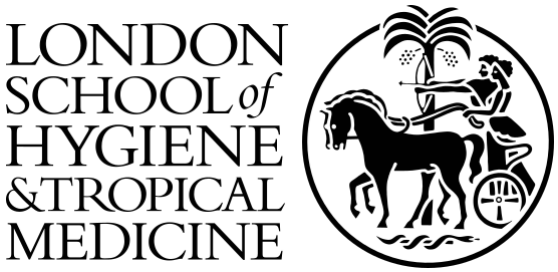


THE ANALYSIS OF VERY SMALL SAMPLES OF REPEATED MEASUREMENTS

SIMON SCOTT SKENE

2008

A thesis submitted in conformity with the requirements for the degree of PhD

London School of Hygiene and Tropical Medicine

University of London

Declaration

I have read and understood the School's definition of plagiarism and cheating given in the Research degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

Signed: 

Date: 16/07/2008

Full name: Simon Scott Skene

Abstract

The statistical analysis of repeated measures or longitudinal data always requires the accommodation of the covariance structure of the repeated measurements at some stage in the analysis. The general linear mixed model is often used for such analyses, and allows for the specification of both a mean model and a covariance structure. Often the covariance structure itself is not of direct interest, but only a means to producing valid inferences about the response. This thesis considers methods for the analysis of repeated measurements which arise from very small samples.

In Part 1, existing methods of analysis are shown to be inadequate for very small samples. More precisely, statistical measures of goodness of fit are not necessarily the right measure of the appropriateness of a covariance structure and inferences based on conventional Wald type procedures (with small sample adjustments) do not approximate sufficiently well their nominal properties when data are unbalanced or incomplete.

In Part 2, adaptive-estimation techniques are considered for the sample covariance matrix which smooth between unstructured and structured forms; 'direct' smoothing, a weighted average of the unstructured and structured estimates, and an estimate chosen via penalised likelihood. Whilst attractive in principle, these approaches are shown to have little success in practice, being critically dependent on the 'correct' choice of smoothing structure.

Part 3 considers methods which are less dependent on the covariance structure. A generalisation of a small sample adjustment to the empirical sandwich estimator is developed which accounts for its inherent bias and increased variance. This has nominal properties but lacks power. Also, a modification to Box's correction, an ANOVA F-statistic which accounts for departures from independence, is given which has both nominal properties and acceptable power.

Finally, Part 4 recommends the adoption of the modified Box statistic for repeated measurements data where the sample size is very small.

Dedication

To Fiona, Calum and Cora.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Mike Kenward, for his careful and patient advice throughout an extended period of research. I am also indebted to the following people, for kindly providing me with the data sets used to illustrate the statistical methods investigated throughout this thesis: Professor Emmanuel Lessafre of the Catholic University of Leuven for the Cardiac Enzyme data used throughout Part II of this thesis, and analysed in Section 10.2; Dr Nicky Wright of QinetiQ for the CRT data analysed in Section 10.3; and, Dr Richard Brammer of Huntingdon Life Sciences for the GPPM data analysed in Section 10.4.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Background

1.1 Introduction

A review is given of methods for dealing with estimation and testing in the analysis of repeated measures and longitudinal data. Our interest lies in procedures which are appropriate for such data which arise from very small samples.

1.2 Modelling Covariance Structures

Repeated measures and longitudinal data typically involve correlation between observations made on the same subject. It is also commonly found that variances are not constant across repeated observations. Where such patterns exist they must be effectively modelled to ensure the validity of inferences about the mean structure. If the data are unbalanced with a lack of common time points for responses, or not extensive enough to allow very general or completely unstructured covariance models to be used, a sufficiently well-fitting and parsimonious model must be found. Identification of an appropriate structure can improve the efficiency of the inferences made.

The general linear mixed effects model, developed following Laird and Ware (1982), is generally adopted for the analysis of repeated measures and longitudinal data.

