

# **Parameter structures in the exponential class**

by

**Hendrina Fredrika Strydom**

**Submitted in partial fulfilment**

**of the requirements for the degree**

**Philosophiae Doctor (Mathematical Statistics)**

**in the Faculty of Natural & Agricultural Sciences**

**University of Pretoria**

**Pretoria**

**October 2011**

# Declaration of originality

I, Hendrina Fredrika Strydom, declare that the thesis, which I hereby submit for the degree Philosophiae Doctor (Mathematical Statistics) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature \_\_\_\_\_

Date \_\_\_\_\_

# Acknowledgements

I wish to express my sincere appreciation towards my supervisor, Prof. N.A.S. Crowther, for his guidance, advice, patience and time set aside to support me in doing this research. It was a privilege to learn from him and to have him as supervisor.

I also wish to express my gratitude to colleagues for moral support.

Finally I would like to thank my family for their love, support and encouragement.

# Summary

Maximum likelihood estimation of parameter structures in the exponential class is considered by using a procedure for maximum likelihood estimation under constraints. This procedure does not require maximization of the likelihood function or derivation of the likelihood equations. Constraints are imposed on the expected values of the canonical statistics and not specifically on the parameters themselves. It provides a solution to maximum likelihood estimation in a restricted parameter space.

Maximum likelihood estimation under constraints provides a flexible procedure for the simultaneous modelling of specific structures of means and variances - in both the univariate and multivariate case. In the univariate case, it includes linear models for means and variances, heteroskedastic linear regression, analysis of variance (the balanced and unbalanced case where the assumption of homoscedasticity is not met), gamma regression and linear mixed models.

Heteroscedastic and/or autocorrelated disturbances in econometric applications are dealt with by treating correlated observations as a single multivariate observation.

The same procedure is used for maximum likelihood estimation of parameter structures in the case of multivariate normal samples. The procedure provides a new statistical methodology for the analysis of specific structures in mean vectors and covariance matrices - including the case where the sample size is small relative to the dimension of the observations. Special cases include different variations of the Behrens-Fisher problem, proportional covariance matrices and proportional mean vectors.

Maximum likelihood estimation under constraints as a general approach to estimation in a variety of problems concerning covariance matrices, is also considered. Estimation of Wishart matrices is a special case of the general procedure.

Specific structures are illustrated with real data examples.

Maximum likelihood estimation under constraints provides a convenient method to obtain maximum likelihood estimates of mean and (co)variance structures directly - also in problematic situations where the likelihood equations need to be solved numerically.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| <b>2</b> | <b>MLE under Constraints</b>  | <b>6</b>  |
| <b>3</b> | <b>MLE for the Linear Model with Heterogeneous Variances</b>                          | <b>10</b> |
| 3.1      | Introductory Example . . . . .  | 11        |
| 3.2      | MLE for Linear Models under Normality . . . . .                                       | 13        |
| 3.2.1    | Heteroskedastic linear regression . . . . .   | 15        |
| 3.2.2    | Two-way analysis of variance: unbalanced data . . . . .                               | 16        |
| 3.3      | MLE for (Non)Linear Models in the case of Non-Normal Observations . . . . .           | 19        |
| 3.3.1    | Exponentially distributed observations . . . . .                                      | 19        |
| 3.3.2    | Gamma distributed observations and nonlinear models . . . . .                         | 20        |
| 3.4      | MLE for Linear Mixed Models . . . . .   | 24        |
| 3.4.1    | Two-way crossed classification interaction mixed model with unbalanced data . . . . . | 25        |
| <b>4</b> | <b>MLE for Nonspherical Disturbances</b>  | <b>28</b> |
| 4.1      | MLE for the Generalized Regression Model . . . . .                                    | 29        |
| 4.2      | Introductory Example . . . . .  | 30        |
| 4.3      | MLE for Heteroskedastic Disturbances . . . . .  | 32        |
| 4.3.1    | Groupwise heteroskedasticity . . . . .  | 34        |
| 4.3.2    | Modelling structural change . . . . .   | 37        |

|          |   |           |
|----------|---|-----------|
| 4.4      | MLE for Autocorrelated Disturbances . . . . .   | 40        |
| 4.4.1    | AR(1) covariance structure with heteroskedasticity . . . . .  | 41        |
| 4.4.2    | Toeplitz(2) covariance structure with heteroskedasticity . . . . .  | 44        |
| 4.4.3    | Modelling in terms of explanatory variables . . . . .   | 46        |
| <b>5</b> | <b>MLE for Multivariate Normal Samples</b>  | <b>49</b> |
| 5.1      | Introductory Example . . . . .  | 49        |
| 5.2      | Independent Multivariate Normal Samples . . . . .   | 52        |
| 5.3      | Two Large Independent Multivariate Normal Samples . . . . .   | 54        |
| 5.3.1    | $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ . . . . .          | 55        |
| 5.3.2    | $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$ . . . . .  | 55        |
| 5.3.3    | $H_0 : \rho_1 = \rho_2$ . . . . .   | 56        |
| 5.3.4    | $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1, H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$ . . . . . | 57        |
| 5.4      | More Than Two Relatively Small Multivariate Normal Samples . . . . .  | 58        |
| 5.4.1    | Proportional covariance matrices . . . . .  | 60        |
| 5.4.2    | Behrens-Fisher with proportional covariance matrices . . . . .  | 61        |
| 5.4.3    | Proportional mean vectors and proportional covariance matrices . . . . .  | 62        |
| 5.4.4    | Proportional mean vectors and proportional covariance matrices with Toeplitz<br>correlation structure . . . . .         | 65        |
| <b>6</b> | <b>MLE for Wishart Matrices</b>   | <b>68</b> |
| 6.1      | Estimation for Wishart Matrices under Constraints . . . . .   | 69        |
| 6.2      | Covariance Matrices with Given Structure . . . . .  | 70        |
| 6.2.1    | Linearly patterned covariance matrices . . . . .  | 70        |
| 6.2.2    | Covariance matrices with zeros . . . . .  | 71        |
| 6.3      | Dependent Multivariate Normal Samples . . . . .   | 74        |
| 6.3.1    | Test for independence . . . . .   | 75        |
| 6.3.2    | Homogeneity of covariance matrices . . . . .  | 76        |
| 6.3.3    | An equivalent approach . . . . .  | 77        |

|          |                                      |           |
|----------|--------------------------------------|-----------|
| <b>7</b> | <b>Concluding remarks</b>            | <b>80</b> |
| <b>8</b> | <b>References</b>                    | <b>82</b> |
| <b>A</b> | <b>Appendix</b>                      | <b>86</b> |
| A.1      | SAS Programs for Chapter 3 . . . . . | 86        |
| A.1.1    | UNIV.SAS . . . . .                   | 86        |
| A.1.2    | REGRESSION.SAS . . . . .             | 87        |
| A.1.3    | ANOVA.SAS . . . . .                  | 89        |
| A.1.4    | EXPONENTIAL.SAS . . . . .            | 91        |
| A.1.5    | GAMMA.SAS . . . . .                  | 92        |
| A.1.6    | MIXED.SAS . . . . .                  | 94        |
| A.2      | SAS Programs for Chapter 4 . . . . . | 97        |
| A.2.1    | MULTIV1OBS.SAS . . . . .             | 97        |
| A.2.2    | GWHETEROSK.SAS . . . . .             | 98        |
| A.2.3    | SPLINES.SAS . . . . .                | 102       |
| A.2.4    | AR1.SAS . . . . .                    | 105       |
| A.2.5    | TOEPLITZ.SAS . . . . .               | 108       |
| A.3      | SAS Programs for Chapter 5 . . . . . | 111       |
| A.3.1    | MULTIV.SAS . . . . .                 | 111       |
| A.3.2    | LMVNORMAL.SAS . . . . .              | 113       |
| A.3.3    | SMVNORMAL1.SAS . . . . .             | 117       |
| A.3.4    | SMVNORMAL2.SAS . . . . .             | 123       |
| A.4      | SAS Programs for Chapter 6 . . . . . | 130       |
| A.4.1    | ZEROCOV.SAS . . . . .                | 130       |
| A.4.2    | WISHART.SAS . . . . .                | 132       |
| A.4.3    | NORMAL.SAS . . . . .                 | 133       |



# List of Tables

|     |  |    |
|-----|--|----|
| 3.1 | Hypothetical income and consumption data. . . . .  | 15 |
| 3.2 | Estimated parameters for heteroskedastic linear regression model. . . . .                              | 16 |
| 3.3 | Amount of time TV watched. . . . .   | 18 |
| 3.4 | Estimated parameters for two-way ANOVA with unbalanced data. . . . .                                   | 18 |
| 3.5 | Exponentially distributed observations with estimated model parameters. . . . .                        | 20 |
| 3.6 | Gamma distributed data with estimated parameters. . . . .  | 23 |
| 3.7 | Data for a two-way crossed classification. . . . .   | 26 |
| 3.8 | Estimated parameters for the 2-way crossed classification interaction mixed model. . .                 | 27 |
| 4.1 | Mle's of structured variances based on a single bivariate observation. . . . .                         | 32 |
| 4.2 | Mle's of a groupwise heteroskedastic model with different variance structures. . . . .                 | 36 |
| 4.3 | Mle's of spline regression models with different variance structures. . . . .                          | 39 |
| 4.4 | Mle's of AR(1) covariance structure with heteroskedasticity. . . . .                                   | 43 |
| 4.5 | Mle's of Toeplitz(2) covariance structure with heteroskedasticity. . . . .                             | 45 |
| 4.6 | Mle's of AR(1) covariance structure with heteroskedasticity in terms of explanatory variables. . . . . | 47 |
| 4.7 | Comparison of lse's and mle's. . . . .   | 47 |
| 4.8 | Comparison of predictions based on least squares estimation (LSE) and MLE. . . . .                     | 48 |
| 5.1 | Double iteration process illustrated for the example in Section 5.1. . . . .                           | 51 |
| 5.2 | A Behrens-Fisher problem for large multivariate samples. . . . .                                       | 57 |
| 5.3 | Sample mean vectors, covariance and correlation matrices. . . . .                                      | 59 |

|     |  |    |
|-----|--|----|
| 5.4 | Proportional covariance matrices. . . . .  | 60 |
| 5.5 | Behrens-Fisher and proportional covariance matrices. . . . .                                 | 61 |
| 5.6 | Asymptotic standard errors of the estimated parameters. . . . .                              | 62 |
| 5.7 | Proportional mean vectors and proportional covariance matrices. . . . .                      | 64 |
| 5.8 | Proportional mean vectors & covariance matrices with Toeplitz correlation structure. . . . . | 67 |

# List of Figures

|     |   |    |
|-----|---|----|
| 4.1 | Spline regression. . . . .  | 39 |
| 5.1 | Sample means. . . . .   | 58 |
| 5.2 | Proportional means assuming proportional covariance matrices. . . . . | 64 |
| 5.3 | Proportional submatrices. . . . .                                     | 65 |

# Chapter 1

## Introduction

Maximum likelihood estimation (MLE) of mean and covariance structures in the presence of heterogeneous variances or correlated data is problematic in the sense that explicit likelihood estimates often do not exist. In many cases the likelihood equation(s) are maximized through a numeric maximization routine. Furthermore, different parameterizations often require derivation of a new set of likelihood equations specific to the parameterization under consideration. A procedure for MLE under constraints by Matthews & Crowther (1995) had proved to provide a very flexible method for MLE in the analysis of categorical data (Matthews & Crowther 1995), generalized linear models (Matthews & Crowther, 1998) and linear models for grouped data (Crafford & Crowther, 2009).

In this thesis, this methodology is utilized for MLE of more general parameter structures in the exponential class. The traditional approach to MLE is simplified since no derivation of likelihood equations are required. It provides a method for the direct modelling of the mean vector(s) and covariance structure(s) of random variables belonging to the exponential class. MLE under constraints provides a general framework within which provision is easily made for heterogeneous variances, longitudinal and correlated data. Constraints are imposed on the expected values of the canonical statistics and not specifically on the parameters themselves. This allows for a much wider class of statistical problems to be addressed than only those that imply constraints on parameters. Maximum likelihood estimates (mle's) are determined from an exact theoretical expression - it does not result from some numeric maximization routine. The latter is current practice in statistical software such as PROC MIXED (SAS Institute, 2004). This is generally used for modelling in the case of some applications considered in this

thesis.

The MIXED procedure in SAS is used widely to model data which exhibit correlation and nonconstant variability. However, "experience shows that both the ideas behind the techniques and their software implementation are not at all straightforward, and users from applied backgrounds often encounter difficulties in using the methodology effectively" (Verbeke & Molenberghs, 2009). The primary assumptions underlying the analyses performed by PROC MIXED (SAS, 2004) are as follows:

- The data are normally distributed.
- The means (expected values) of the data are linear in terms of a certain set of parameters.
- The variances and covariances of the data are in terms of a different set of parameters, and they exhibit a structure matching one of those available in PROC MIXED.

PROC MIXED does not allow for structural relations between means, variances, and covariances of different groups. Covariance structures are limited. It cannot be used when only the sufficient statistics are available. Modelling multivariate observations becomes very cumbersome.

In this thesis, MLE under constraints is introduced as an alternative (and the recommended) method of MLE:

1. for the linear model in the presence of heterogeneous variances (Chapter 3),
2. in the case of nonspherical disturbances in econometric applications (Chapter 4),
3. in the case of several multivariate normal populations with unknown and possibly unequal covariance matrices (Chapter 5)
4. for parameter structures in the case of Wishart distributions (Chapter 6).

The likelihood function, formulated in terms of a general set of parameters, is expressed in its canonical form and the canonical statistics used as basis for this procedure. Models are specified in terms of constraints on the expected values of these canonical statistics. Derivation of likelihood equations is not necessary.

Chapters 3, 4, 5, and 6 (together with the basic theory given in Chapter 2) have been submitted for publication in various journals. Chapter 5 has already been accepted for publication by the South African Statistical Journal.

In this thesis an alternative approach is introduced - not only to MLE where the traditional is problematic, but also in standard estimation applications. Apart from solving problematic estimation problems, the procedure provides an attractive alternative to existing procedures. The elegance of the procedure lies in its simplicity and flexibility. This is illustrated with a wide variety of applications. It can easily be adapted to provide for testing and estimation where structural relationships exist amongst mean vectors and/or covariance matrices, e.g. certain variance/covariance structures such as proportional variances and covariances within a population, or proportional (and possibly structured) covariance matrices of several independent populations. MLE under constraints is also proposed as a convenient method to obtain the mle's in the multivariate case where the sample size is small relative to the dimension of the observations. The objective of this thesis is not to make a comparison with competing procedures (in the case where there are alternative procedures available). The examples considered illustrate sufficiently how complicated problems can be solved.

A wide variety of applications is considered in this thesis - all with the same basic point of departure, namely specifying the model in terms of implied constraints on the expected value of the vector of canonical statistics. In Chapter 2 the underlying theory is outlined. An algorithm is given which summarizes the estimation process for the parameter structures considered in this thesis. In Chapter 3, 4, 5 and 6 specific applications are described. In each chapter it is shown how the theory, as provided in Chapter 2, fits into the specific theoretical framework of the models considered in that particular chapter.

In Chapter 3 the MLE of parameters of the linear model in the presence of heterogeneous variances is considered. A very simple introductory example is given to illustrate the basic estimation procedure as applied within the univariate context. The theory for linear models under the assumption of normality, using the method for maximum likelihood under constraints, is given and illustrated with applications such as heteroskedastic linear regression and two-way analysis of variance in the case of unbalanced data. Here it is also shown how linear hypotheses on both means and variances are imposed by specification of the constraints in a straightforward manner. Maximum likelihood estimation for linear and non-linear models in the case of non-normal observations, e.g. exponentially and gamma distributed observations, are discussed. The linear mixed model is considered next. To illustrate, the maximum likelihood estimates for parameters under the constraints implied by a two-way crossed classification

interaction model with unbalanced data are obtained.

In Chapter 4 econometric applications are considered. The generalized regression model is formulated in terms of the general approach followed in this thesis. An introductory example is again utilized to illustrate the procedure here - specifically for correlated disturbances which are treated as a single multivariate observation from a multivariate normal distribution. A general model for heteroskedastic disturbances is given and illustrated in the case of groupwise heteroskedasticity and for modelling structural change. MLE in the presence of autocorrelated disturbances is discussed and illustrated with examples based on some important South African economic time series. Applications illustrate the flexibility of the methodology for simultaneous modelling of the mean and specifically complex covariance structures. Simple regression models with AR(1) or Toeplitz(2) covariance structure respectively, with heteroskedasticity, are also considered. Mle's are compared to OLS (ordinary least squares) in the case of a multiple regression model with heteroskedastic and autocorrelated disturbances.

Parameter structures in the case of several multivariate normal samples are considered in Chapter 5. The general procedure for MLE under constraints in the multivariate case is illustrated with a simple numeric example. The general theory is given for independent multivariate normal samples, contextualized in terms of the approach followed throughout this thesis. The multivariate Behrens-Fisher problem, with several variations of assumptions, is considered. Estimation of the mean vectors or common mean vector in the presence of homogeneous covariance matrices, proportional covariance matrices and equal correlation coefficients are described and results compared. A model for proportional mean vectors and proportional covariance matrices is also formulated and illustrated. A longitudinal data set is used to illustrate the flexibility of the procedure in the case of two small multivariate normal samples. The model and implied constraints are essentially the same as the model presented in Chapter 4 for the generalized regression model. However, in this case the model makes provision for estimation based on a few, but more than one, independent multivariate observations. Proportional mean vectors and/or proportional covariance matrices with and without a Toeplitz correlation structure imposed additionally, are considered.

In Chapter 6, estimation for Wishart matrices under constraints, as a special case of the general procedure described in Chapter 2, is considered. Covariance matrices with given structure, e.g. patterned covariance matrices and covariance matrices with zeros, are discussed and illustrated with

numerical examples. The problem of testing for the homogeneity of covariance matrices of dependent multivariate normal samples is considered next. The theoretical basis is given and then illustrated with a numerical example. Finally, the test for independence, as well as a test for the homogeneity of covariance matrices of several dependent multivariate normals, is illustrated. It is shown how the latter can be adjusted to make provision for patterns in the covariance matrices at the same time.

In the Appendix, programs written in PROC IML (SAS Institute, 2004) are provided for every numeric example considered in this thesis.

To summarize, MLE of parameters and parametric functions in the case of even standard continuous distributions, belonging to the exponential class, is in certain cases difficult to perform or leads to intractable procedures, which often imply that the likelihood function has to be solved numerically. In this thesis a simple and elegant procedure is described which determines the exact mle's directly. This is done by considering the likelihood function as a restricted case of a broader class of likelihood functions. The procedure provides a new statistical methodology for MLE of parameter structures in the exponential class. Complicated maximization problems can be dealt with in a unified way.



# Chapter 2

## MLE under Constraints

The random vector  $\mathbf{t} : k \times 1$  belongs to the canonical exponential family if its probability density function is of the form (Barndorff-Nielsen (1982) or Brown (1986))

$$p(\mathbf{t}, \boldsymbol{\theta}) = a(\boldsymbol{\theta})b(\mathbf{t})\exp\{\boldsymbol{\theta}'\mathbf{t} - \kappa(\boldsymbol{\theta})\}, \quad \mathbf{t} \in R^k, \quad \boldsymbol{\theta} \in \aleph \quad (2.1)$$

where  $\boldsymbol{\theta} : k \times 1$  the canonical (natural) parameter,  $\mathbf{t} : k \times 1$  the canonical statistic and  $\aleph$  the natural parameter space for the canonical parameter  $\boldsymbol{\theta}$ . The function  $\kappa(\boldsymbol{\theta}) = -\ln a(\boldsymbol{\theta})$  is referred to as the cumulant generating function or the log Laplace transform. It is important to note that the statistic  $\mathbf{t}$  in (2.1) is a canonical statistic. If  $\mathbf{t}$  is a sufficient statistic in the regular exponential class, it can be transformed to canonical form.

The mean vector and covariance matrix of  $\mathbf{t}$  are given by  $E(\mathbf{t}) = \frac{\partial}{\partial \boldsymbol{\theta}} \kappa(\boldsymbol{\theta}) = \mathbf{m}$  and  $Cov(\mathbf{t}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \kappa(\boldsymbol{\theta}) = \mathbf{V}$ . In this case the maximum likelihood estimate (mle) of  $\mathbf{m}$  without any constraints is  $\widehat{\mathbf{m}} = \mathbf{t}$ .

The mle's of  $\mathbf{m}$  may in general not exist, except under constraints which are implied by a particular model. In this thesis, the mle's are not obtained from maximizing a likelihood function in terms of parameters. The mle's  $\widehat{\mathbf{m}}$  of  $\mathbf{m}$  are obtained under constraints imposed on the expected values of the canonical statistics according to the procedure described by Matthews & Crowther (1995,1998). Consider the constraints  $\mathbf{g}(\mathbf{m}) = \mathbf{0}$ , where  $\mathbf{g}$  is a vector valued function of  $\mathbf{m}$ . Then the mle of  $\mathbf{m}$  under the constraints can be obtained from

$$\widehat{\mathbf{m}} = \mathbf{t} - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^* \mathbf{g}(\mathbf{t}) \quad (2.2)$$

where  $\mathbf{G}_m = \frac{\partial \mathbf{g}(m)}{\partial m}$ ,  $\mathbf{G}_t = \frac{\partial \mathbf{g}(m)}{\partial m} \Big|_{m=t}$  and  $(\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^*$  is a generalized inverse of  $(\mathbf{G}_t \mathbf{V} \mathbf{G}_m')$ .

In general, the iterative procedure implies a double iteration over  $t$  and  $m$ . The first iteration stems from the Taylor series linearization of  $\mathbf{g}(t)$  and the second from the fact that  $\mathbf{V}$  may be a function of  $m$ . If  $\mathbf{g}(t)$  is a linear function, no first iteration is necessary. Similarly, if  $\mathbf{V}$  and  $\mathbf{G}_m$  are not functions of  $m$ , no second iteration over  $m$  is necessary. The procedure is initialized with the observed canonical statistic as the starting value for both  $t$  and  $m$ . Convergence is attained first over  $t$  and then over  $m$ . The converged value of  $t$  is used as the next approximation for  $m$ , with iteration over  $m$  starting at the observed  $t$ . The covariance matrix  $\mathbf{V}$  may be a function of  $m$ , in which case it is recalculated for each new value of  $m$  in the iterative procedure. Convergence over  $m$  yields  $\widehat{m}$ , the mle of  $m$  under the constraints. Note that in some instances the notation  $t$  is used specifically to represent the canonical statistic and in others to represent the converged value of  $t$  in the iterative process.

For the mle's to exist, the number of independent constraints,  $\nu = \text{rank}(\mathbf{G}_m \mathbf{V} \mathbf{G}_m')$ , should equal the difference between the number of elements of  $m$  and the number of unknown parameters of the constrained model. Constraints need not be independent (Matthews & Crowther, 1998). Care should be taken that the constraints are correctly formulated. As with any iterative procedure, the choice of starting values may influence convergence or rate of convergence - specifically (and only) in the multivariate case where the sample size is small relative to the dimension of the observations.

A generic algorithm for obtaining the mle's under constraints are outlined below:

**Algorithm 1 Obtaining mle's under constraints.**

- |          |  |
|----------|--|
| Step 1:  | Specify $t_0$ , the vector of observed canonical statistics.   |
| Step 2:  | Let $t = t_0$ .  |
| Step 3A: | Let $m = t$ , $t = t_0$ .<br>Calculate $\mathbf{G}_m$ , $\mathbf{V}$ .   |
| Step 3B: | Let $t_p = t$ .<br>Calculate $\mathbf{g}(t)$ , $\mathbf{G}_t$ .<br>Calculate $t = t - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^* \mathbf{g}(t)$ .<br>If $(t - t_p)'(t - t_p) < \epsilon$ (a small positive number determining the accuracy)<br>then go to Step 3A,<br>else repeat Step 3B. |
| Step 4   | If $(m - t)'(m - t) < \epsilon$<br>then convergence is attained.   |

The exact mle's are given by  $m$ .

The asymptotic covariance matrix of  $\widehat{\mathbf{m}}$  is given by

$$Cov(\widehat{\mathbf{m}}) = \mathbf{V} - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_m \mathbf{V} \mathbf{G}_m')^* \mathbf{G}_m \mathbf{V} \quad (2.3)$$

which is estimated by replacing  $\mathbf{m}$  with  $\widehat{\mathbf{m}}$ . Estimates of parameters and parameter structures implied by different model(s) are obtained as functions of the elements of  $\widehat{\mathbf{m}}$ , say  $h(\widehat{\mathbf{m}})$ . Using the multivariate delta method, it follows that

$$Cov(h(\widehat{\mathbf{m}})) \cong \left( \frac{\partial h(\mathbf{m})}{\partial \mathbf{m}} \right) Cov(\widehat{\mathbf{m}}) \left( \frac{\partial h(\mathbf{m})}{\partial \mathbf{m}} \right)'. \quad (2.4)$$

The standard errors of the estimators are given by  $\sigma(h(\widehat{\mathbf{m}}))$ , the square root of the vector of diagonal elements of  $Cov(h(\widehat{\mathbf{m}}))$ .

The Wald statistic

$$W = (\mathbf{g}(\mathbf{t}))' (\mathbf{G}_t \mathbf{V} \mathbf{G}_t')^* \mathbf{g}(\mathbf{t}) \quad (2.5)$$

is a measure of goodness of fit under the hypothesis  $\mathbf{g}(\mathbf{m}) = \mathbf{0}$ . It is asymptotically  $\chi^2$  distributed with  $\nu$  degrees of freedom.

Different models require different canonical statistics. In general, the appropriate canonical statistics and corresponding covariance matrix are special cases of the following theoretical result and is utilized repeatedly throughout this thesis:

Suppose that  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  represent  $n$  independent observations of a random sample from a  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution with  $\bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$  the sample mean vector and  $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i'$  the matrix of mean sums of squares and products. The likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \det(2\pi\boldsymbol{\Sigma})^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \left[ \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})' \right] \right\} \\ &= \exp \left\{ n \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \right) - \frac{n}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' \right) - \frac{n}{2} \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{n}{2} \ln[\det(2\pi\boldsymbol{\Sigma})] \right\} \\ &= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})] \end{aligned}$$

with vector of canonical statistics,  $\mathbf{t}$ , and vector of corresponding canonical parameters,  $\boldsymbol{\theta}$ , given by:

$$\mathbf{t} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \\ \text{vec} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' \right) \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{y}} \\ \text{vec}(\mathbf{S}) \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} n \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ -\frac{n}{2} \text{vec}(\boldsymbol{\Sigma}^{-1}) \end{pmatrix}$$

where the matrix operation  $\text{vec}$  creates a column vector from the columns of its matrix argument.

Then

$$E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu} \\ \text{vec}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \text{ and } \text{Cov}(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \text{ where:}$$

$$\begin{aligned} \mathbf{V}_{11} &= \frac{1}{n}\boldsymbol{\Sigma} \\ \mathbf{V}_{21} &= \frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma}) \quad (\text{Lemma 1}) \\ \mathbf{V}_{12} &= \mathbf{V}_{21}' \end{aligned} \tag{2.6}$$

$$\mathbf{V}_{22} = \frac{1}{n}(I_{p^2} + \mathbf{K})[\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \boldsymbol{\Sigma}] \quad (\text{Muirhead (1982, page 518)})$$

where the matrix  $\mathbf{K}$  is given by  $\mathbf{K} = \sum_{i,j=1}^p (\mathbf{H}_{ij} \otimes \mathbf{H}'_{ij})$  and  $\mathbf{H}_{ij} : p \times p$  with  $h_{ij} = 1$  and all other elements equal to zero (Muirhead (1982, page 90)).

**Lemma 1**  $\mathbf{V}_{21} = \frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma})$ .

**Proof.** Let  $\mathbf{z}_i = \mathbf{y}_i - \boldsymbol{\mu}$  where  $\mathbf{y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

$$\begin{aligned} \mathbf{V}_{21} &= \text{Cov}[\text{vec}(\mathbf{S}), \bar{\mathbf{y}}'] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Cov}[\text{vec}(\mathbf{y}_i \mathbf{y}_i'), \mathbf{y}_i'] \\ &= \frac{1}{n} \text{Cov}[(\mathbf{y}_i \otimes \mathbf{y}_i), \mathbf{y}_i'] \\ &= \frac{1}{n} \text{Cov}[(\mathbf{z}_i + \boldsymbol{\mu}) \otimes (\mathbf{z}_i + \boldsymbol{\mu}), (\mathbf{z}_i + \boldsymbol{\mu})'] \\ &= \frac{1}{n} [\text{Cov}(\mathbf{z}_i \otimes \boldsymbol{\mu}, \mathbf{z}_i') + \text{Cov}(\boldsymbol{\mu} \otimes \mathbf{z}_i, \mathbf{z}_i')] \text{ since } \text{Cov}(\mathbf{z}_i \otimes \mathbf{z}_i, \mathbf{z}_i') = 0 \\ &= \frac{1}{n} [\text{Cov}(\mathbf{z}_i, \mathbf{z}_i') \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \text{Cov}(\mathbf{z}_i, \mathbf{z}_i')]. \blacksquare \end{aligned}$$

Note that in specific applications the roles of  $n$  and  $p$  are interchanged.

# Chapter 3

## MLE for the Linear Model with Heterogeneous Variances

Standard estimation procedures in the case of simple linear regression and analysis of variance fail when the assumption of constant variance is not met (Bowermann & O'Connell, 1990). Variance-equalizing transformations and weighted least squares are methods proposed to overcome this problem. Often models in the presence of non-constant variance are considered as special cases of the linear mixed model (Littell et. al., 2006). Within the mixed model framework, mle's of parameters are obtained using a numeric maximization procedure such as Newton-Raphson or the EM algorithm (McLachlan & Krishnan, 1996). Very often, calculation of likelihood-based inference is difficult (McCulloch & Searle, 2001).

In this chapter MLE under constraints is discussed as a unified approach to estimation of parameters and parametric functions in the exponential class, permitting heterogeneity of variance(s). Heterogeneity is automatically dealt with by simultaneously modelling the means and variances. It is important to note that the procedure presented in this chapter is based on maximum likelihood under constraints but should not be confused with REML (Restricted MLE). The examples considered in this chapter are chosen to illustrate the application of the procedure and its potential for generalizations in a wide variety of problems with heterogeneous variances. In some instances these problems are known to be solved in specific and distinctive ways, in which case its inclusion here is for illustrative purposes. Simple computer programs for each numeric example considered, are given in the Appendix with the objective of demonstrating the basic procedure.

Linear regression and analysis of variance (ANOVA) in the presence of non-constant variance are

illustrated in the case of normally distributed observations. Unbalanced data in the case of ANOVA are automatically dealt with. The linear mixed model features as a special case in the realm of MLE under constraints. Examples are also included which illustrate linear regression with gamma distributed dependent observations, using the same procedure.

### 3.1 Introductory Example

As an illustration of the estimation procedure, let the elements of

$\mathbf{y}' = (2, 3, 5, 6, 9)$  be an observed sample from a univariate  $n(\mu, \sigma^2)$  distribution. The mle's are  $\hat{\mu} = 5$  and  $\hat{\sigma}^2 = 6$ . If it is however known that  $\mu = 4$ , then direct calculation shows that the mle of  $\sigma^2$  is  $\hat{\sigma}^2 = 7$ . To illustrate the procedure in (2.2), the likelihood function is (Brown, 1986)

$$\begin{aligned} L(\mu, \sigma^2) &= \exp \left\{ \frac{n\mu}{\sigma^2} \left( \frac{1}{n} \sum y_i \right) - \frac{n}{2\sigma^2} \left( \frac{1}{n} \sum y_i^2 \right) - \left( \frac{n\mu^2}{2\sigma^2} + n \ln \sigma + \frac{n}{2} \ln 2\pi \right) \right\} \\ &= \exp \{ \boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta}) \} \end{aligned}$$

where  $\boldsymbol{\theta} = \begin{pmatrix} \frac{n\mu}{\sigma^2} \\ -\frac{n}{2\sigma^2} \end{pmatrix}$  and  $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum y_i \\ \frac{1}{n} \sum y_i^2 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ s^2 \end{pmatrix}$ . In this case  $\mathbf{t}$  is the canonical

statistic. (The vector  $\begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum (y_i - \bar{y})^2 \end{pmatrix}$  is not a canonical statistic.)

Consequently,  $E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \mu \\ \sigma^2 + \mu^2 \end{pmatrix}$  with

$$\mathbf{V} = \text{cov}(\mathbf{t}) = \begin{pmatrix} \frac{1}{n}\sigma^2 & \frac{2}{n}\mu\sigma^2 \\ \frac{2}{n}\mu\sigma^2 & \frac{2}{n}\sigma^4 + \frac{4}{n}\mu^2\sigma^2 \end{pmatrix} \text{ (cf. (2.6) with } p = 1\text{)}.$$

The restriction is  $\mu = 4$ , or  $g(\mathbf{m}) = \mu - 4 = m_1 - 4$ .

Then  $\mathbf{G}_m = \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}} = (1 \ 0) = \mathbf{G}_t = \mathbf{C}$ . Consequently,  $(\mathbf{C}\mathbf{V})' = \frac{1}{n} \begin{pmatrix} \sigma^2 \\ 2\mu\sigma^2 \end{pmatrix}$  and

$\mathbf{C}\mathbf{V}\mathbf{C}' = \frac{\sigma^2}{n}$  so that  $(\mathbf{C}\mathbf{V})'(\mathbf{C}\mathbf{V}\mathbf{C}')^{-1} = \begin{pmatrix} 1 \\ 2m_1 \end{pmatrix}$ . The starting value for both  $\mathbf{m}$  and  $\mathbf{t}$  is

$$\mathbf{t} = \begin{pmatrix} \bar{y} \\ s^2 \end{pmatrix} = \begin{pmatrix} 5 \\ 31 \end{pmatrix}.$$

In the (single) iterative procedure the mean vector under the constraint  $\mu = 4$  is calculated using:

$$\begin{aligned}\widehat{\mathbf{m}} &= \mathbf{t} - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^{-1} \mathbf{g}(\mathbf{t}) \\ &= \mathbf{t} - \begin{pmatrix} 1 \\ 2m_1 \end{pmatrix} (t_1 - 4).\end{aligned}$$

Iteration is performed only over  $\mathbf{m}$  since  $g(\mathbf{m}) = m_1 - 4$  is linear in  $\mathbf{m}$ :

$$\text{Iteration 1: } \begin{pmatrix} 5 \\ 31 \end{pmatrix} - \begin{pmatrix} 1 \\ 10 \end{pmatrix} 1 = \begin{pmatrix} 4 \\ 21 \end{pmatrix}$$

$$\text{Iteration 2: } \begin{pmatrix} 5 \\ 31 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \end{pmatrix} 1 = \begin{pmatrix} 4 \\ 23 \end{pmatrix}.$$

Since  $g(\mathbf{t}) = t_1 - 4 = 0$ , convergence is obtained. The restricted mle of  $\sigma^2$  follows as:

$$\widehat{\sigma}^2 = \widehat{m}_2 - \widehat{m}_1^2 = 23 - 16 = 7.$$

An equivalent way of specifying the restriction  $\mu = 4$ , would be  $g(\mathbf{m}) = \sqrt{m_1} - 2 = 0$ . In this case

$$\mathbf{G}_m = \begin{pmatrix} \frac{1}{2\sqrt{m_1}} & 0 \end{pmatrix} \text{ and } \mathbf{G}_t \mathbf{V} \mathbf{G}_m' = \frac{\sigma^2}{2n\sqrt{t_1 m_1}}$$

and the iterative procedure is

$$\widehat{\mathbf{m}} = \mathbf{t} - 2t_1 \left(1 - \frac{2}{\sqrt{t_1}}\right) \begin{pmatrix} 1 \\ 2m_1 \end{pmatrix}.$$

This would imply a double iteration over  $\mathbf{t}$  and  $\mathbf{m}$ . The first iteration over  $\mathbf{t}$  with starting value for  $\mathbf{m} = \begin{pmatrix} 5 \\ 31 \end{pmatrix}$ , results in:

$$\text{Iteration 1: } \begin{pmatrix} 5 \\ 31 \end{pmatrix} - 2 \times 5 \left(1 - \frac{2}{\sqrt{5}}\right) \begin{pmatrix} 1 \\ 10 \end{pmatrix} = \begin{pmatrix} 3.94427 \\ 20.44272 \end{pmatrix}$$

$$\text{Iteration 2: } \begin{pmatrix} 3.94427 \\ 20.44272 \end{pmatrix} - 2 \times 3.94427 \left(1 - \frac{2}{\sqrt{3.94427}}\right) \begin{pmatrix} 1 \\ 10 \end{pmatrix} = \begin{pmatrix} 3.99980 \\ 20.99805 \end{pmatrix}$$

$$\text{Iteration 3: } \begin{pmatrix} 3.99980 \\ 20.99805 \end{pmatrix} - 2 \times 3.99980 \left(1 - \frac{2}{\sqrt{3.99980}}\right) \begin{pmatrix} 1 \\ 10 \end{pmatrix} = \begin{pmatrix} 4 \\ 21 \end{pmatrix}$$

The second iteration over  $\mathbf{t}$  with  $\mathbf{m} = \begin{pmatrix} 4 \\ 21 \end{pmatrix}$ , proceeds as follows:

$$\begin{aligned}
\text{Iteration 1: } & \begin{pmatrix} 5 \\ 31 \end{pmatrix} - 2 \times 5 \left(1 - \frac{2}{\sqrt{5}}\right) \begin{pmatrix} 1 \\ 2 \times 4 \end{pmatrix} = \begin{pmatrix} 3.94427 \\ 22.55416 \end{pmatrix} \\
\text{Iteration 2: } & \begin{pmatrix} 3.94427 \\ 22.55416 \end{pmatrix} - 2 \times 3.94427 \left(1 - \frac{2}{\sqrt{3.94427}}\right) \begin{pmatrix} 1 \\ 2 \times 4 \end{pmatrix} = \begin{pmatrix} 3.99980 \\ 22.99844 \end{pmatrix} \\
\text{Iteration 3: } & \begin{pmatrix} 3.99980 \\ 22.99844 \end{pmatrix} - 2 \times 3.99980 \left(1 - \frac{2}{\sqrt{3.99980}}\right) \begin{pmatrix} 1 \\ 2 \times 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 23 \end{pmatrix}.
\end{aligned}$$

Since  $g(t) = \sqrt{t_1} - 2 = 0$ , convergence has been obtained. This results in  $\widehat{\mathbf{m}} = \begin{pmatrix} 4 \\ 23 \end{pmatrix}$ , the same result obtained with  $g(\mathbf{m}) = m_1 - 4$ .

SAS program: UNIV.SAS

### 3.2 MLE for Linear Models under Normality

Consider a random sample  $y_i$ ,  $i = 1, 2, \dots, n$  from independent normal distributions with  $E(y_i) = \mu_i$  and  $\text{var}(y_i) = \sigma_i^2$ . The joint density function is given by

$$\begin{aligned}
f(y_1, \dots, y_n) &= \prod_{i=1}^n \exp \left\{ \frac{\mu_i}{\sigma_i^2} y_i - \frac{1}{2\sigma_i^2} y_i^2 - \left( \frac{\mu_i^2}{2\sigma_i^2} + \ln \sigma_i + \frac{1}{2} \ln 2\pi \right) \right\} \\
&= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})]
\end{aligned} \tag{3.1}$$

where

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu}/\boldsymbol{\sigma}^2 \\ -\mathbf{1}/(2\boldsymbol{\sigma}^2) \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mathbf{y} \\ \mathbf{y}^2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{y}^2 = \begin{pmatrix} y_1^2 \\ \vdots \\ y_n^2 \end{pmatrix}, \quad \boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_n^2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

and  $\boldsymbol{\mu}/\boldsymbol{\sigma}^2 = \begin{pmatrix} \frac{\mu_1}{\sigma_1^2} \\ \vdots \\ \frac{\mu_n}{\sigma_n^2} \end{pmatrix}$ ,  $\mathbf{1}/\boldsymbol{\sigma}^2 = \begin{pmatrix} \frac{1}{\sigma_1^2} \\ \vdots \\ \frac{1}{\sigma_n^2} \end{pmatrix}$  indicate elementwise division. (This notation is repeatedly used in the sequel.)

$$\text{Consequently, } E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\sigma}^2 + \boldsymbol{\mu}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}. \tag{3.2}$$

The covariance matrix is given by (2.6) with  $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma}^2)$ :

$$\text{Cov}(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \text{diag}(\boldsymbol{\sigma}^2/n) & 2\text{diag}(\boldsymbol{\mu}\boldsymbol{\sigma}^2/n) \\ 2\text{diag}(\boldsymbol{\mu}\boldsymbol{\sigma}^2/n) & 2\text{diag}[(\boldsymbol{\sigma}^4 + 2\boldsymbol{\mu}^2\boldsymbol{\sigma}^2)/n] \end{pmatrix}. \tag{3.3}$$



This model also represents a generalization of the univariate case considered in Section 3.1. The parameters  $\boldsymbol{\mu}$  and  $\sigma^2$  may be modelled simultaneously. In terms of linear models, this reduces to

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ \sigma^2 &= \mathbf{Z}\boldsymbol{\alpha}\end{aligned}\tag{3.4}$$

where  $\mathbf{X}$  and  $\mathbf{Z}$  are appropriate design matrices. In order to determine the mle of  $\mathbf{m}$  under the model (3.4), the restrictions may be formulated as  $\mathbf{g}(\mathbf{m}) = \begin{pmatrix} g_1(\mathbf{m}) \\ g_2(\mathbf{m}) \end{pmatrix}$  with

$$g_1(\mathbf{m}) = \mathbf{Q}_X\boldsymbol{\mu} = \mathbf{Q}_X\mathbf{m}_1 \quad , \quad g_2(\mathbf{m}) = \mathbf{Q}_Z\sigma^2 = \mathbf{Q}_Z(\mathbf{m}_2 - \mathbf{m}_1^2)\tag{3.5}$$

where  $\mathbf{Q}_X$  and  $\mathbf{Q}_Z$  specify the contrast matrices or corresponding projection matrices implied by (3.4).

In many situations the columns of  $\mathbf{X}$  and  $\mathbf{Z}$  generate the same vector space and  $\mathbf{Q}_X = \mathbf{Q}_Z$ . The corresponding derivatives with respect to  $\mathbf{m}$  are

$$\mathbf{G}_{1m} = (\mathbf{Q}_X, \mathbf{0}), \quad \mathbf{G}_{2m} = (-2\mathbf{Q}_Z \text{diag}(\mathbf{m}_1), \mathbf{Q}_Z), \quad \mathbf{G}_m = \begin{pmatrix} \mathbf{G}_{1m} \\ \mathbf{G}_{2m} \end{pmatrix}.\tag{3.6}$$

The mle of  $\mathbf{m}$  subject to the constraints  $\mathbf{g}(\mathbf{m}) = \mathbf{0}$ , is obtained iteratively from (2.2):

$$\widehat{\mathbf{m}} = \mathbf{t} - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_m \mathbf{V} \mathbf{G}_m')^* \mathbf{g}(\mathbf{t}).$$

In order to obtain the asymptotic variances of  $\widehat{\boldsymbol{\mu}}$ ,  $\widehat{\sigma}^2$ ,  $\widehat{\boldsymbol{\beta}}$  and  $\widehat{\boldsymbol{\alpha}}$ , let

$$\mathbf{h}(\mathbf{m}) = \begin{pmatrix} \boldsymbol{\mu} \\ \sigma^2 \\ \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 - \mathbf{m}_1^2 \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{m}_1 \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\sigma^2 \end{pmatrix}.\tag{3.7}$$

Using the multivariate delta method, it follows that

$$\text{Cov}(\mathbf{h}(\widehat{\mathbf{m}})) \approx \left( \frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} \right)' \text{Cov}(\widehat{\mathbf{m}}) \left( \frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} \right)\tag{3.8}$$

where

$$\frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -2\text{diag}(\mathbf{m}_1) & \mathbf{I} \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & \mathbf{0} \\ -2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\text{diag}(\mathbf{m}_1) & (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix}.\tag{3.9}$$

The asymptotic variances,  $\widehat{\sigma}^2(\mathbf{h}(\widehat{\mathbf{m}}))$ , of the elements of  $\mathbf{h}(\widehat{\mathbf{m}})$ , are given by the diagonal elements of  $\text{Cov}(\mathbf{h}(\widehat{\mathbf{m}}))$ .

The following two examples are used to illustrate estimation in case of both the ordinary linear model and the linear model with heterogeneous variances. It is shown how complex parameter structures can be dealt with through a straightforward adjustment in the basic procedure. McCulloch & Searle (2001) give a comprehensive overview of likelihood equations and their solutions for similar models using the traditional maximum likelihood approach.

### 3.2.1 Heteroskedastic linear regression

Polasek, Liu & Neudecker (1999) illustrate the modelling of heteroskedasticity using a set of hypothetical income and consumption data. They followed a three stage approach to obtain a generalized least squares solution. The same data (Gujarati, 2003), given in Table 3.1, are used here to illustrate the estimation of parameters using (2.2) which yields the results directly. Consumption,  $y$ , is assumed to be normally distributed.

Table 3.1: Hypothetical income and consumption data.

|                  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Consumption, $y$ | 55  | 70  | 75  | 65  | 74  | 80  | 84  | 79  | 90  | 98  | 95  | 108 | 113 | 110 | 125 |
| Income, $x$      | 80  | 85  | 90  | 100 | 105 | 110 | 115 | 120 | 125 | 130 | 140 | 145 | 150 | 160 | 165 |
| Consumption, $y$ | 115 | 130 | 135 | 120 | 140 | 144 | 152 | 140 | 137 | 145 | 175 | 189 | 180 | 178 | 191 |
| Income, $x$      | 180 | 185 | 190 | 200 | 205 | 210 | 220 | 225 | 230 | 240 | 245 | 250 | 260 | 265 | 270 |

Means and variances are modelled simultaneously using (3.2) to (3.6). The results for different models, are summarized in Table 3.2. Different models only require specification of the appropriate design matrices. Note that Model 1, assuming equal variances, yields the least squares solution. For Model 2, linearity in both the expected consumption and the variance in consumption is assumed. Model 3 gives the mle's corresponding to the model estimates obtained by Polasek, Liu & Neudecker (1999). A parabolic trend on the variance is allowed for Model 4 (through the origin), as well as Model 5. Asymptotic standard deviations,  $\hat{\sigma}$ , are obtained using (3.7) to (3.9). Compared to Model 1, the other models show a drastic increase in goodness of fit as implied by the Wald statistic,  $W$  (cf. (2.5)).

