega (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) \otimes \Pi_A$$

4.

$$E(Y_4) = 0$$

$$Var(Y_4) = (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \Omega (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) \otimes (I_n - \Pi_A)$$

5.

$$E(Y_5) = 0$$

$$Var(Y_5) = (I_{qr} - \Pi_{P \otimes I_r})^T \Omega (I_{qr} - \Pi_{P \otimes I_r}) \otimes I_n$$

Proof: It is easy to show that the variables are uncorrelated since for example

$$Cov(Y_2, Y_4) = (\Pi^T_{P_1 \otimes I_r} \Omega (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})) \otimes (I_n - \Pi_A))$$

$$= (\Omega \Pi_{P_1 \otimes I_r} (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})) \otimes (I_n - \Pi_A))$$

$$= (\Omega (\Pi_{P_1 \otimes I_r} - \Pi_{P_1 \otimes I_r})) \otimes (I_n - \Pi_A)) = 0$$

The rest of the theorem follows easily since the transformed variables $Y_1, Y_2, \ldots, Y_5$ are linear functions of $vec(Y)$ which is normally distributed.

It can easily be seen from the above theorem that the variance matrices of $Y_1 \ldots, Y_5$ are singular. Following Mardia, Kent and Bibby (1982) we have to find corresponding generalized inverses in order to define the appropriate likelihoods. These are given in the following theorem:

Theorem 4 *Under Definition 2 and with* $\Omega = I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2$, *the following results hold:*

$$[\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \otimes \Pi_A]^- = \Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi_{P_1 \otimes I_r}^T \otimes \Pi_A$$

$$[\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \otimes (I_n - \Pi_A)]^- = \Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi_{P_1 \otimes I_r}^T \otimes (I_n - \Pi_A)$$

$$[(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \Omega (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) \otimes \Pi_A]^-$$

$$= (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \otimes \Pi_A$$

$$[(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \Omega (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) \otimes (I_n - \Pi_A)]^-$$

$$= (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \otimes (I_n - \Pi_A)$$

$$[(I_{qr} - \Pi_{P \otimes I_r})^T \Omega (I_{qr} - \Pi_{P \otimes I_r}) \otimes I_n]^-$$

$$= (I_{qr} - \Pi_{P \otimes I_r})(I_q \otimes \Sigma_1^{-1})(I_{qr} - \Pi_{P \otimes I_r})$$

Proof: The proof is based on the following property:

$$\Omega \Pi_{P_1 \otimes I_r} = \Pi_{P_1 \otimes I_r}^T \Omega \tag{3.31}$$

We shall prove the first case only since the others arise in a similar way.

By definition, the generalized inverse of a matrix A, A^-, should be such that

$$AA^-A = A$$

Then we can see for example, that in the case of $\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \otimes \Pi_A$, we have,

$$(\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \otimes \Pi_A)(\Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi_{P_1 \otimes I_r}^T \otimes \Pi_A)(\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \otimes \Pi_A) =$$

$$= (\Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi_{P_1 \otimes I_r}^T \Omega \Pi_{P_1 \otimes I_r}) \otimes \Pi_A =$$

$$= (\Pi^T_{P_1\otimes I_r}\Pi^T_{P_1\otimes I_r}\Omega\Omega^{-1}\Pi^T_{P_1\otimes I_r}\Omega\Pi_{P_1\otimes I_r}) \otimes \Pi_A =$$

$$(\Pi^T_{P_1\otimes I_r}\Omega\Pi_{P_1\otimes I_r}) \otimes \Pi_A$$

which means that $\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi^T_{P_1\otimes I_r} \otimes \Pi_A$ is a generalized inverse and the proof is complete.

Using these results the joint log-likelihood of the tranformed variables $Y_1, Y_2, \ldots, Y_5$ becomes:

$$c - \frac{n}{2}\log|\Omega|$$

$$-\frac{1}{2}\{[(\Pi_{P_1\otimes I_r} \otimes \Pi_A)vec(Y) - (\Pi_{P_1\otimes I_r}(P^T \otimes I_r) \otimes A)vec(\xi^*)]^T$$

$$(\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi^T_{P_1\otimes I_r} \otimes \Pi_A)$$

$$[(\Pi_{P_1\otimes I_r} \otimes \Pi_A)vec(Y) - (\Pi_{P_1\otimes I_r}(P^T \otimes I_r) \otimes \Pi_A)vec(\xi^*)]+$$

$$+[(\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))vec(Y)]^T$$

$$(\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))$$

$$[(\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))vec(Y)]+$$

$$+[((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)vec(Y) - ((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T(P^T \otimes I_r) \otimes A)vec(\xi^*)]^T$$

$$((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)$$

$$[((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)vec(Y) - ((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T(P^T \otimes I_r) \otimes A)vec(\xi^*)]+$$

$$+[((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))vec(Y)]^T$$

$$((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))$$

$$[((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))vec(Y)]+$$

$$+[((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)vec(Y)]^T$$

$$((I_{qr} - \Pi_{P\otimes I_r})(I_q \otimes \Sigma_1^{-1})(I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)$$

$$[((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)vec(Y)]\} \qquad (3.32)$$

We shall use property (A11) of the Appendix in order to simplify the above likelihood. According to this

$$vec(ABC) = (C^T \otimes A)vec(B) \tag{3.33}$$

and consequently,

$$\begin{aligned}
(\Pi^T_{P_1\otimes I_r} \otimes \Pi_A)vec(Y) &= vec(\Pi_A Y \Pi_{P_1\otimes I_r}) \\
(\Pi^T_{P_1\otimes I_r}(P^T \otimes I_r) \otimes A)vec(\xi^*) &= vec(A\xi^*(P \otimes I_r)\Pi_{P_1\otimes I_r}) \\
(\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))vec(Y) &= vec((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r}) \\
((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)vec(Y) &= vec(\Pi_A Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})) \\
((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T(P^T \otimes I_r) \otimes A)vec(\xi^*) &= vec(A\xi^*(P \otimes I_r)(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})) \\
((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))vec(Y) &= vec((I_n - \Pi_A)Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})) \\
((I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)vec(Y) &= vec(Y(I_{qr} - \Pi_{P\otimes I_r}))
\end{aligned}$$

and thus, the log-likelihood becomes

$$c - \frac{n}{2}\log|\Omega|$$

$$-\frac{1}{2}\{[vec(\Pi_A Y \Pi_{P_1\otimes I_r} - A\xi_1^*(P_1 \otimes I_r))]^T$$

$$(\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi^T_{P_1\otimes I_r} \otimes \Pi_A)$$

$$[vec(\Pi_A Y \Pi_{P_1\otimes I_r} - A\xi_1^*(P_1 \otimes I_r))]+$$

$$+[vec((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r})]^T$$

$$(\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi^T_{P_1\otimes I_r} \otimes (I_n - \Pi_A))[vec((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r})]+$$

$$+[vec(\Pi_A Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))]^T$$

$$((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes \Pi_A)$$

$$[vec(\Pi_A Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))]+$$

$$+[vec((I_n - \Pi_A)Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}))]^T$$

$$((\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T \otimes (I_n - \Pi_A))$$

$$[vec((I_n - \Pi_A)Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}))]^T +$$

$$+[vec(Y(I_{qr} - \Pi_{P\otimes I_r}))]^T$$

$$(I_{qr} - \Pi_{P\otimes I_r})((I_q \otimes \Sigma_1^{-1})(I_{qr} - \Pi_{P\otimes I_r})^T \otimes I_n)$$

$$[vec(Y(I_{qr} - \Pi_{P\otimes I_r}))]\} \tag{3.34}$$

where $\xi^* = (\xi_1^*, \xi_2^*)$ and ξ_1^* $(m \times tr)$ and ξ_2^* $(m \times (p-t)r)$. The log-likelihood can be further simplified using property (A12) in the Appendix which states that for matrices A, B, C and D for which multiplications are legitimate, we have

$$trABCD = \left(vec(B^T)\right)^T (C \otimes A)vec(D) \tag{3.35}$$

Thus, the log-likelihood (3.34) is also written as

$$c - \frac{n}{2}\log|\Omega|$$

$$-\frac{1}{2}\{tr\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T(\Pi_A Y\Pi_{P_1\otimes I_r} - A\xi_1^*(P_1 \otimes I_r))^T$$

$$\Pi_A(\Pi_A Y\Pi_{P_1\otimes I_r} - A\xi_1^*(P_1 \otimes I_r)) +$$

$$+tr\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r})^T(I_n - \Pi_A)((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r}) +$$

$$+tr(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T$$

$$(\Pi_A Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))^T$$

$$\Pi_A(\Pi_A Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}) - A\xi_2^*(P_2 \otimes I_r)) +$$

$$+tr(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})^T$$

$$((I_n - \Pi_A)Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r}))^T(I_n - \Pi_A)((I_n - \Pi_A)Y(\Pi_{P\otimes I_r} - \Pi_{P_1\otimes I_r})) +$$

$$+ tr(I_q \otimes \Sigma_1^{-1})(I_{qr} - \Pi_{P\otimes I_r})^T(Y(I_{qr} - \Pi_{P\otimes I_r}))^T(Y(I_{qr} - \Pi_{P\otimes I_r})) \tag{3.36}$$

$$= c - \frac{n}{2}\log|\Omega|$$

$$+tr\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r})^T(I_n - \Pi_A)((I_n - \Pi_A)Y\Pi_{P_1\otimes I_r}) +$$

$$+tr[(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})(I_q\otimes\Sigma_1^{-1})(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})^T$$

$$((I_n-\Pi_A)Y(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r}))^T(I_n-\Pi_A)((I_n-\Pi_A)Y(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r}))+$$

$$+(I_q\otimes\Sigma_1^{-1})(I_{qr}-\Pi_{P\otimes I_r})^T(Y(I_{qr}-\Pi_{P\otimes I_r}))^T(Y(I_{qr}-\Pi_{P\otimes I_r}))]+$$

$$+tr[\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))^T$$

$$\Pi_A(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))+$$

$$+(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})(I_q\otimes\Sigma_1^{-1})(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})^T$$

$$(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))^T$$

$$\Pi_A(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))]= \tag{3.37}$$

$$=c-\frac{n}{2}\log|\Omega|$$

$$+tr\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^TY^T(I_n-\Pi_A)Y+$$

$$+tr(I_q\otimes\Sigma_1^{-1})[(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})^TY^T(I_n-\Pi_A)Y(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})+$$

$$+(I_{qr}-\Pi_{P\otimes I_r})^TY^TY(I_{qr}-\Pi_{P\otimes I_r})]+$$

$$+tr[\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))^T$$

$$(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))+$$

$$+(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})(I_q\otimes\Sigma_1^{-1})(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})^T$$

$$(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))^T$$

$$(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))] \tag{3.38}$$

3.3 Derivation of sufficient statistics

The form of the log-likelihood as is given in (3.38) will help us deriving the sufficient statistics for Σ_1 and $\Sigma_1 + \Sigma_2$. The following two theorems are of great help:

Theorem 5 *For Ω as defined in (3.28) and P semi-orthogonal*

$$(P \otimes I_r)\Omega^{-1}(P^T \otimes I_r) = \begin{pmatrix} (I_t \otimes (\Sigma_1 + \Sigma_2))^{-1} & 0 \\ 0 & (I_{p-t} \otimes \Sigma_1)^{-1} \end{pmatrix} \tag{3.39}$$

Proof: From the definition of Ω we see that

$$\Omega^{-1} = (I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2)^{-1}$$

$$= I_q \otimes \Sigma_1^{-1} - (P_1^T \otimes \Sigma_1^{-1})(I_{tr} + (P_1 P_1^T \otimes \Sigma_2 \Sigma_1^{-1}))^{-1}(P_1 \otimes \Sigma_2 \Sigma_1^{-1})$$

$$= I_q \otimes \Sigma_1^{-1} - (P_1^T \otimes I_r)(I_t \otimes \Sigma_1 + P_1 P_1^T \otimes \Sigma_2)^{-1}(P_1 \otimes \Sigma_2 \Sigma_1^{-1}$$

$$= I_q \otimes \Sigma_1^{-1} - (P_1^T \otimes I_r)(I_t \otimes \Sigma_1 + P_1 P_1^T \otimes \Sigma_2)^{-1}$$

$$((I_q \otimes \Sigma_1 + P_1 P_1^T \otimes \Sigma_2)((P_1 P_1^T)^{-1} P_1 \otimes \Sigma_1^{-1}) - (P_1 P_1^T)^{-1} P_1 \otimes I_r)$$

$$= (I_q - \Pi_{P_1}) \otimes \Sigma_1^{-1} + (P_1^T \otimes I_r)(I_t \otimes \Sigma_1 + P_1 P_1^T \otimes \Sigma_2)^{-1}((P_1 P_1^T)^{-1} P_1 \otimes I_r) \tag{3.40}$$

Thus, using (3.40)

$$\begin{aligned} (P_1 \otimes I_r)\Omega^{-1}(P_1^T \otimes I_r) &= (P_1 P_1^T \otimes I_r)(I_t \otimes \Sigma_1 + P_1 P_1^T \otimes \Sigma_2)^{-1} \\ &= (I_t \otimes (\Sigma_1 + \Sigma_2))^{-1} \end{aligned}$$

and

$$\begin{aligned} (P_2 \otimes I_r)\Omega^{-1}(P_2^T \otimes I_r) &= P_2 P_2^T \otimes \Sigma_1^{-1} \\ &= (I_{p-t} \otimes \Sigma_1)^{-1} \end{aligned}$$

In the same way, the off diagonal elements of this partitioned matrix are 0 since P is sub-orthogonal.

Theorem 6 *Projections* $\Pi_{P_1 \otimes I_r}$ *and* $\Pi_{P \otimes I_r}$ *are unweighted and thus,*

$$\Pi_{P_1 \otimes I_r} = \Pi_{P_1} \otimes I_r \quad \text{and} \quad \Pi_{P \otimes I_r} = \Pi_P \otimes I_r$$

Proof:

$$\Pi_{P_1 \otimes I_r} = \Omega^{-1}(P_1^T \otimes I_r)((P_1 \otimes I_r)\Omega^{-1}(P_1^T \otimes I_r))^{-1}(P_1 \otimes I_r)$$

where Ω is as defined in (3.28). Using (3.40) we can easily see that

$$\begin{aligned} \Pi_{P_1 \otimes I_r} &= (P_1^T \otimes I_r)(I_t \otimes (\Sigma_1 + \Sigma_2)^{-1})(I_t \otimes (\Sigma_1 + \Sigma_2))(P_1 \otimes I_r) \\ &= P_1^T P_1 \otimes I_r \\ &= \Pi_{P_1} \otimes I_r \end{aligned}$$

Next, from the definition of $\Pi_{P \otimes I_r}$ we have

$$\Pi_{P \otimes I_r} = \Omega^{-1}(P^T \otimes I_r)((P \otimes I_r)\Omega^{-1}(P^T \otimes I_r))^{-}(P \otimes I_r)$$

Using the expansion $P^T = (P_1^T, P_2^T)$ and theorem 5 we can see that this projection is actually

$$(P_1^T \otimes (\Sigma_1 + \Sigma_2)^{-1}, P_2^T \otimes \Sigma_1^{-1}) \begin{pmatrix} I_t \otimes (\Sigma_1 + \Sigma_2)^{-1} & 0 \\ 0 & I_{p-t} \otimes \Sigma_1^{-1} \end{pmatrix}^{-} \begin{pmatrix} P_1 \otimes I_r \\ P_2 \otimes I_r \end{pmatrix}$$

$$= (P_1^T \otimes (\Sigma_1 + \Sigma_2)^{-1}, P_2^T \otimes \Sigma_1^{-1}) \begin{pmatrix} I_t \otimes (\Sigma_1 + \Sigma_2) & 0 \\ 0 & I_{p-t} \otimes \Sigma_1 \end{pmatrix} \begin{pmatrix} P_1 \otimes I_r \\ P_2 \otimes I_r \end{pmatrix}$$

$$= (P_1^T P_1 \otimes I_r + P_2^T P_2 \otimes I_r) = \Pi_P \otimes I_r$$

Now, using the above two results, we further modify the log-likelihood (3.38), thus taking

$$= c - \frac{n}{2} \log |\Omega|$$

$$+tr(\Pi_{P_1} \otimes I_r)\Omega^{-1}(\Pi_{P_1} \otimes I_r)Y^T(I_n - \Pi_A)Y +$$

$$+tr[(I_n - \Pi_A)Y((\Pi_P - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((\Pi_P - \Pi_{P_1}) \otimes I_r)Y^T$$

$$+Y((I_q - \Pi_P) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((I_q - \Pi_P) \otimes I_r)Y^T] +$$

$$+tr[\Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi^T_{P_1 \otimes I_r} (\Pi_A Y \Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r))^T$$

$$\Pi_A(\Pi_A Y \Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r)) +$$

$$+(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T$$

$$(\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))^T$$

$$\Pi_A(\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))] \tag{3.41}$$

The following lemma will give us a convenient form for the first two traces of (3.41) and sufficient statistics will be easily derived.

Lemma 3

$$tr(\Pi_{P_1} \otimes I_r)\Omega^{-1}(\Pi_{P_1} \otimes I_r)Y^T(I_n - \Pi_A)Y$$

$$= tr[I_t \otimes (\Sigma_1 + \Sigma_2)^{-1}](P_1 \otimes I_r)Y^T(I_n - \Pi_A)Y(P_1^T \otimes I_r)$$

and

$$tr[(I_n - \Pi_A)Y((\Pi_P - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((\Pi_P - \Pi_{P_1}) \otimes I_r)Y^T$$

$$+Y((I_q - \Pi_P) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((I_q - \Pi_P) \otimes I_r)Y^T]$$

$$= tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

Proof: The first part is derived easily by noting that

$$\Pi_{P_1} \otimes I_r = P_1^T P_1 \otimes I_r$$

and from Theorem 5. As regards the other part, with some algebra we see that

$$tr[(I_n - \Pi_A)Y((\Pi_P - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((\Pi_P - \Pi_{P_1}) \otimes I_r)Y^T$$

$$+Y((I_q - \Pi_P) \otimes I_r)(I_q \otimes \Sigma_1^{-1})((I_q - \Pi_P) \otimes I_r)Y^T]$$

$$= tr[Y(\Pi_P \otimes \Sigma_1^{-1})Y^T - Y(\Pi_{P_1} \otimes \Sigma_1^{-1})Y^T - Y(\Pi_P \otimes \Sigma_1^{-1})Y^T \Pi_A$$

$$+Y(\Pi_{P_1} \otimes \Sigma_1^{-1})Y^T \Pi_A + Y(I_q \otimes \Sigma_1^{-1})Y^T - Y(\Pi_P \otimes \Sigma_1^{-1})Y^T]$$

$$= tr[Y(\Pi_{P_1} \otimes \Sigma_1^{-1})Y^T\Pi_A + Y(I_q \otimes \Sigma_1^{-1})Y^T$$

$$-Y(\Pi_{P_1} \otimes \Sigma_1^{-1})Y^T - Y(\Pi_P \otimes \Sigma_1^{-1})Y^T\Pi_A]$$

$$= tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

which some simple algebra can show.

Lemma 3 shows that the log-likelihood (3.41) can be rewritten as

$$= c - \frac{n}{2}\log|\Omega|$$

$$+tr[I_t \otimes (\Sigma_1 + \Sigma_2)^{-1}](P_1 \otimes I_r)Y^T(I_n - \Pi_A)Y(P_1^T \otimes I_r)$$

$$+tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

$$+tr[\Pi_{P_1 \otimes I_r}\Omega^{-1}\Pi_{P_1 \otimes I_r}^T(\Pi_A Y\Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r))^T$$

$$(\Pi_A Y\Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r))+$$

$$+(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T$$

$$(\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))^T$$

$$(\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))] \tag{3.42}$$

Let us analyze a little bit further the first two traces of the log-likelihood (3.42). From the definition of the data matrix Y we can immediately see that

$$tr[I_t \otimes (\Sigma_1 + \Sigma_2)^{-1}](P_1 \otimes I_r)Y^T(I_n - \Pi_A)Y(P_1^T \otimes I_r)$$

$$+tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes I_r)(I_q \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

$$= trY^T(I_n - \Pi_A)Y(P_1^T P_1 \otimes (\Sigma_1 + \Sigma_2)^{-1})$$

$$+tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

$$= tr(Y - \Pi_A Y)(P_1^T P_1 \otimes (\Sigma_1 + \Sigma_2)^{-1})(Y - \Pi_A Y)^T$$

$$+tr(Y - \Pi_A Y(\Pi_P \otimes I_r))((I_q - \Pi_{P_1}) \otimes \Sigma_1^{-1})(Y - \Pi_A Y(\Pi_P \otimes I_r))^T$$

$$= tr \begin{pmatrix} (vec(Y_{11}))^T - \sum_{i=1}^{n_1}(vec(Y_{i1}))^T/n_1 \\ \vdots \\ (vec(Y_{n_m m}))^T - \sum_{i=1}^{n_m}(vec(Y_{im}))^T/n_m \end{pmatrix}$$

$$\left(P_1^T P_1 \otimes (\Sigma_1 + \Sigma_2)^{-1}\right) \begin{pmatrix} (vec(Y_{11}))^T - \sum_{i=1}^{n_1}(vec(Y_{i1}))^T/n_1 \\ \vdots \\ (vec(Y_{n_m m})^T - \sum_{i=1}^{n_m}(vec(Y_{im}))^T/n_m \end{pmatrix}^T$$

$$+tr \left[\begin{pmatrix} (vec(Y_{11})^T \\ \vdots \\ (vec(Y_{n_m m})^T \end{pmatrix} - \Pi_A \begin{pmatrix} (vec(Y_{11}))^T(\Pi_P \otimes I_r) \\ \vdots \\ (vec(Y_{n_m m})^T(\Pi_P \otimes I_r) \end{pmatrix} \right]$$

$$\left((I_q - \Pi_{P_1}) \otimes \Sigma_1^{-1}\right) \left[\begin{pmatrix} (vec(Y_{11})^T \\ \vdots \\ (vec(Y_{n_m m})^T \end{pmatrix} - \Pi_A \begin{pmatrix} (vec(Y_{11})^T(\Pi_P \otimes I_r) \\ \vdots \\ (vec(Y_{n_m m})^T(\Pi_P \otimes I_r) \end{pmatrix} \right]^T$$

Using now property (A11) of the Appendix, the above sum becomes

$$tr \begin{pmatrix} (vec(Y_{11}))^T - (vec(\sum_{i=1}^{n_1} Y_{i1}/n_1))^T \\ \vdots \\ (vec(Y_{n_m m}))^T - (vec(\sum_{i=1}^{n_m} Y_{im}/n_m))^T \end{pmatrix}$$

$$\left(P_1^T P_1 \otimes (\Sigma_1 + \Sigma_2)^{-1}\right) \begin{pmatrix} (vec(Y_{11}))^T - (vec(\sum_{i=1}^{n_1} Y_{i1}/n_1))^T \\ \vdots \\ (vec(Y_{n_m m}))^T - (vec(\sum_{i=1}^{n_m} Y_{im}/n_m))^T \end{pmatrix}^T$$

$$+tr \left[\begin{pmatrix} (vec(Y_{11}))^T \\ \vdots \\ (vec(Y_{n_m m}))^T \end{pmatrix} - \Pi_A \begin{pmatrix} (vec(Y_{11}\Pi_P))^T \\ \vdots \\ (vec(Y_{n_m m}\Pi_P))^T \end{pmatrix} \right]$$

$$\left((I_q - \Pi_{P_1}) \otimes \Sigma_1^{-1}\right) \left[\begin{pmatrix} (vec(Y_{11}))^T \\ \vdots \\ (vec(Y_{n_m m}))^T \end{pmatrix} - \Pi_A \begin{pmatrix} (vec(Y_{11}\Pi_P))^T \\ \vdots \\ (vec(Y_{n_m m}\Pi_P))^T \end{pmatrix} \right]^T$$

$$= tr \begin{pmatrix} \left[vec\left(Y_{11} - \frac{\sum_{i=1}^{n_1} Y_{i1}}{n_1}\right)\right]^T \\ \vdots \\ \left[vec\left(Y_{n_m m} - \frac{\sum_{i=1}^{n_m} Y_{im}}{n_m}\right)\right]^T \end{pmatrix}$$

$$\left(P_1^T P_1 \otimes (\Sigma_1 + \Sigma_2)^{-1}\right) \begin{pmatrix} \left[vec\left(Y_{11} - \frac{\sum_{i=1}^{n_1} Y_{i1}}{n_1}\right)\right]^T \\ \vdots \\ \left[vec\left(Y_{n_m m} - \frac{\sum_{i=1}^{n_m} Y_{im}}{n_m}\right)\right]^T \end{pmatrix}^T$$

$$+tr\begin{pmatrix} \left[vec\left(Y_{11}-\frac{\sum_{i=1}^{n_1}Y_{i1}\Pi_P}{n_1}\right)\right]^T \\ \vdots \\ \left[vec\left(Y_{n_m m}-\frac{\sum_{i=1}^{n_m}Y_{im}\Pi_P}{n_m}\right)\right]^T \end{pmatrix}$$

$$((I_q-\Pi_{P_1})\otimes\Sigma_1^{-1})\begin{pmatrix} \left[vec\left(Y_{11}-\frac{\sum_{i=1}^{n_1}Y_{i1}\Pi_P}{n_1}\right)\right]^T \\ \vdots \\ \left[vec\left(Y_{n_m m}-\frac{\sum_{i=1}^{n_m}Y_{im}\Pi_P}{n_m}\right)\right]^T \end{pmatrix}^T$$

$$=\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left[vec\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)\right]^T(P_1^TP_1\otimes(\Sigma_1+\Sigma_2)^{-1})\left[vec\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)\right]$$

$$+\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left[vec\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)\right]^T((I_q-\Pi_{P_1})\otimes\Sigma_1^{-1})\left[vec\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)\right]$$