(See for example Verbeke and Molenberghs (2000)). It has the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1.2.1}$$

where $\boldsymbol{\beta}$ are the fixed effect parameters and $\boldsymbol{\gamma}$ are random effects. This leads to $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. i.e. $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are uncorrelated Gaussian random variables with zero means and covariance matrices $\mathbf{G}$ and $\mathbf{R}$. Hence, the variance of the data is modelled by the appropriate specifying of any random effects and the structure of the covariance matrices $\mathbf{G}$ and $\mathbf{R}$. If $\mathbf{R} = \sigma^2\mathbf{I}$ and $\mathbf{Z} = \mathbf{0}$, so that there are no random effects, then the mixed model reduces to the general linear model. The fitting of such models has become routine with the arrival of procedures such as *PROC MIXED* in SAS (Sas Institute, 1999).

Generally $\boldsymbol{\Sigma}$ is chosen to maximise the REML log-likelihood, Patterson and Thompson (1971),

$$l_R(\mathbf{G}, \mathbf{R}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^T\mathbf{H}_{\boldsymbol{\Sigma}}\mathbf{y} - \textit{const.} \tag{1.2.2}$$

where $\mathbf{H}_{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}$, and the REML estimator of $\boldsymbol{\beta}$ is the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \tag{1.2.3}$$

The residual or restricted likelihood approach (REML) is preferred to maximum likelihood as it takes account of the loss of degrees of freedom from estimating the fixed effects. This is achieved by replacing $\mathbf{y}$ in the usual likelihood expression by the $(n - r)$ error contrasts $\mathbf{A}^T\mathbf{y}$, where $\mathbf{A}$ is a full rank matrix whose columns are orthogonal to the columns of $\mathbf{X}$. It follows that $\mathrm{Var}(\mathbf{A}^T\mathbf{y}) = \mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A}$ does not depend on $\boldsymbol{\beta}$, and also the REML estimator $\hat{\boldsymbol{\Sigma}}$ does not depend on the choice of $\mathbf{A}$.

There is no general satisfactory approach to finding the correct covariance structure. Possible structures include unstructured and stationary models such as compound symmetry and autoregressive. However, the obvious non-stationarity of most repeated measurements has led to the introduction of non-stationary models such as antependence (Kenward (1987)). Wolfinger (1996) reviews a large number of such proposals and suggests the following strategy based on Diggle (1988) for model selection.

(1) Use graphical analysis to select an initial mean model.

(2) Select initial covariance structures by any relevant means.

(3) Use formal statistical techniques to compare and select an appropriate structure. Assuming the chosen covariance structure, reduce the mean model if necessary.

A number of criteria can be used to compare the fit of such covariance structures. For instance, (reduced) likelihood ratio tests can be used for nested models to test whether the additional parameters of the more complicated covariance structure give a significant improvement to the fit of the model. That is,

$$2\{l_R(\boldsymbol{\Sigma}_1) - l_R(\boldsymbol{\Sigma}_2)\} \sim \chi^2(q_1 - q_2) \tag{1.2.4}$$

where $\boldsymbol{\Sigma}_2$ is nested within $\boldsymbol{\Sigma}_1$ and q_i are the corresponding number of covariance parameters for each model. Also, Akaike's information criterion (AIC, Akaike (1974)) and Schwarz's Bayesian criterion (BIC, Schwarz (1978)) are both likelihood measures penalized by the number of parameters, which can be used to make a direct comparison between models which fit the same fixed effects. *PROC MIXED* uses the following 'smaller is better' formulations based on the REML log-likelihood,

$$\begin{aligned} \text{AICR} &: -2l_R + 2q \\ \text{BICR} &: -2l_R + q\log(m) \end{aligned} \tag{1.2.5}$$

where q is the number of covariance parameters in the selected model, and m is the number of effective subjects.

Graphical and descriptive methods can point to likely classes of covariance structure, but comparison of the fit of a number of models is usually necessary for any given problem. This can be time consuming, and commonly the choice of structure is not consistent among similar types of data set.