SAS program: REGRESSION.SAS

Table 3.2: Estimated parameters for heteroskedastic linear regression model.

| Model |   | $\hat{\beta}$ | $\hat{\sigma}_{\hat{\beta}}$ | $\hat{\alpha}$ | $\hat{\sigma}_{\hat{\alpha}}$ | $W$              | $\nu$ |
|-------|---|---------------|------------------------------|----------------|-------------------------------|------------------|-------|
| 1     | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$               | 9.29          | 5.05                         | 78.71          | 20.32                         | 59.67            | 57    |
|       | $\mathbf{Z} = (\mathbf{1}_n)$                           | 0.64          | 0.03                         |                |                               | ( $p = 0.3789$ ) |       |
| 2     | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$               | 10.09         | 3.89                         | -19.55         | 37.94                         | 47.17            | 56    |
|       | $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x})$               | 0.63          | 0.03                         | 0.54           | 0.30                          | ( $p = 0.7936$ ) |       |
| 3     | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$               | 10.29         | 3.89                         | 12.91          | 20.33                         | 45.85            | 56    |
|       | $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}^2)$             | 0.63          | 0.03                         | 0.002          | 0.8315                        | ( $p = 0.8315$ ) |       |
| 4     | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$               | 10.28         | 3.65                         | 0.003          | 0.001                         | 45.45            | 57    |
|       | $\mathbf{Z} = (\mathbf{x}^2)$                           | 0.63          | 0.03                         |                |                               | ( $p = 0.8646$ ) |       |
| 5     | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$               | 10.40         | 4.16                         | 171.02         | 138.16                        | 44.27            | 55    |
|       | $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{x}^2)$ | 0.63          | 0.03                         | -2.41          | 1.94                          | ( $p = 0.8495$ ) |       |
|       |   |               |                              | 0.01           | 0.01                          |                  |       |

### 3.2.2 Two-way analysis of variance: unbalanced data

Where the unbalanced case usually requires the derivation of its own set of likelihood equations, it is automatically dealt with using the procedure for MLE under constraints.

Let  $y_{ijk}$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ ,  $k = 1, \dots, n_{ij}$  represent the observations in  $IJ$  independent observed samples from  $n(\eta_{ij}, \sigma_{ij}^2)$  distributed populations. Consider the model

$$E(y_{ijk}) = \eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \text{ and } \text{var}(y_{ijk}) = \sigma_{ij}^2 \quad (3.10)$$

with restrictions  $\sum_i v_i \alpha_i = 0$ ,  $\sum_j w_j \beta_j = 0$ ,  $\sum_i v_i \gamma_{ij} = 0$ ,  $\sum_j w_j \gamma_{ij} = 0$  where the  $\{v_i\}$  and  $\{w_j\}$  represent different sets of weights such that  $\sum v_i = 1$ ,  $\sum w_j = 1$  and  $\sum_i \sum_j n_{ij} = n$ . The restrictions and  $n_{ij}$  should be such that the parameters for the model considered, are estimable. More complicated restrictions may be necessary in order to ensure estimability. The observation  $y_{ijk}$  is the  $k$ -th observation at the  $i$ -th level of factor  $A$  and  $j$ -th level of factor  $B$ .

The variance in each cell can similarly be decomposed in terms of variance effects associated with the levels of each factor. This can be modelled as

$$\sigma_{ij}^2 = \sigma^2 + \delta_i + \zeta_j + \tau_{ij} \quad (3.11)$$

$$\sum_i v_i \delta_i = 0, \sum_j w_j \zeta_j = 0, \sum_i v_i \tau_{ij} = 0, \sum_j w_j \tau_{ij} = 0$$

where, in terms of the variances,  $\sigma^2$  represents an overall mean effect,  $\delta_i$  represents the effect due to the  $i$ -th level of Factor A,  $\zeta_j$  the effect due to the  $j$ -th level of Factor B and  $\tau_{ij}$  the effect due to the interaction between the  $i$ -th level of Factor A and the  $j$ -th level of Factor B.

In this case the canonical statistic is:

$$\mathbf{t} = \begin{pmatrix} \mathbf{y}_{11} \\ \vdots \\ \mathbf{y}_{IJ} \\ \mathbf{y}_{11}^2 \\ \vdots \\ \mathbf{y}_{IJ}^2 \end{pmatrix}, \mathbf{y}_{ij} = \begin{pmatrix} y_{ij1} \\ \vdots \\ y_{ijn_{ij}} \end{pmatrix}, \begin{matrix} i = 1, \dots, I \\ j = 1, \dots, J \end{matrix} \text{ with } \mathbf{m} = \begin{pmatrix} \eta_{11} \\ \vdots \\ \eta_{IJ} \\ \sigma_{11}^2 \mathbf{I} + \eta_{11}^2 \\ \vdots \\ \sigma_{IJ}^2 \mathbf{I} + \eta_{IJ}^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\eta} \\ \sigma^2 + \boldsymbol{\eta}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}$$

and from (2.6) with  $\boldsymbol{\Sigma} = \text{diag}(\sigma^2)$  and  $n = 1$ , it follows that

$$\text{Cov}(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \text{diag}(\sigma^2) & 2\text{diag}(\boldsymbol{\eta}\sigma^2) \\ 2\text{diag}(\boldsymbol{\eta}\sigma^2) & 2\text{diag}[(\sigma^4 + 2\boldsymbol{\eta}^2\sigma^2)] \end{pmatrix}.$$

The means and variances are modelled as:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \text{ and } \sigma^2 = \mathbf{X}\boldsymbol{\alpha} \quad (3.12)$$

where  $\mathbf{X} = (\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3)$ , and each submatrix  $\mathbf{X}_i$  corresponds to an effect:  $\mathbf{X}_0 : n \times 1$  to the overall mean,  $\mathbf{X}_1 : n \times (I - 1)$  to Factor A,  $\mathbf{X}_2 : n \times (J - 1)$  to Factor B and  $\mathbf{X}_3 : n \times (I - 1)(J - 1)$  to the interaction between Factor A and Factor B. The mle's for any model of the form (3.12) are obtained by specifying the implied constraints using the appropriate projection matrices (cf. (3.5)). In similar fashion, linear and non-linear hypotheses may be tested by specifying additional constraints.

The following example of a heterogeneous variance model is discussed by Littell et. al. (2006). A study was conducted to evaluate the effect of age and gender on the amount of time the TV was watched by 27 children in a play area of a day-care facility. The data are given in Table 3.3. The means and variances are modelled simultaneously. The results obtained using MLE under constraints for four different models are given in Table 3.4. The mle's for the saturated model are given first (Model 1). Model 2 gives the mle's under the null hypothesis of no interaction effect on the means, assuming equal variances. In Model 3, the results for the same model but without the assumption of

Table 3.3: Amount of time TV watched.

|        |   | Age               |                |             |
|--------|---|-------------------|----------------|-------------|
|        |   | 2                 | 3              | 4           |
| Gender | F | 12 8 9 7          | 19 20 16 11 14 | 22 32 39 12 |
|        | M | 10 13 14 11 11 15 | 17 23 22 16    | 24 31 37 49 |

equal variances are reported. Using PROC MIXED and REML, Littell et. al. (2006) fitted the saturated model (for the means) with unequal variances for each age group. The corresponding mle's using the procedure for maximum likelihood under constraints are the last reported in the table (Model 4). These mle's agree with the mle's obtained using PROC MIXED with METHOD=ML. From the Wald statistic (cf. (2.5)) it follows that this model fits much better than Models 1 and 2.

SAS program: ANOVA.SAS

Table 3.4: Estimated parameters for two-way ANOVA with unbalanced data.

| Model                           |   | Parameters |            |           |           |               |               |
|---------------------------------|---|------------|------------|-----------|-----------|---------------|---------------|
|                                 |   | $\mu$      | $\alpha_1$ | $\beta_1$ | $\beta_2$ | $\gamma_{11}$ | $\gamma_{12}$ |
| 1                               | $\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$        | 19.72      | -2.64      | -9.06     | -1.97     | 0.97          | 0.89          |
|                                 | $\sigma_{ij}^2 = \sigma^2 + \delta_i + \zeta_j + \tau_{ij}$ | 35.86      | 3.64       | -32.50    | -25.83    | -3.50         | -2.86         |
| $W = 56.24 \nu = 42 (p = 0.07)$ |   |            |            |           |           |               |               |
| 2                               | $\eta_{ij} = \mu + \alpha_i + \beta_j$                      | 19.69      | -2.55      | -9.20     | -1.85     | 0             | 0             |
|                                 | $\sigma_{ij}^2 = \sigma^2$                                  | 34.13      | 0.00       | 0.00      | 0.00      | 0.00          | 0.00          |
| $W = 193.40 \nu = 50 (p = 0)$   |   |            |            |           |           |               |               |
| 3                               | $\eta_{ij} = \mu + \alpha_i + \beta_j$                      | 19.82      | -1.75      | -9.17     | -2.07     | 0             | 0             |
|                                 | $\sigma_{ij}^2 = \sigma^2 + \delta_i + \zeta_j + \tau_{ij}$ | 38.42      | 4.22       | -35.06    | -28.40    | -4.08         | -3.45         |
| $W = 65.47 \nu = 45 (p = 0.02)$ |   |            |            |           |           |               |               |
| 4                               | $\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$        | 19.72      | -2.64      | -9.06     | -1.97     | 0.97          | 0.89          |
|                                 | $\sigma_{ij}^2 = \sigma^2 + \zeta_j$                        | 35.88      | -32.54     | -25.77    | 0         | 0             | 0             |
| $W = 57.27 \nu = 46 (p = 0.12)$ |   |            |            |           |           |               |               |

### 3.3 MLE for (Non)Linear Models in the case of Non-Normal Observations

Models for the analysis of non-normal data using nonlinear models fall within the generalized linear models framework (McCulloch & Searle, 2001). In this section it is illustrated how non-normal observations can be accommodated in the linear model without the usual complexities associated with the traditional maximum likelihood approach. In addition, the second example illustrates how easily a nonlinear model may be fitted using the same approach.

#### 3.3.1 Exponentially distributed observations

To illustrate application of the procedure to estimate a regression function in the case of non-normal observations in the exponential class, let the elements  $y_i$ ,  $i = 1, \dots, n$  of  $\mathbf{y}$  be a random sample from an  $\exp(\beta_i)$  distribution. Suppose  $\mathbf{y} \sim \exp(\boldsymbol{\beta})$  where  $\mathbf{m} = E(\mathbf{y}) = \mathbf{X}\boldsymbol{\delta} = \boldsymbol{\beta}$  with covariance matrix  $\mathbf{V} = \text{diag}(\boldsymbol{\beta}^2) = \mathbf{D}_{\boldsymbol{\beta}^2}$ . The mle's of the parameters are obtained by specifying  $g(\mathbf{m}) = \mathbf{Q}_X \mathbf{m}$ ,  $\mathbf{m} = \mathbf{m}_1$  using the general notation in (3.2), and  $\mathbf{G}_m = \mathbf{Q}_X$  where  $\mathbf{Q}_X$  projects orthogonal to the columns of  $\mathbf{X}$ . The (single) iterative procedure now entails

$$\widehat{\mathbf{m}} = \mathbf{y} - (\mathbf{Q}_X \mathbf{D}_{\boldsymbol{\beta}^2})' (\mathbf{Q}_X \mathbf{D}_{\boldsymbol{\beta}^2} \mathbf{Q}_X)^* \mathbf{Q}_X \mathbf{t} \quad (3.13)$$

The mle's with corresponding standard errors obtained for a hypothetical example are given in Table 3.5. For illustrative purposes a straight line model (Model 1), as well as a parabola (Model 2), was fitted to the data. The approximate standard errors are obtained using (2.3) and (2.4) where

$$\mathbf{h}(\mathbf{m}) = \boldsymbol{\delta} = \left( (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{m} \right) \text{ and } \frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} = \left( (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right).$$

The standard errors are the square roots of the diagonal elements of  $\text{Cov}(\mathbf{h}(\widehat{\mathbf{m}}))$ .

|                              |
|------------------------------|
| SAS program: EXPONENTIAL.SAS |
|------------------------------|

Table 3.5: Exponentially distributed observations with estimated model parameters.

| Sample  | $x$  | 2  | 4    | 6     | 8     | 10    | 12    | 14    | 16    |
|---|--|--|------|-------|-------|-------|-------|-------|-------|
|   | $y$  | 3  | 5    | 6     | 21    | 30    | 35    | 50    | 60    |
| Model 1 $\mathbf{X} = \begin{pmatrix} \mathbf{1}_8 & \mathbf{x} \end{pmatrix}$                | $\hat{\beta}$  | 2.56   | 8.60 | 14.63 | 20.67 | 26.70 | 32.73 | 38.77 | 44.80 |
|   | $\hat{\sigma}_{\hat{\beta}}$   | 2.49   | 3.10 | 5.28  | 7.81  | 10.44 | 13.11 | 15.80 | 18.50 |
| $W = 0.77 \nu = 6 (p = 0.993)$  | $\hat{\boldsymbol{\delta}}' = \begin{pmatrix} -3.47, & 3.02 \end{pmatrix}$         | $\hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\delta}}} = \begin{pmatrix} 4.20, & 1.36 \end{pmatrix}$         |      |       |       |       |       |       |       |
| Model 2 $\mathbf{X} = \begin{pmatrix} \mathbf{1}_8 & \mathbf{x} & \mathbf{x}^2 \end{pmatrix}$ | $\hat{\beta}$  | 2.97   | 4.75 | 9.07  | 15.93 | 25.35 | 37.31 | 51.81 | 68.87 |
|   | $\hat{\sigma}_{\hat{\beta}}$   | 2.91   | 3.50 | 5.29  | 6.92  | 10.39 | 16.98 | 26.69 | 39.34 |
| $W = 0.27 \nu = 5 (p = 0.998)$  | $\hat{\boldsymbol{\delta}}' = \begin{pmatrix} 3.75, & -1.02, & 0.32 \end{pmatrix}$ | $\hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\delta}}} = \begin{pmatrix} 8.26, & 3.85, & 0.34 \end{pmatrix}$ |      |       |       |       |       |       |       |

### 3.3.2 Gamma distributed observations and nonlinear models

Consider an observed sample  $\mathbf{y}' = (y_1, y_2, \dots, y_n)$  where the  $y_i$  are independently  $gam(\beta_i, \alpha)$  distributed, with  $\alpha$  the shape parameter and  $\boldsymbol{\beta}' = (\beta_1, \beta_2, \dots, \beta_n)$  the scale parameters. Under linear regression, it is assumed that

$$E(\mathbf{y}) = \alpha\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\delta}. \quad (3.14)$$

Note that the shape parameter  $\alpha$  is assumed to be constant over  $i$  and

$$cov(\mathbf{y}) = diag(\alpha\boldsymbol{\beta}^2) = \frac{1}{\alpha} diag[\mathbf{X}\boldsymbol{\delta}]^2. \quad (3.15)$$

The likelihood function is

$$L(\alpha, \boldsymbol{\beta}) = \prod_{i=1}^n \left\{ [\beta_i^\alpha \Gamma(\alpha)]^{-1} y_i^{\alpha-1} e^{-y_i/\beta_i} \right\} = exp[\boldsymbol{\theta}'\mathbf{t} - \kappa(\boldsymbol{\theta})] \quad (3.16)$$

with

$$\mathbf{t} = \begin{pmatrix} \mathbf{y} \\ \overline{\ln \mathbf{y}} \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \text{ and } \mathbf{m} = E(\mathbf{t}) = \begin{pmatrix} \mathbf{X}\boldsymbol{\delta} \\ \psi(\alpha) - \ln(\alpha) + \overline{\ln(\mathbf{m}_1)} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \quad (3.17)$$

where  $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$  is the logarithmic derivative of the gamma function (see Abramowitz & Stegun (1972, page 258)),  $\overline{\ln \mathbf{y}}$  the average of the natural logarithm of the elements of  $\mathbf{y}$  and  $E[\ln(y_i)]$  and consequently  $E(\overline{\ln \mathbf{y}})$ , are obtained from the moment-generating function of  $\ln Y_i$ :

$$M_{\ell n Y_i}(t) = E[e^{(\ell n y_i)t}] = E(y_i^t) = \frac{\Gamma(\alpha + t)}{\Gamma(\alpha)} \beta^t$$

$$\frac{\partial}{\partial t} M_{\ell n Y_i}(t) = \frac{\Gamma(\alpha + t)}{\Gamma(\alpha)} \psi(\alpha + t) \beta^t + \frac{\Gamma(\alpha + t)}{\Gamma(\alpha)} \beta^t \ell n \beta$$

Setting  $t = 0$  yields:

$$E(\ell n y_i) = \psi(\alpha) + \ell n \beta$$

$$\frac{\partial^2}{\partial t^2} M_{\ell n Y_i}(t) =$$

$$\frac{\{\Gamma(\alpha + t)[\psi(\alpha + t)]^2 \beta^t + \Gamma(\alpha + t) \psi'(\alpha + t) \beta^t + 2\Gamma(\alpha + t) \psi(\alpha + t) \beta^t \ell n \beta + \Gamma(\alpha + t) \beta^t (\ell n \beta)^2\}}{\Gamma(\alpha)}$$

Setting  $t = 0$  in the second-order derivative of  $M_{\ell n Y_i}(t)$  yields:

$$E[(\ell n y_i)^2] = [\psi(\alpha) + \ell n \beta]^2 + \psi'(\alpha)$$

$$E[y_i(\ell n y_i)] = \alpha \beta [\psi(\alpha + 1) + \ell n \beta]$$

$$\text{Cov}(y_i, \ell n y_i) = \alpha \beta [\psi(\alpha + 1) - \psi(\alpha)]$$

$$\text{and } \text{Var}(\ell n y_i) = \psi'(\alpha).$$

Furthermore, under the linear regression model (3.14), it is assumed that  $E(y_i) = \alpha \beta_i = (\mathbf{X} \boldsymbol{\delta})_i = f(x_i)$ , where  $f(x_i)$  a linear function of the elements of  $\boldsymbol{\delta}$ . Consequently,  $E(\ell n y_i) = \psi(\alpha) + \ell n \left[ \frac{f(x_i)}{\alpha} \right] = \psi(\alpha) - \ell n(\alpha) + \ell n[f(x_i)]$ .

The covariance matrix of  $\mathbf{t}$  is then given by

$$\mathbf{V} = \begin{pmatrix} \mathbf{D}_{\alpha \beta^2} & \boldsymbol{\beta}/n \\ \boldsymbol{\beta}'/n & \psi'(\alpha)/n \end{pmatrix} \quad (3.18)$$

since  $\alpha[\psi(\alpha + 1) - \psi(\alpha)] = 1$  (cf. the recurrence formula, Abramowitz & Stegun (1972, page 258)), and where  $\mathbf{D}_{\alpha \beta^2} = \text{diag}(\alpha \beta^2)$ .

Since  $m_2 = \psi(\alpha) - \ell n(\alpha) + \overline{\ell n(\mathbf{m}_1)}$ , it follows that

$$\frac{\exp(m_2)}{\exp(\overline{\ell n \mathbf{m}_1})} = \frac{\exp[\psi(\alpha)]}{\alpha} = k, \text{ say.} \quad (3.19)$$

Since  $\frac{\exp[\psi(\alpha)]}{\alpha}$  is a monotone increasing function (between 0 and 1), this relation can be used to determine the value of  $\alpha$  numerically. The model (3.14) implies

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_X \mathbf{m}_1 \\ m_2 - \overline{\ell n \mathbf{m}_1} - \ell n(k) \end{pmatrix} \quad (3.20)$$

where  $\mathbf{Q}_X$  projects orthogonal to the columns of  $\mathbf{X}$ . Then

$$\mathbf{G}_{1m} = \begin{pmatrix} \mathbf{Q}_X & \mathbf{0}_n \end{pmatrix} \text{ and } \mathbf{G}_{2m} = (0, 0, \dots, 1) - \frac{1}{n}(\mathbf{1}'_n / \mathbf{m}'_1, 0). \quad (3.21)$$



The parameter  $\alpha$  may be known, e.g. in the case where an exponential distribution is fitted ( $\alpha = 1$ ).

Nonlinear models can be fitted to the data by imposing linear constraints on the logarithm of the vector of observations. In this case

$$\mathbf{g}(\mathbf{m}) = \left( \mathbf{Q}_X \ln(\mathbf{m}_1) \right) \text{ and } \mathbf{G}_m = \left( \mathbf{Q}_X \text{diag}(1/\mathbf{m}_1), \mathbf{0}_n \right) \quad (3.22)$$

Consider again the hypothetical data used in Section 3.3.1. Assuming gamma-distributed dependent observations with a constant shape parameter yields the same regression estimates as those obtained when assuming exponentially distributed observations (cf. Model 1 in Table 3.6 and Model 1 in Table 3.5). Model 2 in Table 3.6 gives the regression estimates and mle for  $\beta$  under the constraint that  $\alpha = 1$  (cf. Model 1 in Table 3.5). Whatever the choice of  $\alpha$ , the regression function remains invariant. In Model 3 and Model 4 nonlinear curves were fitted to the data by specifying constraints on the natural logarithm of the observations. Model 5 gives the mle's of the parameters where  $y_i \sim \text{gam}(\beta, \alpha), i = 1, 2, \dots, n$ . These results are given in Table 3.6. Again, the standard errors follow from (2.3) and (2.4).

|                        |
|------------------------|
| SAS program: GAMMA.SAS |
|------------------------|

Table 3.6: Gamma distributed data with estimated parameters.

| Sample  | $x_i$   | 2      | 4      | 6      | 8      | 10     | 12     | 14     | 16     |
|---|---|--------|--------|--------|--------|--------|--------|--------|--------|
|   | $y_i$   | 3      | 5      | 6      | 21     | 30     | 35     | 50     | 60     |
| Model 1   | $\widehat{m}_1$   | 2.56   | 8.60   | 14.63  | 20.67  | 26.70  | 32.73  | 38.77  | 44.80  |
| $\mathbf{g} = \mathbf{Q}_X \mathbf{m}_1$  | $\widehat{\sigma}_{\widehat{m}_1}$  | 0.90   | 1.12   | 1.90   | 2.81   | 3.76   | 4.72   | 5.69   | 6.66   |
| $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$   | $\widehat{\alpha}$  | 7.709  |        |        |        |        |        |        |        |
|   | $\widehat{\beta}$   | 0.333  | 1.115  | 1.898  | 2.681  | 3.463  | 4.246  | 5.029  | 5.811  |
| $W = 5.948 \nu = 6 (p = 0.429), \widehat{\boldsymbol{\delta}} = (-3.47, 3.02) \widehat{\boldsymbol{\sigma}}'_{\widehat{\boldsymbol{\delta}}} = (1.51, 0.49)$            |   |        |        |        |        |        |        |        |        |
| Model 2   | $\widehat{m}_1$   | 2.56   | 8.60   | 14.63  | 20.67  | 26.70  | 32.73  | 38.77  | 44.80  |
| $\mathbf{g}' = (g_1, g_2)$  | $\widehat{\sigma}_{\widehat{m}_1}$  | 2.49   | 3.10   | 5.28   | 7.81   | 10.44  | 13.11  | 15.80  | 18.50  |
| $\mathbf{g}_1 = \mathbf{Q}_X \mathbf{m}_1$  | $\alpha$  | 1.000  |        |        |        |        |        |        |        |
| $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$   | $\widehat{\beta}$   | 2.56   | 8.60   | 14.63  | 20.67  | 26.70  | 32.73  | 38.77  | 44.80  |
| $g_2 = m_2 - \overline{\ln m_1}$  | $\widehat{\boldsymbol{\delta}} = (-3.47, 3.02) \widehat{\boldsymbol{\sigma}}'_{\widehat{\boldsymbol{\delta}}} = (4.20, 1.36)$ |        |        |        |        |        |        |        |        |
| $-\psi(1) + \ln(1)$   | $W = 2.761 \nu = 7 (p = 0.906)$   |        |        |        |        |        |        |        |        |
| Model 3   | $\widehat{m}_1$   | 3.53   | 5.55   | 8.75   | 13.78  | 21.71  | 34.21  | 53.89  | 84.89  |
| $\mathbf{g} = \mathbf{Q}_X \ln(\mathbf{m}_1)$   | $\widehat{\sigma}_{\widehat{m}_1}$  | 0.62   | 0.79   | 1.00   | 1.35   | 2.13   | 3.92   | 7.65   | 14.87  |
| $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$   | $\widehat{\alpha}$  | 13.571 |        |        |        |        |        |        |        |
|   | $\widehat{\beta}$   | 0.260  | 0.409  | 0.645  | 1.016  | 1.600  | 2.521  | 3.971  | 6.255  |
| $W = 8.893 \nu = 6 (p = 0.180), \widehat{\boldsymbol{\delta}} = (-19.68, 5.33) \widehat{\boldsymbol{\sigma}}'_{\widehat{\boldsymbol{\delta}}} = (5.03, 0.93)$           |   |        |        |        |        |        |        |        |        |
| Model 4   | $\widehat{m}_1$   | 2.65   | 5.30   | 9.79   | 16.64  | 26.08  | 37.68  | 50.17  | 61.57  |
| $\mathbf{g} = \mathbf{Q}_X \ln(\mathbf{m}_1)$   | $\widehat{\sigma}_{\widehat{m}_1}$  | 0.44   | 0.55   | 0.92   | 1.72   | 2.70   | 3.56   | 5.20   | 10.16  |
| $\mathbf{X} = (\mathbf{1}_n, \mathbf{x}, \mathbf{x}^2)$   | $\widehat{\alpha}$  | 26.039 |        |        |        |        |        |        |        |
|   | $\widehat{\beta}$   | 0.102  | 0.204  | 0.376  | 0.639  | 1.002  | 1.447  | 1.927  | 2.365  |
| $W = 5.223 \nu = 5 (p = 0.389), \widehat{\boldsymbol{\delta}} = (0.44, 0.35, 0.22) \widehat{\boldsymbol{\sigma}}'_{\widehat{\boldsymbol{\delta}}} = (4.49, 1.79, 0.12)$ |   |        |        |        |        |        |        |        |        |
| Model 5   | $\widehat{m}_1$   | 26.25  | 26.25  | 26.25  | 26.25  | 26.25  | 26.25  | 26.25  | 26.25  |
| $\mathbf{g} = \mathbf{Q}_X \mathbf{m}_1$  | $\widehat{\sigma}_{\widehat{m}_1}$  | 8.33   | 8.33   | 8.33   | 8.33   | 8.33   | 8.33   | 8.33   | 8.33   |
| $\mathbf{X} = (\mathbf{1}_n)$   | $\widehat{\alpha}$  | 1.241  |        |        |        |        |        |        |        |
|   | $\widehat{\beta}$   | 21.155 | 21.155 | 21.155 | 21.155 | 21.155 | 21.155 | 21.155 | 21.155 |
| $W = 5.805 \nu = 7 (p = 0.563), \widehat{\boldsymbol{\delta}} = (26.25) \widehat{\boldsymbol{\sigma}}'_{\widehat{\boldsymbol{\delta}}} = (8.33)$                        |   |        |        |        |        |        |        |        |        |

### 3.4 MLE for Linear Mixed Models

The linear mixed model provides for heterogeneous variances through a variance structure based on random effects and their variance components. Algebraic results are much more complicated than in the case of linear models and numerical maximization techniques such as the EM algorithm and Newton-Raphson are used to obtain mle's. Furthermore, the general linear mixed model and the SAS procedure MIXED are discussed at length by Verbeke & Molenberghs (1997, 2009) who emphasize that "the techniques used and their software implementation are not at all straightforward". However, in the case of MLE, where structure is attributed to the covariance matrix, the model can be specified in terms of constraints on the canonical statistics and the estimation process largely simplified without requiring derivation of the likelihood equations. It provides additional flexibility in modelling.

The linear mixed model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \text{ with } \boldsymbol{\Sigma} = \text{Cov}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$$

where  $\mathbf{y} : n \times 1$  the vector of observations,  $\boldsymbol{\beta} : p \times 1$  the fixed effects,  $\mathbf{u} : q \times 1$  the random effects,  $\mathbf{e}$  the random error component,  $\text{Cov}(\mathbf{u}, \mathbf{e}') = \mathbf{0}$ ,  $\mathbf{X}$  a model matrix and  $\mathbf{Z}$  usually an incidence matrix with  $\text{rank}(\mathbf{X}) = p$ ,  $\text{rank}(\mathbf{Z}) = q$ . Consider the special case where  $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$ , and  $\mathbf{Z}\mathbf{u}$  being partitioned as

$$\mathbf{Z}\mathbf{u} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 & \cdots & \mathbf{Z}_c \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_c \end{pmatrix} \quad \text{where } c \text{ represents the number of random factors} \\ \text{and } q = \sum_{i=1}^c q_i.$$

Then  $\mathbf{u}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}_{q_i})$ ,  $i = 1 \cdots c$ ,  $\text{Cov}(\mathbf{u}_i, \mathbf{u}_j') = \mathbf{0}$ ,  $i \neq j$  so that  $\mathbf{D} = \{\sigma_i^2 \mathbf{I}_{q_i}\}_{i=1}^c$ , a diagonal matrix with typical main diagonal element  $\sigma_i^2 \mathbf{I}_{q_i}$ . Furthermore, it is assumed that  $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$  and  $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ . Then  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ ,  $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{y}, \mathbf{y}') = \sigma_e^2 \mathbf{I}_n + \sum_{i=1}^c \sigma_i^2 \mathbf{Z}_i \mathbf{Z}_i'$ . The likelihood function is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= (2\pi)^{-n/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \\ &= \exp\left\{\mathbf{y}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{y} \mathbf{y}' - \frac{1}{2} \boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \boldsymbol{\beta} - \frac{n}{2} \ln \det(2\pi \boldsymbol{\Sigma})\right\} \\ &= \exp\{\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})\} \end{aligned}$$

where

$$t = \begin{pmatrix} \mathbf{y} \\ \text{vec}(\mathbf{y}\mathbf{y}') \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\text{vec}(\boldsymbol{\Sigma}^{-1}) \end{pmatrix}, \quad \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu} \\ \text{vec}(\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}) \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$$

and from (2.6) (the roles of  $p$  and  $n$  interchanged), it follows that

$$\mathbf{V} = \text{Cov}(t) = \left( \begin{array}{c|c} \boldsymbol{\Sigma} & (\boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma})' \\ \hline \boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma} & (\mathbf{I}_{n^2} + \mathbf{K})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \boldsymbol{\Sigma}) \end{array} \right).$$

Consider  $\boldsymbol{\mu} = m_1$  and  $\text{vec}(\boldsymbol{\Sigma}) = m_2 - \text{vec}(m_1 m_1')$ . The means and variances are modelled as:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ \text{vec}(\boldsymbol{\Sigma}) &= \sigma_e^2 \text{vec}(\mathbf{I}_n) + \sigma_1^2 \text{vec}(\mathbf{Z}_1 \mathbf{Z}_1') + \cdots + \sigma_c^2 \text{vec}(\mathbf{Z}_c \mathbf{Z}_c') = \mathbf{C}\boldsymbol{\sigma}^2 \end{aligned} \quad (3.23)$$

with implied constraints:

$$g(\mathbf{m}) = \begin{pmatrix} g_1(\mathbf{m}) \\ g_2(\mathbf{m}) \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_X m_1 \\ \mathbf{Q}_C(m_2 - \text{vec}(m_1 m_1')) \end{pmatrix} \quad (3.24)$$

where  $\mathbf{Q}_C$  the corresponding projection matrix and

$$\mathbf{G}_m = \left( \begin{array}{c|c} \mathbf{Q}_X & \mathbf{0} \\ \hline \mathbf{Q}_C(-\mathbf{I}_n \otimes m_1 - m_1 \otimes \mathbf{I}_n) & \mathbf{Q}_C \end{array} \right), \quad \mathbf{G}_t = \left( \begin{array}{c|c} \mathbf{Q}_X & \mathbf{0} \\ \hline \mathbf{Q}_C(-\mathbf{I}_n \otimes t_1 - t_1 \otimes \mathbf{I}_n) & \mathbf{Q}_C \end{array} \right).$$

The mle's under constraints,  $\widehat{\mathbf{m}}$ , is obtained iteratively from  $\widehat{\mathbf{m}} = t - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^* g(t)$

(cf. (2.2)). The estimated parameters are obtained as a function of  $\mathbf{m}$  namely:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^* \mathbf{X}'\widehat{m}_1,$$

$$\widehat{\boldsymbol{\sigma}}^2 = (\mathbf{C}'\mathbf{C})^* \mathbf{C}'(\widehat{m}_2 - \text{vec}(\widehat{m}_1 \widehat{m}_1')) \text{ and } \text{vec}(\widehat{\boldsymbol{\Sigma}}) = \mathbf{C}\widehat{\boldsymbol{\sigma}}^2.$$

### 3.4.1 Two-way crossed classification interaction mixed model with unbalanced data

This last example illustrates a very simple solution to what Searle et. al. (1992) calls "the embarrassment of having a negative estimate that, by definition, is positive". Consider the two-way crossed classification interaction mixed model with unbalanced data (Hemmerle & Hartley, 1973) in Table 3.7.

Table 3.7: Data for a two-way crossed classification.

| Fixed<br>Factor A | Random<br>Factor B |             |
|-------------------|--------------------|-------------|
|                   | 1                  | 2           |
| 1                 | 237 254 246        | 178 179     |
| 2                 | 208 178 187        | 146 145 141 |
| 3                 | 186 183            | 142 125 136 |

The model equation considered for the means is given by:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad (3.25)$$

for the fixed factor  $\alpha_i$ ,  $i = 1, 2, 3$ , the random factor  $\beta_j$ ,  $j = 1, 2$  and the random interaction factor  $\gamma_{ijk}$ ,  $k = 1, \dots, n_{ij}$ . The model and constraints are given by (3.23) and (3.24) respectively, where the design matrix is formulated according to (3.23) and the number of random factors,  $c$ , is equal to 2.

Results are given in Table 3.8. For this data set, both traditional maximum likelihood and MLE under constraints yield a negative random variance estimate which is then set to zero and the other variances re-estimated (Model 1). Negative variance estimates usually are an indication that the specified model is not appropriate for the data. In this example, a rather large variation for the combination of Factor A, level 2 and Factor B, level 1, is observed in the data in Table 3.7. Estimation under constraints has the further flexibility to make provision for this by adding the constraint that the error variance in this particular cell differs from the rest.

To make provision for different error variances for every level of the two factors, consider a partitioning of the random error component into 6 subvectors corresponding to the combined levels of the two factors where the subscripts denote the level of each factor:

$$\mathbf{e} = \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{32} \end{pmatrix} = \mathbf{I}_n \mathbf{e} = \text{diag}(\mathbf{I}_{11}, \mathbf{I}_{12}, \dots, \mathbf{I}_{32}) \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{32} \end{pmatrix} \text{ and } \mathbf{e}_{ij} \sim (\mathbf{0}, \sigma_{ij}^2 \mathbf{I}_{ij})$$

where  $\mathbf{I}_{ij}$  is the matrix consisting of the  $n_{ij}$  corresponding columns of the  $16 \times 16$  identity matrix  $\mathbf{I}_n$ .

Then

$$\text{Cov}(\mathbf{y}, \mathbf{y}') = \sum_{i=1}^3 \sum_{j=1}^2 \sigma_{ij}^2 \mathbf{I}_{ij} \mathbf{I}'_{ij} + \sum_{j=1}^2 \sigma_j^2 \mathbf{Z}_j \mathbf{Z}'_j$$

and

$$\text{vec}(\boldsymbol{\Sigma}) = \sum_{i=1}^3 \sum_{j=1}^2 \sigma_{ij}^2 \text{vec}(\mathbf{I}_{ij} \mathbf{I}'_{ij}) + \sum_{j=1}^2 \sigma_j^2 \text{vec}(\mathbf{Z}_j \mathbf{Z}'_j) = \mathbf{C} \boldsymbol{\sigma}^2.$$

Manipulation of the appropriate columns of the identity matrix  $\mathbf{I}_{16}$  yields the estimates of the fixed effects, error variances and random component variances given as Model 2 and Model 3. For Model 2, error variances for five of the six combined levels were set equal:  $\sigma_{11}^2 = \sigma_{12}^2 = \sigma_{22}^2 = \sigma_{31}^2 = \sigma_{32}^2$ . The variance  $\sigma_{21}^2$  was allowed to differ from the other five. For Model 3, three sets of error variances were allowed:  $\sigma_{11}^2 = \sigma_{12}^2 = \sigma_{22}^2 = \sigma_{31}^2$ ,  $\sigma_{21}^2$  and  $\sigma_{32}^2$ . With this approach the problem of negative random component variances is automatically solved.

In comparison to earlier applications considered, convergence in the case of mixed models may be slower since a generalized inverse of the potentially large matrix  $\mathbf{G}_t \mathbf{V} \mathbf{G}'_m$  needs to be obtained in the iterative process.

SAS program: MIXED.SAS

Table 3.8: Estimated parameters for the 2-way crossed classification interaction mixed model.

| Parameter        | Model 1 | Model 2 | Model 3 |
|------------------|---------|---------|---------|
| $\mu + \alpha_1$ | 213.5   | 212.8   | 212.6   |
| $\mu + \alpha_2$ | 167.5   | 169.9   | 170.2   |
| $\mu + \alpha_3$ | 159.7   | 159.8   | 159.0   |
| $\sigma_{11}^2$  | 77.53   | 39.8    | 28.1    |
| $\sigma_{21}^2$  |         | 209.4   | 214.3   |
| $\sigma_{32}^2$  |         |         | 65.9    |
| $\sigma_1^2$     | 723.67  | 781.6   | 800.2   |
| $\sigma_2^2$     | 0       | 9.1     | 13.5    |
| $W$              | 496     | 496     | 495.99  |
| $\nu$            | 151     | 151     | 149     |
| $p$              | 0       | 0       | 0       |

# Chapter 4

## MLE for Nonspherical Disturbances

In the case of nonspherical disturbances in econometric applications, the multiple regression model is extended to the generalized regression model which allows for non-constant variance and correlated disturbances (Greene, 2003). In this thesis, correlated disturbances are treated as a single multivariate observation from a multivariate normal  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution. This entails the estimation of  $\boldsymbol{\Sigma}$  based on an initial observed covariance matrix of rank one.

In the application of statistical inferential procedures based on  $n$  vector observations represented by the columns of the random matrix  $\mathbf{Y} : p \times n$ , from a distribution with parameter vector  $\boldsymbol{\theta} : s \times 1$ , it is usually assumed that  $n$  is relatively large. A fundamental problem arises when  $n$  is relatively small. In such cases the mle of  $\boldsymbol{\theta}$  only exists under specific constraints.

Mle's of the mean vector and covariance matrix of multivariate normal random vectors where the dimension  $p$  is larger than the number of observations  $n$ , do not exist. The sample covariance matrix has a singular Wishart distribution. In this chapter, maximum likelihood estimation under constraints is proposed to obtain mle's of parameter structures in the exponential class which exist in a restricted parameter space. While traditional procedures require derivation of the likelihood equations for different parameter structures, the MLE procedure under constraints only requires slight adjustments in the constraints specified. In the presence of nonspherical disturbances, the likelihood equations are generally highly nonlinear and usually require an iterative solution (Greene, 2003).

In the context dealt with in this chapter, the basic iterative procedure can be applied in a straightforward way. The model determines the constraints. For a specific model, only the formulation of  $\mathbf{g}(\mathbf{m})$ , the vector of constraints, together with the corresponding matrix of derivatives  $\mathbf{G}_m$ , are

required.

## 4.1 MLE for the Generalized Regression Model

The generalized regression model is given by

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{\varepsilon}) &= \mathbf{0} \\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') &= \sigma^2\boldsymbol{\Omega} = \boldsymbol{\Sigma} \end{aligned} \quad (4.1)$$

where  $\mathbf{y} : n \times 1$  the vector of observations, the elements of  $\boldsymbol{\beta}$  are the regression parameters,  $\mathbf{X}$  the design matrix and  $\boldsymbol{\varepsilon}$  the vector of disturbances. The likelihood function for a single multivariate normal observation is given by:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \exp \left\{ \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - \frac{1}{2}\text{tr}\boldsymbol{\Sigma}^{-1}\mathbf{y}\mathbf{y}' - \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\ell\text{ndet}(2\pi\boldsymbol{\Sigma}) \right\} \\ &= \exp \left\{ \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - \frac{1}{2}[\text{vec}(\boldsymbol{\Sigma}^{-1})]' \text{vec}(\mathbf{y}\mathbf{y}') - \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\ell\text{ndet}(2\pi\boldsymbol{\Sigma}) \right\} \\ &= \exp [\boldsymbol{\theta}'\mathbf{t} - \kappa(\boldsymbol{\theta})] \quad \text{where} \end{aligned} \quad (4.2)$$

$$\mathbf{t} = \begin{pmatrix} \mathbf{y} \\ \text{vec}(\mathbf{y}\mathbf{y}') \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\text{vec}(\boldsymbol{\Sigma}^{-1}) \end{pmatrix} \quad (4.3)$$

are the vectors of canonical statistics and canonical parameters respectively.

It follows that

$$E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu} \\ \text{vec}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \quad (4.4)$$

with covariance matrix

$$\text{Cov}(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \quad (4.5)$$

given by (2.6) where the roles of  $p$  and  $n$  are interchanged:

$$\begin{aligned} \mathbf{V}_{11} &= \boldsymbol{\Sigma} \\ \mathbf{V}_{21} &= \boldsymbol{\Sigma} \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\Sigma} \\ \mathbf{V}_{12} &= \mathbf{V}'_{21} \\ \mathbf{V}_{22} &= (I_{n^2} + \mathbf{K}) [\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \boldsymbol{\Sigma}]. \end{aligned} \quad (4.6)$$



Constraints are implied by the structure modelled for the mean vector and covariance matrix in the generalized regression model (4.1). In general, this is specified as

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\mu}) \\ \mathbf{g}_2[\text{vec}(\boldsymbol{\Sigma})] \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}_1) \\ \mathbf{g}_2[\mathbf{m}_2 - \text{vec}(\mathbf{m}_1 \mathbf{m}_1')] \end{pmatrix} \quad (4.7)$$

with corresponding first order partial derivatives

$$\begin{aligned} \mathbf{G}_m = \frac{\partial \mathbf{g}(\mathbf{m})}{\partial \mathbf{m}} &= \begin{pmatrix} \frac{\partial \mathbf{g}_1(\mathbf{m}_1)}{\partial \mathbf{m}_1} & \Big| & \frac{\partial \mathbf{g}_1(\mathbf{m}_1)}{\partial \mathbf{m}_2} \\ \frac{\partial \mathbf{g}_2[\text{vec}(\boldsymbol{\Sigma})]}{\partial \text{vec}(\boldsymbol{\Sigma})} \times \frac{\partial [\text{vec}(\boldsymbol{\Sigma})]}{\partial \mathbf{m}_1} & \Big| & \frac{\partial \mathbf{g}_2[\mathbf{m}_2 - \text{vec}(\mathbf{m}_1 \mathbf{m}_1')]}{\partial \mathbf{m}_2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial \mathbf{g}_1(\mathbf{m}_1)}{\partial \mathbf{m}_1} & \Big| & \mathbf{0} \\ \frac{\partial \mathbf{g}_2[\text{vec}(\boldsymbol{\Sigma})]}{\partial \mathbf{m}_2} (-\mathbf{I} \otimes \mathbf{m}_1 - \mathbf{m}_1 \otimes \mathbf{I}) & \Big| & \frac{\partial \mathbf{g}_2(\mathbf{m}_2)}{\partial \mathbf{m}_2} \end{pmatrix}. \end{aligned} \quad (4.8)$$

Models can be specified in a variety of ways. Often, constraints implied by a specific model can also be formulated in many different ways. Consequently, for illustration purposes, different approaches are followed in the applications considered.

## 4.2 Introductory Example

The essence of the procedure is described by the simplest case where only one observation is assumed, namely vector  $\mathbf{y} : 2 \times 1 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$  distributed as multivariate normal with  $\boldsymbol{\mu} = \mathbf{0}$  and  $\boldsymbol{\Sigma} : 2 \times 2$  a positive definite covariance matrix  $\boldsymbol{\Sigma}$ , i.e.  $\mathbf{y} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} > \mathbf{0}$ .

The likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{y} \mathbf{y}' - \frac{1}{2} \ln \det(2\pi \boldsymbol{\Sigma}) \right\} \\ &= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})] \quad \text{where} \\ \mathbf{t} = \text{vec}(\mathbf{y} \mathbf{y}') &= \begin{pmatrix} y_1^2 \\ y_1 y_2 \\ y_1 y_2 \\ y_2^2 \end{pmatrix}, \quad E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{12} \\ \sigma_{22} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} \quad \text{say, and} \end{aligned}$$

$$\mathbf{V} = (I_{n^2} + \mathbf{K}) [\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}] = \begin{pmatrix} 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2 \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2 \end{pmatrix} \quad (\text{cf. (2.6)}).$$

Suppose the variances are proportional in the sense that  $\sigma_{11} = 2\sigma_{22}$ .

Then  $g(\mathbf{m}) = \sigma_{11} - 2\sigma_{22} = (1, 0, 0, -2)\mathbf{m} = \mathbf{C}\mathbf{m}$ , say, with derivative  $\mathbf{G}_m = \mathbf{C} = \mathbf{G}_t$ .