Using now property (A12) the above sum becomes

$$tr(\Sigma_1+\Sigma_2)^{-1}\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)P_1^TP_1\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)^T$$

$$+tr\Sigma_1^{-1}\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)(I_q-\Pi_{P_1})\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)^T$$

After this analysis, the log-likelihood (3.42) becomes

$$=c-\frac{n}{2}\log|\Omega|$$

$$+tr(\Sigma_1+\Sigma_2)^{-1}\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)P_1^TP_1\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}}{n_j}\right)^T$$

$$+tr\Sigma_1^{-1}\sum_{i=1}^{n_j}\sum_{j=1}^{m}\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)(I_q-\Pi_{P_1})\left(Y_{ij}-\frac{\sum_{i=1}^{n_j}Y_{ij}\Pi_P}{n_j}\right)^T$$

$$+tr[\Pi_{P_1\otimes I_r}\Omega^{-1}\Pi_{P_1\otimes I_r}^T(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))^T$$

$$(\Pi_AY\Pi_{P_1\otimes I_r}-A\xi_1^*(P_1\otimes I_r))+$$

$$+(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})(I_q\otimes\Sigma_1^{-1})(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})^T$$

$$(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))^T$$

$$(\Pi_AY(\Pi_{P\otimes I_r}-\Pi_{P_1\otimes I_r})-A\xi_2^*(P_2\otimes I_r))] \qquad (3.43)$$

The final form of the log-likelihood will be given by noting that

$$\begin{aligned} |I_q \otimes \Sigma_1 + P_1^T P_1 \otimes \Sigma_2| &= |I_q \otimes \Sigma_1||I_t \otimes (I_r + \Sigma_2 \Sigma_1^{-1})| \\ &= |\Sigma_1|^q |I_t \otimes (\Sigma_1 + \Sigma_2)||\Sigma_1|^{-t} \\ &= |\Sigma_1|^{q-t} |I_t \otimes (\Sigma_1 + \Sigma_2)|. \end{aligned}$$

Thus, this final form is

$$\begin{aligned} c - \frac{n}{2}(\log|\Sigma_1|^{q-t} - \log|I_t \otimes (\Sigma_1 + \Sigma_2)|) - \frac{1}{2}[tr(\Sigma_1 + \Sigma_2)^{-1} G + tr\Sigma_1^{-1} H] \\ + tr[\Pi_{P_1 \otimes I_r} \Omega^{-1} \Pi_{P_1 \otimes I_r}^T (\Pi_A Y \Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r))^T \\ (\Pi_A Y \Pi_{P_1 \otimes I_r} - A\xi_1^*(P_1 \otimes I_r)) + \\ + (\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})(I_q \otimes \Sigma_1^{-1})(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r})^T \\ (\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))^T \\ (\Pi_A Y(\Pi_{P \otimes I_r} - \Pi_{P_1 \otimes I_r}) - A\xi_2^*(P_2 \otimes I_r))] \end{aligned} \tag{3.44}$$

where

$$G = \sum_{i=1}^{n_j} \sum_{j=1}^{m} \left(Y_{ij} - \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j} \right) P_1^T P_1 \left(Y_{ij} - \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j} \right)^T \tag{3.45}$$

and

$$H = \sum_{i=1}^{n_j} \sum_{j=1}^{m} \left(Y_{ij} - \frac{\sum_{i=1}^{n_j} Y_{ij} \Pi_P}{n_j} \right) (I_q - \Pi_{P_1}) \left(Y_{ij} - \frac{\sum_{i=1}^{n_j} Y_{ij} \Pi_P}{n_j} \right)^T \tag{3.46}$$

The form of the log-likelihood in (3.44) suggests that G, H, $\Pi_A Y(P_1^T P_1 \otimes I_r)$ and $\Pi_A Y(P_2^T P_2 \otimes I_r)$ are sufficient statistics for $\Sigma_1 + \Sigma_2$, Σ_1, ξ_1^* and ξ_2^* respectively. Note that statistics G and H coincide with S_λ and S_e of Section 3 in Reinsel (1982) in the special case where all random effects are used (that is, $P_1 = P$). The model adopted provides also a generalization to the case considered by Anderson, Anderson and Olkin (1986). There, the authors had considered replicated measurements on time and they used one random effect to explain the variation between replicates. Here, the variables (replications) were correlated and we used an arbitrary number of random effects to explain the variation on time. The difference with Reinsel's approach is that he used

the fact that the sufficient statistics come from independent transformations of the data and therefore are independent, and he obtained unbiased estimates. However, for the case of Σ_2, his estimate may not be positive definite while in the approach which will be developed in the next section ML and REML estimates of Σ_2 are obtained under the restriction that they are positive definite.

3.4 ML estimation

In this section, the form of the log-likelihood (3.44) will be exploited in order to obtain ML estimates of the covariance matrices Σ_1 and Σ_2. The null hypothesis will assume t random effects and that the rank of Σ_2 is at most s ($s \leq r$). We are seeking estimates for Σ_1 and Σ_2 such that both are at least positive semi-definite and this is the condition under which the log-likelihood of the observations will be maximized. The approach will be similar to that of Anderson, Anderson and Olkin (1986) and using ideas of Theobald (1975), Amemiya and Fuller (1984) and Amemiya (1985), the differences arising from the adoption of arbitrary number of mixed effects and from the introduction of constants k_1 and k_2 as follows:

Let us consider the following canonical value decompositions for Σ_1, Σ_2, H and G,

$$\Sigma_1 = \Gamma\Gamma^T, \quad \Sigma_1 + \Sigma_2 = \Gamma\Delta\Gamma^T \tag{3.47}$$

and

$$k_2 H = BB^T \quad k_1 G = BTB^T \tag{3.48}$$

where Γ and B are nonsingular matrices of orders $r \times r$ and Δ and T are diagonal matrices containing the eigenvalues of $\Sigma_1^{-1}(\Sigma_1 + \Sigma_2)$ and $\frac{k_1}{k_2}H^{-1}G$ respectively. Constants k_1 and k_2 can be the inverse numbers of the corresponding degrees of freedom or any other constants such that multiplication with G and H respectively will provide estimators of $\Sigma_1 + \Sigma_2$ and of Σ_1. Note also that the elements of Δ should satisfy

$$\delta_1 \geq \ldots \geq \delta_s > \delta_{s+1} = \ldots = \delta_r = 1 \tag{3.49}$$

We shall use a variation of von Neumann's theorem for the maximization of the log-likelihood. Consider the following (Anderson, Anderson and Olkin (1986)):

Theorem 7 *For Q $(p \times p)$ orthogonal and D and E diagonal ($d_1 \geq \ldots \geq d_p$ $e_1 \geq \ldots \geq e_p$)*

$$\min trD^{-1}QEQ^T = trD^{-1}E$$

and a minimizing value of Q is I_p.

Consider the log-likelihood (3.44). This is maximized with respect to ξ_1^* and ξ_2^* when we set

$$A\hat{\xi}_1^*(P_1 \otimes I_r) = \Pi_A Y(P_1^T P_1 \otimes I_r) \tag{3.50}$$

and

$$A\hat{\xi}_2^*(P_2 \otimes I_r) = \Pi_A Y(P_2^T P_2 \otimes I_r) \tag{3.51}$$

Apart then from a constant the concentrated log-likelihood of the observations is written as

$$-\frac{n}{2}\log|\Sigma_1|^{q-t} - \frac{n}{2}\log|I_t \otimes (\Sigma_1 + \Sigma_2)| - \frac{1}{2}[tr(\Sigma_1 + \Sigma_2)^{-1}G + tr\Sigma_1^{-1}H] \tag{3.52}$$

where G and H are given by (3.45) and (3.46). Considering the canonical forms for Σ_1, Σ_2, H and G defined above we see that this log-likelihood is

$$-n(q-t)\log|\Gamma| - nt\log|\Gamma| - \frac{nt}{2}\log|\Delta|$$

$$-\frac{1}{2k_1}tr(\Gamma\Delta\Gamma^T)^{-1}BTB^T - \frac{1}{2k_2}tr\Gamma\Gamma^T BB^T$$

$$= -nq\log|\Gamma| - \frac{nt}{2}\log|\Delta|$$

$$-\frac{1}{2k_1}tr\Delta^{-1}(\Gamma^{-1}B)T(B^T(\Gamma^T)^{-1}) - \frac{1}{2k_2}tr(\Gamma^{-1}B)(B^T(\Gamma^T)^{-1})$$

$$= -nq\log|\Gamma| - \frac{nt}{2}\log|\Delta|$$

$$-\frac{1}{2k_1}tr\Delta^{-1}(\Gamma^{-1}BT^{1/2})(\Gamma^{-1}BT^{1/2})^T - \frac{1}{2k_2}tr(\Gamma^{-1}BT^{1/2})T^{-1}(\Gamma^{-1}BT^{1/2})^T$$

This log-likelihood should be maximized with respect to Δ and Γ. We consider the singular value decomposition of

$$\Gamma^{-1}BT^{1/2} = CDZ \tag{3.53}$$

where C and Z are orthogonal and D diagonal. Then, $\Gamma = CDZT^{-1/2}B^{-1}$ and

$$|\Gamma|^{-1} = |D||T|^{-1/2}|B|^{-1} \tag{3.54}$$

since C and Z are orthogonal and thus $|C| = |Z| = 1$. Substituting into the log-likelihood we can see that it is divided in two parts: one which is free of unknown parameters and not of any interest and one which depends on C, D, Z, and Δ. The latter has the form

$$\begin{aligned} & nq\log|D| - \frac{nt}{2}\log|\Delta| \\ & -\frac{1}{2k_2}trCDZT^{-1}Z^TDC^T - \frac{1}{2k_1}tr\Delta^{-1}CDZZ^TDC^T \\ & = nq\log|D| - \frac{nt}{2}\log|\Delta| \\ & -\frac{1}{2k_2}trD^2ZT^{-1}Z^T - \frac{1}{2k_1}tr\Delta^{-1}CD^2C^T \end{aligned} \tag{3.55}$$

We apply Theorem 7 in order to maximize expression (3.55) with respect to Z and C. This will be accomplished with the minimization of $trD^2ZT^{-1}Z^T$ and of $tr\Delta^{-1}CD^2C^T$. In the first trace, Z is the orthogonal matrix required by the theorem and a minimizing value is $Z = I_r$. The other trace needs some more analysis. Consider the following notation for Δ^{-1}, C and D^2:

$$\Delta^{-1} = \begin{pmatrix} H_1 & 0 \\ 0 & I_{r-s} \end{pmatrix} \quad C = \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} \quad D^2 = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

where H_1 is a $s \times s$ diagonal matrix, C_1 and C_2 are orthogonal matrices of orders $s \times s$ and $(r-s) \times (r-s)$ respectively and D_1 and D_2 are diagonal matrices of the same order as C_1 and C_2 respectively. Some simple algebra shows that

$$tr\Delta^{-1}CD^2C^T = trH_1C_1D_1C_1^T + trC_2D_2C_2^T$$

However, the second expression on the right hand side shows that $trC_2D_2C_2^T$ remains invariant under any choice of the orthogonal matrix C_2 since

$$trC_2D_2C_2^T = trC_2^TC_2D_2 = trD_2,$$

while $trH_1C_1D_1C_1^T$ does not. This means that we should only minimize $trH_1C_1D_1C_1^T$ for which a minimizing C_1 is I_s (as Theorem 7 certifies) and finally the minimizing C for $tr\Delta^{-1}CD^2C^T$ will be of the form

$$\begin{pmatrix} I_s & 0 \\ 0 & C_2 \end{pmatrix},$$

where C_2 is a $(r-s) \times (r-s)$ orthogonal matrix. Returning again back to the maximization problem we see that the maximum of log-likelihood with respect to Z and C is

$$nq \log |D| - \frac{nt}{2} \log |\Delta|$$

$$-\frac{1}{2k_2} trD^2T^{-1} - \frac{1}{2k_1} tr\Delta^{-1}D^2$$

$$= \frac{1}{2}\left[2nq\sum_{i=1}^{r} \log d_i - nt\sum_{i=1}^{r} \log \delta_i - \frac{1}{k_2}\sum_{i=1}^{r}\frac{d_i^2}{\tau_i} - \frac{1}{k_1}\sum_{i=1}^{r}\frac{d_i^2}{\delta_i}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{r}\left[2nq \log d_i - nt \log \delta_i - d_i^2\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right)\right]$$

We maximize this expression with respect to d_i by taking its first derivative $\forall\ i = 1, \ldots, r$. This derivative has the form

$$\frac{1}{2}\left[\frac{2nq}{d_i} - 2d_i\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right)\right] = 0 \quad \forall\, i = 1, \ldots, r$$

Solution of this equation gives that $\forall\, i = 1, \ldots, r$

$$d_i^2 = nq\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right)^{-1} \tag{3.56}$$

Substituting into the log-likelihood we get that the new expression is

$$\frac{1}{2}\sum_{i=1}^{r}\left[nq \log nq - nq \log\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right) - nt \log \delta_i - nq\right] \tag{3.57}$$

$$= \frac{1}{2}\sum_{i=1}^{r}\left[c - nq\log\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right) - nt\log\delta_i\right]$$

where c is a constant. Taking again the first derivative with respect to δ_i $\forall i = 1,\ldots,r$ and equating it to zero, we get

$$\frac{1}{2}\left[-nq\left(\frac{1}{k_2\tau_i} + \frac{1}{k_1\delta_i}\right)^{-1}\left(-\frac{1}{k_1\delta_i^2}\right) - \frac{nt}{\delta_i}\right] = 0 \Rightarrow$$

$$\hat{\delta}_i = \tau_i\frac{(q-t)k_2}{tk_1}$$

Recall that the log-likelihood will be maximized with respect to the restriction that

$$\delta_1 \geq \ldots \geq \delta_s > \delta_{s+1} = \ldots = \delta_r = 1$$

By monotonicity of the likelihood to the right of the local maximum in δ_i, the maximum restricted to $\delta_i \geq 1$ is at $\hat{\delta}_i = 1$ for $i = s+1,\ldots,r$. Thus, the ML estimates of δ are

$$\begin{cases} \hat{\delta}_i = \tau_i(q-t)k_2/tk_1 & i = 1,\ldots,r_M^* \\ \hat{\delta}_i = 1 & i = r_M^*+1,\ldots,s \\ \hat{\delta}_i = 1 & i = s+1,\ldots,r \end{cases} \tag{3.58}$$

where r_M^* is the number of τ_i's which satisfy

$$\tau_i > \frac{t}{q-t}\frac{k_1}{k_2} \tag{3.59}$$

Returning now to the discussion after (3.55) recall that a maximizing Z is $\hat{Z}_M = I_r$ and a maximizing C is

$$\hat{C}_M = \begin{pmatrix} I_{s_M^*} & 0 \\ 0 & C_2 \end{pmatrix} \tag{3.60}$$

where M stands for ML, C_2 is a $(r - s_M^*) \times (r - s_M^*)$ orthogonal matrix and $s_M^* = min(r_M^*, s)$. Furthermore, if we assume that matrix T is written as

$$T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \tag{3.61}$$

where $T_1 = diag\{\tau_1,\ldots,\tau_{s_M^*}\}$ and $T_2 = diag\{\tau_{s_M^*+1},\ldots,\tau_r\}$ it is easily seen that

$$\hat{\Delta}_M = \begin{pmatrix} (q-t)k_2/tk_1T_1 & 0 \\ 0 & I_{r-s_M^*} \end{pmatrix} \tag{3.62}$$

With this formulation we get that

$$\hat{D}_M^2 = nq\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}^{-1}\right)^{-1} \tag{3.63}$$

Since now from (3.53)

$$T^{-1/2}B^{-1}\Gamma = Z^T D^{-1} C^T \Rightarrow$$

$$\Gamma = BT^{1/2}Z^T D^{-1} C^T \Rightarrow$$

$$\hat{\Gamma}_M = BT^{1/2}\hat{Z}_M^T \hat{D}_M^{-1}\hat{C}_M^T$$

and the ML estimate of Σ_1 is readily obtained as

$$\hat{\Sigma}_1^M = BT^{1/2}\hat{Z}_M^T \hat{D}_M^{-1}\hat{C}_M^T \hat{C}_M \hat{D}_M^{-1}\hat{Z}_M T^{1/2}B^T$$

$$= BT^{1/2}\hat{D}_M^{-2}T^{1/2}B^T \tag{3.64}$$

The ML estimate of $(\Sigma_1 + \Sigma_2)$ (denoted by $(\Sigma_1 + \Sigma_2)_M$) is also easily obtained as:

$$(\Sigma_1 + \Sigma_2)_M = \hat{\Gamma}_M \hat{\Delta}_M \hat{\Gamma}_M^T$$

$$= BT^{1/2}\hat{D}_M^{-1}\hat{C}_M^T \hat{\Delta}_M \hat{C}_M \hat{D}_M^{-1}T^{1/2}B^T \tag{3.65}$$

Subtraction of the two estimates gives

$$\hat{\Sigma}_2^M = BT^{1/2}\hat{D}_M^{-1}(\hat{C}_M^T \hat{\Delta}_M \hat{C}_M - I_r)\hat{D}_M^{-1}T^{1/2}B^T \tag{3.66}$$

Note that by the construction of $\hat{C}_M$ and $\hat{\Delta}_M$, we can see that

$$\hat{C}_M^T \hat{\Delta}_M \hat{C}_M = \begin{pmatrix} I_{s_M^*} & 0 \\ 0 & C_1^T \end{pmatrix} \hat{\Delta}_M \begin{pmatrix} I_{s_M^*} & 0 \\ 0 & C_1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{(q-t)k_2}{tk_1}T_1 & 0 \\ 0 & I_{r-s_M^*} \end{pmatrix}$$

Thus, condition (3.59) certifies us that $\hat{C}_M^T \hat{\Delta}_M \hat{C}_M - I_r$ is positive semi-definite of rank s_M^* and the same holds for $\hat{\Sigma}_2$ something which was required by the null hypothesis.

The discussion of this section can be summarized by the following theorem:

Theorem 8 *The maximum likelihood estimates of Σ_1 and Σ_2 under the assumption that they are both positive semidefinite and Σ_2 is at most of rank s are given by*

$$\hat{\Sigma}_1^M = BT^{1/2}\hat{D}_M^{-2}T^{1/2}B^T \tag{3.67}$$

$$\hat{\Sigma}_2^M = BT^{1/2}\hat{D}_M^{-1}(\hat{C}_M^T\hat{\Delta}_M\hat{C}_M - I_r)\hat{D}_M^{-1}T^{1/2}B^T \tag{3.68}$$

where

$$\hat{D}_M^2 = \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}_M^{-1}\right)^{-1}$$

$$\hat{\Delta}_M = \begin{pmatrix} \frac{[(q-t)k_2}{tk_1]T_1} & 0 \\ 0 & I_{r-s_M^*} \end{pmatrix}$$

and

$$\hat{C}_M = \begin{pmatrix} I_{s_M^*} & 0 \\ 0 & C_2 \end{pmatrix}$$

and C_2 is a $(r-s_M^*)\times(r-s_M^*)$ orthogonal matrix and $s_M^* = min(r_M^*, s)$ with r_M^*, the number of τ_i's $i = 1,\ldots,r$ which satisfy

$$\tau_i > \frac{t}{q-t}\frac{k_1}{k_2}$$

3.5 REML estimates

This section is devoted to the derivation of REML estimates of the variance matrices Σ_1 and Σ_2. In Chapter 2, it was shown that the REML log-likelihood is the same as in the ML case with the addition of the term

$$-\frac{m}{2}\log|P\Sigma^{-1}P^T|$$

where P is the matrix which represents the within subject changes in mean response. From (3.22) it can be deduced that the corresponding term in the log-likelihood of the model has the form

$$-\frac{m}{2}\log|(P\otimes I_r)(I_q\otimes\Sigma_1 + P_1^TP_1\otimes\Sigma_2)^{-1}(P^T\otimes I_r)|$$

$$= -\frac{m}{2}\log\begin{vmatrix} I_t\otimes(\Sigma_1+\Sigma_2)^{-1} & 0 \\ 0 & I_{p-t}\otimes\Sigma_1^{-1} \end{vmatrix}$$

$$= -\frac{m}{2}\log|I_t \otimes (\Sigma_1 + \Sigma_2)|^{-1} - \frac{m}{2}\log|I_{p-t} \otimes \Sigma_1|^{-1}$$

by using Theorem 5. But using (A17) in Appendix, we deduce that this is equal to

$$-\frac{m}{2}\log|\Sigma_1 + \Sigma_2|^{-t} - \frac{m}{2}\log|\Sigma_1|^{-(p-t)}$$

$$= \frac{mt}{2}\log|\Sigma_1 + \Sigma_2| + \frac{(p-t)m}{2}\log|\Sigma_1|.$$

This term modifies the concentrated log-likelihood (3.52) of the observations and makes it

$$-\frac{n(q-t) - m(p-t)}{2}\log|\Sigma_1| - \frac{(n-m)t}{2}\log|\Sigma_1 + \Sigma_2|$$

$$-\frac{1}{2}\left[tr(\Sigma_1 + \Sigma_2)^{-1}G + tr\Sigma^{-1}H\right]$$

where G and H are as defined in (3.45) and (3.46) respectively. Following the same steps as before, and defining Γ, Δ, B and T as in (3.47) and (3.48), we wish finally to maximize the log-likelihood

$$-(nq - mp)\log|\Gamma| - \frac{(n-m)t}{2}\log|\Delta|$$

$$-\frac{1}{2k_1}tr\Delta^{-1}(\Gamma^{-1}B)T(B^T(\Gamma^T)^{-1}) - \frac{1}{2k_2}tr(\Gamma^{-1}B)(B^T(\Gamma^T)^{-1})$$

$$= -(nq - mp)\log|\Gamma| - \frac{(n-m)t}{2}\log|\Delta|$$

$$-\frac{1}{2k_1}tr\Delta^{-1}(\Gamma^{-1}BT^{1/2})(\Gamma^{-1}BT^{1/2})^T - \frac{1}{2k_2}tr(\Gamma^{-1}BT^{1/2})T^{-1}(\Gamma^{-1}BT^{1/2})^T$$

Again as in the previous section this log-likelihood should be maximized with respect to Δ and Γ. Following the same steps as before and defining matrices C, D and Z and applying Theorem 7 we get that the maximum is given by

$$(nq - mp)\log|D| - \frac{(n-m)t}{2}\log|\Delta|$$

$$-\frac{1}{2k_2}trD^2T^{-1} - \frac{1}{2k_1}tr\Delta^{-1}D^2$$

and maximizing Z and C are (see also the discussion after equation (3.55) in the previous section)

$$\hat{Z}_R = I_r \quad \text{and} \quad \hat{C}_R = \begin{pmatrix} I_s & 0 \\ 0 & C_2 \end{pmatrix}, \tag{3.69}$$

where R stands for REML and C_2 is a $(r-s)\times(r-s)$ orthogonal matrix. Some simple algebra shows that this log-likelihood is

$$\frac{1}{2}\left[2(nq-mp)\sum_{i=1}^{r}\log d_i-(n-m)t\sum_{i=1}^{r}\log\delta_i-\frac{1}{k_2}\sum_{i=1}^{r}\frac{d_i^2}{\tau_i}-\frac{1}{k_1}\sum_{i=1}^{r}\frac{d_i^2}{\delta_i}\right]$$

$$=\frac{1}{2}\sum_{i=1}^{r}\left[2(nq-mp)\log d_i-(n-m)t\log\delta_i-d_i^2\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)\right]$$

We, again, maximize this expression with respect to d_i by taking its first derivative $\forall\, i=1,\ldots,r$. This derivative has the form

$$\frac{1}{2}\left[\frac{2(nq-mp)}{d_i}-2d_i\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)\right]=0\quad\forall\, i=1,\ldots,r$$

Solution of this equation gives the estimate for $d_i,\ \forall\, i=1,\ldots,r$ which is

$$d_i^2=(nq-mp)\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)^{-1}$$

Substituting into the log-likelihood we get that the new expression is

$$\frac{1}{2}\sum_{i=1}^{r}\left[(nq-mp)\log nq-mp-(nq-mp)\log\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)\right.$$
$$\left.-(n-m)t\log\delta_i-(nq-mp)\right]$$
$$=\frac{1}{2}\sum_{i=1}^{r}\left[c-(nq-mp)\log\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)-(n-m)t\log\delta_i\right]$$

where c is a constant not related to the maximization procedure. Taking again the first derivative with respect to $\delta_i\ \ \forall i=1,\ldots,r$ and equating it to zero, we get

$$\frac{1}{2}\left[-(nq-mp)\left(\frac{1}{k_2\tau_i}+\frac{1}{k_1\delta_i}\right)^{-1}\left(-\frac{1}{k_1\delta_i^2}\right)-\frac{(n-m)t}{\delta_i}\right]=0\Rightarrow$$

$$\hat{\delta}_i=\tau_i\frac{[n(q-t)-m(p-t)]k_2}{(n-m)tk_1}$$

Using a similar argument now with the corresponding ML case, the REML estimates of δ_i under the restriction that

$$\delta_1>\ldots>\delta_s>\delta_{s+1}=\ldots=\delta_r=1$$

are

$$\begin{cases} \hat{\delta}_i = \tau_i \frac{[n(q-t)-m(p-t)]k_2}{(n-m)tk_1} & i = 1, \ldots, r_R^* \\ \hat{\delta}_i = 1 & i = r_R^* + 1, \ldots, s \\ \hat{\delta}_i = 1 & i = s+1, \ldots, r \end{cases} \tag{3.70}$$

where r_R^* is the number of τ_i's which satisfy

$$\tau_i > \frac{(n-m)t}{n(q-t)-m(p-t)} \frac{k_1}{k_2} \tag{3.71}$$

Thus $\hat{\Delta}_R$ is now written as

$$\hat{\Delta}_R = \begin{pmatrix} \frac{[n(q-t)-m(p-t)]k_2}{(n-m)tk_1} T_1 & 0 \\ 0 & I_{r-s_R^*} \end{pmatrix} \tag{3.72}$$

with $s_R^* = min(s, r_R^*)$ and T_1 now a $s_R^* \times s_R^*$ diagonal submatrix of T corresponding to the first s_R^* rows and columns of T. Similarly,

$$\hat{D}_R^2 = (nq - mp)\left(\frac{1}{k_2} T^{-1} + \frac{1}{k_1} \hat{\Delta}_R^{-1}\right)^{-1} \tag{3.73}$$