Diggle (1988) advocates the use of the variogram

$$\gamma(u) = \frac{1}{2}\mathrm{E}\Big[\{Y(t) - Y(t-u)\}^2\Big], u \geq 0 \tag{1.2.6}$$

to identify a plausible covariance structure for longitudinal data which is stationary. He proposes a three component model for covariances, comprising a random intercept, measurement error and serial correlation. This model has variogram $\gamma(u) = \tau^2 + \sigma^2\{1 - \rho(u)\}$, which corresponds to the covariance structure

$$\mathbf{\Sigma} = \nu^2\mathbf{J} + \sigma^2\mathbf{H} + \tau^2\mathbf{I} \tag{1.2.7}$$

where $\mathbf{H} = \Big(\rho(u)\Big)$ is the matrix of correlations of the form $\rho(u) = \exp(-\alpha u^c)$, and ν^2 and τ^2 are variance components corresponding to the random intercepts and measurement error respectively. Identity, compound symmetry and autoregressive structures are special cases of this model. The parameters can be estimated from the sample variogram, which is a scatterplot of squared residuals from a saturated model against the underlying time differences between observations. Verbeke *et al.* (1998) extend the use of the variogram approach to models for non-stationary data by including random effects other than intercepts.

Verbeke and Molenberghs (2000) suggest the use of residuals from an overparameterised mean model fitted using ordinary least squares to help determine an appropriate covariance structure. Dependence among repeated measures can be assessed from a smoothed average trend of the residuals. For example, a constant variance function would indicate stationarity in the data so that no random effects beyond intercepts should be included in the model. When considering further random effects, Verbeke and Molenberghs suggest only covariates which are included in the fixed part of the model or are linear combinations of these need be used, since the random effects are deemed to have zero mean. Having selected an appropriate set of random effects, the covariance matrix $\mathbf{R}$ for the error terms $\boldsymbol{\varepsilon}$ can be chosen.

An alternative route to the development of covariance structures is to model directly the behaviour of subjects through a random coefficients model, and to use the structure induced by this. (See for example Longford (1990)). While attractive in principle, the success of such an approach is critically dependent on the appropriate choice of subject level model, and this can be difficult in practice.

Verbyla *et al.* (1999) introduce the cubic smoothing spline as an additional level of random effects in the mixed model. This imposes an implicit covariance structure on the model which consists of linear random coefficients, an individual random spline and a constant error term. The spline models departure from the linear model, and thus accounts for the serial correlation between successive measurements. (See for example Ruppert *et al.* (2003)).

Variances and covariances commonly change smoothly with changes in time and time lags. Diggle and Verbyla (1998) make use of this in smoothing the components of the variogram to provide an estimated covariance matrix using the variogram cloud as the input data for a two-dimensional non-parametric estimator for the variogram and the squared residuals as the input data for an estimator for the variance function. These functions are separately smoothed using a kernel weighted local linear regression with bandwidths chosen to minimise a cross-validation criterion, and combined to give a smoothed covariance estimate. This approach is difficult

to implement and does not guarantee an estimate which is postive-definite, so has limited use in inference. However, it is suggested as a diagnostic tool for helping to specify a plausible parametric structure. It could also be useful in providing an estimated covariance structure for data which is highly unbalanced or in which observation times are not common to all subjects, and alternative approaches are not available.

Shrinkage estimation in which parameter estimates are pushed towards pre-determined or believable values has a natural setting in Bayesian statistics where such shrinking is a natural consequence of reliance on informative prior distributions. Many authors following Stein (1975) have sought to shrink the eigenvalues of the sample covariance matrix towards more plausible values (often a common value) resulting in a more robust estimate. Others have sought to achieve such stability through a decomposition of the covariance matrix and placing prior distributions on the separate elements. Daniels and Kass (1999, 2001) review a number of these methods and propose using an estimated covariance matrix which is a compromise between an unstructured and a parametric form. They achieve this through the introduction of two hierarchical priors for the covariance matrix based on two different matrix decompositions.