Consequently,

$$\mathbf{C}\mathbf{V} = (2\sigma_{11}^2 - 4\sigma_{12}^2, 2\sigma_{11}\sigma_{12} - 4\sigma_{12}\sigma_{22}, 2\sigma_{11}\sigma_{12} - 4\sigma_{12}\sigma_{22}, 2\sigma_{12}^2 - 4\sigma_{22}^2)$$

and

$$\mathbf{C}\mathbf{V}\mathbf{C}' = 2\sigma_{11}^2 + 8\sigma_{22}^2 - 8\sigma_{12}^2.$$

In the iterative procedure the covariance matrix subject to the constraint  $\sigma_{11} = 2\sigma_{22}$  is calculated using:

$$\widehat{\mathbf{m}} = \mathbf{t} - (\mathbf{C}\mathbf{V})' (\mathbf{C}\mathbf{V}\mathbf{C}')^{-1} g(\mathbf{t}) = \mathbf{t} - \frac{t_1 - 2t_4}{2m_1^2 + 8m_4^2 - 8m_2^2} \begin{pmatrix} 2m_1^2 - 4m_2^2 \\ 2m_1m_2 - 4m_2m_4 \\ 2m_1m_2 - 4m_2m_4 \\ 2m_2^2 - 4m_4^2 \end{pmatrix}.$$

To illustrate, let  $\mathbf{y} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ . An initial value of  $\boldsymbol{\Sigma} = \mathbf{I}_2$  is used in  $\mathbf{V}$  since the single observed vector yields a singular covariance matrix. Consequently, in the first iteration we have  $\mathbf{C}\mathbf{V}\mathbf{C}' = 10$ ,

$\mathbf{C}\mathbf{V} = (2, 0, 0, -4)$  with  $\mathbf{t} = (4^2, 4 \times 2, 4 \times 2, 2^2)$  and  $g(\mathbf{t}) = 8$ . Since  $g(\mathbf{m}) = (1, 0, 0, -2)\mathbf{m}$  is

linear in  $\mathbf{m}$ , iteration is performed only over  $\mathbf{m}$ :

$$\text{Iteration 1: } \begin{pmatrix} 16 \\ 8 \\ 8 \\ 4 \end{pmatrix} - \frac{8}{10} \begin{pmatrix} 2 \\ 0 \\ 0 \\ -4 \end{pmatrix} = \begin{pmatrix} 14.4 \\ 8 \\ 8 \\ 7.2 \end{pmatrix}$$

$$\text{Iteration 2: } \begin{pmatrix} 16 \\ 8 \\ 8 \\ 4 \end{pmatrix} - \frac{8}{317.44} \begin{pmatrix} 158.72 \\ 0 \\ 0 \\ -79.36 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \\ 8 \\ 6 \end{pmatrix}.$$

In Table 4.1 the results obtained under different constraints  $g(\mathbf{m})$  when  $y_1 = 4$  and  $y_2 = 2$  are given.

The mle's are all found under the constraints  $g(\mathbf{m}) = \mathbf{0}$ .

SAS program: MULTIV1OBS.SAS

Table 4.1: Mle's of structured variances based on a single bivariate observation.

| $g(\mathbf{m})$  | Mle's           |                |                |        |
|--|-----------------|----------------|----------------|--------|
|  | $\sigma_{11}$   | $\sigma_{12}$  | $\sigma_{22}$  | $\rho$ |
| $\sigma_{12}$  | 16              | 0              | 4              | 0      |
| $\sigma_{11} - \sigma_{22}$  | 10              | 8              | 10             | 0.8    |
| $\sigma_{11} - 2\sigma_{22}$   | 12              | 8              | 6              | 0.943  |
| $\sigma_{11} - 3\sigma_{22}$   | 14              | 8              | $4\frac{2}{3}$ | 0.990  |
| $\begin{pmatrix} \sigma_{12} \\ \sigma_{11} - \sigma_{22} \end{pmatrix}$ | 10              | 0              | 10             | 0      |
| $\rho - 0.5$   | $10\frac{2}{3}$ | $2\frac{2}{3}$ | $2\frac{2}{3}$ | 0.5    |
| $\begin{pmatrix} \rho - 0.5 \\ \sigma_{11} - 2\sigma_{22} \end{pmatrix}$ | 8.46            | 2.99           | 4.23           | 0.5    |

### 4.3 MLE for Heteroskedastic Disturbances

Heteroskedastic disturbances imply

$$\sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} \omega_{11} & 0 & \cdots & 0 \\ 0 & \omega_{22} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \omega_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \quad (4.9)$$

In this case, the likelihood function simplifies to:

$$\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \exp \left\{ \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} [\text{vec}(\boldsymbol{\Sigma}^{-1})]' \text{vec}(\mathbf{y}\mathbf{y}') - \frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \ln \det(2\pi \boldsymbol{\Sigma}) \right\} \\
&= \exp \left\{ (\boldsymbol{\mu}/\boldsymbol{\sigma}^2)' \mathbf{y} - (1/2\boldsymbol{\sigma}^2)' \mathbf{y}^2 - \frac{1}{2} (\boldsymbol{\mu}/\boldsymbol{\sigma}^2)' \boldsymbol{\mu} - \frac{1}{2} \ln \det(2\pi \text{diag}(\boldsymbol{\sigma}^2)) \right\} \\
&= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})]
\end{aligned} \tag{4.10}$$

where

$$\boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_n^2 \end{pmatrix}, \mathbf{t} = \begin{pmatrix} \mathbf{y} \\ \mathbf{y}^2 \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu}/\boldsymbol{\sigma}^2 \\ -\mathbf{1}/(2\boldsymbol{\sigma}^2) \end{pmatrix} \text{ and } \boldsymbol{\mu}/\boldsymbol{\sigma}^2 \text{ indicates elementwise division.}$$

Consequently,

$$E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\sigma}^2 + \boldsymbol{\mu}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \tag{4.11}$$

where  $\boldsymbol{\mu}^2 = \boldsymbol{\mu}\boldsymbol{\mu}$ , the elementwise product (Hadamard product). The covariance matrix is given by (cf.

(2.6) where  $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma}^2)$  with  $p = n$  and  $n = 1$ ): .

$$\text{Cov}(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \text{diag}(\boldsymbol{\sigma}^2) & 2\text{diag}(\boldsymbol{\mu}\boldsymbol{\sigma}^2) \\ 2\text{diag}(\boldsymbol{\mu}\boldsymbol{\sigma}^2) & 2\text{diag}[(\boldsymbol{\sigma}^4 + 2\boldsymbol{\mu}^2\boldsymbol{\sigma}^2)] \end{pmatrix} \tag{4.12}$$

where  $\boldsymbol{\mu}\boldsymbol{\sigma}^2$  indicates elementwise multiplication. The parameter vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}^2$  may be modelled simultaneously. In terms of linear models, this reduces to

$$\begin{aligned}
\boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\
\boldsymbol{\sigma}^2 &= \mathbf{Z}\boldsymbol{\alpha}
\end{aligned} \tag{4.13}$$

where  $\mathbf{X}$  and  $\mathbf{Z}$  are appropriate design matrices. In order to determine the mle of  $\mathbf{m}$  under the model

(4.13), the restrictions may be formulated as  $\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}) \\ \mathbf{g}_2(\mathbf{m}) \end{pmatrix}$  with

$$\mathbf{g}_1(\mathbf{m}) = \mathbf{Q}_X \boldsymbol{\mu} = \mathbf{Q}_X \mathbf{m}_1, \quad \mathbf{g}_2(\mathbf{m}) = \mathbf{Q}_Z \boldsymbol{\sigma}^2 = \mathbf{Q}_Z (\mathbf{m}_2 - \mathbf{m}_1^2) \tag{4.14}$$

where  $\mathbf{Q}_X$  and  $\mathbf{Q}_Z$  specify the contrast matrices or corresponding projection matrices implied by (4.13).

In many situations the columns of  $\mathbf{X}$  and  $\mathbf{Z}$  generate the same vector space and  $\mathbf{Q}_X = \mathbf{Q}_Z$ . The corresponding derivatives with respect to  $\mathbf{m}$  are

$$\mathbf{G}_{1m} = (\mathbf{Q}_X, \mathbf{0}), \quad \mathbf{G}_{2m} = (-2\mathbf{Q}_Z \text{diag}(\mathbf{m}_1), \mathbf{Q}_Z), \quad \mathbf{G}_m = \begin{pmatrix} \mathbf{G}_{1m} \\ \mathbf{G}_{2m} \end{pmatrix}. \tag{4.15}$$

The mle of  $\mathbf{m}$  subject to the constraints  $\mathbf{g}(\mathbf{m}) = \mathbf{0}$ , is obtained iteratively from (2.2):

$$\widehat{\mathbf{m}} = \mathbf{t} - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^* \mathbf{g}(\mathbf{t}).$$

In order to obtain the asymptotic variances of  $\widehat{\boldsymbol{\mu}}$ ,  $\widehat{\sigma}^2$ ,  $\widehat{\boldsymbol{\beta}}$  and  $\widehat{\boldsymbol{\alpha}}$ , let

$$\mathbf{h}(\mathbf{m}) = \begin{pmatrix} \boldsymbol{\mu} \\ \sigma^2 \\ \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 - \mathbf{m}_1^2 \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{m}_1 \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\sigma^2 \end{pmatrix}. \quad (4.16)$$

Using (2.4) with

$$\frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -2\text{diag}(\mathbf{m}_1) & \mathbf{I} \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & \mathbf{0} \\ -2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\text{diag}(\mathbf{m}_1) & (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix} \quad (4.17)$$

the asymptotic variances,  $\widehat{\sigma}^2(\mathbf{h}(\widehat{\mathbf{m}}))$ , of the elements of  $\mathbf{h}(\widehat{\mathbf{m}})$ , are given by the diagonal elements of  $\text{Cov}(\mathbf{h}(\widehat{\mathbf{m}}))$ .

This framework can be generalized to  $k$  independent populations. It provides a convenient point of departure to model under different heteroskedastic specifications, systematically varying parameters, seasonality and splines. To illustrate some of these, modelling when variances are constant within subgroups of observations, as well as a piecewise regression model, is considered in the following two sections.

### 4.3.1 Groupwise heteroskedasticity

In the case of constant variances within subgroups of observations, the  $n$  observations are grouped into  $g$  groups, each with  $n_i$  observations such that  $\sum_{i=1}^g n_i = n$ . The groupwise heteroskedastic regression can be written as (Judge, et.al., 1985):

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_g \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_g \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_g \end{pmatrix} \quad (4.18)$$

where  $\mathbf{y}_i : n_i \times 1$ ,  $\mathbf{X}_i : n_i \times p$ ,  $E(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i') = \sigma_i^2 \mathbf{I}$ ,  $E(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j') = 0, i \neq j$ .

The slope vector is the same in all groups and there is no correlation between disturbances in different equations.

Greene (2003) considered a groupwise heteroskedastic model for modelling the natural logarithm of the total cost of production,  $\mathbf{y}$ , (in \$1000), in the US airline industry as a function of the following explanatory variables:

$\ell n(\text{output})$ , denoted by  $\mathbf{x}_1$  (which is measured in "revenue passenger miles"),

$\ell n(\text{fuel price})$ , denoted by  $\mathbf{x}_2$ , and

a load factor, denoted by  $\mathbf{x}_3$ , (which is the average capacity utilization of the fleet)

for a sample of six different airlines observed annually for 15 years.

The model is given by:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_6 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ & & & \vdots & & & \\ \mathbf{1}_{n_6} & \mathbf{x}_{61} & \mathbf{x}_{62} & \mathbf{x}_{63} & \mathbf{0}_{n_6} & \cdots & \mathbf{1}_{n_6} \end{pmatrix} \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_2 \\ \vdots \\ \gamma_6 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_6 \end{pmatrix} \quad (4.19)$$

so that in accordance with (4.13):

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{A})\boldsymbol{\beta}, \text{ say, where}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_5 \end{pmatrix} \text{ is the matrix of five dummy variables denoting the six airlines} \quad (4.20)$$

and the group variances are modelled as

$$\boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_6^2 \end{pmatrix} = \mathbf{Z}\boldsymbol{\alpha} \text{ where } \mathbf{Z} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ & \vdots & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1}_{n_6} \end{pmatrix}, \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_6 \end{pmatrix}. \quad (4.21)$$

Using the procedure for MLE under constraints, the constraints (4.14), with design matrices given in (4.20) and (4.21), together with derivatives (4.15) used in the procedure (2.2) yield the mle's

given by Greene (2003). Results are given in Table 4.2. The least squares solution is obtained when  $\sigma^2 = \mathbf{Z}_0\alpha$  where  $\mathbf{Z}_0 = \mathbf{1}_n$  and  $\alpha = \alpha_0$ . The group variances can be modelled separately in terms of explanatory variables and are also illustrated in Table 4.2. For this purpose, let

$$\sigma^2 = \mathbf{Z}_1\alpha = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{x}_{2,n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{x}_{2,n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{n_6} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_{2,n_6} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_6 \end{pmatrix}. \quad (4.22)$$

where  $\mathbf{x}_{2,n_i} : n_i \times 1$  represents the  $\ln(\text{fuel price})$  values for group  $i$ ,  $i = 1, \dots, 6$ ,  $\alpha_0$  a common variance effect, and  $\alpha_i$  the rate at which the variance for each group  $i$  increases with  $\mathbf{x}_2$ ,  $\ln(\text{fuel price})$ .

SAS program: GWHETEROSK.SAS

Table 4.2: Mle's of a groupwise heteroskedastic model with different variance structures.

| Model      | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{A})$<br>$\mathbf{Z} = \mathbf{Z}_0 = (\mathbf{1}_n)$ (OLS) |            | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{A})$<br>$\mathbf{Z} = \text{block}(\mathbf{1}_{n_i})$ (cf. (4.21)) |            | $\mathbf{X} = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{A})$<br>$\mathbf{Z} = \mathbf{Z}_1 = (\mathbf{1}_n, \text{block}(\mathbf{x}_{2,n_i}))$ (cf.(4.22)) |            |
|------------|---|------------|---|------------|---|------------|
|            | Estimate  | Std. error | Estimate  | Std. error | Estimate  | Std. error |
| $\mu$      | 9.706   | 0.183      | 10.056984   | 0.1343384  | 10.153864   | 0.1209745  |
| $\beta_1$  | 0.919   | 0.0284     | 0.9282934   | 0.0227183  | 0.9403975   | 0.0230146  |
| $\beta_2$  | 0.417   | 0.0144     | 0.3999571   | 0.0107755  | 0.3899792   | 0.0109017  |
| $\beta_3$  | -1.070  | 0.191      | -1.289203   | 0.1637509  | -1.243551   | 0.1617307  |
| $\gamma_2$ | -0.041  | 0.0239     | -0.048675   | 0.0236821  | -0.049952   | 0.0241293  |
| $\gamma_3$ | -0.209  | 0.0406     | -0.19958  | 0.0307597  | -0.185664   | 0.032192   |
| $\gamma_4$ | 0.185   | 0.0576     | 0.1921397   | 0.049882   | 0.2096478   | 0.0521716  |
| $\gamma_5$ | 0.024   | 0.0758     | 0.0418611   | 0.0594169  | 0.0787153   | 0.0626633  |
| $\gamma_6$ | 0.087   | 0.0799     | 0.0963358   | 0.0630786  | 0.1371855   | 0.0669744  |
| $\alpha_0$ | 0.00325   | 0.000485   |   |            | -0.004843   | 0.0029237  |
| $\alpha_1$ |   |            | 0.0008349   | 0.0003049  | 0.0004368   | 0.0002465  |
| $\alpha_2$ |   |            | 0.0062143   | 0.0022691  | 0.0008963   | 0.000298   |
| $\alpha_3$ |   |            | 0.0017819   | 0.0006507  | 0.0005328   | 0.0002402  |
| $\alpha_4$ |   |            | 0.0090712   | 0.0033123  | 0.0011199   | 0.0003548  |
| $\alpha_5$ |   |            | 0.0014184   | 0.0005179  | 0.0005094   | 0.0002395  |
| $\alpha_6$ |   |            | 0.0023931   | 0.0008738  | 0.0005719   | 0.0002434  |
|            | $W=90 (p=1), \nu = 170$<br>$R^2 = 0.997$  |            | $W=90(p=1), \nu = 165$<br>$R^2 = 0.989$   |            | $W=90(p=1), \nu = 164$<br>$R^2 = 0.974$   |            |

The case of groupwise heteroskedasticity may also be treated as an extension of the approach followed in the next subsection for modelling structural change.

### 4.3.2 Modelling structural change

A piecewise linear regression model allows for changes in slope, with the constraint that the lines being estimated be continuous, i.e. it consists of two or more straight line segments which are continuous at every point of structural break. In times series data structural change in the relationship between the dependent variable and the explanatory variables often occurs, i.e. the values of the parameters of the model do not remain the same through the entire time period. Usually, it is assumed that the variance remains constant after a structural change. MLE under constraints provides a simple framework within which structural change can be handled, simultaneously allowing for heteroskedastic variances.

Gujarati (2003) considers data on disposable personal income and personal savings, in billions of dollars, observed for the USA for the period 1970-1995. A structural change in savings (as a function of disposable income) is believed to have occurred between 1981 and 1982 (cf Figure 4.1). The same data are used here to illustrate MLE under constraints as a means to model a spline regression. The model follows as a natural extension of (4.11), treating the time periods before and after the structural change as independent samples linked in a particular way. The two time periods 1970-1981 ( $t_1$ ) and 1982-1995 ( $t_2$ ) are considered. The canonical statistic is

$$t = \begin{pmatrix} y_1 \\ y_1^2 \\ y_2 \\ y_2^2 \end{pmatrix} \text{ with } E(t) = \begin{pmatrix} \mu_1 \\ \mu_1^2 + \sigma_1^2 \\ \mu_2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix} = \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} \text{ and} \quad (4.23)$$

$$Cov(t) = \begin{pmatrix} diag(\sigma_1^2) & 2diag(\mu_1\sigma_1^2) & 0 & 0 \\ 2diag(\mu_1\sigma_1^2) & 2diag[(\sigma_1^4 + 2\mu_1^2\sigma_1^2)] & 0 & 0 \\ 0 & 0 & diag(\sigma_2^2) & 2diag(\mu_2\sigma_2^2) \\ 0 & 0 & 2diag(\mu_2\sigma_2^2) & 2diag[\sigma_2^4 + 2\mu_2^2\sigma_2^2] \end{pmatrix} \quad (4.24)$$

The parameter vectors  $\mu_i$  and  $\sigma_i^2$ ,  $i = 1, 2$ , may be modelled similar to (4.13):

$$\begin{aligned} \mu_1 &= X_1\beta_1 & \mu_2 &= X_2\beta_2 \\ \sigma_1^2 &= Z_1\alpha_1 & \sigma_2^2 &= Z_2\alpha_2 \end{aligned} \quad (4.25)$$



where  $\mathbf{X}_i$  and  $\mathbf{Z}_i$ ,  $i = 1, 2$  are appropriate design matrices. Corresponding constraints are:

$$\begin{aligned} g_{11}(\mathbf{m}_{11}) &= \mathbf{Q}_{X_1} \mathbf{m}_{11}, & g_{12}(\mathbf{m}_{21}) &= \mathbf{Q}_{X_2} \mathbf{m}_{21}, \\ g_{21}(\sigma_1^2) &= \mathbf{Q}_{Z_1} (\mathbf{m}_{12} - \mathbf{m}_{11}^2) & g_{22}(\sigma_2^2) &= \mathbf{Q}_{Z_2} (\mathbf{m}_{22} - \mathbf{m}_{21}^2) \end{aligned} \quad (4.26)$$

where  $\mathbf{Q}_{X_i}$  and  $\mathbf{Q}_{Z_i}$  specify the contrast matrices. The derivatives are:

$$\begin{aligned} \mathbf{G}_{11m} &= \frac{\partial \mathbf{Q}_{X_1} \mathbf{m}_{11}}{\partial \mathbf{m}} = (\mathbf{Q}_{X_1}, & \mathbf{0}_{n_1 \times n_1}, & \mathbf{0}_{n_1 \times n_2}, & \mathbf{0}_{n_1 \times n_2}) \\ \mathbf{G}_{12m} &= \frac{\partial \mathbf{Q}_{X_2} \mathbf{m}_{21}}{\partial \mathbf{m}} = (\mathbf{0}_{n_2 \times n_1}, & \mathbf{0}_{n_2 \times n_1}, & \mathbf{Q}_{X_2}, & \mathbf{0}_{n_2 \times n_2}) \\ \mathbf{G}_{21m} &= \frac{\partial \mathbf{Q}_{Z_1} (\mathbf{m}_{12} - \mathbf{m}_{11}^2)}{\partial \mathbf{m}} = (-2\mathbf{Q}_{Z_1} \text{diag}(\mathbf{m}_{11}), & \mathbf{Q}_{Z_1}, & \mathbf{0}_{n_1 \times n_2}, & \mathbf{0}_{n_1 \times n_2}) \\ \mathbf{G}_{22m} &= \frac{\partial \mathbf{Q}_{Z_2} (\mathbf{m}_{22} - \mathbf{m}_{21}^2)}{\partial \mathbf{m}} = (\mathbf{0}_{n_2 \times n_1}, & \mathbf{0}_{n_2 \times n_1}, & -2\mathbf{Q}_{Z_2} \text{diag}(\mathbf{m}_{21}), & \mathbf{Q}_{Z_2}) \end{aligned} \quad (4.27)$$

The flexibility of the procedure is illustrated by the variety of hypotheses that can be tested by imposing appropriate constraints. These include tests for equality of slope parameters (cf. 4.28) and/or variances (cf. 4.29) or specification of a trend on the separate variances (cf. 4.26). In addition, it allows for the requirement that two regressions are linked (cf. 4.30).

Constraints and corresponding derivatives required to model equal slope parameters are given by:

$$\begin{aligned} g_3(\mathbf{m}) &= \beta_1 - \beta_2 = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{m}_{11} - (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{m}_{21} \text{ with} \\ \mathbf{G}_{3m} &= ((\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1, \mathbf{0}_{p \times n_1}, -(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2, \mathbf{0}_{p \times n_2}). \end{aligned} \quad (4.28)$$

To fit the model (4.25) assuming homoscedastic variances, let  $\mathbf{Z}_i = \mathbf{1}_{n_i}$ . The constraints (4.26) and derivatives (4.27) are required, together with the first elements of  $\sigma_1$  and  $\sigma_2$  respectively, which are set equal:

$$\begin{aligned} g_4(\mathbf{m}) &= \sigma_1[1,] - \sigma_2[1,] = (\mathbf{m}_{12} - \mathbf{m}_{11}^2)[1,] - (\mathbf{m}_{22} - \mathbf{m}_{21}^2)[1,] \\ \mathbf{G}_{4m} &= (-2\text{diag}(\mathbf{m}_{11})[1,], \mathbf{I}_{n_1}[1,], 2\text{diag}(\mathbf{m}_{21})[1,], -\mathbf{I}_{n_2}[1,]). \end{aligned} \quad (4.29)$$

Requiring that two regressions pass through a common point  $\mathbf{x}_0$ , implies the following constraint with corresponding derivative:

$$\begin{aligned} g_5(\mathbf{m}) &= \beta'_1 \mathbf{x}_0 - \beta'_2 \mathbf{x}_0 = \mathbf{m}'_{11} \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{x}_0 - \mathbf{m}'_{21} \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{x}_0 \text{ with} \\ \mathbf{G}_{5m} &= (\mathbf{x}'_0 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1, \mathbf{1}_{n_1}, -\mathbf{x}'_0 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2, \mathbf{1}_{n_2}) \end{aligned} \quad (4.30)$$

where the common point  $\mathbf{x}_0$  may be chosen to be equal to the average of the vectors of observations before and after the time period at which the change occurred, e.g. the average of the last row of  $\mathbf{X}_1$  and the first row of  $\mathbf{X}_2$ .

In Table 4.3, the mle's for the parameters for a variety of hypotheses, are given. The matrices  $\mathbf{X}_{i,t_i}$ ,  $i = 1, 2$  represent the explanatory variables observed at time period  $t_i$ . Furthermore,  $R_i^2, i = 1, 2$  denote the coefficient of determination for each line segment. In Figure 4.1 the data and fitted spline (reported in the middle column of Table 4.3) are illustrated graphically.

SAS program: SPLINES.SAS

Figure 4.1: Spline regression.

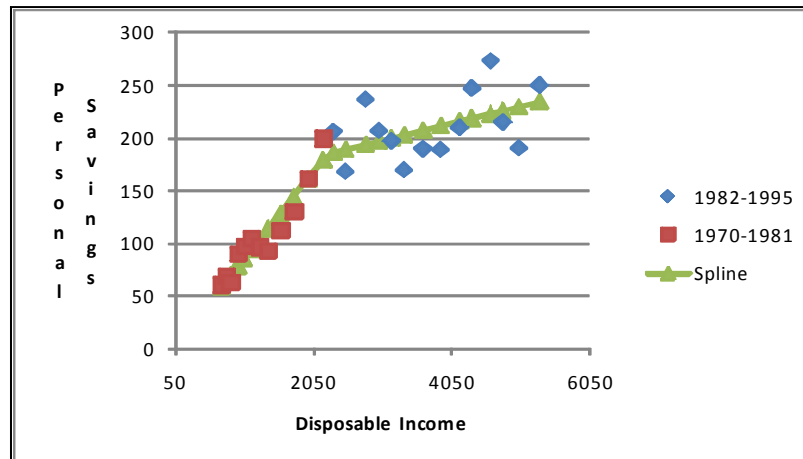


Table 4.3: Mle's of spline regression models with different variance structures.

| Model      | $\mathbf{X}_i = (\mathbf{1}_{n_i}, \mathbf{X}_{i,t_i})$<br>$\mathbf{Z}_i = (\mathbf{1}_{n_i})$ |            | $\mathbf{X}_i = (\mathbf{1}_{n_i}, \mathbf{X}_{i,t_i})$<br>$\mathbf{Z}_i = (\mathbf{1}_{n_i})$<br>Knot between 1981 and 1982 |            | $\mathbf{X}_i = (\mathbf{1}_{n_i}, \mathbf{X}_{i,t_i})$<br>$\mathbf{Z}_i = (\mathbf{1}_{n_i}, t_i)$<br>Knot between 1981 and 1982 |            |
|------------|--|------------|--|------------|---|------------|
|            | Estimate   | Std. error | Estimate   | Std. error | Estimate  | Std. error |
| $\beta_1$  | 1.016  | 3.067      | 0.0487   | 2.838      | 3.063   | 0.971      |
|            | 0.0803   | 0.002      | 0.0812   | 0.00193    | 0.0797  | 0.00134    |
| $\beta_2$  | 153.495  | 8.094      | 148.106  | 4.828      | 147.208   | 5.134      |
|            | 0.0149   | 0.002      | 0.0161   | 0.00144    | 0.0163  | 0.00154    |
| $\alpha_1$ | 148.753  | 17.531     | 148.958  | 17.555     | -26.035   | 3.205      |
|            |  |            |  |            | 26.035  | 3.205      |
| $\alpha_2$ | 714.659  | 72.191     | 716.438  | 72.371     | 218.601   | 331.828    |
|            |  |            |  |            | 25.529  | 17.697     |
|            | $W=340 (p=0), \nu = 46$<br>$R_1^2 = 0.902, R_2^2 = 0.207$                                      |            | $W = 339.31 (p=0), \nu = 47$<br>$R_1^2 = 0.926, R_2^2 = 0.0633$  |            | $W = 325.63 (p=0), \nu = 45$<br>$R_1^2 = 0.979, R_2^2 = 0.0199$   |            |

## 4.4 MLE for Autocorrelated Disturbances

Time series are often homoscedastic so that in the presence of autocorrelated disturbances the covariance matrix may be

$$\sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{pmatrix}. \quad (4.31)$$

The values that appear off the diagonal depend on the model used for the disturbance.

A variety of covariance structures such as AR(1), heterogeneous AR(1), compound symmetry, heterogeneous compound symmetry and Toeplitz structures can be fitted. Software such as the SAS program PROC MIXED can be used for this purpose. However, using maximum likelihood under constraints, models can be fitted by imposing appropriate constraints on the expected value of the canonical statistics (cf. (4.3) and (4.4)). Likelihood equations need not be derived for specific parameter structures. A covariance matrix with AR(1) disturbances is given by:

$$\sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} 1 & r^1 & \dots & r^{n-1} \\ r^1 & 1 & \dots & r^{n-2} \\ & & \vdots & \\ r^{n-1} & r^{n-2} & \dots & 1 \end{pmatrix} \quad (4.32)$$

Using maximum likelihood under constraints, one can additionally provide for heteroskedasticity in the covariance structure, e.g. a covariance structure with AR(1) disturbances with a variance trend:

$$\sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} c^0 & c^0 r^1 & \dots & c^0 r^{n-1} \\ c^0 r^1 & c^1 r^0 & \dots & c^0 r^{n-2} \\ & & \vdots & \\ c^0 r^{n-1} & c^0 r^{n-2} & \dots & c^{n-1} r^0 \end{pmatrix} \quad (4.33)$$

No reparameterization of the likelihood function is required. The new structure only impacts on the constraints implied by the model and is illustrated in the next subsection.

In fitting such a covariance structure it is important that the covariance matrix of the canonical statistics used in the iterative procedure is initialized with values which exhibit the hypothesized structure. The initial matrix is required to be a proper covariance matrix. Since the covariance matrix is a function of the mean vector, care should be taken that the covariance matrix remains positive definite.

This is accomplished by expressing  $vec(\Sigma) = \mathbf{m}_2 - vec(\mathbf{m}_1 \mathbf{m}'_1)$  as

$$vec(\Sigma) = \mathbf{m}_2 - [\mathbf{I}_n \otimes \mathbf{I}_n - \mathbf{Q}_X \otimes \mathbf{Q}_X] vec(\mathbf{m}_1 \mathbf{m}'_1) = \mathbf{m}_2 - \mathbf{Q}_Q vec(\mathbf{m}_1 \mathbf{m}'_1), \text{ say} \quad (4.34)$$

with  $\mathbf{Q}_X$  the projection matrix given in (4.14).

#### 4.4.1 AR(1) covariance structure with heteroskedasticity

In the following example a variety of complex covariance structures are fitted to a time series consisting of an investment index (Real Gross Fixed Capital Index of South Africa) from 1999 to 2008 (South African Reserve Bank, 2009). A quadratic trend in time was fitted to this index, given a covariance matrix with heteroskedastic AR(1) structure. Typical elements of the mean vector and covariance matrix are modelled as

$$\begin{aligned} \mu_i &= a + bt_i + dt_i^2 \\ \sigma_{ij} &= \sigma^2 c^{i-1} r^{|i-j|} \end{aligned} \quad (4.35)$$

where  $t_i$  denotes time  $i, i = 1, \dots, n$ . For the usual AR(1) structure  $c = 1$  in the expression for  $\sigma_{ij}$  above.

In matrix notation the model is expressed as  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  with  $\mathbf{X} = (\mathbf{1} \ t \ t^2)$  and  $\mathbf{t}' = (1 \ 2, \dots, n)$ . Let  $\mathbf{Q}_X$  be the matrix which projects orthogonal to the columns of  $\mathbf{X}$ . Then the model (4.35) implies the constraints:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}_1) \\ \mathbf{g}_2(vec(\Sigma)) \end{pmatrix} \quad (4.36)$$

where  $\mathbf{g}_1(\mathbf{m}_1) = \mathbf{Q}_X \mathbf{m}_1$  with corresponding matrix of derivatives  $\mathbf{G}_{1m} = (\mathbf{Q}_X \ \mathbf{0}_{n \times n^2})$ .

The elements of the constraint  $\mathbf{g}_2(vec(\Sigma)) : \frac{1}{2}n(n+1) \times 1$  are given by

$$\sigma_{ij} - \sigma^2 c^{i-1} r^{j-i}, \quad i = 1, \dots, n, \quad j = i, \dots, n$$

where  $\sigma_{ij}$  the  $\frac{1}{2}n(n+1)$  distinct elements of  $\Sigma$ . The corresponding matrix of derivatives,  $\mathbf{G}_{2m}$ , is given by

$$\frac{\partial \mathbf{g}_2(vec(\Sigma))}{\partial \mathbf{m}} = \frac{\partial \mathbf{g}_2(vec(\Sigma))}{\partial vec(\Sigma)} \times \frac{\partial vec(\Sigma)}{\partial \mathbf{m}}$$

where (cf.  $\mathbf{Q}_Q$  in (4.34))

$$\frac{\partial vec(\Sigma)}{\partial \mathbf{m}} = [\mathbf{Q}_Q(-\mathbf{I}_n \otimes \mathbf{m}_1 - \mathbf{m}_1 \otimes \mathbf{I}_n), \ \mathbf{I}_{n^2}].$$

The rows of  $\frac{\partial g_2(\text{vec}(\boldsymbol{\Sigma}))}{\partial \text{vec}(\boldsymbol{\Sigma})}$  are given by:

$$\frac{\partial(\sigma_{ij} - \sigma^2 c^{i-1} r^{j-i})}{\partial \text{vec}(\boldsymbol{\Sigma})} = \mathbf{I}_n[i, \cdot] \otimes \mathbf{I}_n[\cdot, j] - \frac{\partial \sigma^2}{\partial \text{vec}(\boldsymbol{\Sigma})} c^{(i-1)} r^{(j-i)} \\ - \sigma^2 (i-1) c^{(i-2)} \frac{\partial c}{\partial \text{vec}(\boldsymbol{\Sigma})} r^{(j-i)} - \sigma^2 c^{i-1} (j-i) r^{j-i-1} \frac{\partial r}{\partial \text{vec}(\boldsymbol{\Sigma})}$$

where:

$\sigma^2 = \sigma_{11}$ , the first element of  $\text{vec}(\boldsymbol{\Sigma})$ ,

$\frac{\partial \sigma^2}{\partial \text{vec}(\boldsymbol{\Sigma})} = \mathbf{I}_{n^2}[1, \cdot]$ , the first row of the  $n^2 \times n^2$  identity matrix,

$$c = \frac{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_2 \text{vec}(\boldsymbol{\Sigma})]}{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]},$$

i.e.  $c$  the coefficient of the projection of  $\mathbf{C}_2 \text{vec}(\boldsymbol{\Sigma})$  onto  $\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})$ ,

$\mathbf{C}_1$  selects the first  $n - 1$  diagonal elements of  $\boldsymbol{\Sigma}$ ,

$\mathbf{C}_2$  selects the last  $n - 1$  diagonal elements of  $\boldsymbol{\Sigma}$ ,

$$r = \frac{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_3 \text{vec}(\boldsymbol{\Sigma})]}{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]},$$

i.e.  $r$  the coefficient of the projection of  $\mathbf{C}_3 \text{vec}(\boldsymbol{\Sigma})$  onto  $\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})$ ,

$\mathbf{C}_3$  selects from  $\text{vec}(\boldsymbol{\Sigma})$  the set of  $n - 1$  first-order covariances.

The derivative of  $c$  and  $r$  are respectively given by:

$$\frac{\partial c}{\partial \text{vec}(\boldsymbol{\Sigma})} = \frac{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})] [\text{vec}(\boldsymbol{\Sigma})]' (\mathbf{C}'_1 \mathbf{C}_2 + \mathbf{C}'_2 \mathbf{C}_1) - 2 [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_2 \text{vec}(\boldsymbol{\Sigma})] [\text{vec}(\boldsymbol{\Sigma})]' (\mathbf{C}'_1 \mathbf{C}_1)}{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]^2}$$

$$\frac{\partial r}{\partial \text{vec}(\boldsymbol{\Sigma})} = \frac{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})] [\text{vec}(\boldsymbol{\Sigma})]' (\mathbf{C}'_1 \mathbf{C}_3 + \mathbf{C}'_3 \mathbf{C}_1) - 2 [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_3 \text{vec}(\boldsymbol{\Sigma})] [\text{vec}(\boldsymbol{\Sigma})]' (\mathbf{C}'_1 \mathbf{C}_1)}{[\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]' [\mathbf{C}_1 \text{vec}(\boldsymbol{\Sigma})]^2}.$$

In Table 4.4 the data, together with estimated parameters and measure of fit under model (4.35), are given. The observed and fitted indices are very close. The model fits quite well. In comparison to earlier applications considered, convergence in the case of autocorrelated disturbances may be slower since a generalized inverse of the potentially large matrix  $\mathbf{G}'_t \mathbf{V} \mathbf{G}'_m$  needs to be obtained in the iterative process.

SAS program: AR1.SAS

Table 4.4: Mle's of AR(1) covariance structure with heteroskedasticity.

| Real Gross Fixed Capital Index of South Africa: 1999-2008 |            |           |                  |           |            |       |       |       |       |
|---|------------|-----------|------------------|-----------|------------|-------|-------|-------|-------|
| 95.9  | 100.0      | 103.5     | 107.3            | 117.1     | 127.6      | 140.6 | 159.2 | 185.1 | 203.9 |
| Parameter values  |            |           |                  |           |            |       |       |       |       |
| $a=100.684$   | $b=-4.043$ | $d=1.444$ | $\sigma^2=2.506$ | $c=1.080$ | $r=-0.113$ |       |       |       |       |
| Estimated mean values                                     |            |           |                  |           |            |       |       |       |       |
| 98.1  | 98.4       | 101.5     | 107.6            | 116.6     | 128.4      | 143.1 | 160.8 | 181.3 | 204.7 |
| Estimated (co)variances and correlations*                 |            |           |                  |           |            |       |       |       |       |
| 2.51  | -0.28      | 0.01      | 0                | 0         | 0          | 0     | 0     | 0     | 0     |
| -0.11   | 2.71       | -0.31     | 0.03             | 0         | 0          | 0     | 0     | 0     | 0     |
| 0.01  | -0.11      | 2.92      | -0.33            | 0.03      | 0          | 0     | 0     | 0     | 0     |
| 0   | 0.01       | -0.11     | 3.16             | -0.36     | 0.04       | 0     | 0     | 0     | 0     |
| 0   | 0          | 0.01      | -0.11            | 3.41      | -0.39      | 0.04  | 0     | 0     | 0     |
| 0   | 0          | 0         | 0.01             | -0.11     | 3.68       | -0.42 | 0.05  | -0.01 | 0     |
| 0   | 0          | 0         | 0                | 0.01      | -0.11      | 3.97  | -0.45 | 0.05  | -0.01 |
| 0   | 0          | 0         | 0                | 0         | 0.01       | -0.11 | 4.29  | -0.49 | 0.05  |
| 0   | 0          | 0         | 0                | 0         | 0          | 0.01  | -0.11 | 4.64  | -0.52 |
| 0   | 0          | 0         | 0                | 0         | 0          | 0     | 0.01  | -0.11 | 5.01  |
| $W=57.16 (p=0.54), \nu = 59$                              |            |           |                  |           |            |       |       |       |       |

\*Variances on diagonal, covariances above diagonal, correlations below diagonal

#### 4.4.2 Toeplitz(2) covariance structure with heteroskedasticity

A covariance structure with Toeplitz(2) disturbances implies equal variances, equal first-order covariances and is specified by:

$$\begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_2 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \sigma_2 & \sigma_1 & \sigma^2 \end{pmatrix}. \quad (4.37)$$

Let  $\mathbf{D}$  be the matrix which selects from  $\text{vec}(\boldsymbol{\Sigma})$  the distinct elements of  $\boldsymbol{\Sigma}$ : first the  $n$  variances on the main diagonal, then the  $n - 1$  first-order covariances, then the  $n - 2$  second-order covariances, etc. This enables one to model each set of covariances separately. One can allow for a trend in the variances over time, another trend in first-order covariances over time, etc.

In Table 4.5 the results are given for a model which specifies a quadratic trend in time on the mean structure but with a covariance structure which allows for separate trends over time in the variances and first-order covariances. Typical elements of the mean vector and covariance matrix (cf (4.46)) are modelled as

$$\begin{aligned} \mu_i &= a + bt_i + dt_i^2 \\ \sigma_{ii} &= ft_i \\ \sigma_{i,i+1} &= gt_i \\ \sigma_{i,j} &= 0, \quad j > i + 1 \end{aligned} \quad (4.38)$$

or in matrix notation:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} = (\mathbf{1} \quad \mathbf{t} \quad \mathbf{t}^2)\boldsymbol{\beta} \\ \boldsymbol{\sigma}_1 &= \mathbf{Z}_1\boldsymbol{\alpha}_1 = \mathbf{t}\boldsymbol{\alpha}_1 \\ \boldsymbol{\sigma}_2 &= \mathbf{Z}_2\boldsymbol{\alpha}_2 = \mathbf{t}[1 : n - 1,]\boldsymbol{\alpha}_2 \\ \boldsymbol{\sigma}_3 &= \mathbf{0}_{\frac{1}{2}n(n+1)-2n+1 \times 1} \end{aligned} \quad (4.39)$$

where  $\mathbf{t}' = (1, 2, \dots, n)$ . Let  $\mathbf{P}_{Z_1}$  and  $\mathbf{P}_{Z_2}$  be the projection matrices corresponding to  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  respectively.

Table 4.5: Mle's of Toeplitz(2) covariance structure with heteroskedasticity.

| Real Gross Fixed Capital Index of South Africa: 1999-2008 |            |           |           |           |       |       |       |       |       |
|---|------------|-----------|-----------|-----------|-------|-------|-------|-------|-------|
| 95.9  | 100.0      | 103.5     | 107.3     | 117.1     | 127.6 | 140.6 | 159.2 | 185.1 | 203.9 |
| Parameter values  |            |           |           |           |       |       |       |       |       |
| $a=97.401$  | $b=-2.389$ | $d=1.291$ | $f=0.950$ | $g=0.415$ |       |       |       |       |       |
| Estimated mean values                                     |            |           |           |           |       |       |       |       |       |
| 96.3  | 97.8       | 101.8     | 108.5     | 117.7     | 129.5 | 143.9 | 160.9 | 180.4 | 202.6 |
| Estimated (co)variances and correlations*                 |            |           |           |           |       |       |       |       |       |
| 0.95  | 0.42       | 0         | 0         | 0         | 0     | 0     | 0     | 0     | 0     |
| 0.31  | 1.90       | 0.83      | 0         | 0         | 0     | 0     | 0     | 0     | 0     |
| 0   | 0.36       | 2.85      | 1.25      | 0         | 0     | 0     | 0     | 0     | 0     |
| 0   | 0          | 0.38      | 3.80      | 1.66      | 0     | 0     | 0     | 0     | 0     |
| 0   | 0          | 0         | 0.39      | 4.75      | 2.08  | 0     | 0     | 0     | 0     |
| 0   | 0          | 0         | 0         | 0.40      | 5.70  | 2.49  | 0     | 0     | 0     |
| 0   | 0          | 0         | 0         | 0         | 0.40  | 6.65  | 2.91  | 0     | 0     |
| 0   | 0          | 0         | 0         | 0         | 0     | 0.41  | 7.60  | 3.32  | 0     |
| 0   | 0          | 0         | 0         | 0         | 0     | 0     | 0.41  | 8.55  | 3.74  |
| 0   | 0          | 0         | 0         | 0         | 0     | 0     | 0     | 0.41  | 9.50  |
| $W=73.46$ ( $p=0.11$ ), $\nu = 60$                        |            |           |           |           |       |       |       |       |       |

\*Variances on diagonal, covariances above diagonal, correlations below diagonal

The implied constraints for this model are given by

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}_1) \\ \mathbf{g}_2(\text{vec}(\boldsymbol{\Sigma})) \end{pmatrix} \text{ where } \mathbf{g}_1(\mathbf{m}_1) = \mathbf{Q}_X \mathbf{m}_1, \mathbf{G}_{1m} = (\mathbf{Q}_X \quad \mathbf{0}_{n \times n^2}) \text{ and}$$

$$\mathbf{g}_2(\text{vec}(\boldsymbol{\Sigma})) = \mathbf{Q}_Z \mathbf{D} \text{vec}(\boldsymbol{\Sigma}), \mathbf{G}_{2m} = \mathbf{Q}_Z \mathbf{D}$$

$$\text{where } \mathbf{Q}_Z = \mathbf{I}_{\frac{1}{2}n(n+1)} - \begin{pmatrix} \mathbf{P}_{Z_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{Z_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{4.40}$$

SAS program: TOEPLITZ.SAS



### 4.4.3 Modelling in terms of explanatory variables

The investment index can additionally be modelled in terms of explanatory variables such as GDP and interest (South African Reserve Bank, 2009), given the same complex AR(1) covariance structure with variance trend considered in Section 4.4.1.

Consider the model

$$\begin{aligned}\mu_i &= a + bt_i + dt_i^2 + fx_{1i} + gx_{2i} \\ \sigma_{ij} &= \sigma^2 c^{i-1} r^{|i-j|}\end{aligned}\tag{4.41}$$

Constraints are similar to (4.36) with  $\mathbf{X} = (\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2)$  where  $\mathbf{X}_0 = (\mathbf{1} \ t \ t^2)$ ,  $\mathbf{X}_1$  the GDP and  $\mathbf{X}_2$  the interest for the same time periods. Results are given in Table 4.6.

The model fits very well without the restrictive (unrealistic) assumption of constant variances.

|                      |
|----------------------|
| SAS program: AR1.SAS |
|----------------------|

In Table 4.7 a comparison of least squares estimates (lse's) and mle's is given.

The substantial influence that the estimation method has on predictions for 2009 to 2013 is illustrated in Table 4.8.