In the same way as in the previous section, minimizers Z and C are

$$\hat{Z}_R = I_{s_R^*} \quad \hat{C}_R = \begin{pmatrix} I_{s_R^*} & 0 \\ 0 & C_2 \end{pmatrix} \tag{3.74}$$

where C_2 is a $(r - s_R^*) \times (r - s_R^*)$ orthogonal matrix and estimates for Σ_1^R, $(\Sigma_1 + \Sigma_2)^R$ (denoted by $(\Sigma_1 + \Sigma_2)_R$) and Σ_2^R are obtained as:

$$\hat{\Sigma}_1^R = B T^{1/2} \hat{D}_R^{-2} T^{1/2} B^T \tag{3.75}$$

$$(\Sigma_1 + \Sigma_2)_R = B T^{1/2} \hat{D}_R^{-1} \hat{C}_R^T \hat{\Delta}_R \hat{C}_R \hat{D}_R^{-1} T^{1/2} B^T \tag{3.76}$$

and

$$\hat{\Sigma}_2^R = B T^{1/2} \hat{D}_R^{-1} (\hat{C}_R^T \hat{\Delta} \hat{C}_R - I_r) \hat{D}_R^{-1} T^{1/2} B^T \tag{3.77}$$

respectively. Thus, we proved the following theorem:

Theorem 9 *The restricted maximum likelihood eatimates of Σ_1 and Σ_2 under the assumption that they are both positive semidefinite and Σ_2 is at most of rank s are given by*

$$\hat{\Sigma}_1^R = B T^{1/2} \hat{D}_R^{-2} T^{1/2} B^T \tag{3.78}$$

$$\hat{\Sigma}_2^R = BT^{1/2}\hat{D}_R^{-1}(\hat{C}_R^T\hat{\Delta}_R\hat{C}_R - I_r)\hat{D}_R^{-1}T^{1/2}B^T \tag{3.79}$$

where

$$\hat{D}_R^{-2} = \frac{1}{nq-mp}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}_R^{-1}\right)$$

$$\hat{\Delta}_R = \begin{pmatrix} \frac{[n(q-t)-m(p-t)]k_2}{(n-m)tk_1}T_1 & 0 \\ 0 & I_{r-s_R^*} \end{pmatrix}$$

and

$$\hat{C}_R = \begin{pmatrix} I_{s_R^*} & 0 \\ 0 & C_2 \end{pmatrix}$$

and C_2 is a $(r-s_R^*)\times(r-s_R^*)$ orthogonal matrix and $s_R^* = min(r_R^*, s)$ where r_R^* is the number of τ_i's $i = 1,\ldots,r$ which satisfy

$$\tau_i > \frac{(n-m)t}{n(q-t)-m(p-t)}\frac{k_1}{k_2}$$

3.6 Relationship between ML and REML estimates of variance

In this section we shall use the previous results for deriving relationships between ML and REML estimates of Σ_1 and of $\Sigma_1 + \Sigma_2$. It is obvious from the last two sections that the rank inferred for Σ_2 when using ML will, in general, be different than the corresponding one inferred when using REML. By noting that $(n-m)(q-t) < n(q-t) - m(p-t)$ we deduce that

$$\frac{(n-m)t}{n(q-t)-m(p-t)}\frac{k_1}{k_2} < \frac{t}{q-t}\frac{k_1}{k_2} \tag{3.80}$$

which, in turn, means that

$$s_M^* \leq s_R^* \tag{3.81}$$

Inequality (3.81) shows that the inferred rank for Σ_2 using REML will always be greater than or equal to the corresponding rank inferred by using ML. We shall study the relationship between the ML and REML estimates of Σ_1 and of $\Sigma_1+\Sigma_2$ considering first the case $s_M^* = s_R^* = s^*$. For this, we can note that

$$\hat{\Delta}_M^{-1} = \begin{pmatrix} \frac{tk_1}{(q-t)k_2}T_1^{-1} & 0 \\ 0 & I_{r-s^*} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{(n-m)tk_1}{[n(q-t)-m(p-t)]k_2} T_1^{-1} & 0 \\ 0 & I_{r-s^*} \end{pmatrix} + \begin{pmatrix} \frac{mt(q-p)k_1}{(q-t)[n(q-t)-m(p-t)]k_2} T_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{3.82}$$

$$= \hat{\Delta}_R^{-1} + \Psi \tag{3.83}$$

with

$$\Psi = \begin{pmatrix} \frac{mt(q-p)k_1}{(q-t)[n(q-t)-m(p-t)]k_2} T_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{3.84}$$

A similar equation for $\hat{\Delta}_M$ can be obtained by noting that

$$\hat{\Delta}_M = \hat{\Delta}_R - \Psi_1 \tag{3.85}$$

where

$$\Psi_1 = \begin{pmatrix} \frac{m(q-p)k_2}{(n-m)tk_1} T_1 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.86}$$

Using now equation (3.83) we can easily relate $\hat{D}_R^{-2}$ and $\hat{D}_M^{-2}$ (equations (3.73)and (3.63)) as:

$$\hat{D}_M^{-2} = \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}_M^{-1}\right)$$

$$= \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}(\hat{\Delta}_R^{-1} + \Psi)\right) = \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}_R^{-1}\right) + \frac{1}{nqk_1}\Psi \Rightarrow$$

$$\hat{D}_M^{-2} = \frac{nq-mp}{nq(nq-mp)}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\hat{\Delta}_R^{-1}\right) + \frac{1}{nqk_1}\Psi \Rightarrow$$

$$\left(\hat{D}_M^{-2} - \frac{1}{nqk_1}\Psi\right)\frac{nq}{nq-mp} = \hat{D}_R^{-2} \tag{3.87}$$

using (3.63) and (3.73). Since now the REML estimate of Σ_1^R, $\hat{\Sigma}_1^R$, is written as

$$BT^{1/2}\hat{D}_R^{-2}T^{1/2}B^T,$$

using (3.87) we can rewrite it as

$$BT^{1/2}\left(\frac{nq}{nq-mp}\hat{D}_M^{-2} - \frac{1}{(nq-mp)k_1}\Psi\right)T^{1/2}B^T \Rightarrow$$

$$\hat{\Sigma}_1^R = \frac{nq}{nq-mp}\hat{\Sigma}_1^M - \frac{mt(q-p)}{(q-t)(nq-mp)[n(q-t)-m(p-t)]k_2}B_1B_1^T \tag{3.88}$$

by using (3.84) and the writing of B as (B_1, B_2) where B_1 is a $r \times s^*$ and B_2 a $r \times (r-s^*)$ matrix. Equation (3.88) shows the REML estimate for Σ_1 is not necessarily

larger that the corresponding ML (in the sense that their difference is positive semi-definite). Certainly, $B_1 B_1^T$ is a positive definite matrix but the inclusion of the constant $nq/(nq - mp)$ which is greater than one makes things more complicated. In the limit (as $n \to \infty$), it is easily seen that

$$\frac{nq}{nq - mp} \to 1$$

but the factor of $B_1 B_1^T$ (the other constant) is $o(n^{-1})$ and if we take into account that k_2 is $o(1)$ then

$$\frac{mt(q-p)}{(q-t)(nq-mp)[n(q-t)-m(p-t)]k_2} B_1 B_1^T$$

becomes $o(1)$. Thus,

$$\hat{\Sigma}_1^R = \hat{\Sigma}_1^M \quad \text{as} \quad n \to \infty$$

To relate now $(\Sigma_1 + \Sigma_2)_M$ and $(\Sigma_1 + \Sigma_2)_R$ is a bit more complicated and we shall make the following conventions:

$$\hat{D}_M^{-1} = \begin{pmatrix} D_{1M}^{-1} & 0 \\ 0 & D_{2M}^{-1} \end{pmatrix}, \quad \hat{\Delta}_M = \begin{pmatrix} \Delta_{1M} & 0 \\ 0 & I_{r-s^*} \end{pmatrix} \tag{3.89}$$

where $\Delta_{1M} = \frac{(q-t)k_2}{tk_1} T_1$ (from (3.62)) and D_{1M}^{-1} $(s^* \times s^*)$, D_{2M}^{-1} $((r-s^*) \times (r-s^*))$ diagonal matrices. The above expansion for $\hat{D}_M^{-1}$ always exist since $\hat{D}_M$ is a diagonal matrix. According now to (3.65) the ML estimate of $\Sigma_1 + \Sigma_2$ is

$$(\Sigma_1 + \Sigma_2)_M = B T^{1/2} \hat{D}_M^{-1} \hat{C}_M^T \hat{\Delta}_M \hat{C}_M \hat{D}_M^{-1} T^{1/2} B^T$$

with $\hat{C}$ defined in (3.60). Using now the above convention it is easily seen after some algebra that

$$\hat{D}_M^{-1} \hat{C}^T \hat{\Delta}_M \hat{C} \hat{D}_M^{-1} = \begin{pmatrix} D_{1M}^{-1} \Delta_{1M} D_{1M}^{-1} & 0 \\ 0 & D_{2M}^{-1} C_1^T C_1 D_{2M}^{-1} \end{pmatrix} \tag{3.90}$$

The fact that D_{1M}^{-1} and Δ_{1M} are diagonal matrices and that C_1 is a $(r-s^*) \times (r-s^*)$ orthogonal matrix, allows us to write (3.90) as

$$\hat{D}_M^{-1} \hat{C}_M^T \hat{\Delta}_M \hat{C}_M \hat{D}_M^{-1} = \begin{pmatrix} \Delta_{1M} D_{1M}^{-2} & 0 \\ 0 & D_{2M}^{-2} \end{pmatrix} \tag{3.91}$$

Consequently,

$$(\Sigma_1 + \Sigma_2)_M = BT^{1/2} \begin{pmatrix} \Delta_{1M} D_{1M}^{-2} & 0 \\ 0 & D_{2M}^{-2} \end{pmatrix} T^{1/2} B^T$$

Since again T and Δ_{1M} are diagonal matrices we can have simple expressions for D_{1M}^{-2} and D_{2M}^{-2} deriving from the definition of $\hat{D}_M^2$ in (3.63). These are

$$D_{1M}^{-2} = \frac{1}{nq} \left(\frac{1}{k_2} T_1^{-1} + \frac{1}{k_1} \Delta_{1M}^{-1} \right) \tag{3.92}$$

and

$$D_{2M}^{-2} = \frac{1}{nq} \left(\frac{1}{k_2} T_2^{-1} + \frac{1}{k_1} I_{r-s^*} \right) \tag{3.93}$$

Thus,

$$\Delta_{1M} D_{1M}^{-2} = \frac{1}{nq} \left(\frac{1}{k_2} \Delta_{1M} T_1^{-1} + \frac{1}{k_1} I_{s^*} \right) \tag{3.94}$$

and $(\Sigma_1 + \Sigma_2)_M$ is written as

$$BT^{1/2} \begin{pmatrix} \frac{1}{nq} \left(\frac{1}{k_2} \Delta_{1M} T_1^{-1} + \frac{1}{k_1} I_s \right) & 0 \\ 0 & \frac{1}{nq} \left(\frac{1}{k_2} T_2^{-1} + \frac{1}{k_1} I_{r-s^*} \right) \end{pmatrix} T^{1/2} B^T \tag{3.95}$$

$$= B \begin{pmatrix} \frac{1}{nq} \left(\frac{1}{k_2} \Delta_{1M} + \frac{1}{k_1} T_1 \right) & 0 \\ 0 & \frac{1}{nq} \left(\frac{1}{k_2} I_{r-s^*} + \frac{1}{k_1} T_2 \right) \end{pmatrix} B^T \Rightarrow$$

$$(\Sigma_1 + \Sigma_2)_M = \frac{1}{nq} B \left(\frac{1}{k_2} \hat{\Delta}_M + \frac{1}{k_1} T \right) B^T \tag{3.96}$$

Note that the change of the subscript M to R make this formula to hold for the REML estimate of $\Sigma_1 + \Sigma_2$ apart from constant $1/nq$. However, from (3.85), $\hat{\Delta}_M = \hat{\Delta}_R - \Psi_1$, where

$$\Psi_1 = \begin{pmatrix} \frac{m(q-p)k_2}{(n-m)tk_1} T_1 & 0 \\ 0 & 0 \end{pmatrix}$$

Thus,

$$\begin{aligned} (\Sigma_1 + \Sigma_2)_M &= \frac{1}{nq} B \left(\frac{1}{k_2} (\hat{\Delta}_R - \Psi_1) + \frac{1}{k_1} T \right) B^T \\ &= \frac{1}{nq} B \left(\frac{1}{k_2} \hat{\Delta}_R + \frac{1}{k_1} T \right) B^T - \frac{1}{nqk_2} B \Psi_1 B^T \\ &= \frac{nq - mp}{nq} (\Sigma_1 + \Sigma_2)_R - \frac{1}{nqk_2} B \Psi_1 B^T \\ &= \frac{nq - mp}{nq} (\Sigma_1 + \Sigma_2)_R - \frac{m(q-p)}{nqt(n-m)k_1} B_1 T_1 B_1^T \Rightarrow \end{aligned}$$

$$(\Sigma_1 + \Sigma_2)_R = \frac{nq}{nq - mp}(\Sigma_1 + \Sigma_2)_M + \frac{m(q-p)}{(nq - mp)(n-m)tk_1} B_1 T_1 B_1^T \qquad (3.97)$$

Equation (3.97) clearly shows that $(\Sigma_1 + \Sigma_2)_R$ is greater than $(\Sigma_1 + \Sigma_2)_M$ since

$$(\Sigma_1 + \Sigma_2)_R - (\Sigma_1 + \Sigma_2)_M > (\Sigma_1 + \Sigma_2)_R - \frac{nq}{nq - mp}(\Sigma_1 + \Sigma_2)_M$$

$$= \frac{m(q-p)}{(nq - mp)(n-m)tk_1} B_1 T_1 B_1^T$$

and the last matrix is positive definite. The inclusion of the constant $nq/nq - mp$ now does not affect the positive definiteness of the difference $(\Sigma_1 + \Sigma_2)_R - (\Sigma_1 + \Sigma_2)_M$. In the limit we can see that again

$$\lim_{n \to \infty} \frac{nq}{nq - mp} = 1$$

and that the factor of $B_1 T_1 B_1^T$

$$\frac{m(q-p)}{nqt(n-m)k_1}$$

tends to zero since k_1 is $o(1)$. Thus, again as in the case of Σ_1 we have

$$(\Sigma_1 + \Sigma_2)_R = (\Sigma_1 + \Sigma_2)_M \quad \text{as} \quad n \to \infty$$

If we now assume that $s_M^* < s_R^*$, the method for deriving the required (and more general) relationship will be somewhat different. We shall first form a relationship between the ML estimates of Σ_1 when s_M^* and s_R^* respectively are the inferred ranks for Σ_2. Then, in this expression we shall substitute the relationship between the ML and REML estimates found previously. The same method will be followed for $\Sigma_1 + \Sigma_2$. To that respect, let us assume that $s_R^* - s_M^* = s_0$ and that the diagonal submatrix T_1 $(s_R^* \times s_R^*)$ of T and matrix B are written as:

$$T_1 = \begin{pmatrix} T_{11} & 0 \\ 0 & T_{12} \end{pmatrix} \quad \text{and} \quad B = (B_{11}, B_{12}, B_2) \qquad (3.98)$$

where T_{11} $(s_M^* \times s_M^*)$ and T_{12} $(s_0 \times s_0)$ are diagonal submatrices and B_{11}, B_{12}, B_2 are $r \times s_M^*$, $r \times s_0$ and $r \times (r - s_R^*)$ matrices respectively. Using now, the definition for

D_M^{-2} and writing $D^{-2}_{M,s_R^*-s_0}$ the D_M matrix which corresponds to inferred rank $s_R^* - s_0$ for Σ_2, we have:

$$D^{-2}_{M,s_R^*-s_0} = \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\begin{pmatrix} \frac{(q-t)k_2}{tk_1}T_{11} & 0 \\ 0 & I_{r-(s_R^*-s_0)} \end{pmatrix}^{-1}\right)$$

$$= \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\begin{pmatrix} \frac{tk_1}{(q-t)k_2}T_{11}^{-1} & 0 & 0 \\ 0 & \frac{tk_1}{(q-t)k_2}T_{12}^{-1} & 0 \\ 0 & 0 & I_{r-s_R^*} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{s_0} - \frac{tk_1}{(q-t)k_2}T_{12}^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}\right)$$

$$= \frac{1}{nq}\left(\frac{1}{k_2}T^{-1} + \frac{1}{k_1}\Delta^{-1}_{s_R^*} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{s_0} - \frac{tk_1}{(q-t)k_2}T_{12}^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) \Rightarrow$$

$$D^{-2}_{M,s_R^*-s_0} = D^{-2}_{s_R^*} + \frac{1}{nq}\Psi_2 \tag{3.99}$$

where

$$\Psi_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{s_0} - \frac{tk_1}{(q-t)k_2}T_{12}^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.100}$$

But it is known from the previous sections that $\Sigma^M_{1,s_R^*-s_0} = BT^{1/2}D^{-2}_{M,s_R^*-s_0}T^{1/2}B^T$. By considering (3.99), this equality is written as

$$BT^{1/2}D^{-2}_{M,s_R^*}T^{1/2}B^T + \frac{1}{nq}BT^{1/2}\Psi_2 T^{1/2}B^T$$

After some simple algebra, this is easily seen to be

$$\Sigma^M_{1,s_R^*-s_0} \equiv \Sigma^M_{1,s_M^*} = \Sigma^M_{1,s_R^*} + \frac{1}{nq}B_{12}\left(\frac{1}{k_1}T_{12} - \frac{t}{(q-t)k_2}I_{s_0}\right)B_{12}^T \tag{3.101}$$

In the same way, we can see that

$$(\Sigma_1 + \Sigma_2)_{M,s_R^*-s_0} = \frac{1}{nq}B\left(\frac{1}{k_2}\Delta_{M,s_R^*-s_0} + \frac{1}{k_1}T\right)B^T$$

$$= \frac{1}{nq}B\left(\frac{1}{k_2}\begin{pmatrix} \frac{(q-t)k_2}{tk_1}T_{11} & 0 & 0 \\ 0 & \frac{(q-t)k_2}{tk_1}T_{12} & 0 \\ 0 & 0 & I_{r-s_R^*} \end{pmatrix}\right.$$

$$\left. + \frac{1}{k_2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{s_0} - \frac{(q-t)k_2}{tk_1}T_{12} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{k_1}T\right)B^T \Rightarrow$$

$$(\Sigma_1+\Sigma_2)_{M,s_R^*-s_0}=\frac{1}{nq}(\Sigma_1+\Sigma_2)_{M,s_R^*}+\frac{1}{nqk_2}B_{12}\left(I_{s_0}-\frac{(q-t)k_2}{tk_1}T_{12}\right)B_{12}^T \quad (3.102)$$

Using now (3.101) and (3.102) along with (3.88) and (3.97) we get after elementary operations that

$$\hat{\Sigma}_{1,s_M^*}^M=\frac{nq-mp}{nq}\hat{\Sigma}_{1,s_R^*}^R+\frac{mt(q-p)}{nq(q-t)[n(q-t)-m(p-t)]k_2}B_1B_1^T$$
$$+\frac{1}{nq}B_{12}\left(\frac{1}{k_1}T_{12}-\frac{t}{(q-t)k_2}I_{s_0}\right)B_{12}^T$$

and

$$(\Sigma_1+\Sigma_2)_{M,s_M^*}=\frac{nq-mp}{nq}(\Sigma_1+\Sigma_2)_{R,s_R^*}-\frac{m(q-p)}{nq(n-m)tk_1}B_1T_1B_1^T$$
$$+\frac{1}{nqk_2}B_{12}\left(I_{s_0}-\frac{(q-t)k_2}{tk_1}T_{12}\right)B_{12}^T$$
$$=\frac{nq-mp}{nq}(\Sigma_1+\Sigma_2)_{R,s_R^*}-\frac{m(q-p)}{nq(n-m)tk_1}B_1T_1B_1^T$$
$$-\frac{q-t}{nqt}B_{12}\left(\frac{1}{k_1}T_{12}-\frac{t}{(q-t)k_2}I_{s_0}\right)B_{12}^T$$

But $\left(\frac{1}{k_1}T_{12}-\frac{t}{(q-t)k_2}I_{s_0}\right)$ is negative definite since T_{12} consists of those elements of T for which

$$\frac{(n-m)t}{n(q-t)-m(p-t)}\frac{k_1}{k_2}<\tau_i<\frac{t}{q-t}\frac{k_1}{k_2}$$

The results clearly show that in the case where $s_M^*<s_R^*$, the REML estimate of $(\Sigma_1+\Sigma_2)$ is not any more necessarily greater than the corresponding ML estimate.

The results of this section are summarized in the following theorem:

Theorem 10 *If s_M^* and s_R^* are the ranks inferred for Σ_2 when using ML and REML respectively, then*

$$s_M^*\leq s_R^*$$

Furthermore, the maximum likelihood and the restricted maximum likelihood estimates of Σ_1 and $\Sigma_1+\Sigma_2$ respectively are related with the following equations:

$$\hat{\Sigma}_{1,s_M^*}^M=\frac{nq-mp}{nq}\hat{\Sigma}_{1,s_R^*}^R+\frac{mt(q-p)}{nq(q-t)[n(q-t)-m(p-t)]k_2}B_1B_1^T$$

$$+\frac{1}{nq}B_{12}\left(\frac{1}{k_1}T_{12}-\frac{t}{(q-t)k_2}I_{s_0}\right)B_{12}^T \tag{3.103}$$

and

$$(\Sigma_1+\Sigma_2)_{M,s_M^*}=\frac{nq-mp}{nq}(\Sigma_1+\Sigma_2)_{R,s_R^*}-\frac{m(q-p)}{nq(n-m)tk_1}B_1T_1B_1^T$$
$$-\frac{q-t}{nqt}B_{12}\left(\frac{1}{k_1}T_{12}-\frac{t}{(q-t)k_2}I_{s_0}\right)B_{12}^T \tag{3.104}$$

where $B=(B_1,B_2)$ *(*$B_1\ (r\times s_R^*)$*,* $B_2\ (r\times(r-s_R^*))$*) and* $B_1=(B_{11},B_{12})$ *(*$B_{11}\ (r\times s_M^*)$*,* $B_{12}\ (r\times s_0)$*) and* $s_R^*=s_M^*+s_0$*.*

It will be instructive to consider a simple example which will clarify the relationship between ML and REML estimates when $r=1$, for Σ_1 only. In this case H and G will be univariate functions of the data matrix Y with

$$k_2H=b^2 \quad \text{and} \quad k_1G=b^2\tau \tag{3.105}$$

where $\tau=\frac{k_1}{k_2}\frac{G}{H}$. The ML estimate of $\Sigma_1\equiv\sigma_1$ (scalar) will be according to (3.67)

$$b^2\tau D_M^{-2} \tag{3.106}$$

where

$$D_M^{-2}=\frac{1}{nq}\left(\frac{1}{k_2\tau}+\frac{tk_1}{k_1(q-t)k_2\tau}\right)$$
$$=\frac{1}{n(q-t)\tau k_2} \quad \text{if} \quad \tau>\frac{t}{q-t}\frac{k_1}{k_2}$$

otherwise

$$D_M^{-2}=\frac{1}{nq}\left(\frac{1}{k_2\tau}+\frac{1}{k_1}\right)$$
$$=\frac{1}{nq}\left(\frac{k_1+k_2\tau}{k_1k_2\tau}\right)$$

In that way, $\hat{\sigma}_1^M$ becomes

$$\hat{\sigma}_1^M=\frac{b^2\tau}{n(q-t)\tau k_2}\Rightarrow$$

$$\hat{\sigma}_1^M=\frac{1}{n(q-t)}H \quad \text{if} \quad \tau>\frac{t}{q-t}\frac{k_1}{k_2} \tag{3.107}$$

otherwise

$$\hat{\sigma}_1^M = \frac{b^2\tau}{nq}\left(\frac{k_1 + k_2\tau}{k_1 k_2 \tau}\right) \Rightarrow$$

$$\hat{\sigma}_1^M = \left(\frac{k_1 + k_2\tau}{nqk_1}\right) H \tag{3.108}$$

The first case coincides with the result found by Lange and Laird (1989). In the case however, where τ does not satisfy the condition in (3.107), the required adjustment on the estimate of σ_1 is provided by (3.108).