Computational problems can arise in sampling from non-conjugate priors, so that Daniels and Kass (2001) adopt asymptotic distributions to simplify the computational aspects of their use. Their approach is to first fit a model by maximum likelihood with an unstructured form for $\boldsymbol{\Sigma}$, and conditional on $\hat{\boldsymbol{\beta}}$ compute the observed information matrix based on one of the two parameterisations. This requires the estimation of only two variance components (shrinkage parameters) from the data in addition to the parameters of the structured model. This method is easily computed using existing statistical software, requiring only a simple macro for the estimation of the shrinkage parameters. The structured covariance matrix which the unstructured matrix is smoothed towards may be chosen according to the usual criteria (AIC, BIC or likelihood ratio tests). However, the approximations and

asymptotic results needed to simplify the computations means that these estimates are less reliable in the small sample situations where they are most likely to be useful. Also, it is not clear if independent modelling of the variances and correlations in the sample covariance matrix is appropriate in a repeated measures or longitudinal data setting.

1.3 Testing of Fixed Effects

The fixed effects parameters obtained from (1.2.3) are asymptotically distributed as $\hat{\boldsymbol{\beta}} \sim \mathbf{N}(\boldsymbol{\beta}, \boldsymbol{\Phi})$, where

$$\begin{aligned} \boldsymbol{\Phi} = \mathrm{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathrm{Var}(\mathbf{y})\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})^{-1}, \end{aligned} \tag{1.3.1}$$

so that inferences about the fixed effects made via the general linear hypothesis, $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{L}$ is an $(l \times r)$ fixed matrix, may be tested using the approximate Wald statistic

$$\mathrm{W} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbf{L}^T(\mathbf{L}\boldsymbol{\Phi}\mathbf{L}^T)^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{1.3.2}$$

where $\mathrm{W} \sim \chi^2(l)$. This chi-squared statistic assumes no variation in the denominator term $\mathrm{Var}(\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{L}\boldsymbol{\Phi}\mathbf{L}^T$, so that often the Wald F-statistic is preferred,

$$\mathrm{F} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbf{L}^T(\mathbf{L}\boldsymbol{\Phi}\mathbf{L}^T)^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{l} \tag{1.3.3}$$

where $\mathrm{F} \sim \mathrm{F}(l, v2)$ and the denominator degrees of freedom $v2$ are estimated from the data. Usually the residual degrees of freedom $n - \mathrm{rank}(\mathbf{XZ})$, which is the number of observations less the number of estimated parameters, are adopted, although there

are a number of alternatives which attempt to account for small sample bias in the Wald Statistic.

Kenward and Roger (1997) suggest replacing $\mathbf{\Phi}$ in (1.3.3) by $\mathbf{\Phi}_A$, the adjusted covariance matrix for the fixed effects parameters suggested by Kackar and Harville (1984) and approximating λ times F as an F-distribution with adjusted denominator degrees of freedom m.

$$\mathrm{F}^* = \lambda \mathrm{F} \sim \mathrm{F}(l, m) \tag{1.3.4}$$

$\mathbf{\Phi}_A = \mathbf{\Phi} + 2\mathbf{\Lambda}$ accounts for the additional variability in $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{\Phi} = (\mathbf{X}^T\mathbf{\Sigma}\mathbf{X})^{-1}$ which is caused by the estimation of the unknown $\mathbf{\Sigma}$ by the inclusion of the additional factor $\mathbf{\Lambda}$, which corrects for the downward small sample bias. Denoting the parameters in $\mathbf{\Sigma}$ as the vector σ, $\mathbf{\Lambda}$ is found based on a Taylor series expansion around σ. Kenward and Roger show how the scale factor λ and degrees of freedom parameter m are found by comparing moments from the Taylor series expansion of $(\mathbf{L}\hat{\mathbf{\Phi}}_A\mathbf{L}^T)^{-1}$ about σ, and hence F^* with those of the approximating F-distribution. The adjustment is included as an option in recent versions of *PROC MIXED*. The Kenward Roger approximation has the advantage that it matches the exact values of λ and m in the special cases of Hotelling's T^2 test or the split-plot ANOVA, where $\mathbf{\Sigma}$ has a compound symmetry structure. The approximation also recovers Satterthwaite's adjusted degrees of freedom for the Wald t-test, where $l = 1$, whence $\lambda = 1$ and the square root of m is taken.