Table 4.6: Mle's of AR(1) covariance structure with heteroskedasticity in terms of explanatory variables.

| Real Gross Fixed Capital Index of South Africa: 1999-2008 |            |           |           |            |       |       |                  |           |            |
|---|------------|-----------|-----------|------------|-------|-------|------------------|-----------|------------|
| 95.9  | 100.0      | 103.5     | 107.3     | 117.1      | 127.6 | 140.6 | 159.2            | 185.1     | 203.9      |
| GDP South Africa: 1999-2008                               |            |           |           |            |       |       |                  |           |            |
| 96.0  | 100.0      | 102.7     | 106.5     | 109.8      | 115.2 | 120.9 | 127.3            | 133.8     | 137.9      |
| Interest South Africa: 1999-2008                          |            |           |           |            |       |       |                  |           |            |
| 9.70  | 8.39       | 5.71      | 2.30      | 3.81       | 8.13  | 4.67  | 3.24             | 0.88      | -2.40      |
| Parameter values  |            |           |           |            |       |       |                  |           |            |
| $a=0.193$   | $b=-6.860$ | $d=1.180$ | $f=1.113$ | $g=-0.425$ |       |       | $\sigma^2=1.527$ | $c=1.111$ | $r=-0.318$ |
| Estimated mean values                                     |            |           |           |            |       |       |                  |           |            |
| 97.3  | 99.0       | 102.2     | 109.2     | 116.0      | 126.3 | 142.6 | 161.2            | 182.6     | 204.1      |
| Estimated (co)variances and correlations*                 |            |           |           |            |       |       |                  |           |            |
| 1.53  | -0.49      | 0.15      | -0.05     | 0.02       | 0     | 0     | 0                | 0         | 0          |
| -0.30   | 1.70       | -0.54     | 0.17      | -0.05      | 0.02  | -0.01 | 0                | 0         | 0          |
| 0.09  | -0.30      | 1.88      | -0.60     | 0.19       | -0.06 | 0.02  | -0.01            | 0         | 0          |
| -0.03   | 0.09       | -0.30     | 2.09      | -0.67      | 0.21  | -0.07 | 0.02             | -0.01     | 0          |
| 0.01  | -0.03      | 0.09      | -0.30     | 2.33       | -0.74 | 0.24  | -0.07            | 0.02      | -0.01      |
| 0   | 0.01       | -0.03     | 0.09      | -0.30      | 2.58  | -0.82 | 0.26             | -0.08     | 0.03       |
| 0   | 0          | 0.01      | -0.03     | 0.09       | -0.30 | 2.87  | -0.91            | 0.29      | -0.09      |
| 0   | 0          | 0         | 0.01      | -0.03      | 0.09  | -0.30 | 3.19             | -1.01     | 0.32       |
| 0   | 0          | 0         | 0         | 0.01       | -0.03 | 0.09  | -0.30            | 3.54      | -1.13      |
| 0   | 0          | 0         | 0         | 0          | 0.01  | -0.03 | 0.09             | -0.30     | 3.94       |
| $W=57.46$ ( $p=0.46$ ), $\nu = 57$                        |            |           |           |            |       |       |                  |           |            |

\*Variances on diagonal, covariances above diagonal, correlations below diagonal

Table 4.7: Comparison of lse's and mle's.

| Parameter | Lse's   | Mle's  |
|-----------|---------|--------|
| $a$       | -14.883 | 0.193  |
| $b$       | -7.192  | -6.860 |
| $d$       | 1.151   | 1.180  |
| $f$       | 1.267   | 1.113  |
| $g$       | -0.366  | -0.425 |

Table 4.8: Comparison of predictions based on least squares estimation (LSE) and MLE.

| $y$   | Year | $t$ | $\hat{\mu}$ (LSE) | $\hat{\mu}$ (MLE) |
|-------|------|-----|-------------------|-------------------|
| 95.9  | 1999 | 1   | 97.2              | 97.3              |
| 100.0 | 2000 | 2   | 99.0              | 99.0              |
| 103.5 | 2001 | 3   | 102.0             | 102.2             |
| 107.3 | 2002 | 4   | 108.9             | 109.2             |
| 117.1 | 2003 | 5   | 115.7             | 116.0             |
| 127.6 | 2004 | 6   | 126.4             | 126.3             |
| 140.6 | 2005 | 7   | 142.7             | 142.6             |
| 159.2 | 2006 | 8   | 161.4             | 161.2             |
| 185.1 | 2007 | 9   | 182.9             | 182.6             |
| 203.9 | 2008 | 10  | 203.9             | 204.1             |
|       | 2009 | 11  | 220.9             | 222.1             |
|       | 2010 | 12  | 240.2             | 242.3             |
|       | 2011 | 13  | 261.8             | 265.0             |
|       | 2012 | 14  | 285.6             | 290.0             |
|       | 2013 | 15  | 311.8             | 317.3             |

# Chapter 5

## MLE for Multivariate Normal Samples

Estimation and testing procedures in the case of several multivariate normal populations with unknown and possibly unequal covariance matrices have been studied extensively in statistical literature. MLE of parameters and parametric functions in this case is often difficult to perform or often implies that the likelihood function has to be solved numerically. In addition to that, the likelihood equations may have several roots. Estimation of the common mean vector of several multivariate normal populations (Rukhin, Biggerstaff & Vangel, 2000), the multivariate Behrens-Fisher problem (Buot, Hosten & Richards, 2007) and estimation of proportional covariance matrices (Jensen & Madsen, 2004) represent such examples.

Although some problems of this nature can be solved numerically within the mixed model framework, the approach proposed in this chapter provides a natural and elegant way of analyzing problems of this kind. A solution is provided to MLE problems where normally the likelihood equations do not exist or are hard to derive and solve. Mean and covariance structures are modelled directly. A wider class of problems can be addressed in this way. One is not restricted to a fixed set of patterned covariance structures. Proportionality of variances and proportionality of covariance matrices are easily accommodated.

### 5.1 Introductory Example

The essence of the procedure can be described by the simple case where we have a sample mean vector  $\bar{\mathbf{y}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$  and sample covariance matrix  $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 8 \end{pmatrix}$  of size  $n = 10$  from a  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution where it is assumed that the variance of each variable is equal to its squared mean. This

cannot be accomplished by PROC MIXED.

The vector of canonical statistics,  $\mathbf{t}$ , and  $Cov(\mathbf{t})$  are given by (2.6):

$$\mathbf{t} = \begin{pmatrix} \bar{\mathbf{y}} \\ vec(\mathbf{S}) \end{pmatrix} \text{ and } Cov(\mathbf{t}) = \mathbf{V} = \begin{pmatrix} \frac{1}{n}\Sigma & \frac{1}{n}(\Sigma \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \Sigma)' \\ \frac{1}{n}(\Sigma \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \Sigma) & \frac{1}{n}(I_{p^2} + \mathbf{K})[\Sigma \otimes \Sigma + \Sigma \otimes \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \otimes \Sigma] \end{pmatrix}.$$

Suppose the variances are assumed to be equal to the squared means, i.e.  $\sigma_{11} = \mu_1^2$  and  $\sigma_{22} = \mu_2^2$ . If

the  $i$ -th element of  $vec(\Sigma)$  is indicated by  $vec(\Sigma)[i]$  and the  $i$ -th element of  $\boldsymbol{\mu}$  by  $\boldsymbol{\mu}[i]$ , then

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}) \\ \mathbf{g}_2(\mathbf{m}) \end{pmatrix} \text{ where}$$

$$\mathbf{g}_1(\mathbf{m}) = \sigma_{11} - \mu_1^2 = vec(\Sigma)[1] - \boldsymbol{\mu}[1]^2 = \{\mathbf{m}_2 - vec(\mathbf{m}_1\mathbf{m}_1')\}[1] - \mathbf{m}_1[1]^2,$$

$$\mathbf{g}_2(\mathbf{m}) = \sigma_{22} - \mu_2^2 = vec(\Sigma)[4] - \boldsymbol{\mu}[2]^2 = \{\mathbf{m}_2 - vec(\mathbf{m}_1\mathbf{m}_1')\}[4] - \mathbf{m}_1[2]^2,$$

with derivative

$$\mathbf{G}_m = \begin{pmatrix} \mathbf{G}_1(\mathbf{m}) \\ \mathbf{G}_2(\mathbf{m}) \end{pmatrix} = \begin{pmatrix} \frac{\partial vec(\Sigma)[1]}{\partial vec(\Sigma)} \times \frac{\partial vec(\Sigma)}{\partial \mathbf{m}} - 2\boldsymbol{\mu}[1] \times \frac{\partial \boldsymbol{\mu}[1]}{\partial \mathbf{m}} \\ \frac{\partial vec(\Sigma)[4]}{\partial vec(\Sigma)} \times \frac{\partial vec(\Sigma)}{\partial \mathbf{m}} - 2\boldsymbol{\mu}[2] \times \frac{\partial \boldsymbol{\mu}[2]}{\partial \mathbf{m}} \end{pmatrix}, \mathbf{G}_t = \left. \frac{\partial \mathbf{g}(\mathbf{m})}{\partial \mathbf{m}} \right|_{\mathbf{m} = \mathbf{t}}.$$

The iteration process proceeds as indicated in Table 5.1.

In each new iteration over  $\mathbf{m}$ , the observed  $\mathbf{t}$  ( $\mathbf{t}_0$ ) is used as initial value for  $\mathbf{t}$ . The mle for  $\Sigma$  under

the restrictions  $\sigma_{11} = \mu_1^2$  and  $\sigma_{22} = \mu_2^2$  are calculated as a function of the elements of  $\widehat{\mathbf{m}}$ , namely:

$$\widehat{\boldsymbol{\mu}}_1 = \widehat{\mathbf{m}}_1 = \begin{pmatrix} 2.060 \\ 2.820 \end{pmatrix} \text{ and } vec(\widehat{\Sigma}) = \widehat{\mathbf{m}}_2 - vec(\widehat{\mathbf{m}}_1\widehat{\mathbf{m}}_1') = \begin{pmatrix} 4.24 \\ 3.67 \\ 3.67 \\ 7.95 \end{pmatrix}.$$

SAS program: MULTIV.SAS

Table 5.1: Double iteration process illustrated for the example in Section 5.1.

|   |   | Iteration over $t$  |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
|   | Iteration over $m$  | Starting Value  | 1   | 2   | 3   |   |   |   |
| 1 | $\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$ | $\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$                 | $\begin{pmatrix} 2.047 \\ 2.800 \\ 8.374 \\ 9.374 \\ 9.374 \\ 15.596 \end{pmatrix}$ | $\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$ | $\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$ |   |   |   |
|   | 2   | $\begin{pmatrix} 2.045 \\ 2.789 \\ 8.363 \\ 9.356 \\ 9.356 \\ 15.562 \end{pmatrix}$ | $\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$                 | $\begin{pmatrix} 2.062 \\ 2.828 \\ 8.494 \\ 9.490 \\ 9.490 \\ 15.931 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$ |   |   |
|   |   | 3   | $\begin{pmatrix} 2.060 \\ 2.821 \\ 8.487 \\ 9.479 \\ 9.479 \\ 15.914 \end{pmatrix}$ | $\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$                 | $\begin{pmatrix} 2.062 \\ 2.827 \\ 8.496 \\ 9.489 \\ 9.489 \\ 15.924 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$ |   |
|   |   |   | 4   | $\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$ | $\begin{pmatrix} 2 \\ 3 \\ 8.5 \\ 9.6 \\ 9.6 \\ 16.2 \end{pmatrix}$                 | $\begin{pmatrix} 2.062 \\ 2.827 \\ 8.496 \\ 9.489 \\ 9.489 \\ 15.924 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$ | $\begin{pmatrix} 2.060 \\ 2.820 \\ 8.489 \\ 9.477 \\ 9.477 \\ 15.907 \end{pmatrix}$ |

## 5.2 Independent Multivariate Normal Samples

The theory and methodology that follow can be applied for any number,  $k$ , of samples which may stem from any experimental design. In view of the examples that follow in the next two sections and for the sake of simplicity, the theoretical results, an extension of the theory given in Chapter 2 (cf. (2.6)), are given for  $k = 4$  multivariate normal samples. The results for general  $k$  follow directly.

Suppose that  $\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in_i}$  represent  $n_i$  independent observations of 4 independent random samples from  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  distributions ( $i = 1, 2, \dots, 4$ ). Let  $\bar{\mathbf{y}}_i$  represent the sample mean vector

$$\bar{\mathbf{y}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \quad (5.1)$$

and  $\mathbf{S}_i$  the matrix of mean sums of squares and products of the  $i$ -th sample

$$\mathbf{S}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \mathbf{y}'_{ij}. \quad (5.2)$$

Note that  $n_i$  is not assumed to be larger than  $p + 1$ . If any of the  $n_i$  is less than  $p + 1$ , the mle's in general do not exist unless certain restrictions on parameters are imposed. The likelihood function for the  $k$  multivariate samples can be expressed in its canonical exponential form as follows:

$$\begin{aligned} & L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4) \\ &= \prod_{i=1}^4 \det(2\pi\boldsymbol{\Sigma}_i)^{-n_i/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \left[ \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)(\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \right] \right\} \\ &= \prod_{i=1}^4 \exp \left\{ n_i \boldsymbol{\mu}'_i \boldsymbol{\Sigma}_i^{-1} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \right) - \frac{n_i}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \mathbf{y}'_{ij} \right) - \frac{n_i}{2} \boldsymbol{\mu}'_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \frac{n_i}{2} \ln[\det(2\pi\boldsymbol{\Sigma}_i)] \right\} \\ &= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})] \quad \text{where} \end{aligned} \quad (5.3)$$

$$\mathbf{t} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{y}_{1j} \\ \text{vec}\left(\frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{y}_{1j} \mathbf{y}'_{1j}\right) \\ \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{y}_{2j} \\ \text{vec}\left(\frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{y}_{2j} \mathbf{y}'_{2j}\right) \\ \frac{1}{n_3} \sum_{j=1}^{n_3} \mathbf{y}_{3j} \\ \text{vec}\left(\frac{1}{n_3} \sum_{j=1}^{n_3} \mathbf{y}_{3j} \mathbf{y}'_{3j}\right) \\ \frac{1}{n_4} \sum_{j=1}^{n_4} \mathbf{y}_{4j} \\ \text{vec}\left(\frac{1}{n_4} \sum_{j=1}^{n_4} \mathbf{y}_{4j} \mathbf{y}'_{4j}\right) \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{y}}_1 \\ \text{vec}(\mathbf{S}_1) \\ \bar{\mathbf{y}}_2 \\ \text{vec}(\mathbf{S}_2) \\ \bar{\mathbf{y}}_3 \\ \text{vec}(\mathbf{S}_3) \\ \bar{\mathbf{y}}_4 \\ \text{vec}(\mathbf{S}_4) \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} n_1 \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 \\ -\frac{n_1}{2} \text{vec}(\boldsymbol{\Sigma}_1^{-1}) \\ n_2 \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2 \\ -\frac{n_2}{2} \text{vec}(\boldsymbol{\Sigma}_2^{-1}) \\ n_3 \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\mu}_3 \\ -\frac{n_3}{2} \text{vec}(\boldsymbol{\Sigma}_3^{-1}) \\ n_4 \boldsymbol{\Sigma}_4^{-1} \boldsymbol{\mu}_4 \\ -\frac{n_4}{2} \text{vec}(\boldsymbol{\Sigma}_4^{-1}) \end{pmatrix} \quad (5.4)$$

$$\text{and } E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \text{vec}(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}'_1) \\ \boldsymbol{\mu}_2 \\ \text{vec}(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}'_2) \\ \boldsymbol{\mu}_3 \\ \text{vec}(\boldsymbol{\Sigma}_3 + \boldsymbol{\mu}_3 \boldsymbol{\mu}'_3) \\ \boldsymbol{\mu}_4 \\ \text{vec}(\boldsymbol{\Sigma}_4 + \boldsymbol{\mu}_4 \boldsymbol{\mu}'_4) \end{pmatrix} = \begin{pmatrix} \mathbf{m}_{11} \\ \mathbf{m}_{12} \\ \mathbf{m}_{21} \\ \mathbf{m}_{22} \\ \mathbf{m}_{31} \\ \mathbf{m}_{32} \\ \mathbf{m}_{41} \\ \mathbf{m}_{42} \end{pmatrix}. \quad (5.5)$$

The covariance matrix of  $\mathbf{t}$  (cf. (2.6)) is given by

$$\mathbf{V} = \text{Cov}(\mathbf{t}) = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{44} \end{pmatrix} \text{ with } \mathbf{V}_{ii} = \begin{pmatrix} \mathbf{V}_{11}^i & \mathbf{V}_{12}^i \\ \mathbf{V}_{21}^i & \mathbf{V}_{22}^i \end{pmatrix} \text{ and} \quad (5.6)$$

$$\begin{aligned} \mathbf{V}_{11}^i &= \frac{1}{n_i} \boldsymbol{\Sigma}_i \\ \mathbf{V}_{21}^i &= \frac{1}{n_i} (\boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i + \boldsymbol{\mu}_i \otimes \boldsymbol{\Sigma}_i) \\ \mathbf{V}_{12}^i &= \mathbf{V}_{21}^{i'} \end{aligned} \quad \text{for } i = 1, 2, 3, 4 \quad (5.7)$$

$$\mathbf{V}_{22}^i = \frac{1}{n_i} (I_{p^2} + \mathbf{K}) [\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i \boldsymbol{\mu}'_i + \boldsymbol{\mu}_i \boldsymbol{\mu}'_i \otimes \boldsymbol{\Sigma}_i].$$



In the procedure (2.2) for estimating the  $\mu_i$  and  $\Sigma_i$  under restrictions, the vector  $\mathbf{t}$  of canonical statistics with corresponding covariance matrix  $\mathbf{V}$  is used as point of departure for any model considered within this framework. The algorithm given in Chapter 2 is then used to obtain the mle's. In each section the specific expressions for any  $g(\mathbf{m})$  and  $\mathbf{G}_m$  are given. Constraints  $g(\mathbf{m})$  specified will always be functions of  $\mu_i$  and  $\Sigma_i$ . Consequently, the corresponding matrix of derivatives  $\mathbf{G}_m = \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}}$  may also be expressed as  $\mathbf{G}_m = \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}_\Sigma} \times \frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}}$  where  $\mathbf{m}'_\Sigma = \left( \text{vec}(\Sigma_1)', \text{vec}(\Sigma_2)', \text{vec}(\Sigma_3)', \text{vec}(\Sigma_4)' \right)$ . Basic derivatives that will be required throughout, are the following:

$$\begin{aligned} \frac{\partial \mu_1}{\partial \mathbf{m}} &= (\mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}) \\ \frac{\partial \mu_2}{\partial \mathbf{m}} &= (\mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}) \\ \frac{\partial \mu_3}{\partial \mathbf{m}} &= (\mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}) \\ \frac{\partial \mu_4}{\partial \mathbf{m}} &= (\mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{0}_{p \times p}, \quad \mathbf{0}_{p \times p^2}, \quad \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}) \\ \frac{\partial \mu}{\partial \mathbf{m}} &= \left( \frac{1}{4} \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \frac{1}{4} \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \frac{1}{4} \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2}, \quad \frac{1}{4} \mathbf{I}_p, \quad \mathbf{0}_{p \times p^2} \right) \end{aligned} \quad (5.8)$$

where  $\mu = (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4$ .

Since  $\frac{\partial \text{vec}(\mathbf{m}_{i1} \mathbf{m}'_{i1})}{\partial \mathbf{m}_{i1}} = \mathbf{I}_p \otimes \mathbf{m}_{i1} + \mathbf{m}_{i1} \otimes \mathbf{I}_p$  and  $\text{vec}(\Sigma_i) = \mathbf{m}_{i2} - \text{vec}(\mathbf{m}_{i1} \mathbf{m}'_{i1})$ , it follows that:

$$\frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}} = \begin{pmatrix} \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\Sigma_2)}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\Sigma_3)}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\Sigma_4)}{\partial \mathbf{m}} \end{pmatrix} = \oplus_{i=1}^4 (-\mathbf{I}_p \otimes \mathbf{m}_{i1} - \mathbf{m}_{i1} \otimes \mathbf{I}_p, \quad \mathbf{I}_{p^2}), \quad (5.9)$$

the direct sum (block diagonal structure) of the submatrices  $\left( \frac{\partial [\mathbf{m}_{i2} - \text{vec}(\mathbf{m}_{i1} \mathbf{m}'_{i1})]}{\partial \mathbf{m}_{i1}}, \frac{\partial [\mathbf{m}_{i2} - \text{vec}(\mathbf{m}_{i1} \mathbf{m}'_{i1})]}{\partial \mathbf{m}_{i2}} \right)$ .

### 5.3 Two Large Independent Multivariate Normal Samples

Johnson & Wichern (1982) and Nel & Van der Merwe (1986) considered two measurements of electrical usage during on-peak and off-peak hours, for homeowners with (Population 1) and without air conditioning (Population 2) respectively. Summary statistics, namely the sample mean vectors  $\bar{\mathbf{y}}_i$ , sample covariance matrices  $\mathbf{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)'$  and sample correlation coefficients  $r_i$ ,  $i = 1, 2$ , for samples of size  $n_1 = 45$  and  $n_2 = 55$  respectively, are given in Table 5.2. (The off-peak consumption is higher than the on-peak consumption because there are more off-peak hours

in a month). The SAS procedure MIXED cannot be used in this case, since the raw data are not available. More importantly, although PROC MIXED provides for a variety of covariance structures, it does not allow for structural relations between covariance matrices. In Table 5.2 the results for a variety of hypotheses on the two mean vectors and covariance matrices are also presented. These hypotheses imply specific constraints but the procedure does not require the maximization of a likelihood function.

### 5.3.1 $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

The usual hypothesis of equal mean vectors with homogeneous covariance matrices implies the constraints:

$$g(\mathbf{m}) = \begin{pmatrix} g_1(\mathbf{m}) \\ g_2(\mathbf{m}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \\ \text{vec}(\boldsymbol{\Sigma}_1) - \text{vec}(\boldsymbol{\Sigma}_2) \end{pmatrix} \text{ with } \mathbf{G}_m = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{m}} - \frac{\partial \boldsymbol{\mu}_2}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}} - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}} \end{pmatrix} \quad (5.10)$$

where

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{m}} &= (\mathbf{I}_p, \mathbf{0}_{p \times p^2}, \mathbf{0}_{p \times p}, \mathbf{0}_{p \times p^2}) \\ \frac{\partial \boldsymbol{\mu}_2}{\partial \mathbf{m}} &= (\mathbf{0}_{p \times p}, \mathbf{0}_{p \times p^2}, \mathbf{I}_p, \mathbf{0}_{p \times p^2}) \end{aligned} \quad (5.11)$$

and

$$\begin{aligned} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}} &= \frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}_\boldsymbol{\Sigma}} \times \frac{\partial \mathbf{m}_\boldsymbol{\Sigma}}{\partial \mathbf{m}} \\ &= (\mathbf{I}_{p^2}, \mathbf{0}_{p^2 \times p^2}) \times (-\mathbf{I}_p \otimes \mathbf{m}_{11} - \mathbf{m}_{11} \otimes \mathbf{I}_p, \mathbf{I}_{p^2}, \mathbf{0}_{p^2 \times p}, \mathbf{0}_{p^2 \times p^2}) \\ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}} &= \frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}_\boldsymbol{\Sigma}} \times \frac{\partial \mathbf{m}_\boldsymbol{\Sigma}}{\partial \mathbf{m}} \\ &= (\mathbf{0}_{p^2 \times p^2}, \mathbf{I}_{p^2}) \times (\mathbf{0}_{p^2 \times p}, \mathbf{0}_{p^2 \times p^2}, -\mathbf{I}_p \otimes \mathbf{m}_{21} - \mathbf{m}_{21} \otimes \mathbf{I}_p, \mathbf{I}_{p^2}) \end{aligned} \quad (5.12)$$

where  $\mathbf{m}'_\boldsymbol{\Sigma} = \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma}_1)', \text{vec}(\boldsymbol{\Sigma}_2)' \end{pmatrix}$ . From Table 5.2 it follows that this model does not fit well.

### 5.3.2 $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$

Proportional covariance matrices and proportional mean vectors seem to be a plausible model for this particular data set. The difficulties and limitations traditionally encountered with MLE in the case of proportional covariance matrices are discussed by authors such as Flury (1986), Eriksen (1987), Jensen & Johansen (1987) and Jensen & Madsen (2004). Consider the hypothesis  $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$  where, for example,  $c = \frac{\sigma_{11}^2}{\sigma_{11}^1}$  and  $\sigma_{11}^i$  denotes the variance of the first measurement in the  $i$ -th population. Using

procedure (2.2), only the following slight adjustment is made to the constraint  $g_2(\mathbf{m})$  in (5.10):

$$g_2(\mathbf{m}) = \text{vec}(\boldsymbol{\Sigma}_2) - c\text{vec}(\boldsymbol{\Sigma}_1) \text{ with derivative } \mathbf{G}_{2m} = \frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}} - c \frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}} - \text{vec}(\boldsymbol{\Sigma}_1) \frac{\partial c}{\partial \mathbf{m}} \quad (5.13)$$

and

$$\frac{\partial c}{\partial \mathbf{m}} = \frac{\partial c}{\partial \mathbf{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \mathbf{m}_{\boldsymbol{\Sigma}}}{\partial \mathbf{m}} \text{ where } \frac{\partial c}{\partial \mathbf{m}_{\boldsymbol{\Sigma}}} = \left( \frac{\sigma_{11}^2}{(\sigma_{11}^1)^2} \mathbf{I}_{p^2}[1,], \frac{1}{\sigma_{11}^1} \mathbf{I}_{p^2}[1,] \right), \quad (5.14)$$

$\mathbf{I}_{p^2}[1,]$  denotes the first row of  $\mathbf{I}_{p^2}$  and  $\frac{\partial \mathbf{m}_{\boldsymbol{\Sigma}}}{\partial \mathbf{m}}$  is obtained from (5.12).

In Table 5.2, the results for the hypothesis of proportional covariance matrices are given, with no constraints imposed on the mean vectors. This model fits well. Note that the proportional covariances imply equal correlation coefficients. This correlation coefficient  $\hat{\rho}_i$  is also reported.

### 5.3.3 $H_0 : \rho_1 = \rho_2$

The correlation coefficient  $\hat{\rho}_i$  which is obtained as a result of covariance matrices being proportional, differs from the estimated common correlation coefficient (cf. Manly & Rayner (1987)) which is obtained when the hypothesis of equal correlation coefficients is imposed as a direct constraint. This constraint, which is less restrictive than proportionality of complete covariance matrices, is given by:

$$\begin{aligned} g(\mathbf{m}) &= \rho_1 - \rho_2 = \frac{\sigma_{12}^1}{\sqrt{\sigma_{11}^1 \sigma_{22}^1}} - \frac{\sigma_{12}^2}{\sqrt{\sigma_{11}^2 \sigma_{22}^2}} \\ &= \frac{\text{vec}(\boldsymbol{\Sigma}_1)[2]}{\sqrt{\text{vec}(\boldsymbol{\Sigma}_1)[1] \text{vec}(\boldsymbol{\Sigma}_1)[p+2]}} - \frac{\text{vec}(\boldsymbol{\Sigma}_2)[2]}{\sqrt{\text{vec}(\boldsymbol{\Sigma}_2)[1] \text{vec}(\boldsymbol{\Sigma}_2)[p+2]}} \end{aligned} \quad (5.15)$$

where  $\rho_i$  and  $\sigma_{12}^i$  the correlation and covariance respectively between on-peak and off-peak electrical usage for population  $i$ , and  $\sigma_{11}^i, \sigma_{12}^i, \sigma_{22}^i$  the element in row 1, 2, and  $p+2$  of  $\text{vec}(\boldsymbol{\Sigma}_i)$  respectively,

$$\begin{aligned} \mathbf{G}_m &= \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}_{\boldsymbol{\Sigma}}} \times \frac{\partial \mathbf{m}_{\boldsymbol{\Sigma}}}{\partial \mathbf{m}} \text{ where } \frac{\partial g(\mathbf{m})}{\partial \mathbf{m}_{\boldsymbol{\Sigma}}} = \frac{\partial(\rho_1 - \rho_2)}{\partial \mathbf{m}_{\boldsymbol{\Sigma}}} = \\ &= \left( \frac{\mathbf{I}_{p^2}[2,]}{\sqrt{\text{vec}(\boldsymbol{\Sigma}_1)[1] \text{vec}(\boldsymbol{\Sigma}_1)[p+2]}} - \frac{\mathbf{I}_{p^2}[1,] \text{vec}(\boldsymbol{\Sigma}_1)[2]}{2\sqrt{\{\text{vec}(\boldsymbol{\Sigma}_1)[1]\}^3 \text{vec}(\boldsymbol{\Sigma}_1)[p+2]}} - \frac{\mathbf{I}_{p^2}[p+2,] \text{vec}(\boldsymbol{\Sigma}_1)[2]}{2\sqrt{\text{vec}(\boldsymbol{\Sigma}_1)[1] \{\text{vec}(\boldsymbol{\Sigma}_1)[p+2]\}^3}}, \mathbf{0}_{1 \times p^2} \right) \\ &- \left( \mathbf{0}_{1 \times p^2}, \frac{\mathbf{I}_{p^2}[2,]}{\sqrt{\text{vec}(\boldsymbol{\Sigma}_2)[1] \text{vec}(\boldsymbol{\Sigma}_2)[p+2]}} - \frac{\mathbf{I}_{p^2}[1,] \text{vec}(\boldsymbol{\Sigma}_2)[2]}{2\sqrt{\{\text{vec}(\boldsymbol{\Sigma}_2)[1]\}^3 \text{vec}(\boldsymbol{\Sigma}_2)[p+2]}} - \frac{\mathbf{I}_{p^2}[p+2,] \text{vec}(\boldsymbol{\Sigma}_2)[2]}{2\sqrt{\text{vec}(\boldsymbol{\Sigma}_2)[1] \{\text{vec}(\boldsymbol{\Sigma}_2)[p+2]\}^3}} \right). \end{aligned} \quad (5.16)$$

The results are also given in Table 5.2. Since the hypothesis cannot be formulated in terms of the parameters of the means or (co)variances, PROC MIXED cannot be used in this case.

### 5.3.4 $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1, H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$

For the hypothesis of proportional mean vectors,  $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1$ , the following adjustment to  $g_1(\mathbf{m})$  is required:

$$g_1(\mathbf{m}) = \boldsymbol{\mu}_2 - \sqrt{c}\boldsymbol{\mu}_1 \text{ with derivative } \mathbf{G}_{1m} = \frac{\partial \boldsymbol{\mu}_2}{\partial \mathbf{m}} - \sqrt{c} \frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{m}} - \frac{1}{2\sqrt{c}} \left( \boldsymbol{\mu}_1 \frac{\partial c}{\partial \mathbf{m}} \right) \quad (5.17)$$

This model seems to be the proper solution to the problem stated by Johnson & Wichern (1982). The results are given in Table 5.2.

SAS program: LMVNORMAL.SAS

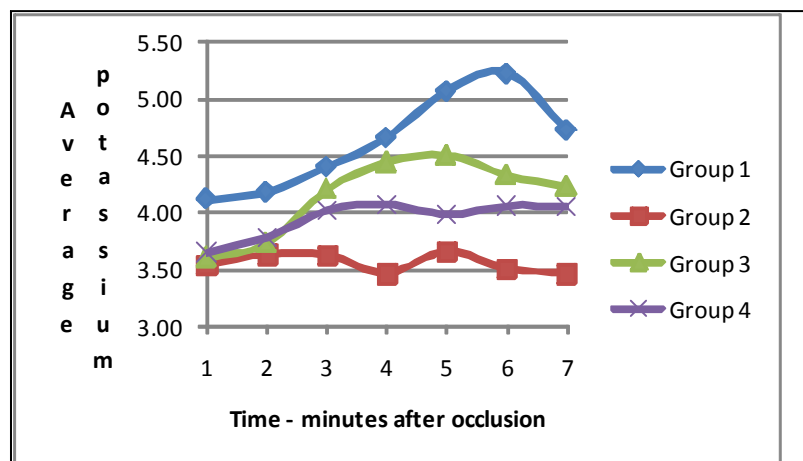
Table 5.2: A Behrens-Fisher problem for large multivariate samples.

|   |  |   |   |
|---|--|---|---|
| $\bar{\mathbf{y}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$   | $\bar{\mathbf{y}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$              | $\mathbf{A}_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}$                     | $\mathbf{A}_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}$                    |
|   |  | $r_1 = 0.75$  | $r_2 = 0.89$  |
| $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$          |  |   |   |
| $\hat{\boldsymbol{\mu}} = \begin{pmatrix} 163.48 \\ 445.72 \end{pmatrix}$                                     |  | $\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 12114.41 & 24787.58 \\ 24787.58 & 72447.12 \end{pmatrix}$    |   |
| $W=27.43, \nu = 5, p\text{-value} < 0.0001$   |  | $\hat{\rho}_1 = \hat{\rho}_2 = 0.8367$  |   |
| $H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$  |  |   |   |
| $\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$                                     | $\hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$        | $\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 16066.45 & 33169.95 \\ 33169.95 & 96892.87 \end{pmatrix}$  | $\hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7501.54 & 15487.27 \\ 15487.27 & 45239.94 \end{pmatrix}$ |
| $W=4.01, \nu = 2, p\text{-value} = 0.1347$  |  | $\hat{\rho}_1 = \hat{\rho}_2 = 0.8407$  |   |
| $H_0 : \rho_1 = \rho_2$   |  |   |   |
| $\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$                                     | $\hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$        | $\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 17118.26 & 33130.55 \\ 33130.55 & 90520.39 \end{pmatrix}$  | $\hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7230.82 & 15495.79 \\ 15495.79 & 46880.11 \end{pmatrix}$ |
| $W=4.08, \nu = 1, p\text{-value} = 0.0434$  |  | $\hat{\rho}_1 = \hat{\rho}_2 = 0.8416$  |   |
| $H_0 : \boldsymbol{\mu}_2 = \sqrt{c}\boldsymbol{\mu}_1, H_0 : \boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$ |  |   |   |
| $\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} 199.9 \\ 545.174 \end{pmatrix}$                                   | $\hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} 132.44 \\ 361.194 \end{pmatrix}$ | $\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 16718.95 & 34657.21 \\ 34657.21 & 101122.67 \end{pmatrix}$ | $\hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7338.72 & 15212.65 \\ 15212.65 & 44387.42 \end{pmatrix}$ |
| $W=4.19, \nu = 4, p\text{-value} = 0.3811$  |  | $\hat{c} = 0.4389, \hat{\rho}_1 = \hat{\rho}_2 = 0.8429$  |   |

## 5.4 More Than Two Relatively Small Multivariate Normal Samples

To further illustrate the flexibility of the procedure, we consider the real data example discussed by Grizzle & Allen (1969), Wang (1998), Zhang & Xu (2009). This longitudinal data set consists of coronary sinus potassium concentrations measured on each of 36 dogs. The seven measurements on each dog were taken every 2 minutes from 1 to 13 minutes after occlusion. These 36 dogs were divided into four treatment groups with  $n_1 = 9$ ,  $n_2 = 10$ ,  $n_3 = 8$  and  $n_4 = 9$  dogs respectively. The sample mean vectors  $\bar{\mathbf{y}}_i$ , sample covariance matrices  $\mathbf{A}_i$ , and sample correlation matrices  $\mathbf{R}_i$ ,  $i = 1, 2, 3, 4$  are given in Table 5.3. Sample mean vectors are represented graphically in Figure 5.1.

Figure 5.1: Sample means.



To make provision for the dependence of observations on the same individual, Grizzle & Allen (1969) analyzed this data set by incorporating into the linear model a set of covariates which are a basis of the set of linear functions of the vectors which span the within individual error space. They assumed a common covariance matrix across individuals. Wang (1998) presented a nested nonparametric mixed effects smoothing spline analysis of variance of the data. Zhang & Xu (2009) transformed this four-sample Behrens-Fisher problem into a one-sample problem by extending Scheffe's transformation

method. The resulting sample consisted of eight observations with dimension 21, to which they then applied an  $L^2$  norm based test.

In Sections 5.4.1, 5.4.2, 5.4.3 and 5.4.4 the method of maximum likelihood under constraints is used to estimate different parameter structures directly. For the sake of illustration only a few of several models considered, are presented in the following sections.

Table 5.3: Sample mean vectors, covariance and correlation matrices.

| $\bar{y}_i$   | $A_i$   | $R_i$  |
|---|---|--|
| $\begin{pmatrix} 4.111 \\ 4.178 \\ 4.400 \\ 4.656 \\ 5.067 \\ 5.222 \\ 4.722 \end{pmatrix}$ | $\begin{pmatrix} 0.20 & 0.12 & 0.01 & -0.01 & 0.05 & 0.22 & 0.21 \\ & 0.08 & 0.01 & -0.00 & 0.05 & 0.14 & 0.16 \\ & & 0.22 & 0.24 & 0.17 & 0.07 & 0.07 \\ & & & 0.41 & 0.36 & 0.20 & 0.13 \\ & & & & 0.42 & 0.29 & 0.27 \\ & & & & & 0.51 & 0.47 \\ & & & & & & 0.57 \end{pmatrix}$ | $\begin{pmatrix} 1 & 0.94 & 0.06 & -0.03 & 0.17 & 0.68 & 0.64 \\ & 1 & 0.05 & -0.02 & 0.26 & 0.71 & 0.72 \\ & & 1 & 0.79 & 0.57 & 0.20 & 0.19 \\ & & & 1 & 0.87 & 0.43 & 0.26 \\ & & & & 1 & 0.62 & 0.56 \\ & & & & & 1 & 0.87 \\ & & & & & & 1 \end{pmatrix}$ |
| $\begin{pmatrix} 3.540 \\ 3.630 \\ 3.620 \\ 3.460 \\ 3.660 \\ 3.500 \\ 3.460 \end{pmatrix}$ | $\begin{pmatrix} 0.29 & 0.22 & 0.26 & 0.26 & 0.33 & 0.24 & 0.26 \\ & 0.19 & 0.20 & 0.20 & 0.25 & 0.18 & 0.21 \\ & & 0.24 & 0.22 & 0.27 & 0.21 & 0.23 \\ & & & 0.33 & 0.36 & 0.23 & 0.27 \\ & & & & 0.46 & 0.27 & 0.32 \\ & & & & & 0.23 & 0.23 \\ & & & & & & 0.26 \end{pmatrix}$   | $\begin{pmatrix} 1 & 0.93 & 0.97 & 0.85 & 0.90 & 0.95 & 0.96 \\ & 1 & 0.94 & 0.81 & 0.83 & 0.87 & 0.92 \\ & & 1 & 0.79 & 0.82 & 0.90 & 0.93 \\ & & & 1 & 0.91 & 0.84 & 0.93 \\ & & & & 1 & 0.85 & 0.92 \\ & & & & & 1 & 0.96 \\ & & & & & & 1 \end{pmatrix}$   |
| $\begin{pmatrix} 3.600 \\ 3.725 \\ 4.200 \\ 4.438 \\ 4.500 \\ 4.325 \\ 4.225 \end{pmatrix}$ | $\begin{pmatrix} 0.18 & 0.16 & 0.23 & 0.29 & 0.19 & 0.07 & 0.08 \\ & 0.17 & 0.29 & 0.28 & 0.17 & 0.07 & 0.14 \\ & & 0.79 & 0.63 & 0.34 & 0.31 & 0.48 \\ & & & 0.71 & 0.48 & 0.36 & 0.35 \\ & & & & 0.46 & 0.36 & 0.29 \\ & & & & & 0.51 & 0.49 \\ & & & & & & 0.66 \end{pmatrix}$   | $\begin{pmatrix} 1 & 0.92 & 0.60 & 0.82 & 0.67 & 0.23 & 0.22 \\ & 1 & 0.79 & 0.82 & 0.62 & 0.25 & 0.42 \\ & & 1 & 0.84 & 0.57 & 0.49 & 0.66 \\ & & & 1 & 0.84 & 0.59 & 0.52 \\ & & & & 1 & 0.75 & 0.52 \\ & & & & & 1 & 0.84 \\ & & & & & & 1 \end{pmatrix}$   |
| $\begin{pmatrix} 3.645 \\ 3.778 \\ 4.011 \\ 4.067 \\ 3.978 \\ 4.056 \\ 4.044 \end{pmatrix}$ | $\begin{pmatrix} 0.13 & 0.11 & 0.10 & 0.15 & 0.13 & 0.15 & 0.09 \\ & 0.16 & 0.15 & 0.17 & 0.19 & 0.19 & 0.09 \\ & & 0.26 & 0.24 & 0.17 & 0.11 & 0.00 \\ & & & 0.30 & 0.28 & 0.28 & 0.15 \\ & & & & 0.47 & 0.51 & 0.36 \\ & & & & & 0.65 & 0.48 \\ & & & & & & 0.41 \end{pmatrix}$   | $\begin{pmatrix} 1 & 0.75 & 0.56 & 0.77 & 0.52 & 0.52 & 0.38 \\ & 1 & 0.74 & 0.77 & 0.69 & 0.60 & 0.36 \\ & & 1 & 0.85 & 0.48 & 0.27 & 0.00 \\ & & & 1 & 0.75 & 0.63 & 0.44 \\ & & & & 1 & 0.93 & 0.82 \\ & & & & & 1 & 0.94 \\ & & & & & & 1 \end{pmatrix}$   |

### 5.4.1 Proportional covariance matrices

Proportional covariance matrices, i.e.  $\Sigma_i = c_i \Sigma_1$ ,  $i = 2, 3, 4$  where  $\Sigma_i = \mathbf{m}_{i2} - \boldsymbol{\mu}_i \boldsymbol{\mu}_i'$ ,  $i = 1, 2, 3, 4$  require the constraints:

$$\mathbf{g}_1(\mathbf{m}) = \begin{pmatrix} \text{vec}(\Sigma_2 - c_2 \Sigma_1) \\ \text{vec}(\Sigma_3 - c_3 \Sigma_1) \\ \text{vec}(\Sigma_4 - c_4 \Sigma_1) \end{pmatrix} \text{ where } c_i = \frac{\sigma_{11}^i}{\sigma_{11}^1} \quad (5.18)$$

and  $\sigma_{11}^i$  the variance of the first measurement in the  $i$ -th sample. Then

$$\mathbf{G}_{1m} = \begin{pmatrix} \frac{\partial \text{vec}(\Sigma_2)}{\partial \mathbf{m}} - c_2 \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \text{vec}(\Sigma_1) \frac{\partial c_2}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\Sigma_3)}{\partial \mathbf{m}} - c_3 \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \text{vec}(\Sigma_1) \frac{\partial c_3}{\partial \mathbf{m}} \\ \frac{\partial \text{vec}(\Sigma_4)}{\partial \mathbf{m}} - c_4 \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \text{vec}(\Sigma_1) \frac{\partial c_4}{\partial \mathbf{m}} \end{pmatrix} \quad (5.19)$$

where  $\frac{\partial c_i}{\partial \mathbf{m}} = \frac{\partial c_i}{\partial \mathbf{m}_\Sigma} \times \frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}}$ ,  $\frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}}$  and  $\frac{\partial \text{vec}(\Sigma_i)}{\partial \mathbf{m}}$  are given in (5.9),

$$\begin{pmatrix} \frac{\partial c_2}{\partial \mathbf{m}_\Sigma} \\ \frac{\partial c_3}{\partial \mathbf{m}_\Sigma} \\ \frac{\partial c_4}{\partial \mathbf{m}_\Sigma} \end{pmatrix} = \begin{pmatrix} \frac{\sigma_{11}^2}{(\sigma_{11}^1)^2} \mathbf{I}_{p^2}[1,] & \frac{1}{\sigma_{11}^1} \mathbf{I}_{p^2}[1,] & \mathbf{0}_{1 \times p^2} & \mathbf{0}_{1 \times p^2} \\ \frac{\sigma_{11}^3}{(\sigma_{11}^1)^2} \mathbf{I}_{p^2}[1,] & \mathbf{0}_{1 \times p^2} & \frac{1}{\sigma_{11}^1} \mathbf{I}_{p^2}[1,] & \mathbf{0}_{1 \times p^2} \\ \frac{\sigma_{11}^4}{(\sigma_{11}^1)^2} \mathbf{I}_{p^2}[1,] & \mathbf{0}_{1 \times p^2} & \mathbf{0}_{1 \times p^2} & \frac{1}{\sigma_{11}^1} \mathbf{I}_{p^2}[1,] \end{pmatrix} \quad (5.20)$$

and  $\mathbf{I}_{p^2}[1,]$  denotes the first row of  $\mathbf{I}_{p^2}$ .

In Table 5.4, the results for the hypothesis of proportional covariance matrices are given, with no constraints imposed on the mean vectors. Note that proportionality of covariance matrices implies equal correlation matrices. One concludes that the proportional model holds for these data.

SAS program: SMVNORMAL1.SAS

Table 5.4: Proportional covariance matrices.

| $\mu_i$ unrestricted   |  |
|--|--|
| $H_0 : \Sigma_2 = c_2 \Sigma_1, \Sigma_3 = c_3 \Sigma_1, \Sigma_4 = c_4 \Sigma_1$  |  |
| $\hat{\Sigma}_1 = \begin{pmatrix} 0.25 & 0.19 & 0.19 & 0.21 & 0.23 & 0.22 & 0.21 \\ & 0.18 & 0.19 & 0.19 & 0.21 & 0.19 & 0.18 \\ & & 0.37 & 0.34 & 0.27 & 0.20 & 0.20 \\ & & & 0.46 & 0.41 & 0.29 & 0.26 \\ & & & & 0.53 & 0.41 & 0.36 \\ & & & & & 0.50 & 0.44 \\ & & & & & & 0.49 \end{pmatrix}$ | $\hat{P}_i = \begin{pmatrix} 1 & 0.88 & 0.62 & 0.62 & 0.64 & 0.62 & 0.60 \\ & 1 & 0.72 & 0.65 & 0.67 & 0.62 & 0.61 \\ & & 1 & 0.81 & 0.62 & 0.46 & 0.48 \\ & & & 1 & 0.84 & 0.61 & 0.54 \\ & & & & 1 & 0.79 & 0.72 \\ & & & & & 1 & 0.89 \\ & & & & & & 1 \end{pmatrix}$ |
| $\hat{c}_2 = 0.61, \hat{c}_3 = 1.29, \hat{c}_4 = 0.87$   | $i = 1, 2, 3, 4$   |
| $W=49.965, \nu = 91, p\text{-value} = 0.9999$  |  |

### 5.4.2 Behrens-Fisher with proportional covariance matrices

To simultaneously estimate the common mean vector and test  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ , the constraints implied and matrix of derivatives are:

$$g_2(\mathbf{m}) = \begin{pmatrix} \mu_2 - \mu \\ \mu_3 - \mu \\ \mu_4 - \mu \end{pmatrix} \quad \mathbf{G}_{2m} = \mathbf{G}_{2t} = \frac{\partial g_2(\mathbf{m})}{\partial \mathbf{m}} = \begin{pmatrix} \frac{\partial \mu_2}{\partial \mathbf{m}} \\ \frac{\partial \mu_3}{\partial \mathbf{m}} \\ \frac{\partial \mu_4}{\partial \mathbf{m}} \end{pmatrix} - \begin{pmatrix} \frac{\partial \mu}{\partial \mathbf{m}} \\ \frac{\partial \mu}{\partial \mathbf{m}} \\ \frac{\partial \mu}{\partial \mathbf{m}} \end{pmatrix} \quad (5.21)$$

where  $\frac{\partial \mu_i}{\partial \mathbf{m}}$  and  $\frac{\partial \mu}{\partial \mathbf{m}}$  are given in (5.8).