In the case of REML estimation now, $\hat{D}_R^{-2}$ has the form (as this is given by Theorem 9)

$$\hat{D}_R^{-2} = \frac{1}{nq - mp}\left(\frac{1}{k_2\tau} + \frac{(n-m)tk_1}{[n(q-t) - m(p-t)]k_1 k_2 \tau}\right)$$

$$= \frac{1}{nq - mp}\left(\frac{n(q-t) - m(p-t) + (n-m)t}{[n(q-t) - m(p-t)]k_2\tau}\right) \Rightarrow$$

$$\hat{D}_R^{-2} = \frac{1}{[n(q-t) - m(p-t)]k_2\tau} \quad \text{if} \quad \tau > \frac{(n-m)t}{n(q-t) - m(p-t)}\frac{k_1}{k_2}$$

otherwise

$$\hat{D}_R^{-2} = \frac{1}{nq - mp}\left(\frac{1}{k_2\tau} + \frac{1}{k_1}\right) \Rightarrow$$

$$\hat{D}_R^{-2} = \frac{k_1 + k_2\tau}{(nq - mp)k_1 k_2 \tau}$$

Consequently, since $\hat{\sigma}_1^R = b^2\tau\hat{D}_R^{-2}$ we get,

$$\hat{\sigma}_1^R = \frac{1}{n(q-t) - m(p-t)}H \quad \text{if} \quad \tau > \frac{(n-m)t}{n(q-t) - m(p-t)}\frac{k_1}{k_2} \tag{3.109}$$

otherwise

$$\hat{\sigma}_1^R = \frac{b^2\tau(k_1 + k_2\tau)}{(nq - mp)k_1 k_2 \tau} \Rightarrow$$

$$\hat{\sigma}_1^R = \frac{k_1 + k_2\tau}{(nq - mp)k_1}H \tag{3.110}$$

It is seen then that σ_1^M and σ_1^R have similar estimates with differences in the multiplicative constants. Recall from the beginning of this section (inequality (3.80)) that

$$\frac{(n-m)t}{n(q-t) - m(p-t)}\frac{k_1}{k_2} < \frac{t}{q-t}\frac{k_1}{k_2}$$

Consequently, the two conditions in (3.107) and in (3.109) inply that we should examine three subcases in order to understand when REML estimates are greater than ML:

- $\tau < \frac{(n-m)t}{n(q-t)-m(p-t)}\frac{k_1}{k_2}$: In this case the multiplicative constant for ML is $\frac{k_1+k_2\tau}{nqk_1}$ and the corresponding one for REML is $\frac{k_1+k_2\tau}{(nq-mp)k_1}$. The latter is always greater than the former, thus, the REML estimate of σ_1 is always greater than the corresponding ML.

- $\frac{(n-m)t}{n(q-t)-m(p-t)}\frac{k_1}{k_2} \leq \tau \leq \frac{t}{q-t}\frac{k_1}{k_2}$: Here the multiplicative constant for ML is as before $\frac{k_1+k_2\tau}{nqk_1}$ and the one for REML is $\frac{1}{n(q-t)-m(p-t)}$. Then,

$$\frac{k_1+k_2\tau}{nqk_1} = \frac{1}{nq} + \frac{\tau}{nq}\frac{k_2}{k_1} \leq \frac{1}{nq} + \frac{t}{nq(q-t)}$$

since $\tau \leq \frac{t}{q-t}\frac{k_1}{k_2}$. But the last expression is equal to

$$\frac{1}{n(q-t)}$$

for which, in turn,

$$\frac{1}{n(q-t)} < \frac{1}{n(q-t)-m(p-t)}$$

Thus, again in this case REML estimate for σ_1 is greater than the corresponding ML.

- $\frac{t}{q-t}\frac{k_1}{k_2} < \tau$: In this final case the multiplicative constants are $\frac{1}{n(q-t)}$ and $\frac{1}{n(q-t)-m(p-t)}$ for ML and REML respectively. However, it is already seen that the latter is always greater than the former.

The example showed clearly that when $r = 1$ then REML estimates of σ_1 are always greater than the corresponding ML but when $r > 1$ such relationship does not exist as Theorem 10 clearly showed.

3.7 Discussion and conclusions

The relationships (3.88) and (3.97) can help us to find a relationship between the REML and ML estimates of the covariance matrix of the estimated coefficients of the

model. From (3.50) and (3.51) we can deduce that

$$Var(\hat{\xi}_1^*) = I_t \otimes (\Sigma_1 + \Sigma_2) \otimes (A^T A)^{-1} \tag{3.111}$$

$$Var(\hat{\xi}_2^*) = I_{p-t} \otimes \Sigma_1 \otimes (A^T A)^{-1} \tag{3.112}$$

while

$$Cov(\hat{\xi}_1^*, \hat{\xi}_2^*) = 0 \tag{3.113}$$

ML or REML estimates of these matrices are obtained if we substitute in the place of $\Sigma_1 + \Sigma_2$ and of Σ_1 their corresponding ML or REML estimates. The discussion in the previous section showed that the REML estimate of the variance matrix of estimated ξ_1^* is larger than the corresponding ML one when the inferred ranks for Σ_2 using both methods are equal, while it is difficult to determine such an explicit relationship when $s_M^* < s_R^*$ or for the estimated variances of ξ_2^*. The reason for this interesting new result can be possibly traced to the randomness of $\hat{\xi}_1^*$ which contributed more variation to the estimator of $\hat{\xi}_1^*$.

Chapter 4

Certain results on the Covariance Adjustment method.

4.1 Introduction

In this chapter, another approach to growth curves will be considered. Although the specifications of the model will remain the same, no specific structure will be considered for the covariance matrix and no maximizing likelihood approach either. The method was introduced by Rao (1965, 1966) as a comment on Potthoff and Roy's approach to growth curves. It has been widely used by experimenters since it requires the minimum of assumptions about the covariance matrix of observations. The mean trend is modelled in the same way as before. However, the method has a drawback since it requires the calculation of the inversion of a matrix which is not positive definite with probability 1 unless the number of residual degrees of freedom $n - m$ is greater than q, the number of dependent variables per individual. This is not the case in all experiments. Undoubtedly, in these cases a likelihood approach with a carefully structured covariance matrix is more suitable.

In section 2 the estimator will be defined and its form will be proved as a GLS estimator. As a consequence, it will be seen that the method consists essentially of the best selection of the estimator's weight covariance matrix from a collection of suitably

defined matrices. In section 3 three methods for the optimal selection of the estimator will be presented along with a newly proposed one, and a comparison will be made between them. Finally, the chapter will close with the conclusions.

4.2 Derivation of estimator

The model assumption will remain the same as in Chapter 2, that is

$$Y = A\xi P + E, \tag{4.1}$$

where Y $(n \times q)$ is a matrix of observations from n individuals at q different times, A $(n \times m)$ and P $(p \times q)$ are known design matrices of ranks m and p respectively, while ξ $(m \times p)$ is a matrix of mp unknown regression parameters. Furthermore, E $(n \times q)$ is a matrix of errors, whose n rows ϵ_i $(i = 1, \ldots, n)$ are independent q-dimensional vectors with

$$\epsilon_i \sim N(0, \Sigma). \tag{4.2}$$

Matrix Σ will be left unrestricted. Let us suppose that we have two matrices H_1 $(q \times p)$ and H_2 $(q \times (q-p))$ such that

$$PH_1 = I_p, \quad PH_2 = 0. \tag{4.3}$$

Consider the transformations $Y_1 = YH_1$ and $Y_2 = YH_2$. Then,

$$E(Y_1) = A\xi \quad E(Y_2) = 0 \tag{4.4}$$

and

$$(Y_1, Y_2) \sim N((A\xi, 0), \begin{pmatrix} H_1^T \\ H_2^T \end{pmatrix} \Sigma (H_1, H_2) \otimes I). \tag{4.5}$$

The inferences are based on the conditional variable $Y_1|Y_2$ which from a well known property of the multinormal distribution (Mardia, Kent, Bibby, 1982) follows a multinormal distribution with

$$E(Y_1|Y_2) = A\xi + Y_2 B \tag{4.6}$$

$$Var(vec(Y_1|Y_2)) = (P\Sigma^{-1}P^T)^{-1} \otimes I_n, \tag{4.7}$$

where $B = (H_2^T \Sigma H_2)^{-1} H_2^T \Sigma H_1$. The reason for which we base our inferences on the mean of the conditional variable $Y_1|Y_2$ arises from the well known equality

$$Var(Y_1) = Var_{Y_2}\left(E(Y_1|Y_2)\right) + E_{Y_2}\left(Var(Y_1|Y_2)\right). \tag{4.8}$$

In that way we see that if Y_1 is an unbiased estimator of ξ (as it is from (4.4)), then $E(Y_1|Y_2)$ will also be an unbiased estimator and since $E_{Y_2}\left(Var(Y_1|Y_2)\right)$ is positive, it will also be more efficient.

The required properties (4.3) suggest that H_1 should have the form

$$H_1 = P^T(PP^T)^{-1} \tag{4.9}$$

and H_2 should be an appropriate multiple of

$$I_q - P^T(PP^T)^{-1}P. \tag{4.10}$$

The reason for this, is that H_1 as in (4.9) is a $q \times p$ matrix but H_2 as in (4.10) is a square matrix which should be modified appropriately in order to be a $q \times (q-p)$ matrix. The appropriate modification will be found in the following way. Let us suppose that we fit the data perfectly (interpolation) without leaving any variation unexplained. This can be done by fitting a polynomial of degree $q-1$ instead of $p-1$. Thus, the mean will have the following expression

$$A(\xi_1, \xi_2)\begin{pmatrix} P \\ P_0 \end{pmatrix}, \tag{4.11}$$

where P_0 is a $(q-p) \times q$ matrix containing the additional polynomial coefficients needed for fitting a $q-1$ degree polynomial. The Ordinary Least Squares (OLS) estimator $\hat{\xi}_{OLS}$ is

$$\begin{aligned} \hat{\xi}_{OLS} &= (A^TA)^{-1}A^TY\left(\begin{pmatrix} P \\ P_0 \end{pmatrix}(P^T, P_0^T)\right)^{-1}\begin{pmatrix} P \\ P_0 \end{pmatrix} && (4.12)\\ &= (A^TA)^{-1}A^TY(\Omega_1, \Omega_2) && (4.13)\\ &= ((A^TA)^{-1}A^TY\Omega_1, (A^TA)^{-1}A^TY\Omega_2), && (4.14) \end{aligned}$$

where

$$\Omega_1 = (I - \Pi_{P_0})P^T(P(I - \Pi_{P_0})P^T)^{-1}$$

$$\Omega_2 = (I - \Pi_P)P_0^T(P_0(I - \Pi_P)P_0^T)^{-1}$$

and $\Pi_P = P^T(PP^T)^{-1}P$, $\Pi_{P_0} = P_0^T(P_0P_0^T)^{-1}P_0$. That is, the estimator of the first p coefficients is a GLS weighted by $I - \Pi_{P_0}$ and the estimator of the remaining $q - p$ is a GLS weighted by $I - \Pi_P$. Note that Ω_2 is a $q \times (q - p)$ matrix and is a multiple of $I - \Pi_P$. That means that Ω_2 satisfies the second property of (4.3) and can be considered as the transformation matrix H_2. From now on, we shall assume that

$$H_2 \equiv V = \Omega_2, \tag{4.15}$$

Without loss of generality we can also assume that

$$V^TV = I_{q-p}, \tag{4.16}$$

since V can also be

$$V = (I - \Pi_P)P_0^T(P_0(I - \Pi_P)P_0^T)^{-1/2}. \tag{4.17}$$

Since now $E(Y_1|Y_2) = A\xi + Y_2B$ and $E_{Y_2}(E(Y_1|Y_2)) = E(Y_1)$ we get that Y_1 actually estimates $A\xi + Y_2B$ whch in turn means that $A\xi$ is estimated by $Y_1 - Y_2B$ and finally ξ is estimated by $(A^TA)^{-1}A^T(Y_1 - Y_2B)$. Direct substitution by (4.9) and by the fact that $B = (H_2^T\Sigma H_2)^{-1}H_2^T\Sigma H_1$ where $H_2 = V$ gives that the covariance adjusted estimator has the form

$$(A^TA)^{-1}A^T(YP^T(PP^T)^{-1} - YV(V^T\Sigma V)^{-1}V^T\Sigma P^T(PP^T)^{-1}).$$

A final substitution of

$$\hat{\Sigma} \equiv \frac{1}{n-m}S = \frac{1}{n-m}Y^T(I - \Pi_A)Y, \tag{4.18}$$

as an estimate of the covariance matrix of observations gives the covariance adjusted estimator as

$$\hat{\xi}_{C.A} = (A^TA)^{-1}A^T(YP^T(PP^T)^{-1} - YV(V^TSV)^{-1}V^TSP^T(PP^T)^{-1}), \tag{4.19}$$

where $V^TV = I_{q-p}$. This is the form also used by Kenward (1986). Rao (1965) claimed that this estimator is better than the GLS estimator used by Potthoff and Roy (1964) because it exploits information which Potthoff and Roy's solution ignores. However, as Baksalary, Corsten and Kala (1978) showed, $\hat{\xi}_{C.A}$ is actually a GLS estimator appropriately weighted. Lemma 1 of Chapter 2 can be used to find this weight:

Theorem 11 *Suppose that we decompose matrix V into $V = (V_1, V_2)$ with $V_1(q \times c)$ and $V_2(q \times (q-p-c))$. Suppose also that the columns of V_1 are used in adjusting the OLS estimator. The form then of the new estimator is a GLS one weighted by $S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}$.*

Proof: Since we use the columns of V_1 as covariates, the estimator (4.19) becomes

$$(A^TA)^{-1}A^TYS^{-1}\left(S - SV_1(V_1^TSV_1)^{-1}V_1^TS\right)P^T(PP^T)^{-1}. \tag{4.20}$$

Consider the matrix $V_1^* = (P^T, V_2)$ and note that $V_1^TV_1^* = 0$. Applying Lemma 1 in Chapter 2, we get that (4.20) becomes

$$\begin{aligned}
&(A^TA)^{-1}A^TYS^{-1}\left[(P^T,V_2)\left[\begin{pmatrix} PS^{-1} \\ V_2^TS^{-1}\end{pmatrix}(P^T,V_2)\right]^{-1}\begin{pmatrix} P \\ V_2^T\end{pmatrix}\right]P^T(PP^T)^{-1} \\
&= (A^TA)^{-1}A^TYS^{-1}\left[(P^T,V_2)\Omega\begin{pmatrix} P \\ V_2^T\end{pmatrix}\right]P^T(PP^T)^{-1}
\end{aligned} \tag{4.21}$$

where

$$\Omega = \begin{pmatrix} A_1^{-1} & -(PS^{-1}P^T)^{-1}PS^{-1}V_2A_2^{-1} \\ -(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}P^TA_1^{-1} & A_2^{-1}\end{pmatrix},$$

and

$$\begin{aligned}
A_1 &= P(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^T \\
A_2 &= V_2^T\left(S^{-1} - S^{-1}P^T(PS^{-1}P^T)^{-1}PS^{-1}\right)V_2.
\end{aligned} \tag{4.22}$$

By noting that

$$\begin{pmatrix} P \\ V_2^T\end{pmatrix}P^T(PP^T)^{-1} = \begin{pmatrix} I \\ 0\end{pmatrix},$$

we get that (4.21) is

$$(A^TA)^{-1}A^TYS^{-1}(P^TA_1^{-1} - V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}P^TA_1^{-1})$$

$$= (A^TA)^{-1}A^TY(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^TA_1^{-1}$$

which, by the definition of A_1, is a GLS estimator weighted by $S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}$. The distribution of this estimator was first studied by Gleser and Olkin (1972) who considered the case where all covariates have been used. The result was later generalized by Kenward (1986) who considered an arbitrary number of covariates. Specifically, the generalized contrast $L\hat{\xi}M$ where $L(l \times m)$, $M(p \times t)$ are known matrices, was expressed as:

$$L\hat{\xi}M = L\xi M + C^{-1}QR^{-1} \tag{4.23}$$

where C $(l \times l)$ is such that $(C^TC)^{-1} = L(A^TA)^{-1}L^T$, R $(t \times t)$ is a unique lower triangular matrix with positive diagonal elements satisfying

$$(RR^T)^{-1} = M^T(PP^T)^{-1}P(\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}M \equiv \Phi$$

and $Q = X_1 - X_2U_{22}^{-1}U_{21}$ where X_1 $(l \times t)$, X_2 $(l \times c)$, U_{22} $(c \times c)$ and U_{21} $(c \times t)$ are distributed according to

$$X = (X_1, X_2) \sim N_l(0, I_{t+c})$$

and

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \sim W_{t+c}(n-m, I_{t+c})$$

and X and U are independent, (Kenward, 1986). Using this expression, he obtained the marginal density of $L\hat{\xi}M$ and showed that it has also an asymptotic normal distribution.

From the work of Rao (1967) and Grizzle and Allen (1969), we can deduce that the unconditional variance of this estimator is

$$\frac{n-m-1}{n-m-c-1}(A^TA)^{-1} \otimes (PP^T)^{-1}P(\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}. \tag{4.24}$$

The key expression for finding an estimate of (4.24) is $\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma$. Let us consider the variable $\begin{pmatrix} V_1^T \\ I_q \end{pmatrix} S(V_1, I_q)$. Since $S \sim W_q(\Sigma, n-m)$ we easily get that

$$\begin{pmatrix} V_1^T \\ I_q \end{pmatrix} S(V_1, I_q) = \begin{pmatrix} V_1^T S V_1 & V_1^T S \\ S V_1 & S \end{pmatrix} \sim W_{q+c}\left(\begin{pmatrix} V_1^T \\ I_q \end{pmatrix} \Sigma(V_1, I_q), n-m\right).$$

Thus, $S - SV_1(V_1^TSV_1)^{-1}V_1^TS \sim W_q(\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma, n-m-c)$ (see also Mardia, Kent and Bibby, 1982). Consequently, an unbiased estimate of (4.24) is

$$\Psi_1 \equiv \frac{n-m-1}{(n-m-c-1)(n-m-c)}$$

$$(A^TA)^{-1} \otimes (PP^T)^{-1}P(S - SV_1(V_1^TSV_1)^{-1}V_1^TS)P^T(PP^T)^{-1}. \qquad (4.25)$$

Exploiting again Lemma 1 of Chapter 2, and with the help of the proof of the previous theorem, we get that the last expression is actually

$$\frac{n-m-1}{(n-m-c-1)(n-m-c)}$$

$$(A^TA)^{-1} \otimes (PP^T)^{-1}P(I_q - V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^TA_1^{-1}$$

$$= \frac{n-m-1}{(n-m-c-1)(n-m-c)}$$

$$(A^TA)^{-1} \otimes \left(P(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^T\right)^{-1}. \qquad (4.26)$$

Another estimate which is likely to perform well and seems reasonable in its definition is obtained by substituting in (4.24), $\frac{1}{n-m}S$ in the place of Σ. This would give an alternative estimator

$$\Psi_2 \equiv \frac{n-m-1}{(n-m-c-1)(n-m)}$$

$$(A^TA)^{-1} \otimes (PP^T)^{-1}P(S - SV_1(V_1^TSV_1)^{-1}V_1^TS)P^T(PP^T)^{-1}. \qquad (4.27)$$

Note now that $(n-m-1)(n-m) > (n-m-c-1)(n-m-c)$ for $c > 0$. This entails the inequalities,

$$\frac{n-m-1}{(n-m-c-1)(n-m-c)} > \frac{n-m-1}{(n-m-c-1)(n-m)} > \frac{1}{n-m}$$

which in turn mean two things: first, that $\Psi_1 > \Psi_2$ and second that for large n the first expression is approximately equal to the second one and both then are approximately

equal to $\frac{1}{n-m}$. This, along with (4.26) shows that when we fit all possible covariates ($V_2 = 0$) and we let n tend to infinity then the covariance matrix of estimated coefficients will be approximately equal to $(A^TA)^{-1} \otimes \frac{1}{n-m}\left(PS^{-1}P^T\right)^{-1}$ which, since S follows a Wishart distribution with mean Σ and degrees of freedom $n-m$, will be approximately

$$(A^TA)^{-1} \otimes \left(P\Sigma^{-1}P^T\right)^{-1}$$

which is the least variance an unbiased estimator of ξ can have. Thus, the inclusion of all possible covariates is asymptotically efficient. In small samples, however, such a strategy will just reduce efficiency and we should look for methods to reduce effectively the number of covariates.

4.3 A comparison between methods of choosing covariates

4.3.1 A description of 3 known methods

The discussion of the previous sections showed clearly that the problem of covariance adjustment as the question of the final choice of the estimator of ξ can be stated, is essentially a problem of selection of the appropriate weight matrix of a GLS estimator (see also Definition 1 in Chapter 2), among the elements of the set which contains all the matrices of the form

$$S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1} \tag{4.28}$$

such that $V = (V_1, V_2)$, V_2 is a $q \times c$ matrix where $c = 0, \ldots, q-p$ and V is defined in (4.17).

Three criteria will be discussed here along with a new one which will be suggested in the next subsection. The criteria will also be compared. The comparison will be made on the basis of the proximity of the estimated variance of the covariance adjusted

estimator to the variance of the Gauss-Markov estimator which is the theoretically optimal one, as well as on the basis of how close to its real variance the estimated variance of the adjusted estimator lies.

The three already known criteria suggested for the optimal choice of covariates have been introduced by Grizzle and Allen (1969), Kenward (1985) and more recently by Fujikoshi and Rao (1991). Verbyla (1986) has also dealt with the subject adopting an approach which exploits the parametric form of the covariance matrix of observations.

Grizzle and Allen (1969) suggested the minimization of the generalized variance of the estimator (Least Generalized Variance method (LGV) from now on) as a reasonable criterion for choosing the appropriate estimator. That is, the minimization of

$$\left|\frac{n-m-1}{n-m-c-1}(A^TA)^{-1}\otimes(PP^T)^{-1}P(\Sigma-\Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}\right|$$

which is equivalent with the minimization of

$$\left(\frac{n-m-1}{n-m-c-1}\right)^{mp}\left|(PP^T)^{-1}P(\Sigma-\Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}\right|.$$

Since in applications Σ is unknown, then either we should use the determinant of Ψ_1 which is

$$\left(\frac{n-m-1}{(n-m-c-1)(n-m-c)}\right)^{mp}\left|(PP^T)^{-1}P(S-SV_1(V_1^TSV_1)^{-1}V_1^TS)P^T(PP^T)^{-1}\right| \tag{4.29}$$

or that of Ψ_2 which is

$$\left(\frac{n-m-1}{(n-m-c-1)(n-m)}\right)^{mp}\left|(PP^T)^{-1}P(S-SV_1(V_1^TSV_1)^{-1}V_1^TS)P^T(PP^T)^{-1}\right|. \tag{4.30}$$

Let us examine more carefully the form $V_1(V_1^TSV_1)^{-1}V_1^T$. Suppose that we fit all possible covariates ($V_1 = V$). then,

$$\begin{aligned}&(V_1,V_2)\left[\begin{pmatrix}V_1^T\\V_2^T\end{pmatrix}S(V_1,V_2)\right]^{-1}\begin{pmatrix}V_1^T\\V_2^T\end{pmatrix}\\&=(V_1,V_2)\begin{pmatrix}D_1 & D_2\\D_2^T & D^{-1}\end{pmatrix}\begin{pmatrix}V_1^T\\V_2^T\end{pmatrix},\end{aligned}$$

where $D_1 = (V_1^T S V_1)^{-1} + (V_1^T S V_1)^{-1} V_1^T S V_2 D^{-1} V_2^T S V_1 (V_1^T S V_1)^{-1}$,
$D_2 = -(V_1^T S V_1)^{-1} V_1^T S V_2 D^{-1}$ and $D = V_2^T S V_2 - V_2^T S V_1 (V_1^T S V_1)^{-1} V_1^T S V_2$. But the last equation is actually equal to

$$V_1 (V_1^T S V_1)^{-1} V_1^T + V_1 (V_1^T S V_1)^{-1} V_1^T S V_2 D^{-1} V_2^T S V_1 (V_1^T S V_1)^{-1} V_1^T$$

$$-V_2 D^{-1} V_2^T S V_1 (V_1^T S V_1)^{-1} V_1^T + V_2 D^{-1} V_2^T - V_1 (V_1^T S V_1)^{-1} V_1^T S V_2 D^{-1} V_2^T \Rightarrow$$

$$V (V^T S V)^{-1} V^T - V_1 (V_1^T S V_1)^{-1} V_1^T$$

$$= (V_1 (V_1^T S V_1)^{-1} V_1^T S - I_q) V_2^T D^{-1} V_2^T (V_1 (V_1^T S V_1)^{-1} V_1^T S - I_q)^T.$$

Since the right hand side matrix is positive definite, it is clear that $V(V^T S V)^{-1} V^T > V_1 (V_1^T S V_1)^{-1} V_1^T$. This, in turn, means that

$$S - S V_1 (V_1^T S V_1)^{-1} V_1^T S > S - S V (V^T S V)^{-1} V^T S.$$

The above relationship shows that the inclusion of all possible covariates would give the least possible variance. Note however, that

$$\left(\frac{n-m-1}{(n-m-c-1)(n-m)} \right)^{mp} < \left(\frac{n-m-1}{(n-m-c-1)(n-m-c)} \right)^{mp}$$

and that

$$\left(\frac{n-m-1}{(n-m-c-1)(n-m-c)} \right)^{mp} < \left(\frac{n-m-1}{(n-m-(q-p)-1)(n-m-(q-p))} \right)^{mp}.$$

Thus, the inclusion of the above constants into expressions (4.29) and (4.30) for the estimated generalized variance of the covariance adjusted estimator makes things more complicated. As was shown in the last paragraph of the previous section and as can be seen from the last three relationships, asymptotically, the inclusion of all possible covariates is the best solution (see also Altham, 1984).

Kenward (1985) introduced a similar technique by fitting orthogonal polynomials and he treated the estimated coefficients separately. The selection was made on the basis of the least estimated variance each estimated coefficient presented (Least Possible Variance (LPV) from now on), using the covariance matrices from the elements of the

set described in (4.28). Working in that way however, different covariates can be used for different coefficients. Using simulation, Kenward pointed out that the variance of the estimated coefficients estimated in that way cannot be much better than the variance of the corresponding ordinary least squares coefficients something which was quite disappointing.