A number of alternatives for calculating the denominator degrees of freedom for the Wald F-statistic are reviewed by Schaalje *et al.* (2003), who recommend the use of the Kenward Roger method. They show it outperforms the other options in a simulation study, but note that when the covariance structure is complicated and the sample size is small this method can lead to inflated type 1 error rates.

Since in the above approaches the standard errors of the fixed effects are based on

known $\boldsymbol{\Sigma}$, inferences about $\boldsymbol{\beta}$ are not robust to misspecification of the covariance structure. An alternative is to base inferences on the robust or empirical variance estimator, Liang and Zeger (1986), also known as the 'sandwich' estimator. That is, use

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \tag{1.3.5}$$

with

$$\mathrm{Var}(\hat{\boldsymbol{\beta}_W}) = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathrm{Var}(\mathbf{y})\mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \tag{1.3.6}$$

where $\mathbf{W}^{-1} = \hat{\boldsymbol{\Sigma}}$ is a 'working' covariance structure and $\mathrm{Var}(\mathbf{y})$ is consistent for $\boldsymbol{\Sigma}$ whatever the true structure. Taking $\mathrm{Var}(\mathbf{y}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T$ uses the observed correlations between the residuals. This is the approach taken to covariance structure modelling via the generalised linear model using generalised estimating equations. The advantage of this estimator is that a poor choice of $\mathbf{W}$ will not affect the validity of inferences about $\boldsymbol{\beta}$. However it should be noted that this approach causes the fixed effects variances to reflect the observed correlations rather than those of any imposed 'structure'. In the simple case that $\mathbf{W}^{-1} = \mathbf{I}$, then we have the ordinary least squares estimate of $\boldsymbol{\beta}$ but with the standard errors of the estimates adjusted to account for the observed correlation structure. Brown and Prescott (1999) suggest the use of this approach whenever the mean response is of primary interest in an analysis and limited time is available to determine an appropriate covariance model, or in pharmaceutical trials where statistical methods have to be specified in advance in the protocol.

A disadvantage of the robust method is that the resulting tests based on the asymptotic Chi-squared distribution are known to be unreliable in small samples. That is, the consistency provided by this estimator comes at the price of increased variability. Diggle *et al.* (1994) suggest that the robust approach is suitable only when the data come from 'many experimental units', although several authors have attempted to

correct for this in certain situations. (See, for example, Kauermann and Carroll (2001)).

Another approach which uses the ordinary least squares estimate of $\boldsymbol{\beta}$ is given by Bellavance *et al.* (1996), who suggest Box's correction (Box (1954a,b)) based on the ANOVA F-statistic in the context of cross-over data as repeated measures. That is, let $\mathbf{X}$ be the $(n \times r)$ design matrix with all terms included, and $\mathbf{X}_R$ $(n \times (r-c))$ have the terms to be tested removed, and define $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and $\mathbf{B} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \mathbf{X}_R(\mathbf{X}_R^T\mathbf{X}_R)^{-1}\mathbf{X}_R^T$. Then,

$$\psi \mathrm{F} = \frac{(n-r)}{c}\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \sim F(v1, v2) \tag{1.3.7}$$

where,

$$\psi = \frac{(n-r)}{c}\frac{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})}{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})}, v1 = \frac{\{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})\}^2}{\mathrm{tr}\{(\mathbf{B}\boldsymbol{\Sigma})^2\}}, \text{ and } v2 = \frac{\{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\}^2}{\mathrm{tr}\{(\mathbf{A}\boldsymbol{\Sigma})^2\}}.$$