The constraints  $g_1(\mathbf{m})$  and  $g_2(\mathbf{m})$  can be imposed simultaneously:

$$g(\mathbf{m}) = \begin{pmatrix} g_1(\mathbf{m}) \\ g_2(\mathbf{m}) \end{pmatrix} \quad \text{and} \quad \mathbf{G}_m = \begin{pmatrix} \mathbf{G}_{1m} \\ \mathbf{G}_{2m} \end{pmatrix}. \quad (5.22)$$

In Table 5.5 Behrens-Fisher is illustrated under the assumption of proportional covariance matrices. The model does not fit well. For this example, for the sake of interest, the standard errors of the estimates,  $\hat{\sigma}(f(\hat{\mathbf{m}}))$  are given in Table 5.6.

SAS program: SMVNORMAL1.SAS

Table 5.5: Behrens-Fisher and proportional covariance matrices.

|   |  |
|---|--|
| $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$   |  |
| $H_0 : \Sigma_2 = c_2 \Sigma_1, \Sigma_3 = c_3 \Sigma_1, \Sigma_4 = c_4 \Sigma_1$   |  |
| $\hat{\boldsymbol{\mu}}' = (3.691, 3.794, 3.975, 4.02, 4.16, 4.125, 3.993)$   |  |
| $\hat{\Sigma}_1 =$  | $\hat{P}_i =$  |
| $\begin{pmatrix} 0.32 & 0.26 & 0.27 & 0.33 & 0.38 & 0.40 & 0.33 \\ & 0.25 & 0.27 & 0.31 & 0.35 & 0.35 & 0.30 \\ & & 0.53 & 0.55 & 0.49 & 0.45 & 0.40 \\ & & & 0.79 & 0.75 & 0.68 & 0.57 \\ & & & & 0.93 & 0.86 & 0.70 \\ & & & & & 1.05 & 0.86 \\ & & & & & & 0.82 \end{pmatrix}$ | $\begin{pmatrix} 1 & 0.90 & 0.65 & 0.65 & 0.69 & 0.67 & 0.64 \\ & 1 & 0.75 & 0.69 & 0.72 & 0.69 & 0.67 \\ & & 1 & 0.85 & 0.69 & 0.59 & 0.60 \\ & & & 1 & 0.88 & 0.74 & 0.70 \\ & & & & 1 & 0.87 & 0.80 \\ & & & & & 1 & 0.92 \\ & & & & & & 1 \end{pmatrix}$ |
| $\hat{c}_2 = 0.60, \hat{c}_3 = 1.10, \hat{c}_4 = 0.71$  | $i = 1, 2, 3, 4$   |
| $W=274.178, \nu = 108, p\text{-value}= 0$   |  |



Table 5.6: Asymptotic standard errors of the estimated parameters.

| $\hat{\sigma}(\hat{\boldsymbol{\mu}}_i)$  | $\hat{\sigma}(\hat{\boldsymbol{\Sigma}}_i)$  |
|---|--|
| $\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$ | $\begin{pmatrix} 0.626 & 0.581 & 0.672 & 0.758 & 0.818 & 0.844 & 0.763 \\ & 0.566 & 0.665 & 0.743 & 0.802 & 0.826 & 0.745 \\ & & 0.872 & 0.935 & 0.952 & 0.954 & 0.875 \\ & & & 1.078 & 1.108 & 1.098 & 0.995 \\ & & & & 1.218 & 1.21 & 1.092 \\ & & & & & 1.289 & 1.163 \\ & & & & & & 1.092 \end{pmatrix}$ |
| $\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$ | $\begin{pmatrix} 0.622 & 0.578 & 0.668 & 0.753 & 0.812 & 0.837 & 0.757 \\ & 0.563 & 0.661 & 0.738 & 0.797 & 0.82 & 0.74 \\ & & 0.863 & 0.926 & 0.943 & 0.945 & 0.868 \\ & & & 1.063 & 1.094 & 1.084 & 0.983 \\ & & & & 1.2 & 1.192 & 1.078 \\ & & & & & 1.267 & 1.145 \\ & & & & & & 1.076 \end{pmatrix}$    |
| $\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$ | $\begin{pmatrix} 0.628 & 0.583 & 0.674 & 0.76 & 0.82 & 0.846 & 0.765 \\ & 0.567 & 0.666 & 0.744 & 0.804 & 0.828 & 0.747 \\ & & 0.875 & 0.939 & 0.955 & 0.957 & 0.878 \\ & & & 1.084 & 1.114 & 1.103 & 0.999 \\ & & & & 1.226 & 1.217 & 1.097 \\ & & & & & 1.298 & 1.17 \\ & & & & & & 1.099 \end{pmatrix}$   |
| $\begin{pmatrix} 0.084 \\ 0.074 \\ 0.108 \\ 0.131 \\ 0.143 \\ 0.152 \\ 0.134 \end{pmatrix}$ | $\begin{pmatrix} 0.623 & 0.579 & 0.669 & 0.754 & 0.813 & 0.839 & 0.759 \\ & 0.563 & 0.662 & 0.739 & 0.798 & 0.822 & 0.742 \\ & & 0.866 & 0.928 & 0.945 & 0.947 & 0.869 \\ & & & 1.067 & 1.097 & 1.088 & 0.986 \\ & & & & 1.204 & 1.197 & 1.081 \\ & & & & & 1.272 & 1.15 \\ & & & & & & 1.08 \end{pmatrix}$  |

### 5.4.3 Proportional mean vectors and proportional covariance matrices

In Figure 5.1, the group means are given. No general trend is observed but the mean vectors may also be modelled as being proportional to an average mean vector,  $\boldsymbol{\mu}$ , but differing with a constant value, i.e.  $\boldsymbol{\mu}_i = d_i \boldsymbol{\mu} + a_i \mathbf{1}$ ,  $i = 1, 2, 3, 4$  where  $a_i = \mu_{i1} - d_i \mu_1$ ,  $d_i = \frac{\mu_{i1} - \mu_{i2}}{\mu_1 - \mu_2}$ ,  $\mu_{ij}$  the  $j$ -th element of the  $i$ -th mean vector  $\boldsymbol{\mu}_i$  and  $\mu_j$  the  $j$ -th element of  $\boldsymbol{\mu}$ .

This is formulated as  $g_1(\mathbf{m}) = \begin{pmatrix} \mu_2 - d_2\mu - a_2\mathbf{1} \\ \mu_3 - d_3\mu - a_3\mathbf{1} \\ \mu_4 - d_4\mu - a_4\mathbf{1} \end{pmatrix}$  with matrix of derivatives

$$\mathbf{G}_{1\mathbf{m}} = \begin{pmatrix} \frac{\partial \mu_2}{\partial \mathbf{m}} - d_2 \frac{\partial \mu}{\partial \mathbf{m}} - \mu \frac{\partial d_2}{\partial \mathbf{m}} - \mathbf{1} \frac{\partial a_2}{\partial \mathbf{m}} \\ \frac{\partial \mu_3}{\partial \mathbf{m}} - d_3 \frac{\partial \mu}{\partial \mathbf{m}} - \mu \frac{\partial d_3}{\partial \mathbf{m}} - \mathbf{1} \frac{\partial a_3}{\partial \mathbf{m}} \\ \frac{\partial \mu_4}{\partial \mathbf{m}} - d_4 \frac{\partial \mu}{\partial \mathbf{m}} - \mu \frac{\partial d_4}{\partial \mathbf{m}} - \mathbf{1} \frac{\partial a_4}{\partial \mathbf{m}} \end{pmatrix} \quad (5.23)$$

where

$$\begin{pmatrix} \frac{\partial d_2}{\partial \mathbf{m}} \\ \frac{\partial d_3}{\partial \mathbf{m}} \\ \frac{\partial d_4}{\partial \mathbf{m}} \end{pmatrix} = \begin{pmatrix} \frac{1}{(\mu_1 - \mu_2)} \frac{\partial}{\partial \mathbf{m}} (\mu_{21} - \mu_{22}) - \frac{(\mu_{21} - \mu_{22})}{(\mu_1 - \mu_2)^2} \frac{\partial}{\partial \mathbf{m}} (\mu_1 - \mu_2) \\ \frac{1}{(\mu_1 - \mu_2)} \frac{\partial}{\partial \mathbf{m}} (\mu_{31} - \mu_{32}) - \frac{(\mu_{31} - \mu_{32})}{(\mu_1 - \mu_2)^2} \frac{\partial}{\partial \mathbf{m}} (\mu_1 - \mu_2) \\ \frac{1}{(\mu_1 - \mu_2)} \frac{\partial}{\partial \mathbf{m}} (\mu_{41} - \mu_{42}) - \frac{(\mu_{41} - \mu_{42})}{(\mu_1 - \mu_2)^2} \frac{\partial}{\partial \mathbf{m}} (\mu_1 - \mu_2) \end{pmatrix} \quad (5.24)$$

and

$$\begin{pmatrix} \frac{\partial a_2}{\partial \mathbf{m}} \\ \frac{\partial a_3}{\partial \mathbf{m}} \\ \frac{\partial a_4}{\partial \mathbf{m}} \end{pmatrix} = \begin{pmatrix} \mathbf{1} \frac{\partial \mu_{21}}{\partial \mathbf{m}} - d_2 \frac{\partial \mu_1}{\partial \mathbf{m}} - \mu_1 \frac{\partial d_2}{\partial \mathbf{m}} \\ \mathbf{1} \frac{\partial \mu_{31}}{\partial \mathbf{m}} - d_3 \frac{\partial \mu_1}{\partial \mathbf{m}} - \mu_1 \frac{\partial d_3}{\partial \mathbf{m}} \\ \mathbf{1} \frac{\partial \mu_{41}}{\partial \mathbf{m}} - d_4 \frac{\partial \mu_1}{\partial \mathbf{m}} - \mu_1 \frac{\partial d_4}{\partial \mathbf{m}} \end{pmatrix} \quad (5.25)$$

with  $\frac{\partial \mu_{ij}}{\partial \mathbf{m}} = \frac{\partial \mu_i}{\partial \mathbf{m}}[j, \cdot]$  the  $j$ -th row of  $\frac{\partial \mu_i}{\partial \mathbf{m}}$  (cf. (5.8)),  $\frac{\partial \mu_j}{\partial \mathbf{m}} = \frac{\partial \mu}{\partial \mathbf{m}}[j, \cdot]$  the  $j$ -th row of  $\frac{\partial \mu}{\partial \mathbf{m}}$  (cf. (5.8)).

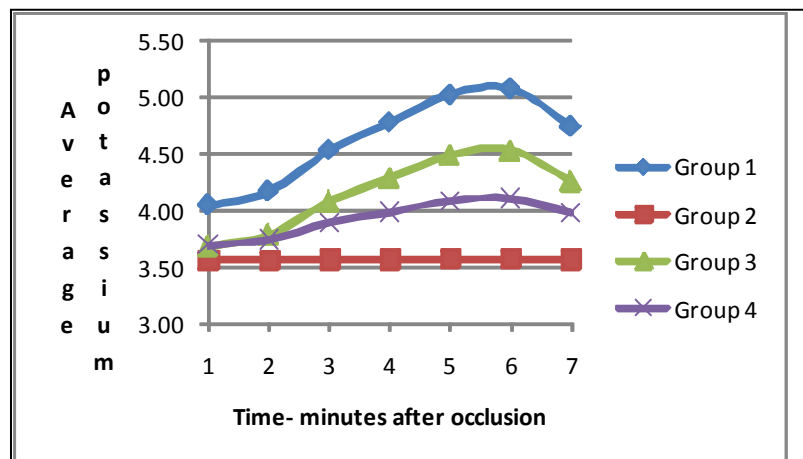
Results for the hypotheses of proportional mean vectors and proportional covariance matrices (cf. (5.18)) are given in Table 5.7. The smoothing effect of this model, which fits quite well, is illustrated in Figure 5.2. This solution seems to provide the best analysis of the data in the sense that it follows the trends observed in the data very well and provides the mle's which are obtained by direct modelling of the mean vectors and covariance matrices.

SAS program: SMVNORMAL1.SAS

Table 5.7: Proportional mean vectors and proportional covariance matrices.

|  |   |
|--|---|
| $H_0 : \mu_2 = d_2\mu + a_2\mathbf{1}, \mu_3 = d_3\mu + a_3\mathbf{1}, \mu_4 = d_4\mu + a_4\mathbf{1}$   |   |
| $H_0 : \Sigma_2 = c_2\Sigma_1, \Sigma_3 = c_3\Sigma_1, \Sigma_4 = c_4\Sigma_1$   |   |
| $\hat{\mu}' = (3.765, 3.829, 4.012, 4.142, 4.304, 4.333, 4.13)$  |   |
| $\hat{d}_2 = 0.186, \hat{d}_3 = 1.294, \hat{d}_4 = 0.562, \hat{a}_2 = 2.875, \hat{a}_3 = -1.218, \hat{a}_4 = 1.673$  |   |
| $\hat{\Sigma}_1 = \begin{pmatrix} 0.24 & 0.17 & 0.16 & 0.19 & 0.21 & 0.21 & 0.19 \\ & 0.17 & 0.17 & 0.17 & 0.19 & 0.17 & 0.16 \\ & & 0.36 & 0.33 & 0.26 & 0.18 & 0.19 \\ & & & 0.46 & 0.39 & 0.28 & 0.26 \\ & & & & 0.50 & 0.38 & 0.34 \\ & & & & & 0.49 & 0.43 \\ & & & & & & 0.48 \end{pmatrix}$ | $\hat{P}_i = \begin{pmatrix} 1.00 & 0.87 & 0.56 & 0.56 & 0.62 & 0.61 & 0.57 \\ & 1.00 & 0.68 & 0.59 & 0.66 & 0.59 & 0.58 \\ & & 1.00 & 0.80 & 0.60 & 0.42 & 0.47 \\ & & & 1.00 & 0.82 & 0.59 & 0.55 \\ & & & & 1.00 & 0.77 & 0.70 \\ & & & & & 1.00 & 0.89 \\ & & & & & & 1.00 \end{pmatrix}$ |
| $\hat{c}_2 = 0.70, \hat{c}_3 = 1.35, \hat{c}_4 = 0.94, \hat{e}_i = 0.71$   |   |
| $i = 1, 2, 3, 4$   |   |
| $W=95.961, \nu = 105, p\text{-value}= 0.7245$  |   |

Figure 5.2: Proportional means assuming proportional covariance matrices.

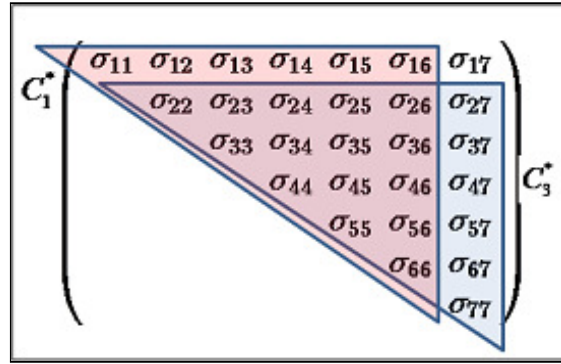


#### 5.4.4 Proportional mean vectors and proportional covariance matrices with Toeplitz correlation structure

As a further illustration of the flexibility of the procedure, proportionality of elements along the diagonals of the covariance matrix within each population is imposed additionally. Let the matrices  $C_1$  and  $C_3$  select from  $vec(\Sigma_i)$  the elements of  $\Sigma_i$  as indicated by the upper and lower triangles ( $C_1^*$  and  $C_3^*$  respectively) in Figure 5.3 and impose the additional constraints:

$$g_3(\mathbf{m}) = \begin{pmatrix} C_3 vec(\Sigma_1) - e_1 C_1 vec(\Sigma_1) \\ C_3 vec(\Sigma_2) - e_2 C_1 vec(\Sigma_2) \\ C_3 vec(\Sigma_3) - e_3 C_1 vec(\Sigma_3) \\ C_3 vec(\Sigma_4) - e_4 C_1 vec(\Sigma_4) \end{pmatrix} \text{ with } e_i = \frac{\sigma_{22}^i}{\sigma_{11}^i}, i = 1, 2, 3, 4. \quad (5.26)$$

Figure 5.3: Proportional submatrices.



The required matrix of derivatives is

$$G_{3m} = \frac{\partial g_3(\mathbf{m})}{\partial \mathbf{m}_\Sigma} \times \frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}} \text{ where } \frac{\partial g_3(\mathbf{m})}{\partial \mathbf{m}_\Sigma} = \begin{pmatrix} C_3 - e_1 C_1 - C_1 vec(\Sigma_1) \frac{\partial e_1}{\partial \mathbf{m}_\Sigma} \\ C_3 - e_2 C_1 - C_1 vec(\Sigma_2) \frac{\partial e_2}{\partial \mathbf{m}_\Sigma} \\ C_3 - e_3 C_1 - C_1 vec(\Sigma_3) \frac{\partial e_3}{\partial \mathbf{m}_\Sigma} \\ C_3 - e_4 C_1 - C_1 vec(\Sigma_4) \frac{\partial e_4}{\partial \mathbf{m}_\Sigma} \end{pmatrix} \quad (5.27)$$

with  $\frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}}$  given in (5.9) and

$$\begin{pmatrix} \frac{\partial e_1}{\partial \mathbf{m}_\Sigma} \\ \frac{\partial e_2}{\partial \mathbf{m}_\Sigma} \\ \frac{\partial e_3}{\partial \mathbf{m}_\Sigma} \\ \frac{\partial e_4}{\partial \mathbf{m}_\Sigma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_{11}^1} \mathbf{I}_{p^2}[p+2,] - \frac{\sigma_{22}^1}{(\sigma_{11}^1)^2} \mathbf{I}_{p^2}[1,] \\ \frac{1}{\sigma_{11}^2} \mathbf{I}_{p^2}[p+2,] - \frac{\sigma_{22}^2}{(\sigma_{11}^2)^2} \mathbf{I}_{p^2}[1,] \\ \frac{1}{\sigma_{11}^3} \mathbf{I}_{p^2}[p+2,] - \frac{\sigma_{22}^3}{(\sigma_{11}^3)^2} \mathbf{I}_{p^2}[1,] \\ \frac{1}{\sigma_{11}^4} \mathbf{I}_{p^2}[p+2,] - \frac{\sigma_{22}^4}{(\sigma_{11}^4)^2} \mathbf{I}_{p^2}[1,] \end{pmatrix}. \quad (5.28)$$

Note that  $\sigma_{22}^i$  is the element in row  $p+2$  of  $\text{vec}(\Sigma_i)$ . This yields a Toeplitz correlation structure common to all four populations, i.e.

$$\mathbf{P}_i = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 & \rho_6 \\ & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 \\ & & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ & & & 1 & \rho_1 & \rho_2 & \rho_3 \\ & & & & 1 & \rho_1 & \rho_2 \\ & & & & & 1 & \rho_1 \\ & & & & & & 1 \end{pmatrix}. \quad (5.29)$$

Proportionality of these submatrices has the implication that proportionality of the elements on the main diagonal and the first row of each population covariance matrix is now sufficient to specify the full set of required constraints. The constraint  $\mathbf{g}_1(\mathbf{m})$  is adjusted by pre-multiplying each  $\text{vec}(\Sigma_i)$  with a matrix  $\mathbf{B}$  which selects only the main diagonal elements  $\sigma_{jj}$  and elements in the first row  $\sigma_{1j}$ ,  $j = 1, 2, \dots, p$  from the respective covariance matrices.

Then

$$\mathbf{g}_1(\mathbf{m}) = \begin{pmatrix} \mathbf{B} \text{vec}(\Sigma_2) - c_2 \mathbf{B} \text{vec}(\Sigma_1) \\ \mathbf{B} \text{vec}(\Sigma_3) - c_3 \mathbf{B} \text{vec}(\Sigma_1) \\ \mathbf{B} \text{vec}(\Sigma_4) - c_4 \mathbf{B} \text{vec}(\Sigma_1) \end{pmatrix}, \quad (5.30)$$

$$\mathbf{G}_{1m} = \begin{pmatrix} \mathbf{B} \frac{\partial \text{vec}(\Sigma_2)}{\partial \mathbf{m}} - c_2 \mathbf{B} \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \mathbf{B} \text{vec}(\Sigma_1) \frac{\partial c_2}{\partial \mathbf{m}} \\ \mathbf{B} \frac{\partial \text{vec}(\Sigma_3)}{\partial \mathbf{m}} - c_3 \mathbf{B} \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \mathbf{B} \text{vec}(\Sigma_1) \frac{\partial c_3}{\partial \mathbf{m}} \\ \mathbf{B} \frac{\partial \text{vec}(\Sigma_4)}{\partial \mathbf{m}} - c_4 \mathbf{B} \frac{\partial \text{vec}(\Sigma_1)}{\partial \mathbf{m}} - \mathbf{B} \text{vec}(\Sigma_1) \frac{\partial c_4}{\partial \mathbf{m}} \end{pmatrix}$$

where  $\frac{\partial \text{vec}(\Sigma_i)}{\partial \mathbf{m}}$  and  $\frac{\partial c_i}{\partial \mathbf{m}} = \frac{\partial c_i}{\partial \mathbf{m}_\Sigma} \times \frac{\partial \mathbf{m}_\Sigma}{\partial \mathbf{m}}$  are given in (5.9) and (5.20) respectively.

The constraints  $\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}) \\ \mathbf{g}_2(\mathbf{m}) \\ \mathbf{g}_3(\mathbf{m}) \end{pmatrix}$  with  $\mathbf{G}_m = \begin{pmatrix} \mathbf{G}_{1m} \\ \mathbf{G}_{2m} \\ \mathbf{G}_{3m} \end{pmatrix}$  are imposed simultaneously.

Table 5.8: Proportional mean vectors &amp; covariance matrices with Toeplitz correlation structure.

|  |  |
|--|--|
| $H_0 : \boldsymbol{\mu}_2 = d_2\boldsymbol{\mu} + a_2\mathbf{1}, \boldsymbol{\mu}_3 = d_3\boldsymbol{\mu} + a_3\mathbf{1}, \boldsymbol{\mu}_4 = d_4\boldsymbol{\mu} + a_4\mathbf{1}$<br>$H_0 : \boldsymbol{\Sigma}_2 = c_2\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_3 = c_3\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_4 = c_4\boldsymbol{\Sigma}_1$<br>$H_0 : \mathbf{C}_3^{i*} = e_i\mathbf{C}_1^{i*}, i = 1, 2, 3, 4$ |  |
| $\hat{\boldsymbol{\mu}}' = (3.744, 3.813, 4.0168, 4.155, 4.291, 4.322, 4.134)$<br>$\hat{d}_2 = 0.0237, \hat{d}_3 = 1.471, \hat{d}_4 = 0.709, \hat{a}_2 = 3.473, \hat{a}_3 = -1.828, \hat{a}_4 = 1.039$   |  |
| $\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.28 & 0.24 & 0.20 & 0.19 & 0.22 & 0.25 & 0.23 \\ & 0.32 & 0.27 & 0.23 & 0.22 & 0.25 & 0.28 \\ & & 0.36 & 0.31 & 0.26 & 0.25 & 0.28 \\ & & & 0.41 & 0.35 & 0.29 & 0.29 \\ & & & & 0.47 & 0.40 & 0.33 \\ & & & & & 0.53 & 0.45 \\ & & & & & & 0.61 \end{pmatrix}$  | $\hat{\mathbf{P}}_i = \begin{pmatrix} 1.00 & 0.80 & 0.63 & 0.57 & 0.60 & 0.64 & 0.57 \\ & 1.00 & 0.80 & 0.63 & 0.57 & 0.60 & 0.64 \\ & & 1.00 & 0.80 & 0.63 & 0.57 & 0.60 \\ & & & 1.00 & 0.80 & 0.63 & 0.57 \\ & & & & 1.00 & 0.80 & 0.63 \\ & & & & & 1.00 & 0.80 \\ & & & & & & 1.00 \end{pmatrix}$ |
| $\hat{c}_2 = 0.58, \hat{c}_3 = 1.29, \hat{c}_4 = 0.78, \hat{e}_i = 1.14$   |  |
| $W=148.441, \nu = 119, p\text{-value} = 0.0349$  |  |

SAS program: SMVNORMAL2.SAS

The model (cf. Table 5.8) does not fit very well as is to be expected on inspection of the sample correlation matrices (cf. Table 5.3). However, it serves to illustrate how more complicated structures can be modelled easily.

# Chapter 6

## MLE for Wishart Matrices

Several methods for MLE of patterned covariance matrices, which are used to model multivariate normal data, are given in the literature. Explicit forms for mle's exist in some cases. Szatrowski (1980) gives necessary and sufficient conditions on linear patterns such that there exists explicit solutions which are obtained in one iteration of the scoring equations from any positive definite starting point. However, patterns in applications often arise where explicit estimates do not exist and an iterative algorithm is required. Such algorithms include the well-known Newton-Raphson algorithm and the method of scoring (Anderson, 1970, 1973) which do not have guaranteed convergence for all patterns. The EM algorithm (Rubin & Szatrowski, 1982) was proposed as an alternative method of estimation in cases where the patterned covariance matrices which do not have explicit mle's can be viewed as submatrices of larger patterned covariance matrices that do have explicit mle's.

To find the mle of a covariance matrix under the constraint that certain covariances are zero, Chaudhuri, Drton & Richardson (2007) proposed an iterative conditional fitting algorithm (with guaranteed convergence properties) under the assumption of multivariate normality. Where the presence of zero covariances can be formulated as a linear hypotheses on the covariance matrix, e.g. covariance graph models (Chaudhuri, Drton & Richardson, 2007), procedures such as Newton-Raphson and the method of scoring (Anderson, 1969,1970,1973) may be used for MLE. However, the algorithm introduced by Chaudhuri, Drton & Richardson (2007) has clearer convergence properties than Anderson's algorithm.

The problem of testing the homogeneity of the covariance matrices of dependent multivariate

normal populations, is another example where no explicit analytical expression of the mle of the covariance matrix (under the null hypothesis) exist. Jiang, Sarkar & Hsuan (1999) proposed a likelihood ratio test and modifications thereof, through an iterative scheme using PROC MIXED in SAS for finding the mle of the covariance matrix - not a straightforward process. An overview of preceding attempts and approaches to variations of this problem is given in the introduction of the paper by Jiang, Sarkar & Hsuan (1999).

In this chapter, the maximum likelihood procedure developed by Matthews & Crowther (1995), is used to estimate the mean vector of the multivariate exponential distribution, subject to constraints on this mean vector. These constraints are implied by the model under consideration and need not be specified in terms of parameters. In particular, the expected value of a Wishart matrix can be estimated enabling applications to patterned covariance matrices where the pattern is not necessarily linear. This procedure lends itself naturally to imposing the constraint that certain covariances are zero. The extension of the procedure for maximum likelihood under constraints to several populations is straightforward. This enables inference regarding covariance matrices of dependent or independent multivariate normals to be accomplished in a straightforward manner. The approach considered in this thesis largely simplifies the philosophy, the modelling and computational aspects involved in the case of MLE of covariance matrices since no derivation of likelihood equations is required. Covariance matrices can be modelled by direct imposition of constraints on specific elements of the covariance matrix.

## 6.1 Estimation for Wishart Matrices under Constraints

Suppose  $\mathbf{y}_1, \dots, \mathbf{y}_N$  is a random sample from a multivariate normal distribution,  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Let  $\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$  denote the sample mean vector and  $\mathbf{A} = \frac{1}{n} \sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$ ,  $n = N - 1$ , the sample covariance matrix. The sample covariance matrix  $\mathbf{A}$ , which is an unbiased estimate of  $\boldsymbol{\Sigma}$  is  $W_p(n, \boldsymbol{\Sigma}/n)$ . The probability density function of  $\mathbf{A}$  is

$$\begin{aligned}
 f(\mathbf{A}) &= \frac{\det(\mathbf{A})^{\frac{n-p-1}{2}} \left(\frac{n}{2}\right)^{\frac{np}{2}}}{\Gamma_p\left(\frac{1}{2}n\right)} \exp\left\{-\frac{n}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{A} - \frac{n}{2} \ln[\det(\boldsymbol{\Sigma})]\right\} \\
 &= \frac{\det(\mathbf{A})^{\frac{n-p-1}{2}} \left(\frac{n}{2}\right)^{\frac{np}{2}}}{\Gamma_p\left(\frac{1}{2}n\right)} \exp\left\{-\frac{n}{2} \left(\frac{N}{n}\right) \text{vec}(\boldsymbol{\Sigma}^{-1})' \text{vec}\left(\frac{n}{N} \mathbf{A}\right) - \frac{n}{2} \ln[\det(\boldsymbol{\Sigma})]\right\} \\
 &= b(\mathbf{t}) \exp[\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})]
 \end{aligned} \tag{6.1}$$



where  $vec(\mathbf{A})$  denotes the stacked columns of the  $p \times p$  matrix  $\mathbf{A}$  and the canonical statistic, corresponding canonical parameter and expected value are respectively given by:

$$\mathbf{t} = \left( vec\left(\frac{n}{N}\mathbf{A}\right) \right), \boldsymbol{\theta} = \left( -\left(\frac{n}{2}\right)\left(\frac{N}{n}\right)vec(\boldsymbol{\Sigma}^{-1}) \right), E(\mathbf{t}) = \mathbf{m} = \left( vec(\boldsymbol{\Sigma}) \right). \quad (6.2)$$

The latter is chosen as the canonical statistic since the mle of  $\mathbf{m}$  without any constraints is  $\widehat{\mathbf{m}} = \mathbf{t}$ .

The covariance matrix of the canonical statistic is given by

$$\mathbf{V} = Cov(\mathbf{t}) = Cov(vec(\mathbf{A})) = (I_{p^2} + \mathbf{K})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})/N \quad (\text{Muirhead, 1982, page 90}). \quad (6.3)$$

A patterned covariance matrix or a covariance matrix with zeros implies specific constraints on the elements of  $\boldsymbol{\Sigma}$ . Consider the general constraint  $\mathbf{g}(\mathbf{m}) = \mathbf{g}[vec(\boldsymbol{\Sigma})]$  with  $\mathbf{G}_m = \frac{\partial \mathbf{g}[vec(\boldsymbol{\Sigma})]}{\partial vec(\boldsymbol{\Sigma})}$ ,  $\mathbf{G}_t = \frac{\partial \mathbf{g}[vec(\boldsymbol{\Sigma})]}{\partial vec(\boldsymbol{\Sigma})}|_{\boldsymbol{\Sigma}=\mathbf{A}}$ . The mle of  $\boldsymbol{\Sigma}$  under the constraint  $\mathbf{g}(vec(\boldsymbol{\Sigma}))$  is obtained iteratively from (2):

$$\widehat{\mathbf{m}} = vec(\mathbf{A}) - (\mathbf{G}_m \mathbf{V})' (\mathbf{G}_t \mathbf{V} \mathbf{G}_m')^* \mathbf{g}[vec(\mathbf{A})]$$

Constraints are imposed only on elements in the upper half or the lower half of the covariance matrix.

Symmetry of the covariance matrix is preserved automatically.

Using the procedure (2.2) for estimating  $\boldsymbol{\Sigma}$  under restrictions, the vector  $\mathbf{t}$  of canonical statistics with corresponding covariance matrix  $\mathbf{V}$  is used as point of departure for any model considered within this framework. The algorithm given in Chapter 2 is then used to obtain the mle's. In each section the model specific expressions for  $\mathbf{g}(\mathbf{m})$  and  $\mathbf{G}_m$  are given.

## 6.2 Covariance Matrices with Given Structure

### 6.2.1 Linearly patterned covariance matrices

Firstly, consider linearly patterned covariance matrices of the general form  $\boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \sum_{g=0}^m \sigma_g \mathbf{C}_g$  (Szatrowski, 1980, Anderson, 1969,1970,1973) where  $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, \dots, \sigma_m)$  are unknown coefficients and  $\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_m$  are known symmetric linearly independent matrices. The linear structure of the covariance matrix implies that

$$\mathbf{m} = vec(\boldsymbol{\Sigma}) = \sum_{g=0}^m \sigma_g vec(\mathbf{C}_g) = (vec\mathbf{C}_0, vec\mathbf{C}_1, \dots, vec\mathbf{C}_m)\boldsymbol{\sigma} = \mathbf{C}\boldsymbol{\sigma}. \quad (6.4)$$

Consequently,

$$\mathbf{g}(\mathbf{m}) = \mathbf{Q}_C \mathbf{C} \boldsymbol{\sigma} = \mathbf{0} \text{ where } \mathbf{Q}_C = \mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^* \mathbf{C}' \quad (6.5)$$

projects orthogonal to the columns of  $\mathbf{C}$ . The matrix of derivatives  $\mathbf{G}_m = \frac{\partial \mathbf{g}(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{Q}_C = \mathbf{G}_t$ .

Structures such as the complete symmetry pattern (with special cases the intraclass correlation pattern and sphericity test), block complete symmetry pattern, compound-, circular- and stationary symmetry pattern can be modelled within this framework.

Consider the  $3 \times 3$  stationary covariance pattern (Rubin & Szatrowski, 1982):

$$\boldsymbol{\Sigma} = \begin{pmatrix} a & b & c \\ b & a & b \\ c & b & a \end{pmatrix}.$$

This may be expressed in linear form (6.4) with corresponding implied constraint (6.5) and matrix of derivatives  $\mathbf{G}_m = \mathbf{C} = \mathbf{G}_t$ . Alternatively, the following three simple constraints can be used to obtain the mle's:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{33} \\ \sigma_{12} - \sigma_{23} \end{pmatrix} = \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma})[1] - \text{vec}(\boldsymbol{\Sigma})[5] \\ \text{vec}(\boldsymbol{\Sigma})[1] - \text{vec}(\boldsymbol{\Sigma})[9] \\ \text{vec}(\boldsymbol{\Sigma})[2] - \text{vec}(\boldsymbol{\Sigma})[6] \end{pmatrix}$$

where  $\text{vec}(\boldsymbol{\Sigma})[i]$  denotes the  $i$ -th element of  $\text{vec}(\boldsymbol{\Sigma})$ . The corresponding matrix of derivatives are given by:

$$\mathbf{G}_m = \mathbf{G}_t = \begin{pmatrix} \mathbf{I}_{p^2}[1, ] - \mathbf{I}_{p^2}[5, ] \\ \mathbf{I}_{p^2}[1, ] - \mathbf{I}_{p^2}[9, ] \\ \mathbf{I}_{p^2}[2, ] - \mathbf{I}_{p^2}[6, ] \end{pmatrix}.$$

Using the procedure (2.2) for MLE under constraints, it is irrelevant whether an explicit mle exist for a specific symmetry structure. The process of estimation follows the same straightforward method for any one of the patterns mentioned.

### 6.2.2 Covariance matrices with zeros

Estimation of the covariance matrix under the constraint that certain covariances are zero, represents another direct application of the procedure (2.2). Consider a covariance matrix which exhibits the following zero pattern:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \sigma_{13} & 0 \\ 0 & \sigma_{22} & 0 & \sigma_{24} \\ \sigma_{31} & 0 & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}.$$

It is estimated by simply specifying the following vector of constraints with corresponding matrix of derivatives:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \sigma_{12} \\ \sigma_{14} \\ \sigma_{23} \end{pmatrix} = \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma})[2] \\ \text{vec}(\boldsymbol{\Sigma})[4] \\ \text{vec}(\boldsymbol{\Sigma})[7] \end{pmatrix}, \mathbf{G}_m = \mathbf{G}_t = \begin{pmatrix} \mathbf{I}_{p^2}[2,] \\ \mathbf{I}_{p^2}[4,] \\ \mathbf{I}_{p^2}[7,] \end{pmatrix}.$$

SAS program: ZERO COV.SAS

Chaudhuri, Drton & Richardson (2007) consider a data example where they focus on  $n = 134$  measurements of  $p = 8$  genes related to galactose used in gene expression data from microarray experiments with yeast strands. They give the marginal correlations and standard deviations of these variables to two decimal places. The sample covariance matrix calculated from this is given below:

$$\mathbf{A} = \begin{pmatrix} 0.1521 & 0.033696 & 0.014664 & -0.11934 & -0.0663 & -0.054756 & -0.050505 & -0.048048 \\ & 0.1296 & 0.038916 & -0.01836 & -0.0612 & 0.033696 & -0.05328 & -0.038808 \\ & & 0.2209 & 0.20774 & 0.22372 & 0.07332 & 0.182595 & 0.188188 \\ & & & 2.89 & 2.5143 & 0.58344 & 2.54745 & 2.27766 \\ & & & & 2.89 & 0.51714 & 2.7676 & 2.40856 \\ & & & & & 0.6084 & 0.7215 & 0.552552 \\ & & & & & & 3.4225 & 2.59259 \\ & & & & & & & 2.3716 \end{pmatrix}$$

Chaudhuri, Drton & Richardson (2007) give mle's of covariances for two different structures, here indicated by  $\boldsymbol{\Sigma}_s$  and  $\boldsymbol{\Sigma}_d$ . The variables are presented in the same order used by Chaudhuri, Drton & Richardson (2007). However, the original variable names/numbers have been dropped:

$$\boldsymbol{\Sigma}_s = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & \sigma_{14} & 0 & \sigma_{16} & 0 & 0 \\ & \sigma_{22} & \sigma_{23} & 0 & 0 & 0 & 0 & 0 \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} & \sigma_{38} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} & \sigma_{48} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} & \sigma_{58} \\ & & & & & \sigma_{66} & \sigma_{67} & \sigma_{68} \\ & & & & & & \sigma_{77} & \sigma_{78} \\ & & & & & & & \sigma_{88} \end{pmatrix} \quad \boldsymbol{\Sigma}_d = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{22} & \sigma_{23} & 0 & 0 & 0 & 0 & 0 \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & 0 & 0 & \sigma_{38} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} & \sigma_{48} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} & \sigma_{58} \\ & & & & & \sigma_{66} & \sigma_{67} & \sigma_{68} \\ & & & & & & \sigma_{77} & \sigma_{78} \\ & & & & & & & \sigma_{88} \end{pmatrix}$$

Using MLE under constraints, covariances can be estimated by simply specifying

$\mathbf{g}(\mathbf{m})$  for the required elements of  $\boldsymbol{\Sigma}$ , where

$$\boldsymbol{\Sigma} \text{ has typical element } \sigma_{ij} = \text{vec}(\boldsymbol{\Sigma})[(i-1)p+j], \quad i, j = 1, \dots, p \quad (6.6)$$

and derivative  $\mathbf{G}_m = \mathbf{G}_t$  with typical element  $\mathbf{I}_{p^2}[(i-1)p+j]$ .

The estimated covariances, which agree with those given by Chaudhuri, Drton & Richardson (2007), are given below:

$$\widehat{\Sigma}_s = \begin{pmatrix} 0.155 & 0.038 & 0 & -0.075 & 0 & -0.063 & 0 & 0 \\ & 0.126 & 0.034 & 0 & 0 & 0 & 0 & 0 \\ & & 0.218 & 0.214 & 0.233 & 0.070 & 0.192 & 0.193 \\ & & & 2.808 & 2.452 & 0.558 & 2.492 & 2.225 \\ & & & & 2.847 & 0.489 & 2.726 & 2.373 \\ & & & & & 0.598 & 0.696 & 0.528 \\ & & & & & & 3.371 & 2.554 \\ & & & & & & & 2.336 \end{pmatrix}$$

$$\widehat{\Sigma}_d = \begin{pmatrix} 0.150 & 0.030 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.126 & 0.036 & 0 & 0 & 0 & 0 & 0 \\ & & 0.218 & 0.061 & 0.086 & 0 & 0 & 0.051 \\ & & & 2.763 & 2.374 & 0.557 & 2.453 & 2.167 \\ & & & & 2.723 & 0.485 & 2.646 & 2.277 \\ & & & & & 0.599 & 0.711 & 0.530 \\ & & & & & & 3.371 & 2.507 \\ & & & & & & & 2.266 \end{pmatrix}$$

The corresponding standard error of the estimates,  $\widehat{\sigma}(\widehat{m})$ , obtained using (2.3), are given below:

$$\widehat{\sigma}(\widehat{\Sigma}_s) = \begin{pmatrix} 0.0178 & 0.0114 & 0 & 0.0259 & 0 & 0.0218 & 0 & 0 \\ & 0.0148 & 0.0135 & 0 & 0 & 0 & 0 & 0 \\ & & 0.0263 & 0.0677 & 0.0690 & 0.0304 & 0.0738 & 0.0621 \\ & & & 0.3388 & 0.3189 & 0.1187 & 0.3382 & 0.2899 \\ & & & & 0.3422 & 0.1171 & 0.3519 & 0.2989 \\ & & & & & 0.0712 & 0.1335 & 0.1092 \\ & & & & & & 0.4081 & 0.3247 \\ & & & & & & & 0.2828 \end{pmatrix}$$

$$\widehat{\sigma}(\widehat{\Sigma}_d) = \begin{pmatrix} 0.0165 & 0.0112 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.0148 & 0.0132 & 0 & 0 & 0 & 0 & 0 \\ & & 0.0247 & 0.0383 & 0.0318 & 0 & 0 & 0.0252 \\ & & & 0.3192 & 0.2961 & 0.1098 & 0.3156 & 0.2705 \\ & & & & 0.3162 & 0.1083 & 0.3251 & 0.2757 \\ & & & & & 0.0665 & 0.1264 & 0.1020 \\ & & & & & & 0.3859 & 0.3028 \\ & & & & & & & 0.2641 \end{pmatrix}$$

The Wald statistic (2.5) for  $\widehat{\Sigma}_s$ ,  $W_s$ , is equal to 9.2767 ( $p = 0.4121$ ) with  $\nu = 9$  and for  $\widehat{\Sigma}_d$  is given by  $W_d = 29.96$ , ( $p = 0.00477$ ) with  $\nu = 13$ .

### 6.3 Dependent Multivariate Normal Samples

Let  $\mathbf{Y} : pk \times 1$  denote a random vector composed of  $k$  groups of  $p$  variables each and  $\mathbf{Y} \sim N_{pk}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with unknown mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  respectively given by:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 : p \times 1 \\ \vdots \\ \boldsymbol{\mu}_k : p \times 1 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{k1} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix} \text{ where } \boldsymbol{\Sigma}_{ij} : p \times p \text{ for } i, j = 1, \dots, k.$$

The hypothesis of interest is  $H_0 : \boldsymbol{\Sigma}_{11} = \cdots = \boldsymbol{\Sigma}_{kk}$  against the alternative hypothesis that at least one of the equalities does not hold. No assumptions are made regarding the off-diagonal covariance matrices. Let  $\mathbf{y}_{ij}$ ,  $i = 1, \dots, q$ ,  $j = 1, \dots, N_i$  denote  $N_i$  observations of the  $pk$ -variate random vector  $\mathbf{Y}$  observed for population  $i$ . The sample covariance matrix is given by

$$\mathbf{A} = \frac{1}{n} \sum_{i=1}^q \sum_{j=1}^{N_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)', \quad \bar{\mathbf{y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{y}_{ij}, \quad n = N - q, \quad N = \sum_{i=1}^q N_i.$$

The density function of  $\mathbf{A}$  is given by (6.1) and expressed in terms of its canonical statistic and corresponding canonical parameter (6.2).

The example considered by Jiang & Sarkar (1998) and Jiang, Sarkar & Hsuan (1999) on the bio-equivalence of two formulations of a drug product is used in the next two subsections to illustrate how a test for independence of groups of variables and a test for the homogeneity of covariance matrices (with or without given structure) can be performed. A standard  $2 \times 2$  crossover experiment was conducted with 25 subjects to compare a new test formulation with a reference formulation. The two treatment periods were separated by a 7-day washout period. Two variables, namely the area under the plasma concentration-time curve (AUC) and the maximum plasma concentration (Cmax), were considered. The natural logarithm of AUC and Cmax are believed to be marginally normally distributed, i.e. the  $2 \times 2$  variate random vector

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N_4 \left[ \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right]. \quad (6.7)$$

The two components of  $\mathbf{y}$ , given by  $\mathbf{y}_{ijk}$ ,  $i = 1, 2$ ,  $j = 1, \dots, N_i$ ,  $k = 1$ (test),  $2$ (reference), represent the  $\log(\text{AUC})$  and  $\log(\text{Cmax})$  of the  $k$ -th formulation of the  $j$ -th subject in the  $i$ -th sequence. Assuming different effects of the period administering formulation  $k$  in sequence  $i$ , the sample covariance matrix

with  $N_1 = 12$ ,  $N_2 = 13$  is given by (Jiang & Sarkar (1998)):

$$\mathbf{A} = \begin{pmatrix} 0.0657739 & 0.01602 & 0.0600567 & -0.002653, \\ 0.01602 & 0.0534908 & 0.0084126 & 0.0125034 \\ 0.0600567 & 0.0084126 & 0.0729088 & -0.000625 \\ -0.002653 & 0.0125034 & -0.000625 & 0.0459961 \end{pmatrix}. \quad (6.8)$$

Three structures for  $\Sigma$ , denoted by  $\Sigma_1$ ,  $\Sigma_2$  and  $\Sigma_3$  respectively, are considered in the next two subsections.