One could claim that both techniques will not be very efficient since they are not based on any probabilistic argument. For example, we may have that a certain number, say l, of covariates give the least possible generalized variance, say z, and that there exists a combination of $l-1$ covariates with the corresponding determinant of the estimated variance as $z+\epsilon$ where ϵ is a positive very small real number. If this is the case, then the method used by Grizzle and Allen would choose as the best combination the combination of l covariates. This may be a mistake, however, since it is natural for the experimenter to include as few covariates as possible in the model. This also can result in a gain of efficiency of the estimator. Recently, Fujikoshi and Rao (1991) developed one procedure (F.R from now on) based on ideas of Laha (1954), Siotani (1957), Mc Kay (1977) and Fujikoshi (1982).

In order to describe this approach, we shall assume that H_2 defined in (4.3) is written as $H_2=(H_{21},H_{22})$ where H_{21}, is a $(q\times c)$ and H_{22} is a $(q\times(q-p-c))$ where c is the number of covariates used in fitting the covariance adjusted estimator. Without loss of generality, we shall assume that H_{21} is the part of the transformation matrix which corresponds to the covariates fitted to the model. Consequently, the transformed data matrices (Y_1,Y_2) will now be written as (Y_1,Y_{21},Y_{22}) and will be defined as

$$Y_1=YH_1,\quad Y_{21}=YH_{21}\quad Y_{22}=YH_{22}. \tag{4.31}$$

Their joint distribution will be normal with mean $(A\xi,0,0)$ and covariance matrix

$$\begin{pmatrix}\Sigma_{yy} & \Sigma_{yh}\\ \Sigma_{hy} & \Sigma_{hh}\end{pmatrix}=\begin{pmatrix}\Sigma_{yy} & \Sigma_{y1} & \Sigma_{y2}\\ \Sigma_{1y} & \Sigma_{11} & \Sigma_{12}\\ \Sigma_{2y} & \Sigma_{21} & \Sigma_{22}\end{pmatrix}=\begin{pmatrix}H_1^T\\ H_{21}^T\\ H_{22}^T\end{pmatrix}\Sigma(H_1,H_{21},H_{22}). \tag{4.32}$$

It is understood from the above notation that Σ_{yy} corresponds to the covariance matrix of Y_1 while Σ_{hh} corresponds to the covariance matrix of Y_2. The latter can be partitioned into submatrices due to the partition of Y_2 into Y_{21} and Y_{22}.

It is now known that all the correlation between two (correlated) random variables X and Y can be expressed in terms of the squared canonical correlations

$$\rho_j^2 = \lambda_j(\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}),$$

where $\lambda_j(A)$ is the jth largest characteristic root of matrix A. A univariate measure, taking into account all the squared canonical correlations, is

$$tr\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}.$$

Thus, Fujikoshi and Rao define their hypothesis as

$$tr\Sigma_{yy}^{-1}\Sigma_{yh}\Sigma_{hh}^{-1}\Sigma_{hy} = tr\Sigma_{yy}^{-1}\Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{1y}.$$

If the above relationship is (statistically) correct, this would mean that the correlation between Y_1 and Y_2 can be summarized by the correlation between Y_1 and Y_{21} and thus, the covariates given by H_{21} are adequate for fitting the covariance adjusted estimator. If we now use a result concerning the inverse of a partitioned matrix (found in the appendix) we get that

$$tr\Sigma_{yy}^{-1}\Sigma_{yh}\Sigma_{hh}^{-1}\Sigma_{hy} = tr\Sigma_{yy}^{-1}\Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{1y}+$$

$$+tr\Sigma_{yy}^{-1}(\Sigma_{y2} - \Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{12})E^{-1}(\Sigma_{2y} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{1y}),$$

where $E = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Thus, the hypothesis above is equivalent with the hypothesis that

$$\Sigma_{y2} - \Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{12} = 0.$$

The authors showed that the likelihood ratio criterion for testing this hypothesis is

$$\lambda^{\frac{2}{n}} = \frac{\begin{vmatrix} W_{yy.1} & W_{y2.1} \\ W_{2y.1} & W_{22.1} \end{vmatrix}}{|W_{yy.1}||W_{22.1}|}$$

where

$$
\begin{aligned}
W_{yy.1} &= H_1^T S H_1 - H_1^T S H_{21} (H_{21}^T S H_{21})^{-1} H_{21}^T S H_1 \\
W_{y2.1} &= H_1^T S H_{22} - H_1^T S H_{21} (H_{21}^T S H_{21})^{-1} H_{21}^T S H_{22} \\
W_{2y.1} &= H_{22}^T S H_1 - H_{22}^T S H_{21} (H_{21}^T S H_{21})^{-1} H_{21}^T S H_1 \\
W_{22.1} &= H_{22}^T S H_{22} - H_{22}^T S H_{21} (H_{21}^T S H_{21})^{-1} H_{21}^T S H_{22}.
\end{aligned}
$$

As a consequence, note that $\lambda^{\frac{2}{n}}$ is invariant under multiplication of S by any constant factor.

The hypothesis of fitting a specific number of covariates can be tested by the fact that $\lambda^{\frac{2}{n}}$ follows asymptotically a χ^2 distribution. However, an alternative and more automatic way can be the use of the Akaike criterion which involves the minimization of the quantity

$$
-n \log(\lambda^{\frac{2}{n}}) - 2p(q-p-c) \tag{4.33}
$$

The combination of covariates which gives the minimum of (4.33) provides also the best solution to the problem.

4.3.2 A new method

From the previous section, we know that the estimated covariance matrix of the estimated covariance adjustment estimator involves the expression:

$$
(PP^T)^{-1} P(S - SV_1(V_1^T S V_1)^{-1} V_1^T S) P^T (PP^T)^{-1}, \tag{4.34}
$$

where $S = Y^T(I_n - A(A^TA)^{-1}A^T)Y$ and V_1 is the matrix which contains the covariates used in fitting the estimator. It is easily seen from the discussion after (4.24) that

$$
(PP^T)^{-1} P(S - SV_1(V_1^T S V_1)^{-1} V_1^T S) P^T (PP^T)^{-1} \sim
$$

$$
W_p\left((PP^T)^{-1} P(\Sigma - \Sigma V_1(V_1^T \Sigma V_1)^{-1} V_1^T \Sigma) P^T (PP^T)^{-1}, n-m-c\right),
$$

where c is the number of covariances fitted.

The philosophy of the new test will be to show that expression (4.34) above is essentially the one derived when we fit all possible covariates (substituting in (4.34), V instead of V_1), since

$$(PP^T)^{-1}P(\Sigma - \Sigma V(V^T\Sigma V)^{-1}V^T\Sigma)P^T(PP^T)^{-1}$$

$$< (PP^T)^{-1}P(\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}.$$

In this way, we hope that instead of using the right hand expression in the above inequality, for the calculation of the variance of the covariance adjusted estimator, we actually use the left hand side which is smaller.

The likelihood ratio test that the above inequality is essentially equality is, by denoting $(PP^T)^{-1}P(\Sigma - \Sigma V_1(V_1^T\Sigma V_1)^{-1}V_1^T\Sigma)P^T(PP^T)^{-1}$ by $\Phi_{V_1}(\Sigma)$:

$$\frac{\dfrac{|\Phi_{V_1}(S)|^{(n-m-c-p-1)/2}\exp -\frac{1}{2}tr\Phi_{V_1}(\Sigma)^{-1}\Phi_{V_1}(S)}{2^{(n-m-c)p/2}|\Phi_{V_1}(\Sigma)|^{(n-m-c)/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-m-c+1-i)\right)}}{\dfrac{|\Phi_{V}(S)|^{(n-m-q-1)/2}\exp -\frac{1}{2}tr\Phi_{V}(\Sigma)^{-1}\Phi_{V}(S)}{2^{(n-m-q+p)p/2}|\Phi_{V}(\Sigma)|^{(n-m-q+p)/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-m-q+p+1-i)\right)}}. \tag{4.35}$$

It is not difficult to prove (see Mardia, Kent and Bibby, 1982) that optimization in both sides of the above fraction appears when $\hat{\Sigma} = S$. The likelihood ratio test then becomes after some operations and using (4.26):

$$l^{2/(p+1)} = \frac{k^{2/(p+1)}|P(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^T|}{|PS^{-1}P^T|}$$

with

$$k = 2^{-(q-p-c)p/2}\frac{\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-m-q+p+1-i)\right)}{\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-m-c+1-i)\right)}.$$

Each combination of covariates can now be tested against the fitting of all possible covariates. Trying to avoid repeated comparisons with corresponding χ^2 values, we can use an AIC-type criterion defined by

$$-(p+1)\log l^{2/(p+1)} - \frac{1}{2}(q-p-c)$$

which compensates for small c's, that is for small number of covariates. Note that the penalty coefficient is different than that of Fujikoshi and Rao. This is done in an attempt to compromise between the observed underfitting when using this criterion with

penalty value 2 and a possible overfitting with smaller penalty values. Unfortunately, we have not been able to find a theoretical explanation for this choice something which, undoubtedly, suggests further research. In the next subsection a simulation study will clarify the performance of the three methods described in subsection 4.3.1, and the one suggested above (which will be called 'Alt.').

4.3.3 Algorithm, simulation scheme and results

A program was written in FORTRAN for the comparison of the four methods described in the two previous subsections. Measurements were derived as pseudo-random normal variables with covariance matrix Σ chosen to be of a change point type as described in Chapter 2 with one change at the fourth time-point. We assumed six different set-ups for the autoregressive parameters ϕ_1 and ϕ_2 as follows:

a) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.6$ and $\phi_2 = 0.2$

b) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.9$ and $\phi_2 = 0.2$

c) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 0.9$ and $\phi_2 = 0.5$

d) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.5$ and $\phi_2 = 1.1$

e) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.9$ and $\phi_2 = 0.8$

f) $\tau^2 = 2.0$, $\nu^2 = 0.5$, $\sigma^2 = 1.5$, $\phi_1 = 1.9$ and $\phi_2 = 1.1$

For each set-up, we assumed cases for which $q = 7$ and $q = 11$ and for each of them, cases where $n = 12$, $n = 20$ and $n = 50$ ($n = 20$ and $n = 50$ when $q = 11$). All these 'individuals' were considered to be assigned to one treatment group ($m = 1$). The mean trend was allowed to be a quadratic polynomial with $\xi_0 = 20.0$, $\xi_1 = 10.0$ and $\xi_2 = 3.0$. Finally, for each combination of n, q and Σ, 100 simulations were run.

For each simulation, unstandardized orthogonal polynomials (i.e. $PP^T = C$, $P_0P_0^T =$

C_0 with C and C_0 being diagonal matrices) were fitted. Consequently, the covariates constructed using (3.16) became

$$V = P_0^T C_0^{-1/2}$$

Before proceeding to the analysis of calculations and the presentation of results we can have an idea about how the F.R, Alt. and LGV criteria choose their best combination set. Figures 4.1 and 4.2 show some typical AIC, AIC-type and generalized variance functions for covariance settings and for different sample sizes. Here, we had $q = 7$ dependent variables and the covariance matrix of observations was taken to have two forms The first was as described in a), (figure 4.1), and the second as in e), (figure 4.2), above. We fitted a second degree orthogonal polynomial ($p = 3$). Consequently, the number of covariates available was 4 (cubic, quartic, quintic and sextic polynomials) and numbers 1-15 correspond to all possible combinations of covariates as given by the following table:

1	$\{1\}$
2	$\{2\}$
3	$\{3\}$
4	$\{4\}$
5	$\{1,2\}$
6	$\{1,3\}$
7	$\{1,4\}$
8	$\{2,3\}$
9	$\{2,4\}$
10	$\{3,4\}$
11	$\{1,2,3\}$
12	$\{1,2,4\}$
13	$\{1,3,4\}$
14	$\{2,3,4\}$
15	$\{1,2,3,4\}$

Both figures should be read rotated clock-wise. In this way, the first row presents typical Akaike functions for F.R ($n = 12, 20, 50$), the second row presents typical

Akaike-type functions for Alt., ($n = 12, 20, 50$) and the third row presends typical generalized variance functions for LGV ($n = 12, 20, 50$). Examination of Figure 4.1 first, shows that F.R and Alt. have the same performance when $n = 12$ and not a very different one in larger samples. It is worth noting that both criteria have a similarity in the covariates they finally choose in all sample sizes. F.R chooses $\{2, 4\}$, $\{1, 4\}$ and $\{4\}$ when $n = 12$, $n = 20$ and $n = 50$ respectively. Quite similarly, Alt. chooses for the same sample size order $\{2, 4\}$, $\{1\}$ and $\{4\}$. LGV chooses $\{4\}$, $\{1\}$ and $\{2\}$. In figure 4.2 we note that Alt. tends as the sample size increases, to favour the choice of few covariates while AIC despite the similarity in small samples tends to choose larger sets when $n = 50$ in order to summarize the greatest part of information. Generalized variances in figure 4.2, tend to concentrate in areas around 0, making the final choice a more difficult task.

The computations in the program involved calculation of the covariance adjusted estimator under all possible combinations of covariates and determination of the optimal combination set under each of the four methods compared. Since however, we noted in Section 4.2 that we could have essentially two estimates of variance of the adjusted estimator and since LPV and LGV both involve minimization of these variances, it is obvious that the optimum combination set determined by using each of these methods depends upon which estimate of variance (Ψ_1 or Ψ_2) we use. Denoting now these optimal sets by OPT_{Ψ_1} and OPT_{Ψ_2} we can immediately see that the real variances of the estimator will in general be different since they depend on OPT_{Ψ_1} and OPT_{Ψ_2}. These are denoted by R_{Ψ_1} and R_{Ψ_2}. This remark does not hold for F.R and Alt. since in the calculations, the criteria do not depend on any constant (see also the note above (4.33)). Thus, $OPT_{\Psi_1} = OPT_{\Psi_2}$ in this case, and the real variance of the estimator is the same whether we use as an estimate of variance Ψ_1 or Ψ_2. Tables 4.1–4.12 contain this information in the following form: The first row presents the mean estimated (unbiased) variances Ψ_1. The other three rows marked by A,B,C correspond to three ratios: $A = \Psi_1/R_{\Psi_1}$, $B = \Psi_2/R_{\Psi_2}$ and $C = \Psi_1/\Psi_2$. A and B give the degree

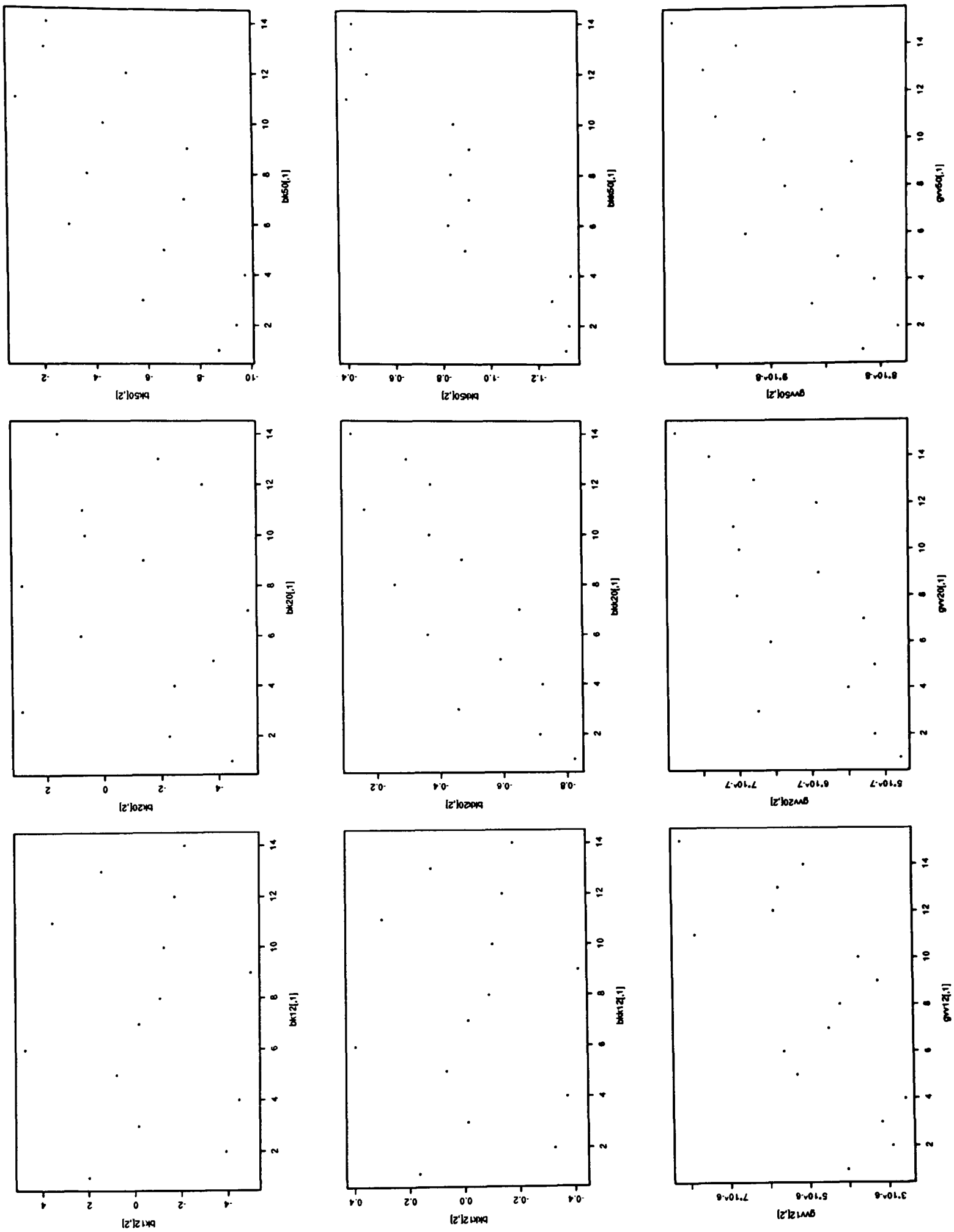


Figure 4.1: Typical values of F.R, Alt. and LGV for three possible sample sizes, when all possible combinations of covariates have been fitted. The numbers of x-axes correspond to every possible combination of covariates as described in the table of p. 101. By rotating the page 90 degrees clock-wise and from left to right: first row: F.R ($n = 12, 20, 50$); second row: Alt. ($n = 12, 20, 50$); third row: LGV. ($n = 12, 20, 50$).)

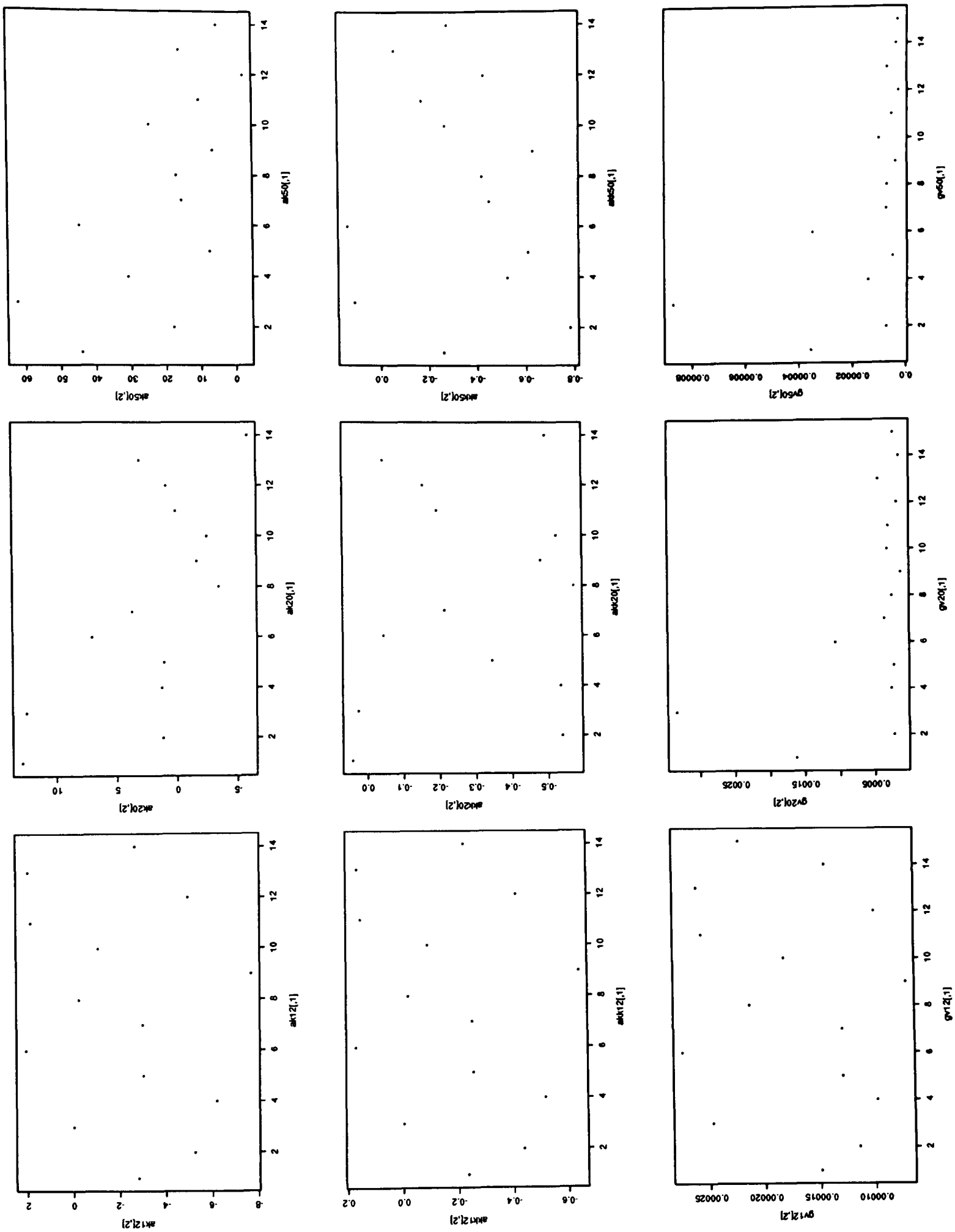


Figure 4.2: Typical values of F.R, Alt. and LGV for three possible sample sizes, when all possible combinations of covariates have been fitted. The numbers of x-axes correspond to every possible combination of covariates as described in the table p. 101. By rotating the page 90 degrees clock-wise and from left to right: first row: F.R ($n = 12, 20, 50$); second row: Alt. ($n = 12, 20, 50$); third row: LGV. ($n = 12, 20, 50$).)

to which the estimated variances Ψ_1 and Ψ_2 underestimate the real variances R_{Ψ_1} and R_{Ψ_2} respectively. C shows how much, estimates Ψ_1 and Ψ_2 differ. The latter will always be greater than 1. The same scheme is repeated for each combination of sample size and dependent variables and for the linear and quadratic coefficients. Under the column marked by 'Gauss–Markov', the theoretical Cramer-Rao bound is presented. As regards the results, the ratios A,B,C are very similar for linear and quadratic coefficients apart the cases where both autoregressive coefficients ϕ_1 and ϕ_2 are greater than 1, something which is related to the way that the variances of observations change with time. These ratios however, change systematically with the increase of sample size reaching gradually values nearer to 1. The results show clearly that the real variance is always or almost always underestimated whether we use Ψ_1 or Ψ_2. Ψ_1 as an unbiased estimate, shows a better degree of approximation something, which is especially clear in small samples ($n = 12$) or even moderate ($n = 20$). When sample size increases, A and B tend to the same value and near to 1 (see also the discussion in the last paragraph of section 4.2). But most important are the differences between the ratios A and B for the four compared methods. As a summary of these differences, Alt. in general gives greater values to A and B and nearer to 1 than all others, especially when $q = 11$, with F.R always very close. This means that Alt. gives estimators whose estimated variance lies nearer its true value than other methods. The differences with the other methods are more acute for B and when sample size is small. LPV method used by Kenward gives substantially lower values and in general the lowest, which show, as Kenward (1985) had pointed out, that LPV cannot, in general, be recommended for the optimal selection of covariates. LGV now has a better performance than LPV and indeed in many cases performs equally well to F.R. and Alt.. When however, the series becomes larger ($q = 11$) LGV can be compared only with F.R..

As regards the approximation to the optimal variance, it is clear that LPV is, in most of the cases, nearer to it, while differences are not so strong as sample size increases. However, one should note that in the covariance set-ups a), b) and c), which

represent milder deviations from stationarity, LPV gives lower estimated variances than the optimal ones, in quadratic coefficients, while in set-ups d), e) and f), in almost all cases, it underestimates the 'Gauss–Markov' values. In contrast with LPV's performance, LGV, F.R and Alt. in general overestimate the optimal variances while OLS can be very close to them (in set–ups a), b) and c)) or extremely inefficient in set–ups d), e) and f). Finally, a similar picture turns up when $q = 11$. This picture certifies that Alt. works really well in many covariance set-ups and in many cases it gives estimated variances which are smaller than those of F.R approximating better the Gauss–Markov value. Because of the latter it seems that Alt. and F.R. (and LGV to some extent), keep a very good balance between the estimated variance and the real variance of the regression coefficients and the optimal 'Gauss–Markov' value. LPV finally, performs relatively poorly and in small samples may give very disappointing results.