This is equivalent to testing the fixed effects based on their ordinary least squares estimates, but adjusting the null distribution to account for the covariance structure. The approximating distribution is derived by considering the quadratic forms $\mathbf{y}^T\mathbf{Q}\mathbf{y}$ of the numerator and denominator to be independently distributed as a constant times a Chi-squared distribution. The constant and the degrees of freedom parameter are chosen by matching the first and second moments of $\mathbf{y}^T\mathbf{Q}\mathbf{y}$ with those of the Chi-squared distribution. In practice, an estimate of $\boldsymbol{\Sigma}$ must be used, but this may be taken to be any consistent estimator of $\mathrm{Var}(\mathbf{y})$. Jones and Kenward (2003) suggest the use of the ordinary least squares covariance estimate is in keeping with the spirit of this approach, but in practice it may be more practical to simply adopt the unstructured REML estimate. Bellavance *et al.* show that this modified F-test approximation gives adequate control over the type 1 error.

The use of various covariance estimates and testing procedures are compared in

Jones and Kenward (2003) in the context of cross-over data as repeated measures. They also consider the use of a randomisation test. Under the null hypothesis of no treatment effect, sequences of responses within each subject should not differ systematically, so that the empirical distribution of the ANOVA F-statistic can be approximated by the repeated re-calculating of the statistic following random reallocation of subjects to sequences. The obtained value of the statistic can then be compared to this distribution. This method can be effective but the permutation of the data needs to be carefully matched to the null hypothesis. For example, the above scheme would not have worked in the presence of any carryover effect. Jones and Kenward use the randomisation test as a robust check of the p-values obtained from the other methods. This type of test is an example of a bootstrap resampling method. (See for example Davison and Hinkley (1997)). Other bootstrap methods may also be suitable, although generally such methods are thought to be unreliable in small samples no matter how many bootstrap samples are taken.

Chapter 2

Assessment of Existing Methods

2.1 Motivation

The need for methods for repeated measures or longitudinal data which adequately deal with the underlying covariance structure, when the sample size is very small, is highlighted in a paper by Brammer (2003).

Brammer analyses the data from two experiments which aim to investigate the effect of a compound on isolated tissues and organs; whether the compound causes electrocardiogram abnormalities in the guinea pig papillary muscle (GPPM), or has an effect on pressure in an isolated lung from a rat. An ascending dose design is adopted in each case, whereby a number of 'control' responses are measured for each assay at the beginning of the experiment and then remeasured at intervals following increasing doses of compound. As there is a limited time in which to experiment with the assays before deterioration, this design ensures the carryover effect at each stage is minimal compared to the increased dosage, as there is no time for a washout period.

These experiments can be considered as repeated measures designs with concentration as the time variable and tissue/animal as the subject, and analysed using a mixed model with an appropriate choice of covariance structure. This model is preferred to either a two-way ANOVA, which assumes a compound symmetry form

that may not be appropriate, or paired t-tests of each concentration against control, which would not make use of the full information from the data. The mixed model would appear to be a sensible approach, but the sample sizes here are very small.

For the GPPM data, Brammer has data from two experiments involving just three assays in each and, as the compounds being tested in each of these experiments are different, these are analysed separately. Since six concentrations of compound are compared with a control, this means that for each experiment a (7×7) covariance matrix is to be estimated from just three subjects. For the isolated lung data, there is data from rats in two strains of six. Here, three concentrations of compound are compared with a control, so that a (4×4) structure is fitted from 12 subjects. These matrices are used to estimate the standard errors of the treatment effects (difference from control).

Brammer chooses the covariance structure by first examining the unstructured form and then assessing the fit of a parametric model based on fewer parameters which appears to have a similar structure. For the GPPM data there were convergence problems when estimating a full unstructured covariance matrix, so he was restricted to comparing the fit of structures using forms with as many parameters as the data would allow. In each case he assesses an AR1 structure for example to fit better than a compound symmetry form, suggesting that the two-way ANOVA is not appropriate, but different structures are preferred for different responses from the same experiment.