### 6.3.1 Test for independence

To test for independence of  $k$  sets of variates the off-diagonal covariance matrices are all set equal to zero:

$$\mathbf{g}(\mathbf{m}) = \mathbf{g}(\text{vec}(\Sigma)) = \begin{pmatrix} \text{vec}(\Sigma_{12}) \\ \dots \\ \text{vec}(\Sigma_{k-1,k}) \end{pmatrix} \text{ with } \mathbf{G}_m = \begin{pmatrix} \frac{\partial \text{vec}(\Sigma_{12})}{\partial \text{vec}(\Sigma)} \\ \dots \\ \frac{\partial \text{vec}(\Sigma_{k-1,k})}{\partial \text{vec}(\Sigma)} \end{pmatrix}.$$

For the example by Jiang, Sarkar & Hsuan (1999), consider  $H_0 : \Sigma_{12} = 0$  where  $\Sigma : 4 \times 4$ , given by (6.7). The constraints implied by the test for independence of the two sets of variates are:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \sigma_{13} \\ \sigma_{14} \\ \sigma_{23} \\ \sigma_{24} \end{pmatrix} = \begin{pmatrix} \text{vec}(\Sigma)[3] \\ \text{vec}(\Sigma)[4] \\ \text{vec}(\Sigma)[7] \\ \text{vec}(\Sigma)[8] \end{pmatrix} = \mathbf{0} \text{ with}$$

$$\mathbf{G}_m = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

SAS program: WISHART.SAS

The mle of  $\Sigma = \Sigma_1 = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix}$  with corresponding standard errors under these constraints are:

$$\hat{\Sigma}_1 = \begin{pmatrix} 0.060512 & 0.014738 & 0 & 0 \\ 0.014738 & 0.049212 & 0 & 0 \\ 0 & 0 & 0.067076 & -0.000575 \\ 0 & 0 & -0.000575 & 0.042316 \end{pmatrix}$$

$$\hat{\sigma}(\hat{\Sigma}_1) = \begin{pmatrix} 0.0063 & 0.0056 & 0 & 0 \\ 0.0056 & 0.0128 & 0 & 0 \\ 0 & 0 & 0.0069 & 0.0052 \\ 0 & 0 & 0.0052 & 0.0111 \end{pmatrix}$$

The Wald statistic,  $W_1$ , is equal to 12.523 ( $p = 0.0139$ ) with  $\nu = 4$ .

### 6.3.2 Homogeneity of covariance matrices

In order to test for the homogeneity of covariance matrices, the  $(k-1)\frac{1}{2}p(p+1)$  constraints implied by the model are simply:

$$g(\mathbf{m}) = g(\text{vec}(\Sigma)) = \begin{pmatrix} \text{vec}(\Sigma_{11} - \Sigma_{22}) \\ \dots \\ \text{vec}(\Sigma_{11} - \Sigma_{kk}) \end{pmatrix} \text{ with } \mathbf{G}_m = \begin{pmatrix} \frac{\partial \text{vec}(\Sigma_{11}) - \text{vec}(\Sigma_{22})}{\partial \text{vec}(\Sigma)} \\ \dots \\ \frac{\partial \text{vec}(\Sigma_{11}) - \text{vec}(\Sigma_{kk})}{\partial \text{vec}(\Sigma)} \end{pmatrix}. \quad (6.9)$$

To fit the model of homogeneity of covariance matrices of dependent multivariate normals to the data provided in the example by Jiang, Sarkar & Hsuan (1999), the null hypothesis,  $H_0 : \Sigma_{11} = \Sigma_{22}, \Sigma_{12} \neq 0$ , is considered. The appropriate constraints in terms of the typical elements  $\sigma_{ij}$  of  $\Sigma$  (cf. (6.7)) are given by:

$$g(\mathbf{m}) = \begin{pmatrix} \sigma_{11} - \sigma_{33} \\ \sigma_{12} - \sigma_{34} \\ \sigma_{22} - \sigma_{44} \end{pmatrix} = \begin{pmatrix} \text{vec}(\Sigma)[1] - \text{vec}(\Sigma)[11] \\ \text{vec}(\Sigma)[2] - \text{vec}(\Sigma)[12] \\ \text{vec}(\Sigma)[6] - \text{vec}(\Sigma)[16] \end{pmatrix} \text{ with}$$

$$\mathbf{G}_m = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

These constraints yield the mle's of  $\Sigma = \Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$  below. These values correspond with the results of Jiang, Sarkar & Hsuan (1999) to five decimal places.

$$\hat{\Sigma}_2 = \begin{pmatrix} 0.064155 & 0.007345 & 0.055672 & 0.003795 \\ 0.007345 & 0.045719 & 0.002148 & 0.011277 \\ 0.0556724 & 0.002148 & 0.064155 & 0.007345 \\ 0.003800 & 0.011277 & 0.007345 & 0.045719 \end{pmatrix} \hat{\sigma}(\hat{\Sigma}_2) = \begin{pmatrix} 0.0159 & 0.0082 & 0.0159 & 0.0086 \\ 0.0082 & 0.0092 & 0.0099 & 0.0094 \\ 0.0159 & 0.0099 & 0.0159 & 0.0082 \\ 0.0086 & 0.0094 & 0.0082 & 0.0092 \end{pmatrix}$$

This model provides improved fit:  $W_2 = 1.915$ , ( $p = 0.5901$ ) with  $\nu = 3$ .

Patterns can be imposed additionally on submatrices  $\Sigma_{ij} : p \times p$  for  $i \leq j = 1, \dots, k$  by specifying the corresponding constraints. To illustrate, consider the sample covariance matrix in (6.8).

It suggests that the dependence relation between the two sets of variates is contained within the sequence. Thus the relation between the two formulations are independent of the sequence in which they were administered. Consequently, a model for homogeneity of covariance matrices of such structurally dependent multivariate normals was fitted, i.e.  $H_0 : \Sigma_{11} = \Sigma_{22}, \Sigma_{12}^* = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$  where  $a$  is the covariance between the test and reference formulations for AUC and  $b$  is the covariance between the test and reference formulations for Cmax. The constraints required are:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \text{vec}(\Sigma)[1] - \text{vec}(\Sigma)[11] \\ \text{vec}(\Sigma)[2] - \text{vec}(\Sigma)[12] \\ \text{vec}(\Sigma)[6] - \text{vec}(\Sigma)[16] \\ \text{vec}(\Sigma)[4] \\ \text{vec}(\Sigma)[7] \end{pmatrix}$$

with matrix of derivatives

$$\mathbf{G}_m = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

SAS program: WISHART.SAS

The mle of  $\Sigma = \Sigma_3 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12}^* \\ \Sigma_{12}^* & \Sigma_{22} \end{pmatrix}$  is:

$$\hat{\Sigma}_3 = \begin{pmatrix} 0.0634099 & 0.0046713 & 0.0548077 & 0 \\ 0.0046713 & 0.0456856 & 0 & 0.0111829 \\ 0.0548077 & 0 & 0.0634099 & 0.0046713 \\ 0 & 0.0111829 & 0.0046713 & 0.0456856 \end{pmatrix} \quad \hat{\sigma}(\hat{\Sigma}_3) = \begin{pmatrix} 0.0156 & 0.0033 & 0.0156 & 0.0000 \\ 0.0033 & 0.0092 & 0.0000 & 0.0091 \\ 0.0156 & 0.0000 & 0.0156 & 0.0033 \\ 0.0000 & 0.0091 & 0.0033 & 0.0092 \end{pmatrix}$$

The Wald statistic,  $W_3$ , is equal to 2.045 ( $p = 0.843$ ) with  $\nu = 5$ . Of the three models considered, this one fits best.

### 6.3.3 An equivalent approach

The likelihood function of independent multivariate normal populations can of course also be used in a straightforward way as point of departure for this problem of testing for homogeneity of covariance matrices of groups of dependent multivariate normal variables.

Theory is given and illustrated for two independent populations - again using the example by Jiang,



Sarkar & Hsuan (1999). Exactly the same results as those given in the previous subsection (cf.  $\widehat{\Sigma}_2$ ), are obtained. Generalization of the process to  $q$  populations follows directly.

Suppose that  $\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in_i}$  represent  $n_i$  independent observations of two independent random samples from  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  distributions ( $i = 1, 2$ ). Let  $\bar{\mathbf{y}}_i$  represent the sample mean vector and  $\mathbf{S}_i$  the matrix of mean sums of squares and products of the  $i$ -th sample:

$$\bar{\mathbf{y}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij}, \quad \mathbf{S}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \mathbf{y}'_{ij}.$$

The likelihood function for the two multivariate samples can be expressed in its canonical exponential form as follows:

$$\begin{aligned} & L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \\ &= \prod_{i=1}^2 \det(2\pi\boldsymbol{\Sigma}_i)^{-n_i/2} \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \left[ \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)(\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \right] \right\} \\ &= \prod_{i=1}^2 \exp \left\{ n_i \boldsymbol{\mu}'_i \boldsymbol{\Sigma}_i^{-1} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \right) - \frac{n_i}{2} \text{tr} \boldsymbol{\Sigma}_i^{-1} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_{ij} \mathbf{y}'_{ij} \right) - \frac{n_i}{2} \boldsymbol{\mu}'_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \frac{n_i}{2} \ln[\det(2\pi\boldsymbol{\Sigma}_i)] \right\} \\ &= \exp [\boldsymbol{\theta}' \mathbf{t} - \kappa(\boldsymbol{\theta})] \quad \text{where} \end{aligned}$$

$$\mathbf{t} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{y}_{1j} \\ \text{vec} \left( \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{y}_{1j} \mathbf{y}'_{1j} \right) \\ \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{y}_{2j} \\ \text{vec} \left( \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{y}_{2j} \mathbf{y}'_{kj} \right) \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{y}}_1 \\ \text{vec}(\mathbf{S}_1) \\ \bar{\mathbf{y}}_2 \\ \text{vec}(\mathbf{S}_2) \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} n_1 \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 \\ -\frac{n_1}{2} \text{vec}(\boldsymbol{\Sigma}_1^{-1}) \\ n_2 \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2 \\ -\frac{n_2}{2} \text{vec}(\boldsymbol{\Sigma}_2^{-1}) \end{pmatrix} \quad (6.10)$$

$$\text{and } E(\mathbf{t}) = \mathbf{m} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \text{vec}(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}'_1) \\ \boldsymbol{\mu}_2 \\ \text{vec}(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}'_2) \end{pmatrix} = \begin{pmatrix} \mathbf{m}_{11} \\ \mathbf{m}_{12} \\ \mathbf{m}_{21} \\ \mathbf{m}_{22} \end{pmatrix}. \quad (6.11)$$

The covariance matrix of  $\mathbf{t}$  (cf.(2.6)) is given by

$$\mathbf{V} = \text{Cov}(\mathbf{t}) = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{22} \end{pmatrix} \quad \text{with } \mathbf{V}_{ii} = \begin{pmatrix} \mathbf{V}_{11}^i & \mathbf{V}_{12}^i \\ \mathbf{V}_{21}^i & \mathbf{V}_{22}^i \end{pmatrix}$$

where for  $i = 1, 2$ :

$$\begin{aligned} \mathbf{V}_{11}^i &= \frac{1}{n_i} \boldsymbol{\Sigma}_i \\ \mathbf{V}_{21}^i &= \frac{1}{n_i} (\boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i + \boldsymbol{\mu}_i \otimes \boldsymbol{\Sigma}_i) \\ \mathbf{V}_{12}^i &= \mathbf{V}_{21}^{i'} \\ \mathbf{V}_{22}^i &= \frac{1}{n_i} (I_{p^2} + \mathbf{K}) [\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i \boldsymbol{\mu}'_i + \boldsymbol{\mu}_i \boldsymbol{\mu}'_i \otimes \boldsymbol{\Sigma}_i]. \end{aligned}$$

To fit the model of homogeneity of two covariance matrices of four multivariate normal variables (cf.

(6.7)), in terms of the canonical statistics (6.10), the appropriate constraints are now given by:

$$\mathbf{g}(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{m}) \\ \mathbf{g}_2(\mathbf{m}) \\ \mathbf{g}_3(\mathbf{m}) \end{pmatrix}, \mathbf{g}_i(\mathbf{m}) = \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma}_i)[1] - \text{vec}(\boldsymbol{\Sigma}_i)[11] \\ \text{vec}(\boldsymbol{\Sigma}_i)[2] - \text{vec}(\boldsymbol{\Sigma}_i)[12] \\ \text{vec}(\boldsymbol{\Sigma}_i)[6] - \text{vec}(\boldsymbol{\Sigma}_i)[16] \end{pmatrix}, i = 1, 2,$$

$$\mathbf{g}_3(\mathbf{m}) = \text{vec}(\boldsymbol{\Sigma}_1) - \text{vec}(\boldsymbol{\Sigma}_2).$$

The corresponding derivatives are given by

$$\mathbf{G}_m = \begin{pmatrix} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[1,] - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[11,] \\ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[2,] - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[12,] \\ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[6,] - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathbf{m}}[16,] \end{pmatrix}, i = 1, 2, \quad \mathbf{G}_3(\mathbf{m}) = \frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}} - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}}$$

where

$$\frac{\partial \text{vec}(\boldsymbol{\Sigma}_1)}{\partial \mathbf{m}} = \frac{\partial[\mathbf{m}_{12} - \text{vec}(\mathbf{m}_{11}\mathbf{m}'_{11})]}{\partial \mathbf{m}} = (-\mathbf{I}_p \otimes \mathbf{m}_{11} - \mathbf{m}_{11} \otimes \mathbf{I}_p, \mathbf{I}_{p^2}, \mathbf{0}_{p^2 \times p}, \mathbf{0}_{p^2 \times p^2})$$

$$\frac{\partial \text{vec}(\boldsymbol{\Sigma}_2)}{\partial \mathbf{m}} = \frac{\partial[\mathbf{m}_{22} - \text{vec}(\mathbf{m}_{21}\mathbf{m}'_{21})]}{\partial \mathbf{m}} = (\mathbf{0}_{p^2 \times p}, \mathbf{0}_{p^2 \times p^2}, -\mathbf{I}_p \otimes \mathbf{m}_{21} - \mathbf{m}_{21} \otimes \mathbf{I}_p, \mathbf{I}_{p^2}).$$

These constraints yield the same results as those in Section 6.3.2 where  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$ .

SAS program: NORMAL.SAS

# Chapter 7

## Concluding remarks

MLE under constraints provides an alternative approach to numerical maximization routines which are traditionally used for linear models with heterogenous variances. Although the characteristics of the procedure have not yet been investigated per se, experience has shown that it meets the three basic characteristics that a good iterative method should have according to Searle, Casella & McCulloch (1992). The procedure converges to a global maximum from a wide range of starting values, computation at each iteration is relatively quick, and it converges in relatively few iterations. The biggest advantage of the procedure lies in the fact that it does not require derivation of the likelihood equations for each formulation of the underlying model. Problem specific parameterizations can be dealt with directly. Generally, the procedure is very flexible without getting too complicated. It is thus evident that this approach opens up possibilities in a very wide field of applications - especially where the traditional maximum likelihood approach is problematic.

MLE under constraints provides a relatively simple methodology for maximum likelihood estimation, without the usual intricacies associated with heteroskedastic disturbances. Furthermore, the mle's obtained are the result of an exact expression for mle's and not the result of a numerical maximization routine. This methodology for MLE in the case of seemingly unrelated regressions may be pursued as a topic for further research.

The proposed methodology for MLE in the case of multivariate normal samples has the potential to provide mle's in cases where the traditional approach through likelihood equations and maximization thereof are intractable and complicated. The potential for extension of this procedure to other

multivariate structures, e.g. MANOVA, multivariate longitudinal data, dependent multivariate normal populations and other distributions in the exponential class is obvious.

MLE under constraints is a very flexible procedure for the analysis of covariance matrices. In addition to its flexibility for modelling a single patterned covariance matrix, it is appropriate for modelling in the case of independent as well as dependent multivariate distributions.

The procedure for MLE under constraints provides a unified approach to solving estimation problems in a wide variety of applications where generally, estimation and hypothesis testing can be problematic. The potential of this methodology for maximum likelihood estimation in the presence of missing observations represents a new field of research. A thorough investigation into the computational properties of this estimation procedure may be worthwhile to pursue.

# Chapter 8

## References

- ABRAMOWITZ, M. & STEGUN, I.A.,(1972). *Handbook of mathematical functions*, 258. New York: Dover publications.
- ANDERSON, T.W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In: Bose, R.C. & Roy, S.N. (Eds), *Essays in Probability and Statistics*, 1-24. Chapel Hill: University of North Carolina Press.
- ANDERSON, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, **1**, 135-141.
- BARNDORFF-NIELSEN, O. (1982). Exponential Families. In: Kotz,S., Johnson, N.L. (Eds.), *Encyclopedia of Statistical Sciences*,**2**,587-596, New York: John Wiley & Sons.
- BOWERMAN, B.L. & O'CONNELL, R.T. (1990). *Linear statistical models: an applied approach*. Belmont, California: Duxbury Press.
- BROWN, L.D. (1986). Fundamentals of exponential families with applications in statistical decision theory. *Institute of Mathematical Statistics, Lecture Notes-Monograph Series*, **9**.
- BUOT, M-L.G., HOSTEN, S. & RICHARDS, D.S.P. (2007). Counting and locating the solutions of polynomial systems of maximum likelihood equations, II: The Behrens-Fisher problem, *Statistica Sinica*, **17**,1343-1354.
- CHAUDHURI, S., DRTON, M., RICHARDSON, T.S. (2007). Estimation of a covariance matrix with zeros. *Biometrika*, **94**, 199-216.

- CRAFFORD, G. & CROWTHER, N.A.S. (2009). Linear models for grouped data. *South African Statistical Journal*, **43**(2), 151-176.
- ERIKSEN, P. S. (1987). Proportionality of Covariance Matrices. *The Annals of Statistics*, **15**(2), 732-748.
- FLURY, B.K. (1986). Proportionality of  $k$  covariance matrices. *Statistics & Probability Letters* **4**, 29-33.
- GREENE, W.H. (2003). *Econometric Analysis*. Upper Saddle River, New Jersey: Prentice Hall Pearson Education International.
- GRIZZLE, J.E. & ALLEN, D.M. (1969). Analysis of growth and dose response curves. *Biometrics*, **25**(2), 357-381.
- GUJARATI, D.N. (2003). *Basic Econometrics*. Avenue of the Americas, New York: McGraw-Hill.
- HEMMERLE, W.J. & HARTLEY, H.O. (1973). Computing Maximum Likelihood Estimates for the Mixed AOV Model Using the W-Transformation, *Technometrics*, **15**, 819-831.
- JENSEN, S. T. & JOHANSEN, S. (1987). Estimation of proportional covariances. *Statistics & Probability Letters* **6**, 83-85.
- JENSEN, S. T. & MADSEN, J. (2004). Estimation of proportional covariances in the presence of certain linear restrictions. *The Annals of Statistics*, **32**(1), 219-232.
- JIANG, G. & SARKAR, S.K. (1998). Some asymptotic tests for the equality of the covariance matrices of two dependent bivariate normals. *Biometrical Journal*, **40**, 205-225.
- JIANG, G., SARKAR, S.K. & HSUAN, F. (1999). A likelihood ratio test and its modifications for the homogeneity of the covariance matrices of dependent multivariate normals. *Journal of Statistical Planning and Inference*, **81**, 95-111.
- JOHNSON, R.A. & WICHERN, D.W. (1982). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, N.J.: Prentice Hall.

- JUDGE, G.G., GRIFFITHS, W.E., HILL, R.C., Lütkepohl, H., Lee, T-C. (1985). *The Theory and Practice of Econometrics*. New York: John Wiley & Sons.
- LITTLE, R.C., MILLIKEN, G.A., STROUP, W.W., WOLFINGER, R.D., SCHABENBERGER, O (2006). *SAS for Mixed Models*. Cary, North Carolina: SAS Institute Inc..
- MANLY, B. F. J. & RAYNER, J. C.W. (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika*, **74**(4), 841-847.
- MATTHEWS, G.B. & CROWTHER, N.A.S. (1995). A maximum likelihood estimation procedure when modelling in terms of constraints. *South African Statistical Journal*, **29**, 29-50.
- MATTHEWS, G.B. & CROWTHER, N.A.S. (1998). A maximum likelihood estimation procedure for the generalized linear model with restrictions. *South African Statistical Journal*, **32**, 119-144.
- MCCULLOCH, C.E. & SEARLE, S.R. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.
- MCLACHLAN, G.J. & KRISHNAN, T. (1996). *The EM Algorithm and Extensions*. New York: John Wiley & Sons.
- MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons, Inc.
- NEL, D.G. & VAN DER MERWE, C.A., (1986). A solution to the multivariate Behrens-Fisher problem. *Communications in Statistics - Theory and Methods*, **15**(12), 3719-3735.
- POLASEK, W., LIU, S. & NEUDECKER, H. (1999). Heteroskedastic Linear Regression. In: Kotz, S., Read, C.B. and Bands, D.L. (Eds.), *Encyclopedia of Statistics*, Update Vol. **3**, 301-309. New York: John Wiley & Sons.
- RUBIN, D.B. & SZATROWSKI, T.H. (1982). Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm. *Biometrika*, **69**(3), 657-660.
- RUKHIN, A.L., BIGGERSTAFF, B.J. & VANGEL, M.G. (2000). Restricted maximum likelihood estimation of a common mean and the Mandel-Paule algorithm. *Journal of Statistical Planning and Inference*, **83**, 319-330.

SAS INSTITUTE INC. (2004). *SAS/STAT User's guide, Version 9.1*, Cary, NC: SAS Institute Inc.

SAS INSTITUTE INC. (2004). *SAS/IML User's guide, Version 9.1*, Cary, NC: SAS Institute Inc.

SEARLE, S.R., CASELLA, G.C. & MCCULLOCH, C.E. (1992). *Variance Components*. New York: John Wiley & Sons.

SOUTH AFRICAN RESERVE BANK. (2009). Statistical and Economic Information. Available at: <http://www.reservebank.co.za> [Accessed 7 July 2009].

SZATROWSKI, T.H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *The Annals of Statistics*, **8**, 802-810.

VERBEKE, G. & MOLENBERGHS, G. (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach*, New York: Springer-Verlag.

VERBEKE, G. & MOLENBERGHS, G. (2009). *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag.

WANG, Y.(1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **60**(1), 159-174.

ZHANG, J.T. & XU, J.F.(2009). On the k-sample Behrens-Fisher problem for high-dimensional data. *Science in China Series A: Mathematics*, **52**(6), 1285-1304.



# Appendix A

## Appendix

### A.1 SAS Programs for Chapter 3

#### A.1.1 UNIV.SAS

```
*Introductory example;

proc iml worksizes=50;
reset nolog;
y={2, 3, 5, 6, 9};
n=nrow(y);

ymean=y[+]/n;
ss=y'*y/n-ymean##2;
s0=(y-4)*(y-4)/n;
s=y'*y/n;
print ymean ss s0 s;

t=ymean//s;
V=J(2,2,0);
verskill=1;
i=0;
t0=t;
do while (verskill>0.00000001);
i=i+1;
m=t;
t=t0;

V[1,1]=(m[2]-m[1]##2)/n;
V[2,2]=2#V[1,1]##2#n+4#V[1,1]#m[1]##2;
V[1,2]=2#V[1,1]#m[1];
V[2,1]=V[1,2];

*Gm={1 0};
Gm1=1/(2#sqrt(m[1]));
Gm2=0;
Gm=Gm1||Gm2;
```

```

verskil=1;
j=0;
do while (verskil>0.0000001);
tv=t;
j=j+1;

*Gt=Gm;
Gt1=1/(2*sqrt(t[1]));
Gt2=0;
Gt=Gt1||Gt2;

*g=t[1]-4;
g=sqrt(t[1])-2;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
print i j t m V g ;
end;
sig=m[2]-m[1]##2;print sig;
verskil1=sqrt((m-t)'*(m-t));
end;

```

## A.1.2 REGRESSION.SAS

```

*Heteroskedastic linear regression;

proc iml worksizes=50;
reset nolog;
y={55,70,75,65,74,80,84,79,90,98,95,108,113,110,125,115,130,
    135,120,140,144,152,140,137,145,175,189,180,178,191};
x0={80,85,90,100,105,110,115,120,125,130,140,145,150,160,165,
    180,185,190,200,205,210,220,225,230,240,245,250,260,265,270};
n=nrow(y);
t=y/(y#y);

verskil1=1;
i=0;
t0=t;

X=J(n,1,1)||x0;
p=ncol(X);
QX=I(n)-X*inv(X'*X)*X'; PX=X*inv(X'*X)*X';
Z=J(n,1,1);
*Z=X;
*Z=J(n,1,1)||x0##2;
*Z=x0##2;
*Z=J(n,1,1)||x0||(x0##2);
QZ=I(n)-Z*inv(Z'*Z)*Z';
*print QX QZ;

do while (verskil1>0.0000001);
i=i+1;
m=t;
m1=m[1:n];m2=m[n+1:2#n];
t=t0;
sig2=m2-(PX*m1)##2;
if i=1 then sig2=J(n,1,1);
V=(diag(sig2)||2#diag(m1#sig2)) //

```

```

((2#diag(m1#sig2))|| (2#diag(sig2##2+2#m1##2#sig2)));
G1m=QX||J(n,n,0);
*G2m=(-2#QZ*diag(m1))||QZ;
G2m=(-2#QZ*diag(PX*m1)*PX)||QZ;
Gm=G1m//G2m;
verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
t1=t[1:n];t2=t[n+1:2#n];
j=j+1;
G1t=QX||J(n,n,0);
G2t=(-2#QZ*diag(PX*t1)*PX)||QZ;
Gt=G1t//G2t;
g1=QX*t1;
g2=QZ*(t2-(PX*t1)#(PX*t1));
g=g1//g2;
t=t-(Gm*V)*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)*(t-tv));
GV=Gt*V;GVG=GV*Gm';
print i j t m g;
end;
verskil1=sqrt((m-t)*(m-t));
end;

beta=ginv(X'*X)*X'*m1;
alpha=ginv(Z'*Z)*Z'*(m2-m1#m1);
print beta alpha;

*R squared;
R2=(beta'*X'*y-y[+,]##2/n)/(y'*y-y[+,]##2/n);
print R2;

*Calculation of variance of estimated parameters;
sigC=V-(Gm*V)*ginv(Gm*V*Gm')*Gm*V; *print sigC;
dfm=((inv(X'*X)*X')||J(p,n,0))//
      ((-2#inv(Z'*Z)*Z'*diag(m1))|| (inv(Z'*Z)*Z'));
;
sigpar=(dfm*sigC*dfm'); *print sigpar;
vsp=vecdiag(sigpar);
vspsqrt=sqrt(vsp);
print vsp vspsqrt;

* Calculate the Wald statistic;
t=t0;
t1=t[1:n];t2=t[n+1:2#n];
G1t=QX||J(n,n,0);
G2t=(-2#QZ*diag(PX*t1)*PX)||QZ;
Gt=G1t//G2t;
g1=QX*t1;
g2=QZ*(t2-(Px*t1)#(Px*t1));
g=g1//g2;

w=(g)*ginv(Gt*V*Gt')*g;

```

```

GVG=Gm*V*Gm';
GVGe=eigval(GVG);
GVGe=GVGe#(GVGe>0.00000001);
lGVGe=loc(GVGe);
rank=ncol(lGVGe);

```

```

prob=1-probchi(w,rank);
print w prob rank;

```

### A.1.3 ANOVA.SAS

```
*ANOVA 2-way classification: unbalanced case;
```

```

proc iml worksizes=50;
reset nolog;
aa=2; bb=3; ab=aa#bb;

```

```

y1={12,8,9,7};
y2={19,20,16,11,14};
y3={22,32,39,12};
y4={10,13,14,11,11,15};
y5={17,23,22,16};
y6={24,31,37,49};

```

```
nij=J(ab,1,0);
```

```

nij[1,]=nrow(y1);
nij[2,]=nrow(y2);
nij[3,]=nrow(y3);
nij[4,]=nrow(y4);
nij[5,]=nrow(y5);
nij[6,]=nrow(y6);

```

```
n=nrow(nij); *print nij n;
```

```

t1=y1//y2//y3//y4//y5//y6;
t2=(y1#y1)//(y2#y2)//(y3#y3)//(y4#y4)//(y5#y5)//(y6#y6);
t=t1//t2;
n=nrow(t1);

```

```
*Matrices required to specify constraints;
```

```

X={1 1 1 0 1 0,
  1 1 1 0 1 0,
  1 1 1 0 1 0,
  1 1 1 0 1 0,
  1 1 0 1 0 1,
  1 1 0 1 0 1,
  1 1 0 1 0 1,
  1 1 0 1 0 1,
  1 1 0 1 0 1,
  1 1 -1 -1 -1 -1,
  1 1 -1 -1 -1 -1,
  1 1 -1 -1 -1 -1,
  1 1 -1 -1 -1 -1,
  1 -1 1 0 -1 0,
  1 -1 1 0 -1 0,
  1 -1 1 0 -1 0,

```

```

1 -1 1 0 -1 0,
1 -1 1 0 -1 0,
1 -1 1 0 -1 0,
1 -1 0 1 0 -1,
1 -1 0 1 0 -1,
1 -1 0 1 0 -1,
1 -1 0 1 0 -1,
1 -1 -1 -1 1 1,
1 -1 -1 -1 1 1,
1 -1 -1 -1 1 1,
1 -1 -1 -1 1 1};

Z=X[,1]||X[,3:4];
*X=X[,1:4];
Cx=I(n)-X*inv(X'*X)*X';
Cz=I(n)-Z*inv(Z'*Z)*Z';

verskil=1;
i=0;
t0=t;

do while (verskil>0.00000001);
i=i+1;

m=t;
m1=m[1:n];m2=m[n+1:2#n];
sig2=m2-m1##2;
t=t0;

if i=1 then sig2=J(n,1,1);
V=(diag(sig2)||((2#diag(m1#sig2)))) //
((2#diag(m1#sig2))||((2#diag(sig2##2+2#m1##2#sig2))));

G1m=Cx||J(nrow(Cx),n,0);
G2m=-2#Cz*diag(m1)||Cz;
Gm=G1m//G2m;

verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
t1=t[1:n];
t2=t[n+1:2#n];
j=j+1;

G1t=Cx||J(nrow(Cx),n,0);
G2t=-2#Cz*diag(t1)||Cz;
Gt=G1t//G2t;

g1=Cx*t1;
g2=Cz*(t2-t1#t1);
g=g1//g2;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
print i j t m g;
end;

```

```

verskil1=sqrt((m-t)^(m-t));
end;

beta=inv(X'*X)*X'*m1;
alpha=inv(Z'*Z)*Z'*sig2;
print m1 sig2 beta alpha;

* Calculate the Wald statistic;
t1=t0[1:n]; t2=t0[n+1:2#n];

G1t=Cx||J(nrow(Cx),n,0);
G2t=-2#Cz*diag(t1)||Cz;
Gt=G1t//G2t;

g1=Cx*t1;
g2=Cz*(t2-t1#t1);
g=g1//g2;

w=(g)'*ginv(Gt*V*Gt')*g;
rank=sum(echelon(echelon(Gt*V*Gt')));
prob=1-probchi(w,rank);
print w prob rank;

```

#### A.1.4 EXPONENTIAL.SAS

```

*Exponentially distributed observations;

proc iml worksizes=50;
reset nolog;
y={3,5,6,21,30,35,50,60};
n=nrow(y);
x={2,4,6,8,10,12,14,16};
t=y;

verskil1=1;
i=0;
t0=t;
x1=J(n,1,1)||x||x##2;
QX=I(n)-x1*inv(x1'*x1)*x1';

do while (verskil1>0.00000001);
i=i+1;
m=t;
t=t0;
par=inv(x1'*x1)*x1'*m;theta=x1*par;
V=diag(theta#theta);
Gm=Qx;
verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
j=j+1;
Gt=Qx;
g=Qx*t;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)^(t-tv));
print i j t0 t m g ;

```

```

print par;
end;
verskill=sqrt((m-t)*(m-t));
end;

*Calculation of variance of estimated parameters;
sigC=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V; *print sigC;
dfm=inv(X1'*X1)*X1';
sigpar=(dfm*sigC*dfm');
vsp=vecdiag(sigpar);
vspsqrt=sqrt(vsp);
print vsp vspsqrt;

* Calculate the Wald statistic;
t=t0;
Gt=QX;
g=QX*t;

w=(g)'*ginv(Gt*V*Gt')*g;

GVG=Gt*V*Gt';
GVGe=eigval(GVG);
GVGe=GVGe#(GVGe>0.00000001);
lGVGe=loc(GVGe);
rank=ncol(lGVGe);

prob=1-probchi(w,rank);
print w prob rank;

```

### A.1.5 GAMMA.SAS

```

*Linear regression in the case of gamma-distributed errors;

proc iml worksizes=50;
*Routine to calculate alpha for given k: eq. (24);
%macro alpha (dig);
k=&dig;
alpha0=0;f=0;
diff=1;alpha0=0;ai=10;
do while (diff>1e-8);
f=0;
ai=ai/10;
do while (f < (k+0.000000001) );
alpha0=alpha0+ai;
ft=f;
f=exp(digamma(alpha0))/alpha0;
end;
alpha0=alpha0-ai;
diff=sqrt((ft-f)**2);
end;
alpha=alpha0;
di=digamma(alpha);
%mend;

reset nolog;
y={3,5,6,21,30,35,50,60}; *Vector of dependent observations;
n=nrow(y);

```

```

x={2,4,6,8,10,12,14,16}; *Vector of independent observations;

t=y/(log(y)[+]/n); *Canonical statistic;
diff1=1;
i=0;
t0=t;
X1=J(n,1,1)||x; *||x##2;
QX=I(n)-X1*inv(X1'*X1)*X1';

do while (diff1>0.00000001); *Iteration over m;
i=i+1;
m=t;
m1=m[1:n];m2=m[n+1];
t=t0;

*Estimate parameters;
factor=exp(m[n+1])/exp(log(m1)[+]/n);
if factor>=0.9999991 then factor=0.9;
%alpha(factor);
beta=m1/alpha;
print factor alpha beta;

*Covariance matrix;
V11=alpha#diag(beta##2);
V12=beta/n;
V21=V12';
V22=trigamma(alpha)/n;
V=(V11||V12)/(V12' ||V22);

*Matrix of derivatives;
G1m=Qx||J(n,1,0);
*G1m=(Qx*diag(1/m1))||J(n,1,0); *Required for non-linear model;
G2m=(J(1,n,0)||1)-((1/m1'/n)||0);
Gm=G1m; *//G2m;

diff=1;
j=0;
do while (diff>0.00000001); *Iteration over t;
t1=t[1:n];t2=t[n+1];
tv=t;
j=j+1;

*Matrix of derivatives;
G1t=Qx||J(n,1,0);
*G1t=(Qx*diag(1/t1))||J(n,1,0); *Required for non-linear model;
G2t=(J(1,n,0)||1)-((1/t1'/n)||0);
Gt=G1t; *//G2t;

*Constraints;
g1=Qx*t1;
*g1=Qx*log(t1); *Required for non-linear model;
g2=t2-(log(t1)[+]/n)-digamma(1)+log(1); * alpha=1;
g=g1; *//g2;

*Eq. (3);
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;

```





```

1 0 1 0,
1 0 1 0,
1 0 1 0,
1 0 1 0,
1 0 1 0,
1 0 1 0,
1 0 1 0,
1 0 0 1,
1 0 0 1,
1 0 0 1,
1 0 0 1,
1 0 0 1};

Z={1 0 1 0 0 0 0 0,
1 0 1 0 0 0 0 0,
1 0 1 0 0 0 0 0,
0 1 0 1 0 0 0 0,
0 1 0 1 0 0 0 0,
1 0 0 0 1 0 0 0,
1 0 0 0 1 0 0 0,
1 0 0 0 1 0 0 0,
0 1 0 0 0 1 0 0,
0 1 0 0 0 1 0 0,
0 1 0 0 0 1 0 0,
1 0 0 0 0 0 1 0,
1 0 0 0 0 0 1 0,
0 1 0 0 0 0 0 1,
0 1 0 0 0 0 0 1,
0 1 0 0 0 0 0 1};

y={237, 254, 246, 178, 179, 208, 178, 187, 146, 145, 141, 186, 183, 142, 125, 136};

const=1;

n=nrow(y);
n2=n##2;
K=j(n2,n2,0);
do i=1 to n;
do j=1 to n;
H=j(n,n,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

QX=I(n)-X*ginv(X'*X)*X';

Z1=Z[,1:2];Z2=Z[,3:8];
ZZt1=Z1*Z1';
ZZt2=Z2*Z2';
*Model 1;
C=colvec(I(n))||colvec(ZZt1);
*Model 2;
*In1=I(n)[,1:5]||I(n)[,9:16];
*In2=I(n)[,6:8];
*C=colvec(In1*In1')||colvec(In2*In2')||colvec(ZZt1)||colvec(ZZt2);
*Model 3;
*In1=I(n)[,1:5]||I(n)[,9:13];

```

```

*In2=I(n) [,6:8];*In3=I(n) [,14:16];
*C=colvec(In1*In1')||colvec(In2*In2')||colvec(In3*In3')||(colvec(ZZt1))||(colvec(ZZt2));
QC=I(n2)-C*ginv(C'*C)*C';
P=X*ginv(X'*X)*X';

YY=colvec((y*y')');
t=y//YY;
verskil1=2;
i=0;
t0=t;
do while (verskil1>1E-4);
i=i+1;
m=t;
t=t0;
m1=m[1:n];
m2=m[n+1:(n+1)#n];
Sig=shape((m2-colvec(m1*m1')),n,n);
if i=1 then Sig=I(n)+const;
Phi=Sig@m1+m1@Sig;

V= (Sig||(Phi'))//((I(n2)+K)*(Sig@Sig+Sig@(m1*m1')+(m1*m1')@Sig)) ;

Gm1=QX||J(n,n2,0);
Gm2=QC*((-I(n)@m1-m1@I(n))||I(n2));
Gm=Gm1//Gm2;

verskil=1;
j=0;

do while (verskil>1E-5);
tv=t;
t1=t[1:n];t2=t[n+1:(n+1)#n];
j=j+1;

Gt1=QX||J(n,n2,0);
Gt2=QC*((-I(n)@t1-t1@I(n))||I(n2));
Gt=Gt1//Gt2;

g1=QX*t1;
g2=QC*(t2-colvec(t1*t1'));
g=g1//g2;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
end;
print i j t m g ;
verskil1=sqrt((m-t)'*(m-t));
end;

print m1;

betah=ginv(X'*X)*X'*m1;
sighat=inv(C'*C)*C'*colvec(Sig);
print betah sighat;

* Calculate the Wald statistic;

```

```

t1=t0[1:n]; t2=t0[n+1:(n+1)#n];

Gt1=QX||J(n,n2,0);
Gt2=QC*((-I(n)@t1-t1@I(n))||I(n2));
Gt=Gt1//Gt2;

g1=QX*t1;
g2=QC*(t2-colvec(t1*t1'));
g=g1//g2;

w=(g)'*ginv(Gt*V*Gt')*g;
rank=sum(echelon(echelon(Gt*V*Gt')));
prob=1-probchi(w,rank);
print w prob rank;

```

## A.2 SAS Programs for Chapter 4

### A.2.1 MULTIV1OBS.SAS

```

*Introductory example;

proc iml worksizes=50;
options pagesize=100;
reset nolog;
y={4,2};
n=nrow(y); n2=n##2; print n;
rho=0.5;
C={1 0 0 -1};
*C={1 0 0 -2};
*C={1 0 0 -3};
*print C1[format=2.0] C3[format=2.0];
K=j(n2,n2,0);
do i=1 to n;
do j=1 to n;
H=j(n,n,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

yy=colvec((y*y')');
Sig=J(n,n,0);
t=yy;
verskil1=1;
i=0;
t0=t;
do while (verskil1>1e-5);
i=i+1;
m=t;
t=t0;
Sig=shape(m,n,n);
if i=1 then Sig=I(2);
V=(I(n2)+K)*(Sig@Sig);
Gm1=I(4)[2,];
Gm2=C;

```

```

Gm3=(rho#sqrt(m[4,]/m[1,])/2)||(-1)
||0|| (rho#sqrt(m[1,]/m[4,])/2);
Gm=Gm1;
verskil=1;
j=0;
do while (verskil>1e-6);
tv=t;
j=j+1;
Gt1=I(4)[2,];
Gt2=C;
Gt3=(rho#sqrt(t[4,]/t[1,])/2)||(-1)
||0|| (rho#sqrt(t[1,]/t[4,])/2);
Gt=Gt1;
g1=t[2,];
g2=C*t;
g3=rho#sqrt(t[1,]#t[4,])-t[2,];
g=g1;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
print i j t m g;
end;
print i j t m g;
verskil1=sqrt((m-t)'*(m-t));
end;
rho=m[2]/sqrt(m[1]#m[4]); print rho;

```

## A.2.2 GWHETEROSK.SAS

\*Linear models on group variances;

\*Additive heteroskedasticity per group;

```

proc iml worksizes=910000000;
options pagesize=100;
reset nolog;
*Greene: lnCost=b1+b2lnOutput+b3LoadFactor+b4lnPriceFuel+Firm (p.237);
*firm time Cost Output PriceFuel LoadFactor;
DATA={
1 1 1140640 .952757 106650 .534487,
1 2 1215690 .986757 110307 .532328,
1 3 1309570 1.091980 110574 .547736,
1 4 1511530 1.175780 121974 .540846,
1 5 1676730 1.160170 196606 .591167,
1 6 1823740 1.173760 265609 .575417,
1 7 2022890 1.290510 263451 .594495,
1 8 2314760 1.390670 316411 .597409,
1 9 2639160 1.612730 384110 .638522,
1 10 3247620 1.825440 569251 .676287,
1 11 3787750 1.546040 871636 .605735,
1 12 3867750 1.527900 997239 .614360,
1 13 3996020 1.660200 938002 .633366,
1 14 4282880 1.822310 859572 .650117,
1 15 4748320 1.936460 823411 .625603,
2 1 569292 .520635 103795 .490851,
2 2 640614 .534627 111477 .473449,
2 3 777655 .655192 118664 .503013,
2 4 999294 .791575 114797 .512501,
2 5 1203970 .842945 215322 .566782,

```

|   |    |         |          |         |          |
|---|----|---------|----------|---------|----------|
| 2 | 6  | 1358100 | .852892  | 281704  | .558133, |
| 2 | 7  | 1501350 | .922843  | 304818  | .558799, |
| 2 | 8  | 1709270 | 1.000000 | 348609  | .572070, |
| 2 | 9  | 2025400 | 1.198450 | 374579  | .624763, |
| 2 | 10 | 2548370 | 1.340670 | 544109  | .628706, |
| 2 | 11 | 3137740 | 1.326240 | 853356  | .589150, |
| 2 | 12 | 3557700 | 1.248520 | 1003200 | .532612, |
| 2 | 13 | 3717740 | 1.254320 | 941977  | .526652, |
| 2 | 14 | 3962370 | 1.371770 | 856533  | .540163, |
| 2 | 15 | 4209390 | 1.389740 | 821361  | .528775, |
| 3 | 1  | 286298  | .262424  | 118788  | .524334, |
| 3 | 2  | 309290  | .266433  | 123798  | .537185, |
| 3 | 3  | 342056  | .306043  | 122882  | .582119, |
| 3 | 4  | 374595  | .325586  | 131274  | .579489, |
| 3 | 5  | 450037  | .345706  | 222037  | .606592, |
| 3 | 6  | 510412  | .367517  | 278721  | .607270, |
| 3 | 7  | 575347  | .409937  | 306564  | .582425, |
| 3 | 8  | 669331  | .448023  | 356073  | .573972, |
| 3 | 9  | 783799  | .539595  | 378311  | .654256, |
| 3 | 10 | 913883  | .539382  | 555267  | .631055, |
| 3 | 11 | 1041520 | .467967  | 850322  | .569240, |
| 3 | 12 | 1125800 | .450544  | 1015610 | .589682, |
| 3 | 13 | 1096070 | .468793  | 954508  | .587953, |
| 3 | 14 | 1198930 | .494397  | 886999  | .565388, |
| 3 | 15 | 1170470 | .493317  | 844079  | .577078, |
| 4 | 1  | 145167  | .086393  | 114987  | .432066, |
| 4 | 2  | 170192  | .096740  | 120501  | .439669, |
| 4 | 3  | 247506  | .141500  | 121908  | .488932, |
| 4 | 4  | 309391  | .169715  | 127220  | .484181, |
| 4 | 5  | 354338  | .173805  | 209405  | .529925, |
| 4 | 6  | 373941  | .164272  | 263148  | .532723, |
| 4 | 7  | 420915  | .170906  | 316724  | .549067, |
| 4 | 8  | 474017  | .177840  | 363598  | .557140, |
| 4 | 9  | 532590  | .192248  | 389436  | .611377, |
| 4 | 10 | 676771  | .242469  | 547376  | .645319, |
| 4 | 11 | 880438  | .256505  | 850418  | .611734, |
| 4 | 12 | 1052020 | .249657  | 1011170 | .580884, |
| 4 | 13 | 1193680 | .273923  | 951934  | .572047, |
| 4 | 14 | 1303390 | .371131  | 881323  | .594570, |
| 4 | 15 | 1436970 | .421411  | 831374  | .585525, |
| 5 | 1  | 91361   | .051028  | 118222  | .442875, |
| 5 | 2  | 95428   | .052646  | 116223  | .462473, |
| 5 | 3  | 98187   | .056348  | 115853  | .519118, |
| 5 | 4  | 115967  | .066953  | 129372  | .529331, |
| 5 | 5  | 138382  | .070308  | 243266  | .557797, |
| 5 | 6  | 156228  | .073961  | 277930  | .556181, |
| 5 | 7  | 183169  | .084946  | 317273  | .569327, |
| 5 | 8  | 210212  | .095474  | 358794  | .583465, |
| 5 | 9  | 274024  | .119814  | 397667  | .631818, |
| 5 | 10 | 356915  | .150046  | 566672  | .604723, |
| 5 | 11 | 432344  | .144014  | 848393  | .587921, |
| 5 | 12 | 524294  | .169300  | 1005740 | .616159, |
| 5 | 13 | 530924  | .172761  | 958231  | .605868, |
| 5 | 14 | 581447  | .186670  | 872924  | .594688, |
| 5 | 15 | 610257  | .213279  | 844622  | .635545, |
| 6 | 1  | 68978   | .037682  | 117112  | .448539, |
| 6 | 2  | 74904   | .039784  | 119420  | .475889, |

```

6   3   83829   .044331   116087   .500562,
6   4   98148   .050245   122997   .500344,
6   5   118449  .055046   194309   .528897,
6   6   133161  .052462   307923   .495361,
6   7   145062  .056977   323595   .510342,
6   8   170711  .061490   363081   .518296,
6   9   199775  .069027   386422   .546723,
6  10   276797  .092749   564867   .554276,
6  11   381478  .112640   874818   .517766,
6  12   506969  .154154  1013170   .580049,
6  13   633388  .186461   930477   .556024,
6  14   804388  .246847   851676   .537791,
6  15  1009500  .304013   819476   .525775
};