4.4 Conclusions

In this chapter we presented another very common approach to the growth curves problem. Given that the estimate of the mean parameters ξ can be considered as a GLS estimator, we proved its form by deriving its weight covariance matrix. As a consequence, it was shown that the optimally chosen estimator would use a weight from a suitably defined set of covariance matrices. Another problem which is essential to the experimenter, was also considered. That is, which is the best method for the optimal choice of the estimator. One new method was suggested and was compared with three already known methods. The comparison was accomplished with respect to the distance of the estimated variance of the estimator from its real value and from its optimal value. It was shown that the proposed method has in general the best ratios A, B and C (especially when $q = 11$) and it also gives good estimates of variance for the estimated coefficients of the mean. This is desirable since values of A and B near to one show an estimate of variance which is near its true value. It was also shown that

LPV has considerably smaller ratios A and B than those of LGV, F.R and Alt. and despite the fact that most of the time its estimated variances approximated well the Gauss-Markov variance, it cannot be recommended as a good method for the choice of covariates. On the other hand, LGV and F.R while overestimating the Gauss–Markov variance, showed, and especially the F.R, a good degree of approximation to their real variance and thus, they could also be a possible choice along with Alt.. Consequently, Alt. and F.R, perform really well and if the experimenter would like to make a better approximation to the optimal variance, it is recommended to use Ψ_2 instead of Ψ_1 which is always smaller and thus, constitutes an acceptable compromise.

There are also other methods which could be considered for the problem of selection of the covariance adjustment estimator of less parametric type. These could be variations of the methods described in Lee (1991) and suggest an interesting subject of further research.

Table 4.1: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.6$ $\phi_2 = 0.2$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.2494		0.2659	0.2402	0.2586	0.2712	0.2712
			A	(1.01)	(0.79)	(0.87)	(0.85)	(0.85)
			B	-	(0.66)	(0.68)	(0.72)	(0.72)
			C	-	(1.14)	(1.14)	(1.18)	(1.18)
	n=20	0.1497		0.1561	0.1505	0.1549	0.1561	0.1556
			A	(0.99)	(0.90)	(0.92)	(0.92)	(0.93)
q=7			B	-	(0.82)	(0.84)	(0.86)	(0.88)
			C	-	(1.07)	(1.07)	(1.07)	(1.06)
	n=50	0.0599		0.0634	0.0623	0.0634	0.0635	0.0635
			A	(1.01)	(0.97)	(0.99)	(0.99)	(0.99)
			B	-	(0.94)	(0.94)	(0.96)	(0.97)
			C	-	(1.03)	(1.03)	(1.03)	(1.02)
Quadratic								
	n=12	3.85-3		3.95-3	3.52-3	3.83-3	4.02-3	4.02-3
			A	(1.01)	(0.78)	(0.87)	(0.85)	(0.85)
			B	-	(0.65)	(0.67)	(0.72)	(0.72)
			C	-	(1.15)	(1.14)	(1.18)	(1.18)
	n=20	2.31-3		2.32-3	2.22-3	2.30-3	2.32-3	2.31-3
			A	(0.99)	(0.89)	(0.92)	(0.92)	(0.93)
q=7			B	-	(0.82)	(0.84)	(0.86)	(0.88)
			C	-	(1.07)	(1.07)	(1.07)	(1.06)
	n=50	0.92-3		0.94-3	0.92-3	0.94-3	0.94-3	0.94-3
			A	(1.01)	(0.96)	(0.99)	(0.99)	(0.99)
			B	-	(0.93)	(0.94)	(0.96)	(0.97)
			C	-	(1.03)	(1.03)	(1.03)	(1.02)

Table 4.2: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.9$ $\phi_2 = 0.2$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.2636		0.3199	0.2694	0.2955	0.3126	0.3126
			A	(1.04)	(0.77)	(0.86)	(0.84)	(0.84)
			B	-	(0.63)	(0.66)	(0.68)	(0.68)
			C	-	(1.16)	(1.18)	(1.23)	(1.23)
	n=20	0.1581		0.1900	0.1723	0.1829	0.1853	0.1824
			A	(1.03)	(0.89)	(0.94)	(0.94)	(0.95)
q=7			B	-	(0.80)	(0.82)	(0.86)	(0.89)
			C	-	(1.09)	(1.11)	(1.09)	(1.06)
	n=50	0.0633		0.0748	0.0701	0.0712	0.0715	0.0729
			A	(1.01)	(0.95)	(0.96)	(0.97)	(0.99)
			B	-	(0.92)	(0.92)	(0.94)	(0.97)
			C	-	(1.03)	(1.04)	(1.03)	(1.02)
Quadratic								
	n=12	4.30-3		4.76-3	3.94-3	4.39-3	4.64-3	4.64-3
			A	(1.04)	(0.76)	(0.86)	(0.84)	(0.84)
			B	-	(0.62)	(0.66)	(0.68)	(0.68)
			C	-	(1.16)	(1.18)	(1.23)	(1.23)
	n=20	2.58-3		2.84-3	2.56-3	2.73-3	2.77-3	2.72-3
			A	(1.03)	(0.88)	(0.94)	(0.94)	(0.95)
q=7			B	-	(0.79)	(0.82)	(0.86)	(0.89)
			C	-	(1.09)	(1.11)	(1.10)	(1.06)
	n=50	1.03-3		1.11-3	1.04-3	1.06-3	1.07-3	1.09-3
			A	(1.01)	(0.95)	(0.96)	(0.97)	(0.99)
			B	-	(0.91)	(0.92)	(0.94)	(0.97)
			C	-	(1.03)	(1.04)	(1.03)	(1.02)

Table 4.3: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.9$ $\phi_2 = 0.5$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.2889		0.3222	0.2917	0.3118	0.3314	0.3314
			A	(1.00)	(0.80)	(0.86)	(0.87)	(0.87)
			B	-	(0.66)	(0.70)	(0.74)	(0.74)
			C	-	(1.15)	(1.13)	(1.18)	(1.18)
	n=20	0.1733		0.1964	0.1814	0.1886	0.1904	0.1913
			A	(1.02)	(0.89)	(0.93)	(0.92)	(0.95)
q=7			B	-	(0.80)	(0.81)	(0.85)	(0.89)
			C	-	(1.08)	(1.09)	(1.09)	(1.06)
	n=50	0.0693		0.0783	0.0752	0.0766	0.0767	0.0773
			A	(1.02)	(0.97)	(0.98)	(0.98)	(1.00)
			B	-	(0.93)	(0.94)	(0.96)	(0.98)
			C	-	(1.03)	(1.04)	(1.03)	(1.02)
Quadratic								
	n=12	4.61-3		4.80-3	4.28-3	4.62-3	4.92-3	4.92-3
			A	(1.00)	(0.78)	(0.86)	(0.87)	(0.87)
			B	-	(0.65)	(0.69)	(0.73)	(0.73)
			C	-	(1.15)	(1.13)	(1.18)	(1.18)
	n=20	2.76-3		2.94-3	2.70-3	2.82-3	2.85-3	2.86-3
			A	(1.02)	(0.88)	(0.93)	(0.92)	(0.95)
q=7			B	-	(0.80)	(0.81)	(0.85)	(0.90)
			C	-	(1.08)	(1.09)	(1.09)	(1.06)
	n=50	1.11-3		1.17-3	1.12-3	1.14-3	1.15-3	1.15-3
			A	(1.02)	(0.97)	(0.98)	(0.99)	(1.00)
			B	-	(0.93)	(0.94)	(0.96)	(0.98)
			C	-	(1.03)	(1.04)	(1.03)	(1.02)

Table 4.4: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.5$ $\phi_2 = 1.1$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.4442		0.3949	0.3770	0.3821	0.4071	0.4071
			A	(1.00)	(0.85)	(0.87)	(0.88)	(0.88)
			B	-	(0.71)	(0.70)	(0.75)	(0.75)
			C	-	(1.14)	(1.16)	(1.18)	(1.18)
	n=20	0.2665		0.2309	0.2245	0.2250	0.2294	0.2289
			A	(0.97)	(0.90)	(0.91)	(0.91)	(0.93)
q=7			B	-	(0.83)	(0.81)	(0.84)	(0.87)
			C	-	(1.08)	(1.09)	(1.09)	(1.06)
	n=50	0.1066		0.0965	0.0911	0.0912	0.0914	0.0924
			A	(1.01)	(0.97)	(0.97)	(0.97)	(0.98)
			B	-	(0.93)	(0.92)	(0.94)	(0.96)
			C	-	(1.04)	(1.04)	(1.03)	(1.02)
Quadratic								
	n=12	4.06-3		4.33-3	3.87-3	4.30-3	4.51-3	4.51-3
			A	(1.00)	(0.80)	(0.89)	(0.89)	(0.89)
			B	-	(0.66)	(0.72)	(0.75)	(0.75)
			C	-	(1.15)	(1.16)	(1.18)	(1.18)
	n=20	2.44-3		2.48-3	2.36-3	2.49-3	2.51-3	2.51-3
			A	(0.96)	(0.86)	(0.91)	(0.91)	(0.93)
q=7			B	-	(0.79)	(0.81)	(0.84)	(0.87)
			C	-	(1.08)	(1.09)	(1.09)	(1.06)
	n=50	0.98-3		1.06-3	0.99-3	1.01-3	1.01-3	1.02-3
			A	(1.02)	(0.95)	(0.98)	(0.97)	(0.98)
			B	-	(0.91)	(0.93)	(0.94)	(0.96)
			C	-	(1.04)	(1.04)	(1.03)	(1.02)

Table 4.5: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.9$ $\phi_2 = 0.8$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.6575		2.0015	0.5996	0.6367	0.6732	0.6732
			A	(1.01)	(0.67)	(0.72)	(0.75)	(0.75)
			B	-	(0.55)	(0.58)	(0.59)	(0.59)
			C	-	(1.24)	(1.27)	(1.28)	(1.28)
	n=20	0.3945		1.1924	0.3583	0.3687	0.3751	0.4015
			A	(1.00)	(0.81)	(0.83)	(0.83)	(0.81)
q=7			B	-	(0.72)	(0.73)	(0.73)	(0.74)
			C	-	(1.16)	(1.17)	(1.14)	(1.08)
	n=50	0.1578		0.4827	0.1379	0.1385	0.1398	0.1770
			A	(1.01)	(0.94)	(0.95)	(0.94)	(0.95)
			B	-	(0.89)	(0.90)	(0.89)	(0.93)
			C	-	(1.07)	(1.07)	(1.06)	(1.02)
Quadratic								
	n=12	8.20-3		30.20-3	8.83-3	9.47-3	10.01-3	10.01-3
			A	(1.01)	(0.66)	(0.72)	(0.75)	(0.75)
			B	-	(0.55)	(0.58)	(0.58)	(0.58)
			C	-	(1.24)	(1.28)	(1.28)	(1.28)
	n=20	4.92-3		18.02-3	5.31-3	5.49-3	5.58-3	5.99-3
			A	(1.00)	(0.81)	(0.83)	(0.83)	(0.80)
q=7			B	-	(0.72)	(0.73)	(0.73)	(0.74)
			C	-	(1.16)	(1.17)	(1.14)	(1.04)
	n=50	1.97-3		7.29-3	2.05-3	2.06-3	2.08-3	2.64-3
			A	(1.01)	(0.94)	(0.95)	(0.94)	(0.95)
			B	-	(0.89)	(0.90)	(0.89)	(0.93)
			C	-	(1.06)	(1.07)	(1.06)	(1.02)

Table 4.6: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.9$ $\phi_2 = 1.1$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=12	0.5603		0.8330	0.5366	0.5120	0.5466	0.5466
			A	(1.01)	(0.79)	(0.79)	(0.84)	(0.84)
			B	-	(0.66)	(0.62)	(0.65)	(0.65)
			C	-	(1.21)	(1.26)	(1.28)	(1.28)
	n=20	0.3362		0.5145	0.3195	0.3039	0.3091	0.3144
			A	(1.04)	(0.87)	(0.91)	(0.92)	(0.88)
q=7			B	-	(0.79)	(0.79)	(0.80)	(0.81)
			C	-	(1.15)	(1.16)	(1.14)	(1.08)
	n=50	0.1345		0.2017	0.1159	0.1119	0.1123	0.1335
			A	(1.02)	(0.96)	(1.00)	(0.98)	(0.94)
			B	-	(0.91)	(0.94)	(0.93)	(0.92)
			C	-	(1.08)	(1.07)	(1.06)	(1.02)
Quadratic								
	n=12	4.16-3		6.71-3	4.40-3	5.35-3	5.61-3	5.61-3
			A	(1.04)	(0.75)	(0.92)	(0.94)	(0.94)
			B	-	(0.64)	(0.70)	(0.73)	(0.73)
			C	-	(1.16)	(1.26)	(1.29)	(1.29)
	n=20	2.50-3		4.14-3	2.83-3	3.10-3	3.13-3	3.04-3
			A	(1.07)	(0.88)	(1.00)	(1.01)	(0.96)
q=7			B	-	(0.80)	(0.86)	(0.88)	(0.89)
			C	-	(1.11)	(1.15)	(1.14)	(1.08)
	n=50	1.00-3		1.58-3	1.05-3	1.08-3	1.07-3	1.19-3
			A	(1.02)	(0.93)	(1.00)	(0.98)	(0.97)
			B	-	(0.90)	(0.93)	(0.93)	(0.95)
			C	-	(1.05)	(1.07)	(1.06)	(1.02)

Table 4.7: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.6$ $\phi_2 = 0.2$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.0350		0.0355	0.0318	0.0355	0.0367	0.0354
			A	(1.01)	(0.82)	(0.91)	(0.89)	(0.94)
			B	-	(0.65)	(0.64)	(0.74)	(0.88)
			C	-	(1.17)	(1.16)	(1.20)	(1.07)
q=11								
	n=50	0.0140		0.0137	0.0132	0.0135	0.0136	0.0136
			A	(0.97)	(0.91)	(0.93)	(0.93)	(0.95)
			B	-	(0.85)	(0.84)	(0.90)	(0.93)
			C	-	(1.05)	(1.08)	(1.04)	(1.02)
Quadratic								
	n=20	2.26-4		2.29-4	2.02-4	2.30-4	2.38-4	2.29-4
			A	(1.00)	(0.79)	(0.91)	(0.87)	(0.94)
			B	-	(0.64)	(0.64)	(0.73)	(0.88)
			C	-	(1.14)	(1.09)	(1.18)	(1.07)
q=11								
	n=50	0.91-4		0.89-4	0.85-4	0.88-4	0.88-4	0.89-4
			A	(0.97)	(0.90)	(0.93)	(0.93)	(0.95)
			B	-	(0.85)	(0.84)	(0.90)	(0.93)
			C	-	(1.04)	(1.02)	(1.04)	(1.02)

Table 4.8: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.9$ $\phi_2 = 0.2$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.0338		0.0346	0.0299	0.0329	0.0352	0.0333
			A	(0.99)	(0.77)	(0.84)	(0.86)	(0.88)
			B	-	(0.61)	(0.60)	(0.69)	(0.82)
			C	-	(1.18)	(1.17)	(1.24)	(1.08)
q=11								
	n=50	0.0135		0.0136	0.0130	0.0135	0.0136	0.0136
			A	(0.97)	(0.90)	(0.94)	(0.94)	(0.96)
			B	-	(0.84)	(0.83)	(0.89)	(0.94)
			C	-	(1.08)	(1.12)	(1.06)	(1.02)
Quadratic								
	n=20	2.21-4		2.19-4	1.86-4	2.10-4	2.25-4	2.12-4
			A	(0.99)	(0.74)	(0.84)	(0.83)	(0.88)
			B	-	(0.60)	(0.60)	(0.68)	(0.81)
			C	-	(1.14)	(1.08)	(1.21)	(1.08)
q=11								
	n=50	0.88-4		0.86-4	0.82-4	0.87-4	0.87-4	0.87-4
			A	(0.96)	(0.89)	(0.93)	(0.93)	(0.96)
			B	-	(0.83)	(0.83)	(0.89)	(0.94)
			C	-	(1.05)	(1.04)	(1.05)	(1.02)

Table 4.9: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 0.9$ $\phi_2 = 0.5$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.0394		0.0442	0.0389	0.0433	0.0455	0.0438
			A	(1.01)	(0.81)	(0.90)	(0.89)	(0.93)
			B	-	(0.65)	(0.64)	(0.73)	(0.86)
			C	-	(1.17)	(1.15)	(1.22)	(1.08)
q=11								
	n=50	0.0158		0.0175	0.0167	0.0174	0.0174	0.0175
			A	(1.00)	(0.93)	(0.97)	(0.97)	(0.99)
			B	-	(0.86)	(0.87)	(0.92)	(0.97)
			C	-	(1.08)	(1.11)	(1.06)	(1.02)
Quadratic								
	n=20	2.74-4		2.85-4	2.47-4	2.82-4	2.96-4	2.85-4
			A	(1.02)	(0.79)	(0.91)	(0.87)	(0.94)
			B	-	(0.64)	(0.64)	(0.73)	(0.87)
			C	-	(1.14)	(1.09)	(1.20)	(1.08)
q=11								
	n=50	1.09-4		1.12-4	1.07-4	1.13-4	1.13-4	1.13-4
			A	(1.00)	(0.93)	(0.97)	(0.97)	(1.00)
			B	-	(0.86)	(0.87)	(0.92)	(0.98)
			C	-	(1.05)	(1.04)	(1.05)	(1.02)

Table 4.10: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.5$ $\phi_2 = 1.1$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.1157		0.1094	0.1037	0.0988	0.1053	0.1032
			A	(0.96)	(0.88)	(0.84)	(0.96)	(0.87)
			B	-	(0.69)	(0.60)	(0.73)	(0.80)
			C	-	(1.36)	(2.02)	(1.33)	(1.09)
q=11								
	n=50	0.0463		0.0450	0.0426	0.0415	0.0422	0.0431
			A	(0.99)	(0.97)	(0.99)	(1.00)	(0.95)
			B	-	(0.86)	(0.84)	(0.88)	(0.83)
			C	-	(1.30)	(1.94)	(1.15)	(1.02)
Quadratic								
	n=20	3.76-4		3.87-4	3.26-4	3.78-4	3.85-4	3.72-4
			A	(0.95)	(0.73)	(0.84)	(0.82)	(0.86)
			B	-	(0.60)	(0.60)	(0.69)	(0.79)
			C	-	(1.14)	(1.12)	(1.18)	(1.08)
q=11								
	n=50	1.50-4		1.59-4	1.45-4	1.55-4	1.55-4	1.57-4
			A	(0.98)	(0.89)	(0.93)	(0.93)	(0.96)
			B	-	(0.83)	(0.84)	(0.88)	(0.94)
			C	-	(1.05)	(1.05)	(1.06)	(1.02)

Table 4.11: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.9$ $\phi_2 = 0.8$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.0935		0.3917	0.0915	0.0977	0.1053	0.1112
			A	(0.96)	(0.58)	(0.63)	(0.69)	(0.71)
			B	-	(0.46)	(0.55)	(0.54)	(0.63)
			C	-	(1.28)	(1.09)	(1.27)	(1.12)
q=11								
	n=50	0.0374		0.1628	0.0422	0.0429	0.0438	0.0569
			A	(0.99)	(0.87)	(0.89)	(0.90)	(0.98)
			B	-	(0.80)	(0.87)	(0.82)	(0.96)
			C	-	(1.13)	(1.04)	(1.10)	(1.02)
Quadratic								
	n=20	6.46-4		26.89-4	6.05-4	6.51-4	7.03-4	7.47-3
			A	(0.95)	(0.56)	(0.62)	(0.68)	(0.70)
			B	-	(0.45)	(0.55)	(0.54)	(0.63)
			C	-	(1.24)	(1.06)	(1.26)	(1.18)
q=11								
	n=50	2.58-4		11.18-4	2.83-4	2.88-4	2.94-4	3.85-4
			A	(0.99)	(0.86)	(0.88)	(0.89)	(0.98)
			B	-	(0.80)	(0.87)	(0.82)	(0.96)
			C	-	(1.11)	(1.01)	(1.09)	(1.02)

Table 4.12: Estimated variances Ψ_1 of linear and quadratic coefficients for the four methods; $\tau^2 = 2.0$ $\nu^2 = 0.5$ $\sigma^2 = 1.5$ $\phi_1 = 1.9$ $\phi_2 = 1.1$. In brackets: A: Ψ_1/R_{Ψ_1}, B: Ψ_2/R_{Ψ_2}, C: Ψ_1/Ψ_2. Expressions of the type E-F mean $E \times 10^{-F}$.

		Gauss–Markov		OLS	LPV	LGV	F.R	Alt.
Linear								
	n=20	0.1666		0.2716	0.1645	0.1431	0.1579	0.1565
			A	(0.99)	(0.87)	(0.85)	(1.16)	(0.79)
			B	-	(0.65)	(0.63)	(0.63)	(0.68)
			C	-	(1.63)	(2.31)	(1.84)	(1.17)
q=11								
	n=50	0.0666		0.1126	0.0613	0.0561	0.0573	0.0728
			A	(1.07)	(1.18)	(1.47)	(1.36)	(0.92)
			B	-	(0.83)	(0.85)	(0.84)	(0.90)
			C	-	(1.87)	(2.15)	(1.62)	(1.02)
Quadratic								
	n=20	4.26-4		7.34-4	4.15-4	5.07-4	5.17-4	5.11-4
			A	(1.00)	(0.68)	(0.85)	(0.82)	(0.83)
			B	-	(0.56)	(0.63)	(0.64)	(0.74)
			C	-	(1.20)	(1.19)	(1.28)	(1.12)
q=11								
	n=50	1.70-4		2.95-4	1.78-4	1.89-4	1.91-4	2.21-4
			A	(1.01)	(0.87)	(0.95)	(0.95)	(0.95)
			B	-	(0.81)	(0.85)	(0.86)	(0.93)
			C	-	(1.08)	(1.05)	(1.10)	(1.02)

Chapter 5

Bounds of efficiency under a misspecified matrix

5.1 Introduction

This final chapter is devoted to the derivation and investigation of bounds of efficiency for estimators of regression coefficients in GMANOVA models. The assumptions concerning the model will remain the same as those of chapter 2, that is the response matrix Y $(n \times q)$ will be written as

$$Y = A\xi P + E \tag{5.1}$$

where A $(n \times m)$ is a design matrix, ξ $(m \times p)$ is a matrix of unknown regression parameters and P $(p \times q)$ is a matrix of known coefficients (usually polynomials of order $p-1$). E $(n \times q)$ is a matrix of errors whose each row is normally distributed with mean 0 and covariance matrix Σ. As has been stressed previously, this matrix Σ is unknown in applications and partial knowledge or good estimation of it is essential for efficient inference about the other parameters of the linear system. In previous chapters we examined different methods of estimation of the covariance matrix of observations. We showed, for example, in Chapter 2 that likelihood based methods can be inefficient when we do not use the correct covariance model. We also showed that covariance adjustment may give estimators of the mean with different degrees of accuracy depending on the

method we use to determine the optimum set of covariates which will adjust the OLS estimator. In this chapter we go one step further by investigating theoretically these matters in a more general framework.

It is known that the minimum variance estimator (MV) of ξ has the form

$$\hat{\xi}_\Sigma = (A^T A)^{-1} A^T Y \Sigma^{-1} P^T (P\Sigma^{-1} P^T)^{-1} \tag{5.2}$$

and its covariance matrix is

$$Var(vec(\hat{\xi}_\Sigma)) = (P\Sigma^{-1} P^T)^{-1} \otimes (A^T A)^{-1} \tag{5.3}$$

Various methods of estimation provide us with an estimate of Σ, $\hat{\Sigma}$ and the estimated MV estimator is a GLS one with $\hat{\Sigma}^{-1}$ as weight matrix, that is,

$$\hat{\xi}_{\hat{\Sigma}} = (A^T A)^{-1} A^T Y \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} \tag{5.4}$$

In the same way as before the covariance matrix of this estimator is

$$Var(vec(\hat{\xi}_{\hat{\Sigma}})) = (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} \otimes (A^T A)^{-1} \tag{5.5}$$

An obvious estimate for $Var(\hat{\xi}_{\hat{\Sigma}})$ is obtained by setting in (5.5) $\Sigma = \hat{\Sigma}$, thus giving

$$\widehat{Var(vec}(\hat{\xi}_{\hat{\Sigma}})) = (P\hat{\Sigma}^{-1} P^T)^{-1} \otimes (A^T A)^{-1} \tag{5.6}$$

It is also known that $Var(\hat{\xi}_{\hat{\Sigma}}) \geq Var(\hat{\xi}_\Sigma)$ in the sense that their difference is a positive definite matrix. From this point of view, bounds of variances and of efficiency are useful because they provide a measure of the error caused by a misspecified covariance matrix.

There is a history of results which deal with bounds of the relative efficiency defined as the ratio of generalized variances of the MV and the OLS estimators. Here, this number will be represented by e' with

$$e' = \frac{|Var(\hat{\xi}_\Sigma)|}{|Var(\hat{\xi}_{I_q})|}$$

$$
\begin{aligned}
&= \frac{|(P\Sigma^{-1}P^T)^{-1} \otimes (A^TA)^{-1}|}{|(PP^T)^{-1}P\Sigma P^T(PP^T)^{-1} \otimes (A^TA)^{-1}|} \\
&= \frac{|(P\Sigma^{-1}P^T)^{-1}|^m}{|(PP^T)^{-2}|^m|P\Sigma P^T|^m} \\
&= \left(\frac{|(PP^T)|^2}{|(P\Sigma^{-1}P^T)||P\Sigma P^T|}\right)^m \qquad (5.7)
\end{aligned}
$$

since for example

$$|(P\Sigma^{-1}P^T)^{-1} \otimes (A^TA)^{-1}| = |(A^TA)^{-1}|^p|(P\Sigma^{-1}P^T)^{-1}|^m,$$

(see also in Appendix). The main results were derived using a version of the so-called Kantorovich inequality (1948) which states that

Let A be a positive definite $n \times n$ matrix with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Then

$$1 \leq (\boldsymbol{x}^TA\boldsymbol{x})(\boldsymbol{x}^TA^{-1}\boldsymbol{x}) \leq \frac{1}{4}\left[\left(\frac{\lambda_1}{\lambda_n}\right)^{1/2} + \left(\frac{\lambda_n}{\lambda_1}\right)^{1/2}\right]^2 \qquad (5.8)$$

for every $\boldsymbol{x} \in R^n$ such that $\boldsymbol{x}^T\boldsymbol{x} = 1$.

Assume that the covariance matrix Σ has eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_q$. Then J. Durbin conjectured that subject to $PP^T = I_p$,

$$e' \geq \left(\prod_{i=1}^{p} \frac{4\lambda_i\lambda_{q-i+1}}{(\lambda_i + \lambda_{q-i+1})^2}\right)^m \qquad (5.9)$$

Bloomfield and Watson (1975) and Knott (1975) gave independently the correct proof of this conjecture. The former authors also introduced another measure of relative performance of estimators. This is given by matrix C, also called commutator matrix defined as

$$C = P^TP\Sigma - \Sigma P^TP$$

and justified by the following result: If the Gauss–Markov estimator coincides with the OLS one, then it has been shown by many authors (see for example, Anderson, 1958) that

$$\Sigma P^T = P^TA$$

with A nonsingular. Thus, $P\Sigma P^T = PP^TA = A$ which means that A is also symmetric and in this case C is 0. The authors gave also upper bounds for trC^TC using (5.8).

If now, instead of the OLS estimator we consider any GLS estimator weighted by $\hat{\Sigma}^{-1}$, then the generalization arises as follows:

Since Σ and $\hat{\Sigma}^{-1}$ are positive definite matrices, they can be written as

$$\Sigma = \Gamma T \Gamma^T \text{ and } \hat{\Sigma}^{-1} = \Gamma\Gamma^T$$

where Γ is a nonsingular $q \times q$ matrix and T is a diagonal matrix with elements the eigenvalues of $\hat{\Sigma}\Sigma$. The relative efficiency e' is written as

$$e' = \left(\frac{|P\hat{\Sigma}^{-1}P^T|^2}{|P\Sigma^{-1}P^T||P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T|} \right)^m$$

Define now $Q = P\Gamma$. Then simple calculations leads us to

$$e' = \left(\frac{|QQ^T|^2}{|Q(\Gamma^T\Sigma\Gamma)^{-1}Q^T||Q(\Gamma^T\Sigma\Gamma)Q^T|} \right)^m$$

which is the same as in (5.7) and the lower bound has the same form as in (5.9) with λ's now being the eigenvalues of $\Gamma^T\Sigma\Gamma$. A similar approach was adopted by Golub (1963). A criticism has been made of this approach (see Strand, 1974), however, because the above mentioned bounds are functions of some eigenvalues of the unknown population covariance matrix Σ and they have to be estimated from the data if we want to have an idea of the error we commit in our inferences when we estimate this matrix. In the following sections a different approach will be presented which will lead to two new results concerning bounds of a relative efficiency.

5.2 Definitions and assumptions

In the following sections certain new results concerning the efficiency of GLS estimators will be presented. For the reasons explained in the previous section a different approach

will be considered here. We assume that we have an estimate $\hat{\Sigma}$ of Σ which is related to Σ with the following identity

$$\Sigma = \hat{\Sigma} + \Delta \tag{5.10}$$

Δ will be a $q \times q$ matrix, necessarily symmetric with Euclidean norm modelled by the following equation:

$$tr(\Sigma - \hat{\Sigma})^2 = tr\Delta^2 \leq \epsilon^2 \tag{5.11}$$

where ϵ is a real number sufficiently small so that every Σ satisfying (5.10) is a positive definite matrix (Strand, 1974). As a univariate measure of efficiency, traces and not generalized variances will be considered. The following two definitions of efficiency will be assumed to hold throughout this chapter:

$$e_1 = trVar(\hat{\xi}_{\hat{\Sigma}}) - trVar(\hat{\xi}_{\Sigma}) \tag{5.12}$$

$$e_2 = trVar(\hat{\xi}_{\hat{\Sigma}}) - \widehat{trVar(\hat{\xi}_{\hat{\Sigma}})} \tag{5.13}$$

e_1 is justified from the fact that it essentially is the mean square error derived from the difference $\hat{\xi}_{\hat{\Sigma}} - \hat{\xi}_{\Sigma}$ that is

$$Etr(\hat{\xi}_{\hat{\Sigma}} - \hat{\xi}_{\Sigma})^T(\hat{\xi}_{\hat{\Sigma}} - \hat{\xi}_{\Sigma})$$

(see, Strand, 1974). It provides also a measure of discrepancy from the optimum variance of a GLS estimator. e_2 on the other hand forms a measure of relative efficiency similar to e_1 but with $(P\hat{\Sigma}^{-1}P^T)^{-1}$ in the place of $(P\Sigma^{-1}P^T)^{-1}$ as an estimate of it.

In the following two sections, bounds for e_1 and e_2 will be presented. One approximate for e_1 and exact upper and lower bounds for e_2. Note that e_1 will always be greater than zero, while e_2 can be positive or negative depending on the data at hand or the form of the covariance matrix, something which we shall not consider here.

5.3 An approximate bound for e_1

The expression for e_1 as defined in (5.12) is, following the equations (5.3) and (5.5)

$$e_1 = tr(A^TA)^{-1}\left(tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} - tr(P\Sigma^{-1}P^T)^{-1}\right) \tag{5.14}$$

Strand (1974) derived an alternative expression for e_1 as it is defined in (5.14). He showed that this expression depends on two matrices P_1 and P_2

$$P_1 = \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-2}P\hat{\Sigma}^{-1} \quad \text{and} \tag{5.15}$$

$$P_2 = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1} \tag{5.16}$$

These $q \times q$ matrices apart from being symmetric, have other desirable properties which are easily proved:

- $P_2P^T = PP_2 = 0$
- $PP_1P^T = I_p$
- $r(P_1) = p$
- $P_2\hat{\Sigma}P_1 = P_1\hat{\Sigma}P_2 = 0$
- $P_2\hat{\Sigma}$ and $\hat{\Sigma}P_2$ are idempotent
- for the special case where $\hat{\Sigma} = I_q$, P_1 is also idempotent

The following lemmas will be used in obtaining a convenient expression for e_1.

Lemma 4 *For every square matrix R for which $I + R$ is nonsingular we have*

$$(I + R)^{-1} = \sum_{j=0}^{k}(-R)^j + T_k \tag{5.17}$$

where $k = 0, 1, 2, \ldots,$ and $T_k = (-R)^{k+1}(I + R)^{-1}$

Proof: The lemma holds for $k = 0$, that is

$$(I + R)^{-1} = I + (-R)(I + R)^{-1}$$

since by postmultiplication by $I + R$ we get

$$I + R - R = I$$

Let us assume that the lemma holds for k, that is,

$$(I + R)^{-1} = \sum_{j=0}^{k}(-R)^j + (-R)^{k+1}(I + R)^{-1}$$

We shall show that it also holds for $k + 1$. This latter is written as

$$\sum_{j=0}^{k+1}(-R)^j + (-R)^{k+2}(I + R)^{-1} = \sum_{j=0}^{k}(-R)^j + (-R)^{k+1} + (-R)^{k+2}(I + R)^{-1}$$

Postmultiplication by $I + R$ gives

$$\sum_{j=0}^{k}(-R)^j(I + R) + (-R)^{k+1}(I - (-R)) + (-R)^{k+2}$$

$$= \sum_{j=0}^{k}(-R)^j(I + R) + (-R)^{k+1}$$

which is I since the lemma holds for k.

Lemma 5 *Under assumption (5.10) and for P_2 defined in (5.16) we have*

$$(I + \hat{\Sigma}^{-1}\Delta)^{-1} = (I + P_2\Delta)^{-1}(I - B) \tag{5.18}$$

where $B = \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta(I + \hat{\Sigma}^{-1}\Delta)^{-1}$

Proof:

$$I + P_2\Delta = I + \hat{\Sigma}^{-1}\Delta - \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta$$

$$= (I - B)(I + \hat{\Sigma}^{-1}\Delta)$$

where

$$B = \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta(I + \hat{\Sigma}^{-1}\Delta)^{-1}$$

That shows that

$$I = (I + P_2\Delta)^{-1}(I - B)(I + \hat{\Sigma}^{-1}\Delta)$$

Thus,

$$(I + P_2\Delta)^{-1}(I - B) = (I + \hat{\Sigma}^{-1}\Delta)^{-1}$$

Theorem 12 *If* $\Sigma = \hat{\Sigma} + \Delta$ *then*

$$tr(A^TA)^{-1}\left(tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} - (P\Sigma^{-1}P^T)^{-1})\right)$$

$$\simeq tr(A^TA)^{-1}trP_1\Delta P_2\Delta \tag{5.19}$$

where

$$P_1 = \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-2}P\hat{\Sigma}^{-1} \quad \text{and} \tag{5.20}$$

$$P_2 = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1} \tag{5.21}$$

Proof [Strand, 1974]:

First, we shall give a convenient expression for $tr(P\Sigma^{-1}P^T)^{-1}$. It is known that the following identity holds provided that A, B and $A + B$ are nonsingular:

$$(A + B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1} \tag{5.22}$$

Since we know that $\Sigma = \hat{\Sigma} + \Delta$ we get

$$tr(P\Sigma^{-1}P^T)^{-1} = tr(P\hat{\Sigma}^{-1}P^T)^{-1}\left((I_q - P\hat{\Sigma}^{-1}\Delta(I + \hat{\Sigma}^{-1}\Delta)^{-1}\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}\right)^{-1} \tag{5.23}$$

By applying Lemma 4 into (5.23) with $k = 0$ and

$$-R \equiv A = P\hat{\Sigma}^{-1}\Delta(I + \hat{\Sigma}^{-1}\Delta)^{-1}\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \tag{5.24}$$

we get

$$tr(P\Sigma^{-1}P^T)^{-1} = tr(P\hat{\Sigma}^{-1}P^T)^{-1} + tr(P\hat{\Sigma}^{-1}P^T)^{-1}A(I - A)^{-1} \tag{5.25}$$

But from Lemma 5 the expression $(I + \hat{\Sigma}^{-1}\Delta)^{-1}$ in (5.24) becomes

$$(I + P_2\Delta)^{-1}(I - B) \tag{5.26}$$

where

$$B = \hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta(I + \hat{\Sigma}^{-1}\Delta)^{-1} \tag{5.27}$$

Substituting (5.26) into A we get

$$A(I - A)^{-1} = P\hat{\Sigma}^{-1}\Delta(I + P_2\Delta)^{-1}\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \tag{5.28}$$

A final substitution to (5.25) gives

$$tr(P\Sigma^{-1}P^T)^{-1} = tr(P_1\hat{\Sigma} + P_1\Delta(I + P_2\Delta)^{-1}) \tag{5.29}$$

Applying again Lemma 4 with $R = P_2\Delta$ and $k = 1$ and assuming that T_1 as defined in Lemma 4 is small enough we get

$$tr(P\Sigma^{-1}P^T)^{-1} \simeq tr(P_1\hat{\Sigma} + P_1\Delta - P_1\Delta P_2\Delta) \Rightarrow \tag{5.30}$$

$$trP_1\Sigma - tr(P\Sigma^{-1}P^T)^{-1} \simeq trP_1\Delta P_2\Delta$$

whence the result of the theorem is proved.

Lemma 6 *Let A and B be matrices of the same order and assume that A has full row rank. Define $C = A^T(AA^T)^{-1}B$. Then*

$$trX^2 \geq 2trC(I - A^+A)C^T + trC^2 \tag{5.31}$$

for every symmetric matrix X satisfying $AX = B$, with equality if and only if $X = C + C^T - CA^T(AA^T)^{-1}A$.

Proof: Consider the Lagrangian function

$$\frac{1}{2}trX^TX - trL^T(AX - B) \tag{5.32}$$

This gives two conditions

$$X = A^T L \tag{5.33}$$

and

$$AX = B \tag{5.34}$$

A third condition is necessary because X must be symmetric. Thus,

$$X = X^T \tag{5.35}$$

The general solution for $AX = B$ since A is of full rank is

$$X = C + (I - \Pi_A)Q$$

(see Magnus and Neudecker, 1988) where $C = A^T(AA^T)^{-1}B$ and $\Pi_A = A^T(AA^T)^{-1}A = A^+A$ (A^+ is the Moore-Penrose inverse of A) and Q arbitrary.

If we had the first two conditions only then the solution of the linear system would give

$$X = C$$

that is $Q = 0$ (since $X = A^T L$). But the additional condition $X = X^T$ means that

$$\frac{X + X^T}{2} = X \Rightarrow$$

$$C + (I - \Pi_A)Q + C^T + Q^T(I - \Pi_A) = 2C + 2(I - \Pi_A)Q \Rightarrow$$

Postmultiplying by A^T we get

$$CA^T + C^T A^T(I - \Pi_A)QA^T = 2CA^T + 2(I - \Pi_A)QA^T \Rightarrow$$

$$C^T A^T - CA^T = QA^T - \Pi_A QA^T$$

A solution to this, is $Q = C^T$. That is, the final solution is

$$X = C + (I - \Pi_A)C^T$$

But since $AB^T = AX^T A^T$, that is AB^T is symmetric, we get that the solution is also written as

$$X = C + C^T - CA^T(AA^T)^{-1}A$$

By substituting this result to $trX^T X = trX^2$ we get the desired result.

Lemma 7 *For any positive semidefinite matrix V, let $\lambda_{max}(V)$ denote its largest characteristic root. Then for any matrix A $(n \times m)$ we have*

$$\lambda_{min}(V) tr AA^T \leq tr AVA^T \leq \lambda_{max}(V) tr AA^T$$

Proof: Since V is positive semidefinite it has an expansion of the form

$$V = \Gamma \Lambda \Gamma^T$$

Then

$$\begin{aligned} tr AVA^T &= tr A\Gamma\Lambda\Gamma^T A^T \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_j \omega_{ij}^2 \end{aligned}$$

But

$$\lambda_{min}(V) \sum_{i=1}^{n}\sum_{j=1}^{m} \omega_{ij}^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_j \omega_{ij}^2 \leq \lambda_{max}(V) \sum_{i=1}^{n}\sum_{j=1}^{m} \omega_{ij}^2 \Rightarrow$$

$$\lambda_{min}(V) tr A\Gamma\Gamma^T A^T \leq \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_j \omega_{ij}^2 \leq \lambda_{max}(V) tr A\Gamma\Gamma^T A^T \Rightarrow$$

$$\lambda_{min}(V) tr AA^T \leq tr AVA^T \leq \lambda_{max}(V) tr AA^T$$

where ω_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, m$ are the elements of $A\Gamma$ and λ_i, $i = 1, \ldots, m$ are the elements of matrix Λ.

Using these results the upper bound for $tr P_1 \Delta P_2 \Delta$ will be given by the following theorem:

Theorem 13 *If P_1 and P_2 are defined as in (5.15) and (5.16) respectively, then*

$$tr P_1 \Delta P_2 \Delta \leq \frac{tr \Delta \hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1}}{2\lambda_{min}(P \hat{\Sigma}^{-1} P^T)} \tag{5.36}$$

where $\lambda_{min}(B)$ is the minimum characteristic root of matrix B

Proof: Since

$$P_1 = \hat{\Sigma}^{-1} P^T (P \hat{\Sigma}^{-1} P^T)^{-2} P \hat{\Sigma}^{-1}$$

and

$$P_2 = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1}$$

It is easy to see that

$$tr P_1 \Delta P_2 \Delta = tr \hat{\Sigma}^{-1/2} \Delta P_1 \Delta \hat{\Sigma}^{-1/2} (I - \Pi)$$

where $\Pi = \hat{\Sigma}^{-1/2} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1/2}$. Note that Π is a projection matrix, that is symmetric and idempotent. Consequently, the equation above can be written as

$$tr P_1 \Delta P_2 \Delta = tr (I - \Pi) \hat{\Sigma}^{-1/2} \Delta P_1 \Delta \hat{\Sigma}^{-1/2} (I - \Pi) \tag{5.37}$$

If we apply Lemma 6 with

$$X = \hat{\Sigma}^{-1/2} \Delta \hat{\Sigma}^{-1/2}$$

and

$$A = (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1/2}$$

we get that

$$B = (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1/2}$$

$$C = \hat{\Sigma}^{-1/2} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1/2}$$

and that

$$(I - \Pi) C^T C (I - \Pi) = (I - \Pi) \hat{\Sigma}^{-1/2} \Delta \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1/2} (I - \Pi) \tag{5.38}$$

But

$$(I - \Pi) \hat{\Sigma}^{-1/2} \Delta P_1 \Delta \hat{\Sigma}^{-1/2} (I - \Pi)$$

$$= (I - \Pi) \hat{\Sigma}^{-1/2} \Delta \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-2} P\hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1/2} (I - \Pi)$$

and

$$\lambda_{min} (P\hat{\Sigma}^{-1} P^T) tr (I - \Pi) \hat{\Sigma}^{-1/2} \Delta P_1 \Delta \hat{\Sigma}^{-1/2} (I - \Pi)$$

$$\leq tr (I - \Pi) \hat{\Sigma}^{-1/2} \Delta \hat{\Sigma}^{-1} P^T (P\hat{\Sigma}^{-1} P^T)^{-1} P\hat{\Sigma}^{-1} \Delta \hat{\Sigma}^{-1/2} (I - \Pi)$$

$$= tr (I - \Pi) C^T C (I - \Pi) = tr C (I - \Pi) C^T$$

$$\leq \frac{1}{2}trX^2 - \frac{1}{2}trC^2$$
$$\leq \frac{1}{2}trX^2$$

using the fact that

$$trC^2 = trC\Pi C^T$$

which means that this trace is positive.

Since

$$trX^2 = tr\Delta\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}$$

we get that

$$\lambda_{min}(P\hat{\Sigma}^{-1}P^T)trP_1\Delta P_2\Delta \leq \frac{tr\Delta\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}}{2} \Rightarrow$$
$$trP_1\Delta P_2\Delta \leq \frac{tr\Delta\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}}{2\lambda_{min}(P\hat{\Sigma}^{-1}P^T)}.$$

Upper bound (5.36) can be extended, thus giving an upper which exploits assumption (5.11). This extension will be derived using the matrix version of Cauchy–Schwarz inequality which is presented as a lemma below.

Lemma 8 *For any two real matrices A and B of the same order, we have*

$$tr(A^TB)^2 \leq tr(A^TA)(B^TB) \tag{5.39}$$

Proof[Magnus and Neudecker, 1988]: Let us assume that we have two matrices X and Y of the same order. Define $\boldsymbol{a} = vec(X)$ and $\boldsymbol{b} = vec(Y)$. Applying the Cauchy–Schwarz inequality for vectors we get:

$$(\boldsymbol{a}^T\boldsymbol{b})^2 \leq (\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) \Rightarrow$$
$$(trX^TY)^2 \leq trX^TXtrY^TY \tag{5.40}$$

by using property (A12) of the Appendix. Assuming now that $X = A^TB$ and that $Y = BA^T$, application of (5.40) gives

$$(trBA^TBA^T)^2 \leq trBA^TAB^TtrAB^TBA^T$$

But $trBA^TAB^T = trAB^TBA^T$. This immediately gives

$$tr(A^TB)^2 \leq tr(A^TA)(B^TB)$$

and the proof is complete.

Theorem 14 *An alternative bound for* $trP_1\Delta P_2\Delta$ *is*

$$trP_1\Delta P_2\Delta \leq \frac{\epsilon^2}{2}\frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(P\hat{\Sigma}^{-1}P^T)} \tag{5.41}$$

Proof: Since

$$trP_1\Delta P_2\Delta \leq \frac{tr\Delta\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}}{2\lambda_{min}(P\hat{\Sigma}^{-1}P^T)}$$

using the matrix version of Cauchy–Schwarz inequality we get that

$$tr\Delta\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1} = tr(\Delta\hat{\Sigma}^{-1})^2$$

$$\leq tr\Delta^2(\hat{\Sigma}^{-1})^2$$

$$= tr\Delta(\hat{\Sigma}^{-2})\Delta \leq \lambda_{max}(\hat{\Sigma}^{-2})tr\Delta^2$$

$$\leq \epsilon^2\lambda_{max}(\hat{\Sigma}^{-2})$$

$$= \epsilon^2\lambda^2_{max}(\hat{\Sigma}^{-1})$$

Thus,

$$trP_1\Delta P_2\Delta \leq \frac{\epsilon^2}{2}\frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(P\hat{\Sigma}^{-1}P^T)}$$

and the proof of the theorem is complete.

(5.41) shows clearly that the approximate upper bound for e_1 is a linear function of ϵ^2. An alternative form of the bound can be given as:

$$tr(A^TA)^{-1}\left(\frac{\epsilon^2}{2}\frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(P\hat{\Sigma}^{-1}P^T)} + O(\epsilon^3)\right) \tag{5.42}$$

An important result can be deduced when matrix P is sub-orthogonal i.e. $PP^T = I_p$ which shows that the bound is independent of the matrix which describes the within subject changes in mean response. The result is a consequence of the well known Poincaré's Separation Theorem (PST) which is given below without proof:

Theorem 15 *Let $A(m \times m)$ be symmetric and $B(m \times k)$ be such that $B^T B = I_k$. Then*

i)

$$\lambda_{m-k+i}(A) \leq \lambda_i(B^T AB) \leq \lambda_i(A) \;\; i = 1, \ldots, k$$

ii) The second equalities are attained iff $B = P_{(k)}T$ where $P_{(k)}$ is the matrix which contains as columns the first k eigenvectors of A and T is unitary.

Corollary 2 *If P sub-orthogonal then the following result holds:*

$$tr P_1 \Delta P_2 \Delta \leq \frac{\epsilon^2}{2} \frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(\hat{\Sigma}^{-1})} \tag{5.43}$$

Proof: Using PST we can see that if we assume that

$$\lambda_1(\hat{\Sigma}^{-1}) \geq \ldots \geq \lambda_q(\hat{\Sigma}^{-1})$$

and that

$$\lambda_1(P\hat{\Sigma}^{-1}P^T) \geq \ldots \geq \lambda_p(P\hat{\Sigma}^{-1}P^T)$$

then

$$\lambda_q(\hat{\Sigma}^{-1}) \leq \lambda_p(P\hat{\Sigma}^{-1}P^T) \leq \lambda_p(\hat{\Sigma}^{-1})$$

That means that

$$\frac{\epsilon^2}{2} \frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(P\hat{\Sigma}^{-1}P^T)} \leq \frac{\epsilon^2}{2} \frac{\lambda^2_{max}(\hat{\Sigma}^{-1})}{\lambda_{min}(\hat{\Sigma}^{-1})}$$

which proves the Corollary.

Two illustrated examples of bounds (5.41) and (5.43) will be presented. These examples concern two well known cases of estimators of covariance matrices in linear models:

Example 1: Suppose that $\hat{\Sigma} = I_q$, that is, Σ lies in a region of the identity matrix. Let us also assume that the rows of P are orthonormal that is,

$$PP^T = I$$

Then, from Corollary 1, we can deduce that

$$trP_1\Delta P_2\Delta \leq \frac{\epsilon^2}{2} + O(\epsilon^3)$$

which is the result found by Strand (1974).

Example 2: Let us consider the covariance adjusted estimator defined in Chapter 4. There, it had been shown that this estimator is a GLS one with weight matrix $S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}$. Thus,

$$\hat{\Sigma}^{-1} = S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1}$$

and

$$(P\hat{\Sigma}^{-1}P^T)^{-1} = (P(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})P^T)^{-1}$$

Using Lemma 1 of Chapter 2 with $B^* = (P^T, V_1)$ and $B = V_2$ we get that the right side is

$$\left[P(P^T, V_1)\begin{pmatrix} PSP^T & PSV_1^T \\ V_1^TSP^T & V_1^TSV_1 \end{pmatrix}^{-1}\begin{pmatrix} P \\ V_1^T \end{pmatrix}P^T\right]^{-1} =$$

$$= \left[(PP^T, 0)\begin{pmatrix} PSP^T & PSV_1^T \\ V_1^TSP^T & V_1^TSV_1 \end{pmatrix}^{-1}\begin{pmatrix} PP^T \\ 0 \end{pmatrix}\right]^{-1}$$

since $PV_1 = 0$. This expression is equal to

$$(PP^T)^{-1}P(S - SV_1(V_1^TSV_1)^{-1}V_1^TSP^T(PP^T)^{-1}$$

$$(PP^T)^{-1}P(S - SV_1(V_1^TSV_1)^{-1}V_1^TSP^T(PP^T)^{-1}$$

By noting that

$$\lambda_{min}(P\hat{\Sigma}^{-1}P^T) = \lambda_{max}^{-1}((P\hat{\Sigma}^{-1}P^T)^{-1})$$

the bound (5.41) becomes

$$\frac{\epsilon^2}{2}\lambda_{max}^2(S^{-1} - S^{-1}V_2(V_2^TS^{-1}V_2)^{-1}V_2^TS^{-1})$$

$$\lambda_{max}\left((PP^T)^{-1}P(S - SV_1(V_1^TSV_1)^{-1}V_1^TSP^T(PP^T)^{-1}\right)$$

since for a matrix A, $\lambda_{max}^{-1}(A) = \lambda_{min}(A^{-1})$.