```

```

y1=log(data[loc(data[,1]=1),3]);
y2=log(data[loc(data[,1]=2),3]);
y3=log(data[loc(data[,1]=3),3]);
y4=log(data[loc(data[,1]=4),3]);
y5=log(data[loc(data[,1]=5),3]);
y6=log(data[loc(data[,1]=6),3]);
*print y1 y2 y3 y4 y5 y6;

```

```

z1=log(data[loc(data[,1]=1),5]);
z2=log(data[loc(data[,1]=2),5]);
z3=log(data[loc(data[,1]=3),5]);
z4=log(data[loc(data[,1]=4),5]);
z5=log(data[loc(data[,1]=5),5]);
z6=log(data[loc(data[,1]=6),5]);
*print z1 z2 z3 z4 z5 z6;

```

```

ng1=nrow(y1);
ng2=nrow(y2);
ng3=nrow(y3);
ng4=nrow(y4);
ng5=nrow(y5);
ng6=nrow(y6);
y=y1//y2//y3//y4//y5//y6;
n=nrow(y); n2=n##2; nc=ncol(y); *print n nc;

```

```

*Xf=block(J(ng1,1,0),J(ng2,1,1),J(ng3,1,1),J(ng4,1,1),J(ng5,1,1),J(ng6,1,1));
Xf=J(ng1,5,0)/(block(J(ng2,1,1),J(ng3,1,1),J(ng4,1,1),J(ng5,1,1),J(ng6,1,1)));
X=J(n,1,1)||log(data[,4])||log(data[,5])||(data[,6])||Xf; *print Xf;
p=ncol(X); *print p;
PX=X*ginv(X'*X)*X';
QX=I(n)-PX;

```

```

*Z=J(n,1,1);*OLS;
Z=block(J(ng1,1,1),J(ng2,1,1),J(ng3,1,1),J(ng4,1,1),J(ng5,1,1),J(ng6,1,1));
*print Z;
*Z=J(n,1,1)||block(z1,z2,z3,z4,z5,z6);
PZ=Z*inv(Z'*Z)*Z';
QZ=I(n)-PZ;
*print PZ[format=4.2] QZ[format=4.2];

```

```

t=y/(y#y);
verskil1=1;
i=0;
t0=t;
do while (verskil1>0.00000001);
i=i+1;
m=t;
m1=m[1:n];m2=m[n+1:2#n];
t=t0;
sig2=m2-m1##2;
if i=1 then sig2=J(n,1,1);
V=(diag(sig2)|| (2#diag(m1#sig2))) //
  ((2#diag(m1#sig2))|| (2#diag(sig2##2+2#m1##2#sig2)));
G1m=QX||J(n,n,0);
G2m=(-2#QZ*diag(m1))||QZ;
Gm=G1m//G2m;
verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
t1=t[1:n];t2=t[n+1:2#n];
j=j+1;
G1t=QX||J(n,n,0);
G2t=(-2#QZ*diag(t1))||QZ;
Gt=G1t//G2t;
g1=QX*t1;
g2=QZ*(t2-t1#t1);
g=g1//g2;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
GV=Gt*V;GVG=GV*Gm';
*print i j t m V g;
end;
verskil1=sqrt((m-t)'*(m-t));
print i j t m g;
end;

beta=ginv(X'*X)*X'*m1;
alpha=ginv(Z'*Z)*Z'*(m2-m1#m1);
print beta alpha;
Sig=block(m2-m1#m1); print Sig;

*R squared;
R2=(beta'*X'*y-y[+,]##2/n)/(y'*y-y[+,]##2/n);
print R2;

*Calculation of variance of estimated parameters;
sigC=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V; *print sigC;
dfm=((ginv(X'*X)*X')||J(p,n,0))//
  ((-2#inv(Z'*Z)*Z'*diag(m1))|| (inv(Z'*Z)*Z'));
;
sigpar=(dfm*sigC*dfm');
vsp=vecdiag(sigpar);
vspsqrt=sqrt(vsp);
print vspsqrt;

* Calculate the Wald statistic;

```



```

t=t0;
t1=t[1:n];t2=t[n+1:2#n];
G1t=QX||J(n,n,0);
G2t=(-2#QZ*diag(t1))||QZ;
Gt=G1t//G2t;
g1=QX*t1;
g2=QZ*(t2-t1#t1);
g=g1//g2;

w=(g1)'*ginv(G1t*V*G1t')*g1;

GVG=Gm*V*Gm';
GVGe=eigval(GVG);
GVGe=GVGe#(GVGe>0.00000001);
lGVGe=loc(GVGe);
rank=ncol(lGVGe);

prob=1-probchi(w,rank);
print w prob rank;

```

### A.2.3 SPLINES.SAS

```

*Gujarati, p.274 Savings=f(Income);
*Year Time Savings Income;
*Spline regression;

proc iml;
data={
1970 1 61      727.1 ,
1971 2 68.6 790.2 ,
1972 3 63.6 855.3 ,
1973 4 89.6  965  ,
1974 5 97.6 1054.2,
1975 6 104.4 1159.2,
1976 7 96.4 1273  ,
1977 8 92.5 1401.4,
1978 9 112.6 1580.1,
1979 10 130.1 1769.5,
1980 11 161.8 1973.3,
1981 12 199.1 2200.2,
1982 13 205.5 2347.3,
1983 14 167      2522.4,
1984 15 235.7 2810  ,
1985 16 206.2 3002  ,
1986 17 196.5 3187.6,
1987 18 168.4 3363.1,
1988 19 189.1 3640.8,
1989 20 187.8 3894.5,
1990 21 208.7 4166.8,
1991 22 246.4 4343.7,
1992 23 272.6 4613.7,
1993 24 214.4 4790.2,
1994 25 189.4 5021.7,
1995 26 249.3 5320.8};

y1=data[1:12,3];
y2=data[13:26,3];

```

```

n1=nrow(y1);
n2=nrow(y2);
*print n1 n2;

X1=J(n1,1,1)||data[1:12,4];
X2=J(n2,1,1)||data[13:26,4];
p=ncol(X1);
x0=((X1[12,]+X2[1,])/2)';
PX1=X1*ginv(X1'*X1)*X1';
QX1=I(n1)-PX1;
PX2=X2*ginv(X2'*X2)*X2';
QX2=I(n2)-PX2;

beta1ls=inv(X1'*X1)*X1'*y1;
beta2ls=inv(X2'*X2)*X2'*y2;
print beta1ls beta2ls;

Z1=J(n1,1,1)||data[1:12,2];
Z2=J(n2,1,1)||data[13:26,2];
pZ=ncol(Z1);
PZ1=Z1*inv(Z1'*Z1)*Z1';
QZ1=I(n1)-PZ1;
PZ2=Z2*inv(Z2'*Z2)*Z2';
QZ2=I(n2)-PZ2;

t=y1//(y1#y1)//y2//(y2#y2);

verskil1=1;
i=0;
t0=t;
do while (verskil1>0.00000001);
i=i+1;
m=t;
m1=m[1:2#n1]; m2=m[2#n1+1:2#n1+2#n2];
m11=m1[1:n1]; m12=m1[n1+1:2#n1];
m21=m2[1:n2]; m22=m2[n2+1:2#n2];

sig1m=m12-m11##2;
sig2m=m22-m21##2;
if i=1 then do; sig1m=J(n1,1,1); sig2m=J(n2,1,1); end;
V11=(diag(sig1m/n1)|| (2#diag(m11#sig1m/n1))) //
((2#diag(m11#sig1m/n1))|| (2#diag(sig1m##2+2#m11##2#sig1m)/n1));
V22=(diag(sig2m/n2)|| (2#diag(m21#sig2m/n2))) //
((2#diag(m21#sig2m/n2))|| (2#diag(sig2m##2+2#m21##2#sig2m)/n2));
V=block(V11,V22);

G11m=QX1||J(n1,n1,0)||J(n1,n2,0)||J(n1,n2,0);
G12m=J(n2,n1,0)||J(n2,n1,0)||QX2||J(n2,n2,0);
G21m=(-2#QZ1*diag(m11))||QZ1||J(n1,n2,0)||J(n1,n2,0);
G22m=J(n2,n1,0)||J(n2,n1,0)||(-2#QZ2*diag(m21))||QZ2;
G3m=(inv(X1'*X1)*X1')||J(p,n1,0)||(-inv(X2'*X2)*X2')||J(p,n2,0);
G4m=(-2#(diag(m11))[1,])|| (I(n1)[1,])|| ((2#diag(m21))[1,])|| (-I(n2)[1,]);
G5m=(x0'*inv(X1'*X1)*X1')||J(1,n1,0)||(-x0'*inv(X2'*X2)*X2')||J(1,n2,0);
Gm=G11m//G12m//G21m//G22m//G5m;

t=t0;

```

```

verskil=1;
j=0;
do while (verskil>0.0000001);
tv=t;
j=j+1;
t1=t[1:2#n1]; t2=t[2#n1+1:2#n1+2#n2];
t11=t1[1:n1]; t12=t1[n1+1:2#n1];
t21=t2[1:n2]; t22=t2[n2+1:2#n2];
sig1t=t12-t11##2;
sig2t=t22-t21##2;

G11t=G11m;
G12t=G12m;
G21t=(-2#QZ1*diag(t11)||QZ1||J(n1,n2,0)||J(n1,n2,0);
G22t=J(n2,n1,0)||J(n2,n1,0)||(-2#QZ2*diag(t21)||QZ2;
G3t=G3m;
G4t=(-2#(diag(t11))[1,])||(I(n1)[1,])|(2#(diag(t21))[1,])|(-I(n2)[1,]);
G5t=G5m;
Gt=G11t//G12t//G21t//G22t//G5t;

g11=QX1*t11;
g12=QX2*t21;
g21=QZ1*(t12-t11#t11);
g22=QZ2*(t22-t21#t21);
g3=inv(X1'*X1)*X1'*t11-inv(X2'*X2)*X2'*t21;
g4=sig1t[1,]-sig2t[1,];
g5=t11'*X1*inv(X1'*X1)*x0-t21'*X2*inv(X2'*X2)*x0;

g=g11//g12//g21//g22//g5;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
*print i j t m V g;
end;
verskil1=sqrt((m-t)'*(m-t));
print i j t m g;
end;

beta1=inv(X1'*X1)*X1'*m11;
alpha1=inv(Z1'*Z1)*Z1'*sig1m;
beta2=inv(X2'*X2)*X2'*m21;
alpha2=inv(Z2'*Z2)*Z2'*sig2m;
print beta1 alpha1 beta2 alpha2;
print sig1m sig2m;
*Sig=block(diag(sig1m),diag(sig2m)); *print Sig;

*R squared;
R21=(beta1'*X1'*y1-y1[+,]##2/n1)/(y1'*y1-y1[+,]##2/n1);
R22=(beta2'*X2'*y2-y2[+,]##2/n2)/(y2'*y2-y2[+,]##2/n2);
print R21 R22;

m01=m11'*X1*inv(X1'*X1)*x0;
m02=m21'*X2*inv(X2'*X2)*x0;
print m01 m02;

*Calculation of variance of estimated parameters;
sigC=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V; *print sigC;

```

```

dfm=
((ginv(X1'*X1)*X1')||J(p,n1,0)||J(p,n2,0)||J(p,n2,0) )//
((-2#inv(Z1'*Z1)*Z1'*diag(m11))||(inv(Z1'*Z1)*Z1')||J(pZ,n2,0)||J(pZ,n2,0) )//
(J(p,n1,0)||J(p,n1,0)||(ginv(X2'*X2)*X2')||J(p,n2,0) )//
(J(pZ,n1,0)||J(pZ,n1,0)||(-2#inv(Z2'*Z2)*Z2'*diag(m21))||(inv(Z2'*Z2)*Z2'));

sigpar=(dfm*sigC*dfm'); *print sigpar;
vsp=vecdiag(sigpar);
vspsqrt=sqrt(vsp);
print vspsqrt;

* Calculate the Wald statistic;
t=t0;
t1=t[1:2#n1]; t2=t[2#n1+1:2#n1+2#n2];
t11=t1[1:n1]; t12=t1[n1+1:2#n1];
t21=t2[1:n2]; t22=t2[n2+1:2#n2];
sig1t=t12-t11##2;
sig2t=t22-t21##2;

G11t=G11m;
G12t=G12m;
G21t=(-2#QZ1*diag(t11))||QZ1||J(n1,n2,0)||J(n1,n2,0);
G22t=J(n2,n1,0)||J(n2,n1,0)||(-2#QZ2*diag(t21))||QZ2;
G3t=G3m;
G4t=(-2#(diag(t11))[1,])||(I(n1)[1,])||(2#(diag(t21))[1,])||(-I(n2)[1,]);
G5t=G5m;
Gt=G11t//G12t//G21t//G22t//G5t;

g11=QX1*t11;
g12=QX2*t21;
g21=QZ1*(t12-t11#t11);
g22=QZ2*(t22-t21#t21);
g3=inv(X1'*X1)*X1'*t11-inv(X2'*X2)*X2'*t21;
g4=sig1t[1,]-sig2t[1,];
g5=t11'*X1*inv(X1'*X1)*x0-t21'*X2*inv(X2'*X2)*x0;

g=g11//g12//g21//g22//g5;

g1=g11//g12; G1t=G11t//G12t;
w=(g1)'*ginv(G1t*V*G1t')*g1;

GVG=Gm*V*Gm';
GVGe=eigval(GVG);
GVGe=GVGe#(GVGe>0.00000001);
lGVGe=loc(GVGe);
rank=ncol(lGVGe);

prob=1-probchi(w,rank);
print w prob rank;

```

## A.2.4 AR1.SAS

```

*Real Gross Capital Index of South Africa: 1999-2008;
*AR(1) with heteroskedastic variances;
*AR(1) with heteroskedastic variances+explanatory variables;

```

```
proc iml worksizes=50;
```

```

options pagesize=100;
reset nolog;
DATA={
  0.95901    0.96011    9.70167    1999,
  1.00000    1.00000    8.38620    2000,
  1.03510    1.02735    5.70660    2001,
  1.07339    1.06504    2.30092    2002,
  1.17118    1.09827    3.81404    2003,
  1.27555    1.15169    8.12716    2004,
  1.40594    1.20887    4.66702    2005,
  1.59163    1.27320    3.23816    2006,
  1.85095    1.33810    0.88436    2007,
  2.03939    1.37908    -2.39917   2008};

y=DATA[,1]#100;
*print y[format=6.2] (data[,1])[format=6.2] (data[,2])[format=6.2] (data[,3])[format=6.2];
n=nrow(y); n2=n##2;
const=0.2;
X=J(n,1,1)||cusum(J(n,1,1))||(cusum(J(n,1,1))##2); *AR(1) with heteroscedasticity;
*Modelling in terms of explanatory variables;
X=J(n,1,1)||cusum(J(n,1,1))||(cusum(J(n,1,1))##2)|| (data[,2]#100)||data[,3];
PX=X*inv(X'*X)*X';
QX=I(n)-PX;
QQ=I(n2)-(I(n)-PX)@(I(n)-PX);

free C1 C2 C3;
do i=1 to n-1;
C1=C1//I(n2)[1+(n+1)#(i-1),];
C2=C2//I(n2)[n+1+1+(n+1)#(i-1),];
C3=C3//I(n2)[2+(n+1)#(i-1),];
print (1+(n+1)#(i-1)) (n+1+1+(n+1)#(i-1)) (2+(n+1)#(i-1));
end;
Cd=C1//I(n2)[n#n,];

*print C1[format=2.0] C3[format=2.0];
K=j(n2,n2,0);
do i=1 to n;
do j=1 to n;
H=j(n,n,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

ym=y[,+];
yy=colvec((y*y')');
Sig=J(n,n,0);
t=ym//yy;
verskil1=1;
i=0;
t0=t;
do while (verskil1>1e-5);
i=i+1;
m=t;
t=t0;
mum=m[1:n]; m2=m[n+1:(n+1)#n]; *nu=PX*mu;
covm=m2-QQ*colvec(mum*mum'); *print "covm" (shape(covm,n,n));

```

```

if i=1 then covm=colvec(I(n)+const);
sigm=shape(covm,n,n); *print sigm;

cm=(covm'*C1'*C2*covm)/(covm'*C1'*C1*covm);
dcm=( (covm'*C1'*C1*covm)#covm'*(C1'*C2+C2'*C1)
      -2#(covm'*C1'*C2*covm)#covm'*(C1'*C1) )/(covm'*C1'*C1*covm)##2;
rm=(covm'*C1'*C3*covm)/(covm'*C1'*C1*covm);
drm=( (covm'*C1'*C1*covm)#covm'*(C1'*C3+C3'*C1)
      -2#(covm'*C1'*C3*covm)#covm'*(C1'*C1) )/(covm'*C1'*C1*covm)##2;
rcm=rm/cm;
drcm=(cm#drm-rm#dcm)/cm##2;
sig2m=covm[1];dsig2m=I(n2)[1,];

print sig2m cm rm;

do ii=1 to n;
do jj=ii to n;
Sig[ii,jj]=Sig2m#cm##(ii-1)#rm##(jj-ii);
Sig[jj,ii]=Sig[ii,jj];
end;end;
print Sig;

Phi=(Sig@mum+mum@Sig);
V= ( (Sig||Phi')//((I(n2)+K)*(Sig@Sig+Sig@(mum*mum')+(mum*mum')@Sig)) );
Gm1=QX||J(n,n2,0);
sigm=shape(covm,n,n);

free Gm2;
do ii=1 to n;
do jj=ii to n;
Gm2=Gm2//( I(n)[ii,]@I(n)[jj,]-dsig2m#cm##(ii-1)#rm##(jj-ii)
           -(ii-1)#sig2m#cm##(ii-2)#rm##(jj-ii)#dcm
           -(jj-ii)#sig2m#cm##(ii-1)#rm##(jj-ii-1)#drm );
end;end;
Gm2=Gm2*(QQ*(-I(n)@mum-mum@I(n))||I(n2));

Gm=Gm1//Gm2;
verskil=1;
j=0;
do while (verskil>1e-6);
tv=t;
mut=t[1:n]; t2=t[n+1:(n+1)#n];
j=j+1;

covt=t2-QQ*colvec(mut*mut');
Gt1=QX||J(n,n2,0);

sigt=shape(covt,n,n); *print sigt;
ct=(covt'*C1'*C2*covt)/(covt'*C1'*C1*covt);
dct=( (covt'*C1'*C1*covt)#covt'*(C1'*C2+C2'*C1)
      -2#(covt'*C1'*C2*covt)#covt'*(C1'*C1) )/(covt'*C1'*C1*covt)##2;
rt=(covt'*C1'*C3*covt)/(covt'*C1'*C1*covt);
drt=( (covt'*C1'*C1*covt)#covt'*(C1'*C3+C3'*C1)
      -2#(covt'*C1'*C3*covt)#covt'*(C1'*C1) )/(covt'*C1'*C1*covt)##2;
rct=rt/ct;

```

```

drct=(ct#drt-rt#dct)/ct##2;
sig2t=covt[1];dsig2t=I(n2)[1,];

print sig2m cm rm sig2t ct rt;

g1=QX*mut;
free g2 Gt2;
do ii=1 to n;
do jj=ii to n;
g2=g2//( sigt[ii,jj]-sig2t#ct##(ii-1)#rt##(jj-ii) );
Gt2=Gt2//( I(n)[ii,]@I(n)[jj,]-dsig2t#ct##(ii-1)#rt##(jj-ii)
          -(ii-1)#sig2t#ct##(ii-2)#rt##(jj-ii)#dct
          -(jj-ii)#sig2t#ct##(ii-1)#rt##(jj-ii-1)#drt );
end;end;
Gt2=Gt2*(QQ*(-I(n)@mut-mut@I(n))||I(n2));

Gt=Gt1//Gt2;
g=g1//g2;

if i=1 & j=1 then do; gb=g1//g2; Gtb=Gt1//Gt2; end;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
*print i j t m g;
end;
print i j t m g;
verskil1=sqrt((m-t)'*(m-t));
end;

Gm=Gm1//Gm2;Gmm=Gm1;GmV=Gm2;
nrg=nrow(g);
eigv=eigval(Gm*V*Gm');
Wald= gb'*ginv(Gtb*V*Gtb')*gb;
vg=sum(echelon(echelon(Gm*V*Gm')));
prob=1-probchi(Wald,vg); print nrg vg Wald prob;
GV=Gm*V;

covt=t2-QQ*colvec(mut*mut');
sig1t=Cd*covt;
rho=((1/sqrt(sig1t))@(1/sqrt(sig1t)))#covt;
rhom=shape(rho,n,n);
par=mut//colvec(Sig);
print t[format=6.2] m[format=6.2];
print rhom[format=6.2] sig[format=6.2];

X=J(n,1,1)||cusum(J(n,1,1))||(cusum(J(n,1,1))##2)|| (data[,2]#100)|| (data[,3]);
mupar=ginv(X'*X)*X'*mut;
print mupar;

```

## A.2.5 TOEPLITZ.SAS

\*Toeplitz(2) with heteroskedasticity;

```

proc iml worksizes=50;
options pagesize=100;
reset nolog;

```

```

DATA={
0.95901    0.96011    9.70167    1999,
1.00000    1.00000    8.38620    2000,
1.03510    1.02735    5.70660    2001,
1.07339    1.06504    2.30092    2002,
1.17118    1.09827    3.81404    2003,
1.27555    1.15169    8.12716    2004,
1.40594    1.20887    4.66702    2005,
1.59163    1.27320    3.23816    2006,
1.85095    1.33810    0.88436    2007,
2.03939    1.37908    -2.39917    2008};

y=DATA[,1]#100;
*print DATA;
*print y[format=6.2] (data[,1])[format=6.2] (data[,2])[format=6.2] (data[,3])[format=6.2];
n=nrow(y); n2=n##2; nc=ncol(y);print n nc;
const=0.2;
x=J(n,1,1)||cusum(J(n,1,1))||(cusum(J(n,1,1))##2)*||(data[25:34,2]#100)||data[25:34,3];
PX=X*inv(X'*X)*X';
QX=I(n)-PX;
QQ=I(n2)-(I(n)-PX)@(I(n)-PX);

do i=1 to n;
do j=1 to n-i+1;
D=D/I(n2)[i+(n+1)#(j-1),];
end;end;
*print D;

z=cusum(J(n,1,1));
z1=z[1:n];
z2=z[1:n-1];
PZ1=z1*inv(z1'*z1)*z1';
PZ2=z2*inv(z2'*z2)*z2';
PZ=block(PZ1,PZ2,J(n#(n+1)/2-2#n+1,n#(n+1)/2-2#n+1,0));
QZ=I(n#(n+1)/2)-PZ;
*print PZ[format=4.2] QZ[format=4.2];

K=j(n2,n2,0);
do i=1 to n;
do j=1 to n;
H=j(n,n,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

ym=y[,+]/nc;
yy=colvec((y*y')')/nc;
Sig=J(n,n,0);
t=ym//yy;
verskil1=1;
i=0;
t0=t;
do while (verskil1>1e-2);
i=i+1;
m=t;
t=t0;

```



```

mu=m[1:n]; m2=m[n+1:(n+1)#n]; * nu=PX*mu;

*if i=1 then covm=colvec(I(n)+const);
covm=m2-QQ*colvec(mu*mu');

dsig=PZ*D*covm; print dsig;
sig[n,n]=dsig[n];
do ii=1 to n-1;
sig[ii,ii]=dsig[ii];
sig[ii,ii+1]=dsig[n+ii];
sig[ii+1,ii]=dsig[n+ii];
end;

print sig;
print "covm" (shape(covm,n,n));

Phi=(Sig@mu+mu@Sig);
V= ( (Sig||Phi')/(Phi||((I(n2)+K)*(Sig@Sig+Sig@(mu*mu')+(mu*mu')@Sig))) )/nc;
Gm1=QX||J(n,n2,0);
sigm=shape(covm,n,n);

Gm2=QZ*D;
Gm2=Gm2*(QQ*(-I(n)@mu-mu@I(n))||I(n2));

Gm=Gm1//Gm2;
verskil=1;
j=0;
do while (verskil>1e-2);
tv=t;
mut=t[1:n]; t2=t[n+1:(n+1)#n];
j=j+1;

covt=t2-QQ*colvec(mut*mut');
Gt1=QX||J(n,n2,0);

*sigt=shape(covt,n,n);*print sigt;

g1=QX*mut;
g2=QZ*D*covt;
Gt2=QZ*D;
Gt2=Gt2*(QQ*(-I(n)@mut-mut@I(n))||I(n2));

Gt=Gt1//Gt2;
g=g1//g2;
if i=1 & j=1 then do; gb=g1//g2; Gtb=Gt1//Gt2; end;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
verskil=sqrt((t-tv)'*(t-tv));
*print i j t m g;
end;
verskil1=sqrt((m-t)'*(m-t));
print i j t m g;
end;

*Wald;
Gm=Gm1//Gm2;
nrg=nrow(g);

```

```

eigv=eigval(Gm*V*Gm');
Wald= gb'*ginv(Gtb*V*Gtb')*gb;
vg=sum(echelon(echelon(Gm*V*Gm')));
prob=1-probchi(Wald,vg);
print eigv nrg Wald vg prob;

*Parameters;
covt=t2-QQ*colvec(mut*mut');
sig1t=(D*covt)[1:n,];
sig2t=(D*covt)[n+1:2#n-1,];
rho=((1/sqrt(sig1t))@(1/sqrt(sig2t)))#covt;
rhom=shape(rho,n,n);
vecsig=vecdiag(Sig);
par=mu//colvec(Sig);
print rhom[format=6.2];
print vecsig[format=6.2] sig[format=6.2];
print t[format=6.2] m[format=6.2] par;

x=J(n,1,1)||cusum(J(n,1,1))||(cusum(J(n,1,1))##2);
mupar=ginv(X'*X)*X'*mut;
print mupar;
z=cusum(J(n,1,1));
z1=z[1:n];
alpha1=ginv(z1'*z1)*z1'*sig1t;
print sig1t alpha1;
z2=z[1:n-1];
alpha2=ginv(z2'*z2)*z2'*sig2t;
print sig2t alpha2;

```

## A.3 SAS Programs for Chapter 5

### A.3.1 MULTIV.SAS

```

*Introductory example;

proc iml worksizes=50;
options pagesize=100 linesize=80;
reset nolog;

n1=10; p=2; p2=p#p;
ym1={2, 3};
A1={5 4,
    4 8}#(n1-1)/n1;
yy1=colvec((A1+ym1*ym1')');

*D selects from vec(Sigma) the elements on the diagonals of Sigma;
%macro mat(D,p);
D=J(p,p##2,0);
kr=0;
do i=1 to 1;
do j=1 to p-i+1;
kr=kr+1;
D[kr,(p+1)#(j-1)+i]=1;
end;end;
%mend mat;
%mat (D,p);*print D[format=2.0];

```

```

K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

t=ym1//yy1;

diff1=1;
i=0;
t0=t;
do while (diff1>0.0001);
i=i+1;
m=t;
t=t0;

mu1m=m[1:p]; M21=m[p+1:p#(p+1)];
dmu1m=I(p)||J(p,p2,0);

cov1m=m21-colvec(mu1m*mu1m');
dcov1m=(-I(p)@mu1m-mu1m@I(p))||I(p2);

Sig1=shape(cov1m,p,p);
Phi1=(Sig1@mu1m+mu1m@Sig1);
V= ( (Sig1||Phi1')//
      (Phi1||((I(p2)+K)*(Sig1@Sig1+Sig1@(mu1m*mu1m')+
      (mu1m*mu1m')@Sig1))) )/n1;

G1m=(dcov1m[1,])-2#mu1m[1]#dmu1m[1,];
G2m=(dcov1m[4,])-2#mu1m[2]#dmu1m[2,];

Gm=G1m//G2m;

diff=1;
j=0;
do while (diff>0.0001);
tv=t;
j=j+1;

mu1t=t[1:p]; t21=t[p+1:p#(p+1)];
dmu1t=I(p)||J(p,p2,0);

cov1t=t21-colvec(mu1t*mu1t');
dcov1t=(-I(p)@mu1t-mu1t@I(p))||I(p2);

g1=cov1t[1,]-mu1t[1]#mu1t[1];
g2=cov1t[4,]-mu1t[2]#mu1t[2];

G1t=(dcov1t[1,])-2#mu1t[1]#dmu1t[1,];
G2t=(dcov1t[4,])-2#mu1t[2]#dmu1t[2,];

Gt=G1t//G2t;
g=g1//g2;

```

```

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
diff=sqrt((t-tv)'*(t-tv));
print i j t m g ;
end;
diff1=sqrt((m-t)'*(m-t));
end;

sig1m=sqrt(D*cov1m);

cov1m=m21-colvec(mu1m*mu1m');
rho1=((1/sig1m)@(1/sig1m))#cov1m;
rhom1=shape(rho1,p,p);
cov1m=shape(cov1m,p,p);

print mu1m cov1m rhom1;

```

### A.3.2 LMVNORMAL.SAS

```

*Two large independent multivariate normal samples;

proc iml worksizes=50;
options pagesize=100 linesize=80;
reset nolog;

n1=45; n2=55; p=2; p2=p#p;
ym1={204.4, 556.6};
A1={13825.3 23823.4,
     23823.4 73107.4}#(n1-1)/n1;
ym2={130.0, 355.0};
A2={8632.0 19616.7,
     19616.7 55964.5}#(n2-1)/n2;
yy1=colvec((A1+ym1*ym1')');
yy2=colvec((A2+ym2*ym2')');

*D selects from vec(Sigma) the elements on the diagonals of Sigma;
%macro mat(D,p);
D=J(p,p##2,0);
kr=0;
do i=1 to 1;
do j=1 to p-i+1;
kr=kr+1;
D[kr,(p+1)#(j-1)+i]=1;
end;end;
%mend mat;
%mat (D,p);*print D[format=2.0];

K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

t=ym1//yy1//ym2//yy2;

```

```

diff1=1;
i=0;
t0=t;
do while (diff1>0.00000001);
i=i+1;
m=t;
t=t0;

mu1m=m[1:p];
mu2m=m[p#(p+1)+1:p#(p+1)+p];
M21=m[p+1:p#(p+1)];
M22=m[p#(p+1)+p+1:2#p#(p+1)];

dmu1m=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2m=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);

cov1m=m21-colvec(mu1m*mu1m');
cov2m=m22-colvec(mu2m*mu2m');

dcov1m=(-I(p)@mu1m-mu1m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov2m=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2m-mu2m@I(p))||I(p2);

cm=cov2m[1]/cov1m[1];
dcm=(-cov2m[1]/cov1m[1]##2#I(p2)[1,])||(1/cov1m[1]#I(p2)[1,]);
dcm=dcm*(dcov1m//dcov2m);

Sig1=shape(cov1m,p,p);
Sig2=shape(cov2m,p,p);

Phi1=(Sig1@mu1m+mu1m@Sig1);
Phi2=(Sig2@mu2m+mu2m@Sig2);

V1=( (Sig1||Phi1')//
      (Phi1||((I(p2)+K)*(Sig1@Sig1+Sig1@(mu1m*mu1m')+
      (mu1m*mu1m')@Sig1))) )/n1;
V2=( (Sig2||Phi2')//
      (Phi2||((I(p2)+K)*(Sig2@Sig2+Sig2@(mu2m*mu2m')+
      (mu2m*mu2m')@Sig2))) )/n2;
V=block(V1,V2);

G1m=dmu1m-dmu2m;
G2m=dcov1m-dcov2m;

G1cm=dmu2m-sqrt(cm)#dmu1m-(mu1m*dcm)/(2#sqrt(cm));
G2cm=dcov2m-cm#dcov1m-cov1m*dcm;

dm=mu2m[1]/mu1m[1];
ddm=(dmu2m[1,]#mu1m[1]-dmu1m[1,]#mu2m[1])/(mu1m[1]##2);
G1dm=dmu2m-mu1m*ddm-dmu1m*dm;

drm=(I(p2)[2,]/sqrt(cov1m[1]#cov1m[p+2])-I(p2)[1,]#cov1m[2]/
      (2#sqrt(cov1m[1]##3#cov1m[p+2]))-I(p2)[p+2,]#cov1m[2]/
      (2#sqrt(cov1m[1]#cov1m[p+2]##3))||J(1,p2,0) )
      -( J(1,p2,0) || (I(p2)[2,]/sqrt(cov2m[1]#cov2m[p+2])-
      I(p2)[1,]#cov2m[2]/(2#sqrt(cov2m[1]##3#cov2m[p+2]))-
      I(p2)[p+2,]#cov2m[2]/(2#sqrt(cov2m[1]#cov2m[p+2]##3) ) ) );
Grm=drm*(dcov1m//dcov2m);

```

```

Gm=Grm;
*Gm=G1dm//G2cm;
*Gm=G1cm//G2cm;
*Gm=G1m//G2m;

diff=1;
j=0;
do while (diff>0.00000001);
tv=t;
j=j+1;
t1=t[1:2#p]; t2=t[2#p+1:2#p#(p+1)];

mu1t=t[1:p]; t21=t[p+1:p#(p+1)];
mu2t=t[p#(p+1)+1:p#(p+1)+p]; t22=t[p#(p+1)+p+1:2#p#(p+1)];

dmu1t=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2t=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);

cov1t=t21-colvec(mu1t*mu1t');
cov2t=t22-colvec(mu2t*mu2t');

dcov1t=(-I(p)@mu1t-mu1t@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov2t=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2t-mu2t@I(p))||I(p2);

ct=cov2t[1]/cov1t[1];
dct=( -cov2t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );
dct=dct*(dcov1t//dcov2t);
g1c=mu2t-sqrt(ct)#mu1t;
g2c=cov2t-ct#cov1t;
G1ct=dmu2t-sqrt(ct)#dmu1t-(mu1t*dct)/(2#sqrt(ct));
G2ct=dcov2t-ct#dcov1t-cov1t*dct;

dt=mu2t[1]/mu1t[1];
ddt=(dmu2t[1,]#mu1t[1]-dmu1t[1,]#mu2t[1])/(mu1t[1]##2);
g1d=mu2t-dt#mu1t;
G1dt=dmu2t-mu1t*ddt-dmu1t*dt;

gr=cov1t[2]/sqrt(cov1t[1]#cov1t[p+2])-
cov2t[2]/sqrt(cov2t[1]#cov2t[p+2]);
drt=(I(p2)[2,]/sqrt(cov1t[1]#cov1t[p+2])-
I(p2)[1,]#cov1t[2]/(2#sqrt(cov1t[1]##3#cov1t[p+2]))-
I(p2)[p+2,]#cov1t[2]/(2#sqrt(cov1t[1]#cov1t[p+2]##3))
||J(1,p2,0) )
-( J(1,p2,0) || (I(p2)[2,]/sqrt(cov2t[1]#cov2t[p+2])-
I(p2)[1,]#cov2t[2]/(2#sqrt(cov2t[1]##3#cov2t[p+2]))-
I(p2)[p+2,]#cov2t[2]/(2#sqrt(cov2t[1]#cov2t[p+2]##3) ) ) );
Grt=drt*(dcov1t//dcov2t);

g1=mu1t-mu2t;
g2=cov1t-cov2t;

G1t=dmu1t-dmu2t;
G2t=dcov1t-dcov2t;

*Gt=G1dt//G2ct;
*Gt=G1ct//G2ct;
*Gt=G1t//G2t;

```

```

Gt=Grt;
*g=g1d//g2c;
*g=g1c//g2c;
*g=g1//g2;
g=gr;

if i=1 & j=1 then do;
g0=g; Gt0=Gt;
Sig10=shape(cov1t,p,p);
Sig20=shape(cov2t,p,p);
Phi10=(Sig10@mu1t+mu1t@Sig10);
Phi20=(Sig20@mu2t+mu2t@Sig20);
V10= ( (Sig10||Phi10')//
        (Phi10||((I(p2)+K)*(Sig10@Sig10+Sig10@(mu1t*mu1t')+
        (mu1t*mu1t')@Sig10))) )/n1;
V20= ( (Sig20||Phi20')//
        (Phi20||((I(p2)+K)*(Sig20@Sig20+Sig20@(mu2t*mu2t')+
        (mu2t*mu2t')@Sig20))) )/n2;
V0=block(V10,V20);
end;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
diff=sqrt((t-tv)'*(t-tv));
print i j t m g ;
end;
diff1=sqrt((m-t)'*(m-t));
end;

Wald=(g0)'*ginv(Gt0*V0*Gt0')*g0;
vg=sum(echelon(echelon(Gm*V*Gm')'));
prob=1-probchi(Wald,vg);
print Wald vg prob;

sig1m=sqrt(D*cov1m);
sig2m=sqrt(D*cov2m);

cov1m=m21-colvec(mu1m*mu1m');
rho1=((1/sig1m)@(1/sig1m))#cov1m;
rhom1=shape(rho1,p,p);
cov1m=shape(cov1m,p,p);

cov2m=m22-colvec(mu2m*mu2m');
rho2=((1/sig2m)@(1/sig2m))#cov2m;
rhom2=shape(rho2,p,p);
cov2m=shape(cov2m,p,p);

print cm dm;
print mu1m mu2m;
print cov1m cov2m;
print rhom1 rhom2;

*Calculation of variance of par's;

Acov=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V;
var=vecdiag(Acov); *print var;

semu1=sqrt(var[1:p]);

```

```

semu2=sqrt(var[p#(p+1)+1:p#(p+1)+p]);

secov1=sqrt(shape(var[p+1: p#(p+1)],p,p));
secov2=sqrt(shape(var[p#(p+1)+p+1: 2#p#(p+1)],p,p));

print semu1 semu2;
print secov1 secov2;

```

### A.3.3 SMVNORMAL1.SAS

```

proc iml worksizes=50;
options pagesize=100;
reset nolog;
Y1={4.0 4.2 4.3 4.2 4.6 3.1 3.7 4.3 4.6,
     4.0 4.3 4.2 4.4 4.4 3.6 3.9 4.2 4.6,
     4.1 3.7 4.3 4.6 5.3 4.9 3.9 4.4 4.4,
     3.6 3.7 4.3 4.9 5.6 5.2 4.8 5.2 4.6,
     3.6 4.8 4.5 5.3 5.9 5.3 5.2 5.6 5.4,
     3.8 5.0 5.8 5.6 5.9 4.2 5.4 5.4 5.9,
     3.1 5.2 5.4 4.9 5.3 4.1 4.2 4.7 5.6};

Y2={3.4 3.0 3.0 3.1 3.8 3.0 3.3 4.2 4.1 4.5,
     3.4 3.2 3.1 3.2 3.9 3.6 3.3 4.0 4.2 4.4,
     3.5 3.0 3.2 3.2 4.0 3.2 3.3 4.2 4.3 4.3,
     3.1 3.0 3.0 3.2 2.9 3.1 3.4 4.1 4.3 4.5,
     3.1 3.1 3.3 3.3 3.5 3.0 3.6 4.2 4.2 5.3,
     3.7 3.2 3.0 3.1 3.5 3.0 3.1 4.0 4.0 4.4,
     3.3 3.1 3.0 3.1 3.4 3.0 3.1 4.0 4.2 4.4};

Y3={3.2 3.3 3.1 3.6 4.5 3.7 3.5 3.9,
     3.3 3.4 3.3 3.4 4.5 4.0 3.9 4.0,
     3.8 3.4 3.2 3.5 5.4 4.4 5.8 4.1,
     3.8 3.7 3.1 4.6 5.7 4.2 5.4 5.0,
     4.4 3.7 3.2 4.9 4.9 4.6 4.9 5.4,
     4.2 3.6 3.1 5.2 4.0 4.8 5.3 4.4,
     3.7 3.7 3.1 4.4 4.0 5.4 5.6 3.9};

Y4={3.1 3.3 3.5 3.4 3.7 4.0 4.2 4.1 3.5,
     3.5 3.2 3.9 3.4 3.8 4.6 3.9 4.1 3.6,
     3.5 3.6 4.7 3.5 4.2 4.8 4.5 3.7 3.6,
     3.2 3.7 4.3 3.3 4.3 4.9 4.7 4.0 4.2,
     3.0 3.7 3.9 3.4 3.6 5.4 3.9 4.1 4.8,
     3.0 4.2 3.4 3.2 3.8 5.6 3.8 4.6 4.9,
     3.2 4.4 3.5 3.4 3.7 4.8 3.7 4.7 5.0};

p=nrow(Y1); p2=p##2; n1=ncol(Y1);n2=ncol(Y2);n3=ncol(Y3);n4=ncol(Y4);
const=0.5;

*Ci selects from vec(Sigma) submatrices from Sigma;
C1=I(p2)[1:p-1,];
do i=2 to p-1;
C1=C1//I(p2)[p#(i-1)+i:p#i-1,];
end;
C3=I(p2)[1+p+1:p-1+p+1,];
do i=2 to p-1;
C3=C3//I(p2)[p#(i-1)+i+p+1:p#i-1+p+1,];

```



```

end;
*print C1[format=2.0] C2[format=2.0] C3[format=2.0];

%macro mat(D,p); *D selects from vec(Sigma) the elements on the diagonals of Sigma;
D=J(p,p##2,0);
kr=0;
do i=1 to 1;
do j=1 to p-i+1;
kr=kr+1;
D[kr,(p+1)#(j-1)+i]=1;
end;end;
%mend mat;
%mat (D,p);*print D[format=2.0];

*K is used in the expression for the cov(Vec(Sigma));*****;
K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

*Canonical statistics;
ym1=Y1[,+]/n1;
yy1=colvec((Y1*Y1')')/n1;
ym2=Y2[,+]/n2;
yy2=colvec((Y2*Y2')')/n2;
ym3=Y3[,+]/n3;
yy3=colvec((Y3*Y3')')/n3;
ym4=Y4[,+]/n4;
yy4=colvec((Y4*Y4')')/n4;

t=ym1//yy1//ym2//yy2//ym3//yy3//ym4//yy4;
diff1=1;
i=0;
t0=t;
do while (diff1>1e-8);
i=i+1;
m=t;
t=t0;

mu1m=m[1:p]; M21=m[p+1:p#(p+1)];
mu2m=m[p#(p+1)+1:p#(p+1)+p]; M22=m[p#(p+1)+p+1:2#p#(p+1)];
mu3m=m[2#p#(p+1)+1:2#p#(p+1)+p]; M23=m[2#p#(p+1)+p+1:3#p#(p+1)];
mu4m=m[3#p#(p+1)+1:3#p#(p+1)+p]; M24=m[3#p#(p+1)+p+1:4#p#(p+1)];
mum=(mu1m+mu2m+mu3m+mu4m)/4;

dmu1m=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2m=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu3m=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu4m=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);
dmum=( I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0) )/4;

d2m=(mu2m[1]-mu2m[2])/(mum[1]-mum[2]);

```

```

d3m=(mu3m[1]-mu3m[2])/(mum[1]-mum[2]);
d4m=(mu4m[1]-mu4m[2])/(mum[1]-mum[2]);

a2m=mu2m[1]-d2m#mum[1];
a3m=mu3m[1]-d3m#mum[1];
a4m=mu4m[1]-d4m#mum[1];

cov1m=m21-colvec(mu1m*mu1m');
cov2m=m22-colvec(mu2m*mu2m');
cov3m=m23-colvec(mu3m*mu3m');
cov4m=m24-colvec(mu4m*mu4m');

dcov1m=(-I(p)@mu1m-mu1m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov2m=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2m-mu2m@I(p))||I(p2)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov3m=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||(-I(p)@mu3m-mu3m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov4m=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu4m-mu4m@I(p))||I(p2);

if i=1 then cov1m=colvec(I(p)+const);
if i=1 then cov2m=colvec(I(p)+const);
if i=1 then cov3m=colvec(I(p)+const);
if i=1 then cov4m=colvec(I(p)+const);

c2m=cov2m[1]/cov1m[1];
c3m=cov3m[1]/cov1m[1];
c4m=cov4m[1]/cov1m[1];

dc2m=( -cov2m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );
dc3m=( -cov3m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );
dc4m=( -cov4m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );

Sig1=shape(cov1m,p,p);
Sig2=shape(cov2m,p,p);
Sig3=shape(cov3m,p,p);
Sig4=shape(cov4m,p,p);

Phi1=(Sig1@mu1m+mu1m@Sig1);
Phi2=(Sig2@mu2m+mu2m@Sig2);
Phi3=(Sig3@mu3m+mu3m@Sig3);
Phi4=(Sig4@mu4m+mu4m@Sig4);

*mu1m=mum;
*mu2m=mum;
*mu3m=mum;
*mu4m=mum;

V1= ( (Sig1||Phi1')//(Phi1||((I(p2)+K)*(Sig1@Sig1+Sig1@(mu1m*mu1m')
+(mu1m*mu1m')@Sig1))) )/n1;
V2= ( (Sig2||Phi2')//(Phi2||((I(p2)+K)*(Sig2@Sig2+Sig2@(mu2m*mu2m')
+(mu2m*mu2m')@Sig2))) )/n2;
V3= ( (Sig3||Phi3')//(Phi3||((I(p2)+K)*(Sig3@Sig3+Sig3@(mu3m*mu3m')
+(mu3m*mu3m')@Sig3))) )/n3;
V4= ( (Sig4||Phi4')//(Phi4||((I(p2)+K)*(Sig4@Sig4+Sig4@(mu4m*mu4m')
+(mu4m*mu4m')@Sig4))) )/n4;