5.4 Exact bounds for e_2

The form of e_2 as defined in (5.13) is

$$trVar(vec(\hat{\xi}_{\hat{\Sigma}})) = tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \otimes (A^TA)^{-1}$$

$$= tr(A^TA)^{-1}tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}$$

$$= tr(A^TA)^{-1}\left(tr(P\hat{\Sigma}^{-1}P^T)^{-1} + tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}\right)$$

$$= trVar(\widehat{vec}(\hat{\xi}_{\hat{\Sigma}})) + tr(A^TA)^{-1}tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \quad (5.44)$$

This equation shows that

$$e_2 = trVar(vec(\hat{\xi}_{\hat{\Sigma}})) - trVar(\widehat{vec}(\hat{\xi}_{\hat{\Sigma}})) \Rightarrow$$

$$e_2 = tr(A^TA)^{-1}tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \quad (5.45)$$

Thus, e_2 will be bounded if we bound

$$tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}$$

This upper bound obviously now depends on the error matrix Δ we make in the estimation of Σ. The following theorem by von Neumann (1937) will be of great help:

Theorem 16 (von Neumann, 1937) *Let A $(n \times m)$ and B $(n \times m)$ have singular value decompositions (s.v.d's) $P\Delta_1Q^T$ and $R\Delta_2S^T$ respectively. Then*

$$|trAUB^TV| \leq tr\Delta_1\Delta_2 \quad (5.46)$$

where U, V are unitary (orthogonal) matrices of orders n and m respectively and the equality is attained when $U = QS^T$ and $V = RP^T$.

Consider (5.45). The aim is to maximize and minimize

$$tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Delta\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1} \quad (5.47)$$

Note that the sign of this number will be either positive or negative depending on two things:

- the number of the positive eigenvalues of Δ,
- the specific values of these (positive or negative) eigenvalues.

If, for example, $\Delta < 0$ then the above trace will always be negative whatever the specific values of the eigenvalues. The inverse will happen when $\Delta > 0$. Here, we are interested in the upper and lower values this trace will take, when a specific number of the eigenvalues of Δ is known to be positive and for all possible variations of these eigenvalues.

(5.47) is now written as

$$tr\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-2}P\hat{\Sigma}^{-1}\Delta$$

Matrix $\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-2}P\hat{\Sigma}^{-1}$ is a $q \times q$ symmetric one of rank p. It is also a positive semidefinite one and it has a s.v.d RQ^2R^T where

$$Q^2 = \begin{pmatrix} \lambda_1^2 & & & & \\ & \ddots & & 0 & \\ & & \lambda_p^2 & & \\ & & & 0 & \\ & 0 & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

where $\lambda_1, \ldots, \lambda_p$ are the singular values of $(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}$. Δ has an s.v.d expansion $\Delta = \Delta_1 M_1 \Delta_1^T$ where

$$M_1 = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_q \end{pmatrix} = \begin{pmatrix} M_{11} & 0 \\ 0 & M_{12} \end{pmatrix}$$

where M_{11} $(s \times s)$ consists of the positive eigenvalues of Δ and M_{12} $((q-s) \times (q-s))$ of the nonpositive ones. Thus,

$$tr\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-2}P\hat{\Sigma}^{-1}\Delta = tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1$$

If we further assume that

$$\Delta_1 = (\Delta_{11}, \Delta_{12})$$

then

$$tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1 = \begin{pmatrix} \Delta_{11}^T R \\ \Delta_{12}^T R \end{pmatrix} Q^2 \left(R^T\Delta_{11}, R^T\Delta_{12}\right) M_1$$
$$= tr\Delta_{11}^T RQ^2R^T\Delta_{11}M_{11} + tr\Delta_{12}^T RQ^2R^T\Delta_{12}M_{12}$$

But the last trace is easily seen to be

$$tr\Delta_{12}^T RQ^2R^T\Delta_{12}M_{12} = tr\Phi^T M_{12}\Phi$$

where Φ is such that $\Phi\Phi^T = \Delta_{12}^T RQ^2R^T\Delta_{12}$. But M_{12} is negative semi-definite, thus the trace $tr\Phi^T M_1^2\Phi$ is also nonpositive. For the same reason $tr\Delta_{11}^T RQ^2R^T\Delta_{11}M_{11}$ is positive. Consequently,

$$tr\Delta_{12}^T RQ^2R^T\Delta_{12}M_{12} \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1 \leq tr\Delta_{11}^T RQ^2R^T\Delta_{11}M_{11}$$

But again

$$tr\Delta_{11}^T RQ^2R^T\Delta_{11}M_{11} = tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1^*$$

and

$$tr\Delta_{12}^T RQ^2R^T\Delta_{12}M_{12} = tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_2^*$$

where

$$M_1^* = \begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$M_2^* = \begin{pmatrix} 0 & 0 \\ 0 & M_{12} \end{pmatrix}$$

Thus,

$$tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_2^* \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1 \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1^*$$

or

$$-tr(\Delta_1^T R)Q^2(R^T\Delta_1)\hat{M}_2^* \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1 \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1^*$$

with $\hat{M}_2^* = -M_2^*$. Application now of Theorem 16 in both bounds shows that

$$tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1^* \leq trQ^2M_1^*$$

and

$$tr(\Delta_1^T R)Q^2(R^T\Delta_1)\hat{M}_2^* \leq trQ^2\hat{M}_2^* \Rightarrow$$

$$tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_2^* = -tr(\Delta_1^T R)Q^2(R^T\Delta_1)\hat{M}_2^*$$

$$\geq -trQ^2\hat{M}_2^* = trQ^2M_2^*$$

and thus

$$trQ^2M_2^* \leq tr(\Delta_1^T R)Q^2(R^T\Delta_1)M_1 \leq trQ^2M_1^*$$

Note that such optimization has been done with respect to all possible variations of the eigenvectors of matrix Δ. The resulting traces will be optimized with respect to variations of μ_i's under the conditions

$$\sum_{i=1}^{p^*} \mu_i^2 = \epsilon_1^2$$

for the upper bound, and

$$\sum_{i=p^*+1}^{p} \mu_i^2 = \epsilon_2^2$$

for the lower bound, with ϵ_1 and ϵ_2 being small constants with $\epsilon_1^2 + \epsilon_2^2 \leq \epsilon^2$ and $p^* = min(p, s)$. The existance of such constants arises from assumption (5.11). Consider now the Lagrangian functions

$$\sum_{i=1}^{p^*} \lambda_i^2\mu_i - \lambda\left(\sum_{i=1}^{p^*} \mu_i^2 - \epsilon_1^2\right) \tag{5.48}$$

$$\sum_{i=p^*+1}^{p} \lambda_i^2\mu_i - \lambda\left(\sum_{i=p^*+1}^{p} \mu_i^2 - \epsilon_2^2\right) \tag{5.49}$$

The second derivative of both functions is -2λ which is negative for $\lambda > 0$ and positive for $\lambda < 0$. The former case corresponds to a concave Lagrangian function and a maximum at the point which solves the first order conditions, while the latter corresponds to a convex Lagrangian function and a minimum. The first order conditions can be found by taking the first derivative of (5.48) and of (5.49) with respect to μ_i $i = 1, \ldots, q$ and to λ. Equating these derivatives to zero we get:

$$\begin{aligned} 2\lambda\mu_i &= \lambda_i^2 \quad i = 1, \ldots, p^* \\ \sum_{i=1}^{p^*} \mu_i^2 &= \epsilon_1^2 \end{aligned}$$

for Lagrangian function (5.48), and

$$\begin{aligned} 2\lambda\mu_i &= \lambda_i^2 \quad i = p^* + 1, \ldots, p \\ \sum_{i=p^*+1}^{p} \mu_i^2 &= \epsilon_2^2 \end{aligned}$$

for function (5.49). The form of these equations suggests that for (5.48) λ should be positive (something which corresponds to a maximum) and for (5.49) λ should be negative (which corresponds to a minimum). Solution of both systems gives

$$\lambda = \frac{1}{2}\sqrt{\frac{\sum_{i=1}^{p^*} \lambda_i^4}{\epsilon_1^2}}$$

and

$$\mu_i = \epsilon_1 \frac{\lambda_i^2}{\sqrt{\sum_{i=1}^{p^*} \lambda_i^4}} \quad i = 1, \ldots, p^*$$

for (5.48) and

$$\lambda = -\frac{1}{2}\sqrt{\frac{\sum_{i=p^*+1}^{p} \lambda_i^4}{\epsilon_2^2}}$$

and

$$\mu_i = -\epsilon_2 \frac{\lambda_i^2}{\sqrt{\sum_{i=p^*+1}^{p} \lambda_i^4}} \quad i = p^* + 1, \ldots, p$$

for (5.49). Finally, by direct substitution, the optimized values of the trace are

$$\epsilon_1 \sqrt{\sum_{i=1}^{p^*} \lambda_i^4} \leq \epsilon \sqrt{\sum_{i=1}^{p^*} \lambda_i^4}$$

and

$$-\epsilon_2 \sqrt{\sum_{i=p^*+1}^{p} \lambda_i^4} \geq -\epsilon \sqrt{\sum_{i=p^*+1}^{p} \lambda_i^4}$$

and the maximized value of e_2 is

$$tr(A^T A)^{-1} \left(\epsilon \sqrt{\sum_{i=1}^{p^*} \lambda_i^4} \right)$$

while the corresponding minimized one is

$$-tr(A^T A)^{-1} \left(\epsilon \sqrt{\sum_{i=p^*+1}^{p} \lambda_i^4} \right)$$

Thus, we proved the following theorem:

Theorem 17 *For e_2 as defined in (5.44) and under the condition that $tr\Delta^2 \leq \epsilon^2$ the forllowing result holds:*

$$-tr(A^TA)^{-1}\left(\epsilon\sqrt{\sum_{i=p^*+1}^{p}\lambda_i^4}\right) \leq e_2 \leq tr(A^TA)^{-1}\left(\epsilon\sqrt{\sum_{i=1}^{p^*}\lambda_i^4}\right) \tag{5.50}$$

where λ_i $i = 1,\ldots,p$ are the singular values of $(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}$ and $p^* = min(p,s)$ where s is the number of positive eigenvalues of Δ.

A very interesting result appears in the case where Δ is a positive definite matrix:

Corollary 3 *In the case where $\Delta > 0$ then exact upper and lower bounds for*

$$tr(A^TA)^{-1}\left[tr(P\Sigma^{-1}P^T)^{-1} - tr(P\hat{\Sigma}^{-1}P^T)^{-1}\right] = \phi$$

are

$$0 \leq \phi \leq k_1$$

where $k_1 = tr(A^TA)^{-1}\left(\epsilon\sqrt{\sum_{i=1}^{p}\lambda_i^4}\right)$

Proof: Let us denote by

$$a = tr(A^TA)^{-1}tr(P\Sigma^{-1}P^T)^{-1}$$

$$b = tr(A^TA)^{-1}tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}$$

$$c = tr(A^TA)^{-1}tr(P\hat{\Sigma}^{-1}P^T)^{-1}$$

Then as was noted in the introduction

$$a - b \leq 0$$

Theorem 17 shows that

$$b - c \leq k_1,$$

where

$$k_1 = tr(A^TA)^{-1}\left(\epsilon\sqrt{\sum_{i=1}^{p}\lambda_i^4}\right)$$

since Δ is positive definite and $p^* = min(p, s) = p$. Addition of the two inequalities gives

$$a - c \leq k_1.$$

The lower bound arises from the fact that $tr(P\Sigma^{-1}P^T)^{-1}$ is an increasing function of Σ in the sense that if

$$\Sigma_1 < \Sigma_2 \Rightarrow tr(P\Sigma_1^{-1}P^T)^{-1} < tr(P\Sigma_2^{-1}P^T)^{-1},$$

where the first inequality means that $\Sigma_1 - \Sigma_2$ is a positive definite matrix. Thus, since $\Sigma - \hat{\Sigma} > 0$ then

$$a - c \geq 0$$

and the proof of the corollary is complete.

If we are more interested now to the ratio of the traces

$$\frac{tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}},$$

which gives another degree to which $Var(vec(\hat{\xi}_{\hat{\Sigma}}))$ approximates $Var(vec(\hat{\xi}_{\hat{\Sigma}}))$, we can see immediately that,

$$\begin{aligned}\psi \equiv \frac{tr(P\hat{\Sigma}^{-1}P^T)^{-1}P\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}P^T(P\hat{\Sigma}^{-1}P^T)^{-1}}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}} &= \frac{e_2 + tr(P\hat{\Sigma}^{-1}P^T)^{-1}}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}} \\ &= 1 + \frac{e_2}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}}\end{aligned}$$

From Theorem 17 then we conclude that

$$1 - \frac{\epsilon\sqrt{\sum_{i=p^*+1}^{p}\lambda_i^4}}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}} \leq \psi \leq 1 + \frac{\epsilon\sqrt{\sum_{i=1}^{p^*}\lambda_i^4}}{tr(P\hat{\Sigma}^{-1}P^T)^{-1}}. \tag{5.51}$$

5.5 Conclusions and further research

For the measures of performance e_1 and e_2 defined in (5.12) and (5.13) respectively, some bounds were derived. The bounds show that the error we commit in our inferences by adopting a certain approximation of the covariance matrix of observations Σ,

depends on the distance this estimate will have from the real covariance matrix and on certain constants which can be large or small. It is evident from (5.50) that the more positive eigenvalues Δ has, the less $\widehat{Var(vec(\hat{\xi}_{\Sigma}))}$ is distant from $Var(vec(\hat{\xi}_{\Sigma}))$. If moreover, $\Delta > 0$ then Corollary 2 shows that $tr(P\hat{\Sigma}^{-1}P^T)^{-1} < tr(P\Sigma^{-1}P^T)^{-1}$ and this results in an inprecise impression of accuracy in our inferences.

In Chapter 2 we saw that when using parametric models for the covariance matrix, the correct model always gives ratios which are close to 1 irrespective of the form of the population covariance matrix. Thus, the results of this chapter and especially expression (5.51) can be used for the derivation of methods which will aim at the computation of weight matrices $\hat{\Sigma}$ which will give the least possible bounds in (5.51) by keeping ϵ constant. However, this is an entirely different subject which will not concern us here.

Chapter 6

Conclusions

This thesis investigates of some methods in the analysis of growth curves. It is accepted that the efficient estimation of the covariance matrix of observations plays one of the most important parts to the subsequent efficient estimation of the regression coefficients. From this point of view two different classes of methods were investigated. Likelihood based and Covariance adjustment. In the first class a comparison was aimed between Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML). It was shown that the general estimated scatter of the estimated regression coefficients is larger for REML than for ML. It was conjectured that this should also be expected for the variances of the specific coefficients which constitute the diagonals of the corresponding estimated variance matrices. Empirical results suggest that this may be true and that REML provides estimates of the regression coefficients whose estimated variance is closer to their real value than ML and at the same time this estimated variance lies close to its optimal value. This along with the above mentioned conjecture and the empirical result of frequent underestimation of the optimal variances of the regression coefficients could lead us to prefer REML from ML for estimation purposes. It is worthwhile to say that these results hold for a wide class of covariance structures and not only for mixed effects models where closed formulae exist.

The generalization to more measured characteristics at a particular time-point of-

fered another opportunity for a comparison between the two methods. Here, the results of Anderson, Anderson and Olkin (1986) were generalized to an arbitrary number of random-effects and the REML estimates of variance were also derived. Since now the variance components are not constants but matrices, the interesting problem of the estimated ranks for the covariance matrices arises. It was shown that REML gives larger estimated rank and that only in the case where the ML and REML give identical estimated ranks, we can explicitly make a comparison between the estimated covariance matrices of the regression coefficients evaluated by these two different methods.

Covariance Adjustment method was also investigated. Its form was given as a generalized least squares estimator and a method was suggested for its optimal choice. The method as well as the one suggested by Fujikoshi and Rao (1991) was shown to give estimates of the variance of the regression coefficients which approximate well the optimal 'Gauss–Markov' values and their estimated variance lies nearer to its true value than the other methods.

Finally, the subject of efficiency was more thoroughly studied by assuming that the real covariance matrix of observations lies into a region of the estimated covariance matrix not greater (according to Euclidean norm) than a constant. Upper and lower bounds were derived for certain measures of efficiency as functions of this constant and of the estimated covariance matrix. Apart from the further possible theoretical work on this idea, the results can possibly suggest applied work in the area, by investigating the possibilities of constructing new algorithms which would aim into the derivation of more efficient estimates of the regression coefficients.

APPENDIX

This part of the thesis is devoted to certain results on linear algebra which are needed for the better understanding of the previous chapters. These results will be given without proofs.

A1: Let A $(n \times n)$ and α any scalar. Then $|\alpha A| = \alpha^n |A|$.

A2: For matrices A and B for which AB and BA are legitimate

$$trAB = trBA$$

A3: If

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

then

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}D^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}D^{-1} \\ -D^{-1}A_{21}A_{11}^{-1} & D^{-1} \end{pmatrix}$$

with $D = A_{22} - A_{21}A_{11}^{-1}A_{12}$ and

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|$$

A4: For any symmetric $n \times n$ matrix A, there exist an orthogonal $n \times n$ matrix Γ and a diagonal matrix Λ containing the eigenvalues of A such that

$$A = \Gamma \Lambda \Gamma^T.$$

A5: Let A be a $n \times n$ square matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$trA = \sum_{i=1}^{n} \lambda_i \quad |A| = \prod_{i=1}^{n} \lambda_i.$$

A6: If A is an idempodent matrix with r eigenvalues then,

$$r(A) = trA = r.$$

A7: Let A be p.d and B be a p.s.d of the same order. Then there exist a non-singular matrix P and a p.s.d diagonal matrix Λ:

$$A = PP^T \quad B = P\Lambda P^T.$$

A8: Let A, B be p.d $n \times n$ matrices. Then $A - B$ is p.d if and only if $B^{-1} - A^{-1}$ is p.d.

A9: Let A, B be p.d such that $A - B$ is p.s.d. Then $|A| \geq |B|$ with equality if and only if $A = B$.

Definition: Let A be a $m \times n$ matrix and B a $p \times q$ matrix. Then the *Kronecker product* of A and B is denoted and defined as:

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ & \dots & \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}.$$

This product has the following properties:

A10: $(A, B) \otimes C = (A \otimes C, B \otimes C)$.

A11: $vec(ABC) = \left(C^T \otimes A\right) vec(B)$.

A12: $trABCD = \left(vec(B^T)\right)^T \left(C \otimes A^T\right) vec(D)$.

A13: $(A \oslash B)(C \oslash D) = AC \oslash BD$.

A14: $(A \otimes B)^T = A^T \otimes B^T$.

A15: $tr(A \otimes B) = (trA)(trB)$.

A16: $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

A17: If $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of the $m \times m$ matrix A and $\mu_1, \ldots, \mu_p$ are the eigenvalues of the $p \times p$ matrix B, then the mp eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$, $i = 1, \ldots, m$, $j = 1, \ldots, p$ and consequently,

$$|A \otimes B| = |A|^p |B|^m \quad r(A \otimes B) = r(A) r(B).$$

Bibliography

[1] Altham, P. M. E. (1984). Improving the precision of estimation by fitting a model. *Journal of Royal Statistical Society, Ser. B*, **46**, 118–119.

[2] Amemiya, Y. (1985). What should be done when an estimated between group covariance matrix in not nonnegative definite. *American Statistician*, **39**, 112–117.

[3] Amemiya, Y. & Fuller, W. (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Annals of Statistics*, **12**, 497–509.

[4] Anderson, B. M., Anderson, T. W. & Olkin, I. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *Annals of Statistics*, **14**, 405–417.

[5] Anderson, T. W. (1958). *An Introduction to Multivariate Analysis.* John Wiley, 1958.

[6] Anderson, T. W. (1969). *Statistical inferences for covariance matrices with linear structure.* In *Multivariate Analysis II*, ed. Krishnaiah, P.R., 55–66.

[7] Azzalini, A. (1984). Estimation and hypothesis testing for collections of autoregressive time-series. *Biometrika*, **71**, 85–90.

[8] Baksalary, J. K., Corsten, L. C. A. & Kala, R. (1978). Reconciliation of two different views on estimation of growth curve parameters. *Biometrika*, **65**, 662–665.

[9] Bjornsson, H. (1978). Analysis of a series of long-term grassland experiments with autocorrelated errors. *Biometrics*, **34**, 645–651.

[10] Bloomfield, P. & Watson, G. S. (1975). The inefficiency of least squares. *Biometrika*, **62**, 121–128.

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

[12] Danford, M. B., Hughes, H. M. & Mc Nee, R. C. (1960). On the analysis of repeated-measurements experiments. *Biometrics*, **16**, 547–565.

[13] Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics*, **44**, 959–971.

[14] Draper, N. & Smith, H. (1981). *Applied Regression analysis.* John Wiley, 1981.

[15] Elston, R. C. & Grizzle, J. E. (1962) Estimation of time response curves and their confidence band. *Biometrics*, **18**, 148–159.

[16] Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis. *Annals of Institute of Statistical Mathematics*, **34**,137–144.

[17] Fujikoshi, Y. & Rao, C. R. (1991). Selection of covariables in the growth curve model. *Biometrika*, **78**, 779–785.

[18] Gabriel, K. R. (1961). The model of ante-dependence for data of biological growth. *Bulletin Institute 33rd Session*(Paris), 253–264.

[19] Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, **33**, 201–212.

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

[21] Gleser, L. J. & Olkin, I. (1972). Estimation for regression model with an unknown covariance matrix. *Proc. Sixth Berkeley Symposium of Mathematical Statistics and Probability*, **1**, 541–568.

[22] Golub, G. H. (1963). Comparison of the variance of minimum variance and weighted least squares regression coefficients. *Annals of Mathematical Statistics*, **34**, 984–991.

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

[24] Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

[25] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Association*, **72**, 320–340.

[26] Johansen, S. (1982). Asymptotic inference in random coefficient regression models. *Scandinavian Journal of Statistics*, **9**, 201–207.

[27] Joreskog, V. G. (1970). A general method for analysis of covariance structures. *Biometrika*, **57**, 239–251.

[28] Kantorovich, L. V. (1948). Functional analysis and applied mathematics. *Uspekhi Matematichekikh Nauk*, **3**, 89–185. Translated from Russian by C. D. Benster, National Bureau of Standards, Report 1509, 7 March 1952.

[29] Kenward, M. G. (1981). Unpublished Ph.D thesis, University of Reading.

[30] Kenward, M. G. (1985). The use of fitted higher-order polynomial coefficients as covariates in the analysis of growth curves. *Biometrics*, **41**, 19–28.

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

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

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

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

[35] Knott, M. (1975). On the minimum efficiency of least squares. *Biometrika*, **62**, 129–132.

[36] Laha, R. G. (1954). On some problems in canonical correlations. *Sankhyā*, **14**, 61–68.

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

[38] Lee, J. C. (1991). Tests and model selection for the general growth curve model. *Biometrics*, **47**, 147–159.

[39] Magnus, J. R. & Neudecker, H. (1988). Matrix differential calculus. John Wiley.

[40] Mansour, H., Nordheim, E. V. & Rutledge, J. J. (1985). Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors. *Biometrics*, **41**, 287–294.

[41] Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā B*, **36**, 115–128.

[42] Mardia, K. V., Kent, J. T. & Bibby, J. M. (1982). Multivariate analysis. Academic Press, London.

[43] McGilchrist, C. A. (1989). Bias of ML and REML estimators in regression models with ARMA errors. *J. Stat. Comput. Simul.*, **32**, 127–136.

[44] McKay, R. J. (1977). Variable selection in multivariate regression: An application of simultaneous test procedures. *Journal of Royal Statistical Society Ser. B*, **39**, 371–380.

[45] McGilchrist, C. A. & Cullis, B. R. (1991). REML estimation for repeated measures analysis *J. Statist. Comput. Simul.*, **38**, 151–163.

[46] Nelder, J. A. (1961). The fitting of a generalization of the logistic curve. *Biometrics*, **17**, 89–110.

[47] von Neumann, J. (1937). Some matrix inequalities and metrization of matrix space. *Tomsk University Review*, **1**, 283–300. Reprinted (1962) in *John von Neumann, Collected works*, Taub, A. H., ed. **4**, 205–219. Pergamon, New York.

[48] Oliver, F. R. (1964) Methods of estimating the logistic growth function. *Applied Statistics*, **13**, 57–66.

[49] Pantula, S. G. & Polock, K. H. (1985). Nested analysis of variance with autocorellated errors. *Biometrics*, **41**, 909–920.

[50] Patterson, H. D. & Lowe, B. I. (1070). The errors of long–term experiments. *Journal of Agricultural Science, Cambridge*, **74**, 53–60.

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

[52] Patterson, H. D. & Thompson, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*, 197–207.

[53] Potthoff, R. F. & Roy, S. N. (1964). A generalized MANOVA model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[54] 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.

[55] Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. In P. R. Krishnaiah, editor, *Multivariate Analysis I*, pages 87–103. New York: Academic Press.

[56] Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In L.M. Le Cam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 355–372. Berkeley: University of California Press.

[57] Reinsel, G. (1982). Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *Journal of American Statistical Association*, **77**, 190–195.

[58] Reinsel, G. (1984). Estimation and prediction in a multivariate random-effects generalized linear model. *Journal of American Statistical Association*, **79**, 406–414.

[59] Reinsel, G. (1985). Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model. *Journal of American Statistical Association*, **80**, 642–650.

[60] Roy, S. N. (1957). *Some Aspects of Multivariate Analysis.* John Wiley, 1957.

[61] Siotani, M. (1957). Effect of the additional variates on the canonical correlation coefficients. *Proceedings of Institute of Statistical Mathematics*, **5**, 52–57.

[62] Strand, O. N. (1974). Coefficient errors caused by using the wrong covariance matrix in the general linear model. *Annals of Statistics*, **2**, 935–949.

[63] Swallow, W. H. & Monahan, J. F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics*, **26**, 47–57.

[64] Theobald, C. M. (1975) An inequality with application to multivariate analysis. *Biometrika*, **62**, 461–466.

[65] Verbyla, A. P. (1986). Conditioning in the growth curve model. *Biometrika*, **75**, 129–138.

[66] Verbyla, A. P. & Cullis, B. R. (1990). Modelling in repeated measures experiments. *Applied Statistics*, **39**, 341–356.

[67] Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, **39**, 95–101.

[68] Wilson, D. P., Hebel, R. J. & Sherwin, R. (1981). Screening and diagnosis when within-individual observations are markov-dependent. *Biometrics*, **37**, 553–565.

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