```

```

V=block(V1,V2,V3,V4);

Gc2m=(J(p2,p2,0)||I(p2))-c2m#( I(p2)||J(p2,p2,0))-I(p2)*cov1m*dc2m;
Gc2m=Gc2m*(dcov1m//dcov2m);
Gc3m=(J(p2,p2,0)||I(p2))-c3m#( I(p2)||J(p2,p2,0))-I(p2)*cov1m*dc3m;
Gc3m=Gc3m*(dcov1m//dcov3m);
Gc4m=(J(p2,p2,0)||I(p2))-c4m#( I(p2)||J(p2,p2,0))-I(p2)*cov1m*dc4m;
Gc4m=Gc4m*(dcov1m//dcov4m);

dd2m=( (mum[1]-mum[2])#(dmu2m[1,]-dmu2m[2,])-(mu2m[1]-mu2m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;
dd3m=( (mum[1]-mum[2])#(dmu3m[1,]-dmu3m[2,])-(mu3m[1]-mu3m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;
dd4m=( (mum[1]-mum[2])#(dmu4m[1,]-dmu4m[2,])-(mu4m[1]-mu4m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;

da2m=dmu2m[1,]-d2m#dmum[1,]-mum[1]#dd2m;
da3m=dmu3m[1,]-d3m#dmum[1,]-mum[1]#dd3m;
da4m=dmu4m[1,]-d4m#dmum[1,]-mum[1]#dd4m;

*Gd2m=dmu2m-dmum;
*Gd3m=dmu3m-dmum;
*Gd4m=dmu4m-dmum;
Gd2m=dmu2m-d2m#dmum-mum*dd2m-J(p,1,1)*da2m;
Gd3m=dmu3m-d3m#dmum-mum*dd3m-J(p,1,1)*da3m;
Gd4m=dmu4m-d4m#dmum-mum*dd4m-J(p,1,1)*da4m;

Gm=Gc2m//Gc3m//Gc4m//Gd2m//Gd3m//Gd4m;

diff=1;
j=0;
do while (diff>1e-8);
tv=t;
j=j+1;

mu1t=t[1:p];          t21=t[p+1:p#(p+1)];
mu2t=t[p#(p+1)+1:p#(p+1)+p];    t22=t[p#(p+1)+p+1:2#p#(p+1)];
mu3t=t[2#p#(p+1)+1:2#p#(p+1)+p]; t23=t[2#p#(p+1)+p+1:3#p#(p+1)];
mu4t=t[3#p#(p+1)+1:3#p#(p+1)+p]; t24=t[3#p#(p+1)+p+1:4#p#(p+1)];
mut=(mu1t+mu2t+mu3t+mu4t)/4;

dmu1t=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2t=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu3t=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu4t=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);
dmut=( I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0) )/4;

cov1t=t21-colvec(mu1t*mu1t');
cov2t=t22-colvec(mu2t*mu2t');
cov3t=t23-colvec(mu3t*mu3t');
cov4t=t24-colvec(mu4t*mu4t');

dcov1t=(-I(p)#mu1t-mu1t#I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov2t=J(p2,p,0)||J(p2,p2,0)||(-I(p)#mu2t-mu2t#I(p))||I(p2)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);

```

```

dcov3t=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
      ||(-I(p)@mu3t-mu3t@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov4t=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
      ||J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu4t-mu4t@I(p))||I(p2);

c2t=cov2t[1]/cov1t[1];
c3t=cov3t[1]/cov1t[1];
c4t=cov4t[1]/cov1t[1];

dc2t=( -cov2t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );
dc3t=( -cov3t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );
dc4t=( -cov4t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );

gc2=cov2t-c2t#cov1t;
gc3=cov3t-c3t#cov1t;
gc4=cov4t-c4t#cov1t;

Gc2t=(J(p2,p2,0)||I(p2))-c2t#( I(p2)||J(p2,p2,0))-I(p2)*cov1t*dc2t;
Gc2t=Gc2t*(dcov1t//dcov2t);
Gc3t=(J(p2,p2,0)||I(p2))-c3t#( I(p2)||J(p2,p2,0))-I(p2)*cov1t*dc3t;
Gc3t=Gc3t*(dcov1t//dcov3t);
Gc4t=(J(p2,p2,0)||I(p2))-c4t#( I(p2)||J(p2,p2,0))-I(p2)*cov1t*dc4t;
Gc4t=Gc4t*(dcov1t//dcov4t);

d2t=(mu2t[1]-mu2t[2])/(mut[1]-mut[2]);
d3t=(mu3t[1]-mu3t[2])/(mut[1]-mut[2]);
d4t=(mu4t[1]-mu4t[2])/(mut[1]-mut[2]);

dd2t=( (mut[1]-mut[2])#(dmu2t[1,]-dmu2t[2,])-(mu2t[1]-mu2t[2])#(dmut[1,]-dmut[2,]) )
      /(mut[1]-mut[2])##2;
dd3t=( (mut[1]-mut[2])#(dmu3t[1,]-dmu3t[2,])-(mu3t[1]-mu3t[2])#(dmut[1,]-dmut[2,]) )
      /(mut[1]-mut[2])##2;
dd4t=( (mut[1]-mut[2])#(dmu4t[1,]-dmu4t[2,])-(mu4t[1]-mu4t[2])#(dmut[1,]-dmut[2,]) )
      /(mut[1]-mut[2])##2;

a2t=mu2t[1]-d2t#mut[1];
a3t=mu3t[1]-d3t#mut[1];
a4t=mu4t[1]-d4t#mut[1];

da2t=dmu2t[1,]-d2t#dmut[1,]-mut[1]#dd2t;
da3t=dmu3t[1,]-d3t#dmut[1,]-mut[1]#dd3t;
da4t=dmu4t[1,]-d4t#dmut[1,]-mut[1]#dd4t;

*gd2=mu2t-mut;
*gd3=mu3t-mut;
*gd4=mu4t-mut;
gd2=mu2t-d2t#mut-a2t#J(p,1,1);
gd3=mu3t-d3t#mut-a3t#J(p,1,1);
gd4=mu4t-d4t#mut-a4t#J(p,1,1);

*Gd2t=dmu2t-dmut;
*Gd3t=dmu3t-dmut;
*Gd4t=dmu4t-dmut;
Gd2t=dmu2t-d2t#dmut-mut*dd2t-J(p,1,1)*da2t;
Gd3t=dmu3t-d3t#dmut-mut*dd3t-J(p,1,1)*da3t;
Gd4t=dmu4t-d4t#dmut-mut*dd4t-J(p,1,1)*da4t;

```

```

Gt=Gc2t//Gc3t//Gc4t//Gd2t//Gd3t//Gd4t;
g=gc2//gc3//gc4//gd2//gd3//gd4;

if i=1 & j=1 then do;
g0=g; Gt0=Gt;
Sig10=shape(cov1t,p,p);
Sig20=shape(cov2t,p,p);
Sig30=shape(cov3t,p,p);
Sig40=shape(cov4t,p,p);
Phi10=(Sig10@mult+mult@Sig10);
Phi20=(Sig20@mu2t+mu2t@Sig20);
Phi30=(Sig30@mu3t+mu3t@Sig30);
Phi40=(Sig40@mu4t+mu4t@Sig40);
V10= ( (Sig10||Phi10')/(Phi10||((I(p2)+K)*(Sig10@Sig10+Sig10@(mult*mult')
+ (mult*mult')@Sig10))) )/n1;
V20= ( (Sig20||Phi20')/(Phi20||((I(p2)+K)*(Sig20@Sig20+Sig20@(mu2t*mu2t')
+ (mu2t*mu2t')@Sig20))) )/n2;
V30= ( (Sig30||Phi30')/(Phi30||((I(p2)+K)*(Sig30@Sig30+Sig30@(mu3t*mu3t')
+ (mu3t*mu3t')@Sig30))) )/n3;
V40= ( (Sig40||Phi40')/(Phi40||((I(p2)+K)*(Sig40@Sig40+Sig40@(mu4t*mu4t')
+ (mu4t*mu4t')@Sig40))) )/n4;
V0=block(V10,V20,V30,V40);
end;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
diff=sqrt((t-tv)'*(t-tv));
*print i j e1m e2m e3m e4m t m g;
end;
diff1=sqrt((m-t)'*(m-t));
end;

Wald=(g0)'*ginv(Gt0*V0*Gt0')*g0;
vg=sum(echelon(echelon(Gm*V*Gm')'));
prob=1-probchi(Wald,vg);
print Wald vg prob;

sig1m=sqrt(D*cov1m);
sig2m=sqrt(D*cov2m);
sig3m=sqrt(D*cov3m);
sig4m=sqrt(D*cov4m);

cov1m=m21-colvec(mu1m*mu1m');
rho1=((1/sig1m)@(1/sig1m))#cov1m;
rhom1=shape(rho1,p,p);
cov1m=shape(cov1m,p,p);

cov2m=m22-colvec(mu2m*mu2m');
rho2=((1/sig2m)@(1/sig2m))#cov2m;
rhom2=shape(rho2,p,p);
cov2m=shape(cov2m,p,p);

cov3m=m23-colvec(mu3m*mu3m');
rho3=((1/sig3m)@(1/sig3m))#cov3m;
rhom3=shape(rho3,p,p);
cov3m=shape(cov3m,p,p);

```

```

cov4m=m24-colvec(mu4m*mu4m');
rho4=((1/sig4m)@(1/sig4m))#cov4m;
rho4=shape(rho4,p,p);
cov4m=shape(cov4m,p,p);

print a2m a3m a4m;
print d2m d3m d4m;
print c2m c3m c4m e1m;
print mu1m mu2m mu3m mu4m mum;
print cov1m cov2m cov3m cov4m;
print rhom1 rhom2 rhom3 rhom4;

*Calculation of variance of par's;

Acov=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V;
var=vecdiag(Acov); *print var;

semu1=sqrt(var[1:p]);
semu2=sqrt(var[p#(p+1)+1:p#(p+1)+p]);
semu3=sqrt(var[2#p#(p+1)+1:2#p#(p+1)+p]);
semu4=sqrt(var[3#p#(p+1)+1:3#p#(p+1)+p]);

secov1=sqrt(shape(var[p+1: p#(p+1)],p,p));
secov2=sqrt(shape(var[p#(p+1)+p+1: 2#p#(p+1)],p,p));
secov3=sqrt(shape(var[2#p#(p+1)+p+1: 3#p#(p+1)],p,p));
secov4=sqrt(shape(var[3#p#(p+1)+p+1: 4#p#(p+1)],p,p));

print semu1 semu2 semu3 semu4;
print secov1 secov2 secov3 secov4;

```

### A.3.4 SMVNORMAL2.SAS

```

proc iml worksizes=50;
options pagesize=100;
reset nolog;
Y1={4.0 4.2 4.3 4.2 4.6 3.1 3.7 4.3 4.6,
     4.0 4.3 4.2 4.4 4.4 3.6 3.9 4.2 4.6,
     4.1 3.7 4.3 4.6 5.3 4.9 3.9 4.4 4.4,
     3.6 3.7 4.3 4.9 5.6 5.2 4.8 5.2 4.6,
     3.6 4.8 4.5 5.3 5.9 5.3 5.2 5.6 5.4,
     3.8 5.0 5.8 5.6 5.9 4.2 5.4 5.4 5.9,
     3.1 5.2 5.4 4.9 5.3 4.1 4.2 4.7 5.6};

Y2={3.4 3.0 3.0 3.1 3.8 3.0 3.3 4.2 4.1 4.5,
     3.4 3.2 3.1 3.2 3.9 3.6 3.3 4.0 4.2 4.4,
     3.5 3.0 3.2 3.2 4.0 3.2 3.3 4.2 4.3 4.3,
     3.1 3.0 3.0 3.2 2.9 3.1 3.4 4.1 4.3 4.5,
     3.1 3.1 3.3 3.3 3.5 3.0 3.6 4.2 4.2 5.3,
     3.7 3.2 3.0 3.1 3.5 3.0 3.1 4.0 4.0 4.4,
     3.3 3.1 3.0 3.1 3.4 3.0 3.1 4.0 4.2 4.4};

Y3={3.2 3.3 3.1 3.6 4.5 3.7 3.5 3.9,
     3.3 3.4 3.3 3.4 4.5 4.0 3.9 4.0,
     3.8 3.4 3.2 3.5 5.4 4.4 5.8 4.1,
     3.8 3.7 3.1 4.6 5.7 4.2 5.4 5.0,
     4.4 3.7 3.2 4.9 4.9 4.6 4.9 5.4,

```

```

4.2 3.6 3.1 5.2 4.0 4.8 5.3 4.4,
3.7 3.7 3.1 4.4 4.0 5.4 5.6 3.9};

Y4={3.1 3.3 3.5 3.4 3.7 4.0 4.2 4.1 3.5,
3.5 3.2 3.9 3.4 3.8 4.6 3.9 4.1 3.6,
3.5 3.6 4.7 3.5 4.2 4.8 4.5 3.7 3.6,
3.2 3.7 4.3 3.3 4.3 4.9 4.7 4.0 4.2,
3.0 3.7 3.9 3.4 3.6 5.4 3.9 4.1 4.8,
3.0 4.2 3.4 3.2 3.8 5.6 3.8 4.6 4.9,
3.2 4.4 3.5 3.4 3.7 4.8 3.7 4.7 5.0};

p=nrow(Y1); p2=p##2; n1=ncol(Y1);n2=ncol(Y2);n3=ncol(Y3);n4=ncol(Y4);
const=0.5;

*Ci selects from vec(Sigma) submatrices from Sigma;
C1=I(p2)[1:p-1,];
do i=2 to p-1;
C1=C1//I(p2)[p#(i-1)+i:p#i-1,];
end;
C3=I(p2)[1+p+1:p-1+p+1,];
do i=2 to p-1;
C3=C3//I(p2)[p#(i-1)+i+p+1:p#i-1+p+1,];
end;
*print C1[format=2.0] C2[format=2.0] C3[format=2.0];

%macro mat(D,p); *D selects from vec(Sigma) the elements on the diagonals of Sigma;
D=J(p,p##2,0);
kr=0;
do i=1 to 1;
do j=1 to p-i+1;
kr=kr+1;
D[kr,(p+1)#(j-1)+i]=1;
end;end;
%mend mat;
%mat (D,p);*print D[format=2.0];

*B selects from vec(Sigma) the main diagonal and first row of Sigma;
B=D[1:p,]//I(p2)[2:p,];

*K is used in the expression for the cov(Vec(Sigma));*****;
K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;
end;

*Canonical statistics;
ym1=Y1[,+]/n1;
yy1=colvec((Y1*Y1')')/n1;
ym2=Y2[,+]/n2;
yy2=colvec((Y2*Y2')')/n2;
ym3=Y3[,+]/n3;
yy3=colvec((Y3*Y3')')/n3;
ym4=Y4[,+]/n4;

```

```

yy4=colvec((Y4*Y4')')/n4;

t=ym1//yy1//ym2//yy2//ym3//yy3//ym4//yy4;
diff1=1;
i=0;
t0=t;
do while (diff1>1e-8);
i=i+1;
m=t;
t=t0;

mu1m=m[1:p];          M21=m[p+1:p#(p+1)];
mu2m=m[p#(p+1)+1:p#(p+1)+p];  M22=m[p#(p+1)+p+1:2#p#(p+1)];
mu3m=m[2#p#(p+1)+1:2#p#(p+1)+p];  M23=m[2#p#(p+1)+p+1:3#p#(p+1)];
mu4m=m[3#p#(p+1)+1:3#p#(p+1)+p];  M24=m[3#p#(p+1)+p+1:4#p#(p+1)];
mum=(mu1m+mu2m+mu3m+mu4m)/4;

dmu1m=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2m=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu3m=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu4m=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);
dmum=( I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0) )/4;

d2m=(mu2m[1]-mu2m[2])/(mum[1]-mum[2]);
d3m=(mu3m[1]-mu3m[2])/(mum[1]-mum[2]);
d4m=(mu4m[1]-mu4m[2])/(mum[1]-mum[2]);

a2m=mu2m[1]-d2m#mum[1];
a3m=mu3m[1]-d3m#mum[1];
a4m=mu4m[1]-d4m#mum[1];

cov1m=m21-colvec(mu1m*mu1m');
cov2m=m22-colvec(mu2m*mu2m');
cov3m=m23-colvec(mu3m*mu3m');
cov4m=m24-colvec(mu4m*mu4m');

dcov1m=(-I(p)@mu1m-mu1m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov2m=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2m-mu2m@I(p))||I(p2)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov3m=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||(-I(p)@mu3m-mu3m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov4m=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu4m-mu4m@I(p))||I(p2);

if i=1 then cov1m=colvec(I(p)+const);
if i=1 then cov2m=colvec(I(p)+const);
if i=1 then cov3m=colvec(I(p)+const);
if i=1 then cov4m=colvec(I(p)+const);

c2m=cov2m[1]/cov1m[1];
c3m=cov3m[1]/cov1m[1];
c4m=cov4m[1]/cov1m[1];

dc2m=( -cov2m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );
dc3m=( -cov3m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );
dc4m=( -cov4m[1]/cov1m[1]##2#I(p2)[1,] )||( 1/cov1m[1]#I(p2)[1,] );

```



```

Sig1=shape(cov1m,p,p);
Sig2=shape(cov2m,p,p);
Sig3=shape(cov3m,p,p);
Sig4=shape(cov4m,p,p);

Phi1=(Sig1@mu1m+mu1m@Sig1);
Phi2=(Sig2@mu2m+mu2m@Sig2);
Phi3=(Sig3@mu3m+mu3m@Sig3);
Phi4=(Sig4@mu4m+mu4m@Sig4);

V1= ( (Sig1||Phi1')/(Phi1|((I(p2)+K)*(Sig1@Sig1+Sig1@(mu1m*mu1m'
+ (mu1m*mu1m')@Sig1))) )/n1;
V2= ( (Sig2||Phi2')/(Phi2|((I(p2)+K)*(Sig2@Sig2+Sig2@(mu2m*mu2m'
+ (mu2m*mu2m')@Sig2))) )/n2;
V3= ( (Sig3||Phi3')/(Phi3|((I(p2)+K)*(Sig3@Sig3+Sig3@(mu3m*mu3m'
+ (mu3m*mu3m')@Sig3))) )/n3;
V4= ( (Sig4||Phi4')/(Phi4|((I(p2)+K)*(Sig4@Sig4+Sig4@(mu4m*mu4m'
+ (mu4m*mu4m')@Sig4))) )/n4;

V=block(V1,V2,V3,V4);

e1m=cov1m[p+2]/cov1m[1];
de1m=(cov1m[1]#I(p2)[p+2,]-cov1m[p+2]#I(p2)[1,])/cov1m[1]##2;
e2m=cov2m[p+2]/cov2m[1];
de2m=(cov2m[1]#I(p2)[p+2,]-cov2m[p+2]#I(p2)[1,])/cov2m[1]##2;
e3m=cov3m[p+2]/cov3m[1];
de3m=(cov3m[1]#I(p2)[p+2,]-cov3m[p+2]#I(p2)[1,])/cov3m[1]##2;
e4m=cov4m[p+2]/cov4m[1];
de4m=(cov4m[1]#I(p2)[p+2,]-cov4m[p+2]#I(p2)[1,])/cov4m[1]##2;

Ge1m=C3-e1m#C1-C1*cov1m*de1m;
Ge1m=Ge1m*dcov1m;
Ge2m=C3-e2m#C1-C1*cov2m*de2m;
Ge2m=Ge2m*dcov2m;
Ge3m=C3-e3m#C1-C1*cov3m*de3m;
Ge3m=Ge3m*dcov3m;
Ge4m=C3-e4m#C1-C1*cov4m*de4m;
Ge4m=Ge4m*dcov4m;

Gc2m=(B*J(p2,p2,0)||B)-c2m#( B|(B*J(p2,p2,0))) -B*cov1m*dc2m;
Gc2m=Gc2m*(dcov1m/dcov2m);
Gc3m=(B*J(p2,p2,0)||B)-c3m#( B|(B*J(p2,p2,0))) -B*cov1m*dc3m;
Gc3m=Gc3m*(dcov1m/dcov3m);
Gc4m=(B*J(p2,p2,0)||B)-c4m#( B|(B*J(p2,p2,0))) -B*cov1m*dc4m;
Gc4m=Gc4m*(dcov1m/dcov4m);

dd2m=( (mum[1]-mum[2])#(dmu2m[1,]-dmu2m[2,])-(mu2m[1]-mu2m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;
dd3m=( (mum[1]-mum[2])#(dmu3m[1,]-dmu3m[2,])-(mu3m[1]-mu3m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;
dd4m=( (mum[1]-mum[2])#(dmu4m[1,]-dmu4m[2,])-(mu4m[1]-mu4m[2])#(dmum[1,]-dmum[2,]))
/(mum[1]-mum[2])##2;

da2m=dmu2m[1,]-d2m#dmum[1,]-mum[1]#dd2m;
da3m=dmu3m[1,]-d3m#dmum[1,]-mum[1]#dd3m;
da4m=dmu4m[1,]-d4m#dmum[1,]-mum[1]#dd4m;

```

```

Gd2m=dmu2m-d2m#dmum-mum*dd2m-J(p,1,1)*da2m;
Gd3m=dmu3m-d3m#dmum-mum*dd3m-J(p,1,1)*da3m;
Gd4m=dmu4m-d4m#dmum-mum*dd4m-J(p,1,1)*da4m;

Gm=Gd2m//Gd3m//Gd4m//Ge1m//Ge2m//Ge3m//Ge4m//Gc2m//Gc3m//Gc4m;

diff=1;
j=0;
do while (diff>1e-8);
tv=t;
j=j+1;

mu1t=t[1:p];          t21=t[p+1:p#(p+1)];
mu2t=t[p#(p+1)+1:p#(p+1)+p];  t22=t[p#(p+1)+p+1:2#p#(p+1)];
mu3t=t[2#p#(p+1)+1:2#p#(p+1)+p];  t23=t[2#p#(p+1)+p+1:3#p#(p+1)];
mu4t=t[3#p#(p+1)+1:3#p#(p+1)+p];  t24=t[3#p#(p+1)+p+1:4#p#(p+1)];
mut=(mu1t+mu2t+mu3t+mu4t)/4;

dmu1t=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2t=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu3t=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu4t=J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);
dmu=( I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0)||I(p)||J(p,p2,0) )/4;

cov1t=t21-colvec(mu1t*mu1t');
cov2t=t22-colvec(mu2t*mu2t');
cov3t=t23-colvec(mu3t*mu3t');
cov4t=t24-colvec(mu4t*mu4t');

dcov1t=(-I(p)@mu1t-mu1t@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov2t=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2t-mu2t@I(p))||I(p2)
||J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0);
dcov3t=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||(-I(p)@mu3t-mu3t@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov4t=J(p2,p,0)||J(p2,p2,0)||J(p2,p,0)||J(p2,p2,0)
||J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu4t-mu4t@I(p))||I(p2);

c2t=cov2t[1]/cov1t[1];
c3t=cov3t[1]/cov1t[1];
c4t=cov4t[1]/cov1t[1];

dc2t=( -cov2t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );
dc3t=( -cov3t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );
dc4t=( -cov4t[1]/cov1t[1]##2#I(p2)[1,] )||( 1/cov1t[1]#I(p2)[1,] );

gc2=B*cov2t-c2t#B*cov1t;
gc3=B*cov3t-c3t#B*cov1t;
gc4=B*cov4t-c4t#B*cov1t;

e1t=cov1t[p+2]/cov1t[1];
de1t=(cov1t[1]#I(p2)[p+2,]-cov1t[p+2]#I(p2)[1,])/cov1t[1]##2;
e2t=cov2t[p+2]/cov2t[1];
de2t=(cov2t[1]#I(p2)[p+2,]-cov2t[p+2]#I(p2)[1,])/cov2t[1]##2;
e3t=cov3t[p+2]/cov3t[1];
de3t=(cov3t[1]#I(p2)[p+2,]-cov3t[p+2]#I(p2)[1,])/cov3t[1]##2;
e4t=cov4t[p+2]/cov4t[1];

```

```

de4t=(cov4t[1]#I(p2)[p+2,]-cov4t[p+2]#I(p2)[1,])/cov4t[1]##2;

ge1=C3*cov1t-e1t#C1*cov1t;
ge2=C3*cov2t-e2t#C1*cov2t;
ge3=C3*cov3t-e3t#C1*cov3t;
ge4=C3*cov4t-e4t#C1*cov4t;

Ge1t=C3-e1t#C1-C1*cov1t*de1t;
Ge1t=Ge1t*dcov1t;
Ge2t=C3-e2t#C1-C1*cov2t*de2t;
Ge2t=Ge2t*dcov2t;
Ge3t=C3-e3t#C1-C1*cov3t*de3t;
Ge3t=Ge3t*dcov3t;
Ge4t=C3-e4t#C1-C1*cov4t*de4t;
Ge4t=Ge4t*dcov4t;

Gc2t=(B*J(p2,p2,0)||B)-c2t#( B|| (B*J(p2,p2,0))) -B*cov1t*dc2t;
Gc2t=Gc2t*(dcov1t//dcov2t);
Gc3t=(B*J(p2,p2,0)||B)-c3t#( B|| (B*J(p2,p2,0))) -B*cov1t*dc3t;
Gc3t=Gc3t*(dcov1t//dcov3t);
Gc4t=(B*J(p2,p2,0)||B)-c4t#( B|| (B*J(p2,p2,0))) -B*cov1t*dc4t;
Gc4t=Gc4t*(dcov1t//dcov4t);

d2t=(mu2t[1]-mu2t[2])/(mut[1]-mut[2]);
d3t=(mu3t[1]-mu3t[2])/(mut[1]-mut[2]);
d4t=(mu4t[1]-mu4t[2])/(mut[1]-mut[2]);

dd2t=( (mut[1]-mut[2])#(dmu2t[1,]-dmu2t[2,])-(mu2t[1]-mu2t[2])#(dmut[1,]-dmut[2,]))
      /(mut[1]-mut[2])##2;
dd3t=( (mut[1]-mut[2])#(dmu3t[1,]-dmu3t[2,])-(mu3t[1]-mu3t[2])#(dmut[1,]-dmut[2,]))
      /(mut[1]-mut[2])##2;
dd4t=( (mut[1]-mut[2])#(dmu4t[1,]-dmu4t[2,])-(mu4t[1]-mu4t[2])#(dmut[1,]-dmut[2,]))
      /(mut[1]-mut[2])##2;

a2t=mu2t[1]-d2t#mut[1];
a3t=mu3t[1]-d3t#mut[1];
a4t=mu4t[1]-d4t#mut[1];

da2t=dmu2t[1,]-d2t#dmut[1,]-mut[1]#dd2t;
da3t=dmu3t[1,]-d3t#dmut[1,]-mut[1]#dd3t;
da4t=dmu4t[1,]-d4t#dmut[1,]-mut[1]#dd4t;

gd2=mu2t-d2t#mut-a2t#J(p,1,1);
gd3=mu3t-d3t#mut-a3t#J(p,1,1);
gd4=mu4t-d4t#mut-a4t#J(p,1,1);

Gd2t=dmu2t-d2t#dmut-mut*dd2t-J(p,1,1)*da2t;
Gd3t=dmu3t-d3t#dmut-mut*dd3t-J(p,1,1)*da3t;
Gd4t=dmu4t-d4t#dmut-mut*dd4t-J(p,1,1)*da4t;

Gt=Gd2t//Gd3t//Gd4t//Ge1t//Ge2t//Ge3t//Ge4t//Gc2t//Gc3t//Gc4t;
g=gd2//gd3//gd4//ge1//ge2//ge3//ge4//gc2//gc3//gc4;

if i=1 & j=1 then do;
g0=g; Gt0=Gt;
Sig10=shape(cov1t,p,p);
Sig20=shape(cov2t,p,p);

```

```

Sig30=shape(cov3t,p,p);
Sig40=shape(cov4t,p,p);
Phi10=(Sig10@mu1t+mu1t@Sig10);
Phi20=(Sig20@mu2t+mu2t@Sig20);
Phi30=(Sig30@mu3t+mu3t@Sig30);
Phi40=(Sig40@mu4t+mu4t@Sig40);
V10= ( (Sig10||Phi10')//((I(p2)+K)*(Sig10@Sig10+Sig10@(mu1t*mu1t')
+ (mu1t*mu1t')@Sig10))) )/n1;
V20= ( (Sig20||Phi20')//((I(p2)+K)*(Sig20@Sig20+Sig20@(mu2t*mu2t')
+ (mu2t*mu2t')@Sig20))) )/n2;
V30= ( (Sig30||Phi30')//((I(p2)+K)*(Sig30@Sig30+Sig30@(mu3t*mu3t')
+ (mu3t*mu3t')@Sig30))) )/n3;
V40= ( (Sig40||Phi40')//((I(p2)+K)*(Sig40@Sig40+Sig40@(mu4t*mu4t')
+ (mu4t*mu4t')@Sig40))) )/n4;
V0=block(V10,V20,V30,V40);
end;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
diff=sqrt((t-tv)'*(t-tv));
print i j e1m e2m e3m e4m t m g;
end;
diff1=sqrt((m-t)'*(m-t));
end;

Wald=(g0)'*ginv(Gt0*V0*Gt0')*g0;
vg=sum(echelon(echelon(Gm*V*Gm')'));
prob=1-probchi(Wald,vg);
print Wald vg prob;

sig1m=sqrt(D*cov1m);
sig2m=sqrt(D*cov2m);
sig3m=sqrt(D*cov3m);
sig4m=sqrt(D*cov4m);

cov1m=m21-colvec(mu1m*mu1m');
rho1=((1/sig1m)@(1/sig1m))#cov1m;
rhom1=shape(rho1,p,p);
cov1m=shape(cov1m,p,p);

cov2m=m22-colvec(mu2m*mu2m');
rho2=((1/sig2m)@(1/sig2m))#cov2m;
rhom2=shape(rho2,p,p);
cov2m=shape(cov2m,p,p);

cov3m=m23-colvec(mu3m*mu3m');
rho3=((1/sig3m)@(1/sig3m))#cov3m;
rhom3=shape(rho3,p,p);
cov3m=shape(cov3m,p,p);

cov4m=m24-colvec(mu4m*mu4m');
rho4=((1/sig4m)@(1/sig4m))#cov4m;
rhom4=shape(rho4,p,p);
cov4m=shape(cov4m,p,p);

print a2m a3m a4m;
print d2m d3m d4m;
print c2m c3m c4m e1m;

```

```

print mu1m mu2m mu3m mu4m mum;
print cov1m cov2m cov3m cov4m;
print rhom1 rhom2 rhom3 rhom4;

*Calculation of variance of par's;

Acov=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V;
var=vecdiag(Acov); *print var;

semu1=sqrt(var[1:p]);
semu2=sqrt(var[p#(p+1)+1:p#(p+1)+p]);
semu3=sqrt(var[2#p#(p+1)+1:2#p#(p+1)+p]);
semu4=sqrt(var[3#p#(p+1)+1:3#p#(p+1)+p]);

secov1=sqrt(shape(var[p+1: p#(p+1)],p,p));
secov2=sqrt(shape(var[p#(p+1)+p+1: 2#p#(p+1)],p,p));
secov3=sqrt(shape(var[2#p#(p+1)+p+1: 3#p#(p+1)],p,p));
secov4=sqrt(shape(var[3#p#(p+1)+p+1: 4#p#(p+1)],p,p));

print semu1 semu2 semu3 semu4;
print secov1 secov2 secov3 secov4;

```

## A.4 SAS Programs for Chapter 6

### A.4.1 ZEROCOV.SAS

```

*Estimation of covariance matrices with zeros;
*Example in Chaudhuri, Drton & Richardson (2007);

proc iml worksizes=50;
reset nolog;
n=134;p=8;p2=p#p;
R={ 1      0.24  0.08 -0.18 -0.10 -0.18 -0.07 -0.08,
    0.24   1      0.23 -0.03 -0.10  0.12 -0.08 -0.07,
    0.08  0.23   1      0.26  0.28  0.20  0.21  0.26,
   -0.18 -0.03  0.26   1      0.87  0.44  0.81  0.87,
   -0.10 -0.10  0.28   0.87  1      0.39  0.88  0.92,
   -0.18  0.12  0.20   0.44  0.39  1      0.50  0.46,
   -0.07 -0.08  0.21   0.81  0.88  0.50   1      0.91,
   -0.08 -0.07  0.26   0.87  0.92  0.46  0.91   1};
sigma={0.39,0.36, 0.47, 1.7, 1.7, 0.78, 1.85, 1.54};
S=J(p,p,0);
do i=1 to p;
do j=1 to i;
S[i,j]=sigma[i,]*sigma[j,]*R[i,j];
S[j,i]=S[i,j];
end;
end;
print S;

t=((n-2)/n)#shape(S,p2,1);

K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);

```

```

H[i,j]=1;
K=H@H'+K;
end;
end;
verskil1=1;
i=0;
t0=t;
do while (verskil1>0.00000001);
i=i+1;
m=t;
t=t0;
Sig=shape(m,p,p); print sig;
V=(I(p2)+K)*(Sig@Sig);
V=V/n;
Gm1=I(p2)[3,]/I(p2)[5,]/I(p2)[7,]/I(p2)[8,]
//I(p2)[12,]/I(p2)[13,]/I(p2)[14,]/I(p2)[15,]
//I(p2)[16,];
Gm2=I(p2)[3,]/I(p2)[4,]/I(p2)[5,]/I(p2)[6,]
//I(p2)[7,]/I(p2)[8,]/I(p2)[12,]/I(p2)[13,]
//I(p2)[14,]/I(p2)[15,]/I(p2)[16,]/I(p2)[22,]
//I(p2)[23,];
Gm=Gm1*Gm2;
verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
j=j+1;
Gt1=Gm1;
Gt2=Gm2;
Gt=Gt1*Gt2;
g1=t[3]/t[5]/t[7]/t[8]/t[12]
//t[13]/t[14]/t[15]/t[16];
g2=t[3]/t[4]/t[5]/t[6]/t[7]/t[8]/t[12]
//t[13]/t[14]/t[15]/t[16]/t[22]/t[23];
g=g1*g2;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
nt=t[+];
verskil=sqrt((t-tv)'*(t-tv));
print i j nt t m g;
end;
verskil1=sqrt((m-t)'*(m-t));
end;
Sighat=shape(t,p,p);
sig=sqrt(vecdiag(Sighat));
Corhat=sqrt(inv(diag(Sighat)))*Sighat*sqrt(inv(diag(Sighat)));
print sig[format=6.2] Sighat[format=6.2] Corhat[format=6.2];

* Wald;
t=t0;
Gt=Gt1*Gt2;
g1=t[3]/t[5]/t[7]/t[8]/t[12]
//t[13]/t[14]/t[15]/t[16];
g2=t[3]/t[4]/t[5]/t[6]/t[7]/t[8]/t[12]
//t[13]/t[14]/t[15]/t[16]/t[22]/t[23];
g=g1*g2;
Sig=shape(t,p,p); *print sig;
V=(I(p2)+K)*(sig@sig)/n;

```

```

wald=(g)'*ginv(Gt*V*Gt')*g;
ng=nrow(g);
prob=1-probchi(wald,ng);
print wald prob ng;

* Asymptotic covariance matrix;
SigmaC=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V;
vsp=vecdiag(SigmaC);
vecvsp=round(vsp,0.000000001);
vspqrt=sqrt(vecvsp); vspmatrix=shape(vspqrt,p,p);
print vspmatrix[format=6.4];

```

## A.4.2 WISHART.SAS

```

*Estimation of Wishart matrices;
*Test for independence of 2 sets of variates;
*Homogeneity of covariance matrices;
*Homogeneity of covariance matrices with dependence structure;
*Example in Jiang&Sarkar (1998), Jiang et al (1999);

```

```

proc iml worksizes=50;
reset nolog;
S={0.0657739  0.01602  0.0600567  -0.002653,
    0.01602  0.0534908  0.0084126  0.0125034,
    0.0600567  0.0084126  0.0729088  -0.000625,
    -0.002653  0.0125034  -0.000625  0.0459961};
print S;
n=25;p=4;p2=p#p;
t=((n-2)/n)#shape(S,16,1);
V=J(4,4,0);
K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;
end;
verskill1=1;
i=0;
t0=t;
do while (verskill1>0.00000001);
i=i+1;
m=t;
t=t0;
Sig=shape(m,p,p); print sig;
V=(I(p2)+K)*(Sig@Sig);
V=V/n;
Gm1={0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0,
      0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0,
      0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0,
      0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0};
Gm2={1 0 0 0 0 0 0 0 0 0 -1 0 0 0 0 0,
      0 1 0 0 0 0 0 0 0 0 -1 0 0 0 0 0,
      0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 -1};
Gm3={1 0 0 0 0 0 0 0 0 0 -1 0 0 0 0 0,
      0 1 0 0 0 0 0 0 0 0 -1 0 0 0 0 0,

```

```

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 -1,
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0};
Gm=Gm1;
verskil=1;
j=0;
do while (verskil>0.00000001);
tv=t;
j=j+1;
Gt=Gm;
g1=(t[3])//(t[4])//(t[7])//(t[8]);
g2=(t[1]-t[11])//(t[2]-t[12])//(t[6]-t[16]);
g3=(t[1]-t[11])//(t[2]-t[12])//(t[6]-t[16])//(t[4])//(t[7]);
g=g1;
t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
nt=t[+];
verskil=sqrt((t-tv)'*(t-tv));
print i j nt t m g;
end;
verskil1=sqrt((m-t)'*(m-t));
end;
Sighat=shape(t,4,4);
Corhat=sqrt(inv(diag(Sighat))*Sighat*sqrt(inv(diag(Sighat))));
print Sighat[format=6.4] Corhat[format=6.4];

* Wald;
t=t0;
g1=(t[3])//(t[4])//(t[7])//(t[8]);
g2=(t[1]-t[11])//(t[2]-t[12])//(t[6]-t[16]);
g3=(t[1]-t[11])//(t[2]-t[12])//(t[6]-t[16])//(t[4])//(t[7]);
g=g1;
Sig=shape(t,p,p); *print sig;
V=(I(p2)+K)*(sig@sig)/n;
wald=(g)'*ginv(Gt*V*Gt')*g;
ng=nrow(g);
prob=1-probchi(wald,ng);
print wald prob ng;

* Asymptotic covariance matrix;
SigmaC=V-(Gm*V)'*ginv(Gm*V*Gm')*Gm*V;
vsp=vecdiag(SigmaC);
vecvsp=round(vsp,0.000000001);
vspqrt=sqrt(vecvsp); vspmatrix=shape(vspqrt,p,p);
print vspmatrix[format=6.4];

```

### A.4.3 NORMAL.SAS

```

*Equivalent approach;
*Test for independence of 2 sets of variates;
*Homogeneity of covariance matrices;
*Homogeneity of covariance matrices with dependence structure;
*Example in Jiang&Sarkar (1998), Jiang et al (1999);

proc iml worksizes=50;
options pagesize=100 linesize=80;
reset nolog;

```



```

data={
155.67    21.83    122.80    18.25,
132.28    28.92    119.12    37.99,
135.28    19.73    121.81    24.09,
134.96    29.25    131.75    36.47,
194.10    31.16    216.52    24.60,
117.77    24.71    132.56    29.25,
118.82    31.25    151.74    28.27,
105.34    36.36    105.88    32.77,
149.31    34.96    142.17    25.79,
105.96    29.59    103.41    32.50,
 93.78    31.28    106.26    32.41,
150.00    22.64    202.19    19.52,
227.44    33.41    226.22    31.13,
140.80    41.05     99.24    21.21,
146.38    47.79    137.09    28.32,
103.56    35.73     83.43    37.66,
100.54    28.48     91.08    21.22,
128.13    27.30    126.09    36.33,
105.38    22.82    117.01    28.20,
196.40    38.62    191.75    33.54,
 86.50    25.99     81.53    29.34,
198.73    29.38    175.79    40.00,
111.31    36.27    102.12    33.26,
148.17    40.59    118.01    27.82,
104.30    19.38    106.24    27.64}
;
Y1=log(data[1:13,]);
Y2=log(data[14:25,]);
n1=nrow(Y1); n2=nrow(Y2);

n=nrow(data); p=ncol(data); p2=p#p; Y=log(data);

ym1=(Y1'*J(n1,1,1)/n1);
ym2=(Y2'*J(n2,1,1)/n2);
yy1=colvec(Y1'*Y1/n1);
yy2=colvec(Y2'*Y2/n2);

*D selects from vec(Sigma) the elements on the diagonals of Sigma;
%macro mat(D,p);
D=J(p,p##2,0);
kr=0;
do i=1 to 1;
do j=1 to p-i+1;
kr=kr+1;
D[kr,(p+1)#(j-1)+i]=1;
end;end;
%mend mat;
%mat (D,p);*print D[format=2.0];

K=j(p2,p2,0);
do i=1 to p;
do j=1 to p;
H=j(p,p,0);
H[i,j]=1;
K=H@H'+K;
end;

```

```

end;

t=ym1//yy1//ym2//yy2; print t;

diff1=1;
i=0;
t0=t;
do while (diff1>1E-9);
i=i+1;
m=t;
t=t0;

mu1m=m[1:p];          M21=m[p+1:p#(p+1)];
mu2m=m[p#(p+1)+1:p#(p+1)+p];  M22=m[p#(p+1)+p+1:2#p#(p+1)];

dmu1m=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2m=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);

cov1m=m21-colvec(mu1m*mu1m');
cov2m=m22-colvec(mu2m*mu2m');

dcov1m=(-I(p)@mu1m-mu1m@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov2m=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2m-mu2m@I(p))||I(p2);

Sig1=shape(cov1m,p,p);
Sig2=shape(cov2m,p,p);

Phi1=(Sig1@mu1m+mu1m@Sig1);
Phi2=(Sig2@mu2m+mu2m@Sig2);

V1= ( (Sig1||Phi1')//
      (Phi1||((I(p2)+K)*(Sig1@Sig1+Sig1@(mu1m*mu1m')+(mu1m*mu1m')@Sig1))) )/n1;
V2= ( (Sig2||Phi2')//
      (Phi2||((I(p2)+K)*(Sig2@Sig2+Sig2@(mu2m*mu2m')+(mu2m*mu2m')@Sig2))) )/n2;
V=block(V1,V2);

Gm11=dcov1m[3,]//dcov1m[4,]//dcov1m[7,]//dcov1m[8,];
Gm12=dcov2m[3,]//dcov2m[4,]//dcov2m[7,]//dcov2m[8,];
Gm21=(dcov1m[1,]-dcov1m[11,])//(dcov1m[2,]-dcov1m[12,])
      //(dcov1m[6,]-dcov1m[16,]);
Gm22=(dcov2m[1,]-dcov2m[11,])//(dcov2m[2,]-dcov2m[12,])
      //(dcov2m[6,]-dcov2m[16,]);
Gm31=(dcov1m[1,]-dcov1m[11,])//(dcov1m[2,]-dcov1m[12,])
      //(dcov1m[6,]-dcov1m[16,])//dcov1m[4,]//dcov1m[7,];
Gm32=(dcov2m[1,]-dcov2m[11,])//(dcov2m[2,]-dcov2m[12,])
      //(dcov2m[6,]-dcov2m[16,])//dcov2m[4,]//dcov2m[7,];
Gm4=dcov1m-dcov2m;

Gm=Gm21//Gm22//Gm4;
*Gm=Gm31//Gm32//Gm4;

diff=1;
j=0;
do while (diff>1E-9);
tv=t;

```

```

j=j+1;

t1=t[1:2#p]; t2=t[2#p+1:2#p#(p+1)];

mu1t=t[1:p];          t21=t[p+1:p#(p+1)];
mu2t=t[p#(p+1)+1:p#(p+1)+p];  t22=t[p#(p+1)+p+1:2#p#(p+1)];

dmu1t=I(p)||J(p,p2,0)||J(p,p,0)||J(p,p2,0);
dmu2t=J(p,p,0)||J(p,p2,0)||I(p)||J(p,p2,0);

cov1t=t21-colvec(mu1t*mu1t');
cov2t=t22-colvec(mu2t*mu2t');

dcov1t=(-I(p)@mu1t-mu1t@I(p))||I(p2)||J(p2,p,0)||J(p2,p2,0);
dcov2t=J(p2,p,0)||J(p2,p2,0)||(-I(p)@mu2t-mu2t@I(p))||I(p2);

g11=(cov1t[3,])/(cov1t[4,])/(cov1t[7,])/(cov1t[8,]);
g12=(cov2t[3,])/(cov2t[4,])/(cov2t[7,])/(cov2t[8,]);
g21=(cov1t[1,]-cov1t[11,])/(cov1t[2,]-cov1t[12,])
    /(cov1t[6,]-cov1t[16,]);
g22=(cov2t[1,]-cov2t[11,])/(cov2t[2,]-cov2t[12,])
    /(cov2t[6,]-cov2t[16,]);
g31=(cov1t[1,]-cov1t[11,])/(cov1t[2,]-cov1t[12,])
    /(cov1t[6,]-cov1t[16,])/(cov1t[4,])/(cov1t[7,]);
g32=(cov2t[1,]-cov2t[11,])/(cov2t[2,]-cov2t[12,])
    /(cov2t[6,]-cov2t[16,])/(cov2t[4,])/(cov2t[7,]);
g4=cov1t-cov2t;

Gt11=dcov1t[3,]/dcov1t[4,]/dcov1t[7,]/dcov1t[8,];
Gt12=dcov2t[3,]/dcov2t[4,]/dcov2t[7,]/dcov2t[8,];
Gt21=(dcov1t[1,]-dcov1t[11,])/(dcov1t[2,]-dcov1t[12,])
    /(dcov1t[6,]-dcov1t[16,]);
Gt22=(dcov2t[1,]-dcov2t[11,])/(dcov2t[2,]-dcov2t[12,])
    /(dcov2t[6,]-dcov2t[16,]);
Gt31=(dcov1t[1,]-dcov1t[11,])/(dcov1t[2,]-dcov1t[12,])
    /(dcov1t[6,]-dcov1t[16,])/(dcov1t[4,])/(dcov1t[7,]);
Gt32=(dcov2t[1,]-dcov2t[11,])/(dcov2t[2,]-dcov2t[12,])
    /(dcov2t[6,]-dcov2t[16,])/(dcov2t[4,])/(dcov2t[7,]);
Gt4=dcov1t-dcov2t;

g=g21/g22/g4;
Gt=Gt21/Gt22/Gt4;
*g=g31/g32/g4;
*Gt=Gt31/Gt32/Gt4;

t=t-(Gm*V)'*ginv(Gt*V*Gm')*g;
diff=sqrt((t-tv)'*(t-tv));
print i j t m g ;
end;
diff1=sqrt((m-t)'*(m-t));
end;

sig1m=sqrt(D*cov1m);
sig2m=sqrt(D*cov2m);
cov1m=m21-colvec(mu1m*mu1m');
rho1=((1/sig1m)@(1/sig1m))#cov1m;
rhom1=shape(rho1,p,p);

```

```
cov1m=shape(cov1m,p,p);
cov2m=m22-colvec(mu2m*mu2m');
rho2=((1/sig2m)@(1/sig2m))#cov2m;
rhom2=shape(rho2,p,p);
cov2m=shape(cov2m,p,p);

print mu1m[format=8.7] mu2m[format=8.7];
print cov1m[format=8.7] cov2m[format=8.7];
print rhom1[format=8.7] rhom2[format=8.7];
```