In each experiment, Brammer compares the standard errors and unadjusted p-values for the treatment differences under each response and chosen covariance structure to indicate that the choice of structure is important. Due to the design, the treatment means for each response are the same regardless of chosen model, but this is not true for an extension study using the isolated lung data, where strain main effects and a strain/treatment interaction are included in the model.

This type of problem is typical in early stage drug development, where such studies

are routinely undertaken using very small samples, due to costs and the desire to minimise unnecessary animal experimentation. A particular example of this is in the area of safety pharmacology, where investigations are undertaken into the effects of test compounds on the vital functions of the central nervous, cardiovascular and respiratory systems prior to trials involving human subjects. Such experiments are carried out under 'good laboratory practice' to give an indication of potentially adverse effects, but can also indicate possible clinical benefits. Typically between four and eight (animal) subjects are used in parallel group or crossover trials involving repeated measurements. Clearly it is of great interest to see whether the usual techniques for analysing repeated measurements, which were outlined in Chapter 1, are appropriate in such small data sets, or if the inferences made are valid. The purpose of this assessment is to consider such issues through a series of simulation studies.

2.2 A Pilot Study

Consider a simple repeated measures experiment, where 16 subjects are randomly allocated to two treatment groups (of equal size), and a response recorded for each subject at each of five time points. Five time points are chosen so that the resulting (5×5) estimates of the covariance structure within subjects will be large enough for differences between models to be adequately determined; and to give a range to the number of covariance parameters estimated by each model. The sample size here may be considered only moderately small, but this design serves as a useful comparison with later studies.

2.2.1 Methodology

1000 data samples are independently generated arising from a Gaussian distribution with zero mean and each of six underlying covariance structures; identity, compound symmetry, AR1 (low and high correlation), first order antedependence, and two 'unstructured' forms. These structures are shown in Figures A.1.1-A.1.6 of Appendix

A. The two unstructured forms are chosen to have patterns far removed from that usually fitted by the structured forms, such as correlations which increase by lag. One of these is found by adoption of a quadratic random effects model with unusual parameter values. The number of covariance parameters to be estimated by each model are then 1 (identity), 2 (compound symmetry, AR1), 11 (antedependence), and 15 (unstructured).

For each data set, a saturated means model is fitted, giving a separate mean parameter for each treatment and time combination (10 parameters). REML estimates of this model using various estimates of the 'true' underlying covariance structure are then used to assess the existing methods of choosing between covariance structures; reduced likelihood ratio tests and AICR/BICR. Also, the properties of the various Wald statistics (χ^2 and F with either residual degrees of freedom or the Kenward Roger adjustment) in making inferences about an appropriate hypothesis are compared. All the simulations are undertaken using SAS *IML*.

2.2.2 Results

Results from the pilot study are shown in Tables 2.2.1 and 2.2.2.

Choice of Covariance Model

Likelihood Ratio Test: Structured Covariance Model v's Identity (independence)

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_I \quad \text{v's} \quad H_1 : \mathbf{\Sigma} = \mathbf{\Sigma}_S$$

This test is of little practical interest since it is known that repeated measurements from a subject will be correlated. However, it is useful to consider it here as part of an investigation of the nominal properties of such tests. That is, to determine whether the actual size of the test, or type 1 error rate, is close to the nominal level. This can be checked by noting the proportion of significant results obtained when a 'structure' is fitted to data generated under the null hypothesis of an underlying identity covariance structure (independence). The actual sizes were checked for each

Underlying 'True' Covariance Structure	Proportion of Significant Reduced Likelihood Ratio Tests						Proportion of Structures chosen by AICR (BICR) (No. of Covariance parameters)				
	'True' Model v's Identity			'True' Model v's Unstr			UN (15)	ID (1)	CS (2)	AR1 (2)	ANTE (11)
	df	at 5%	at 1%	df	at 5%	at 1%					
Identity	—	—	—	14	0.111	0.032	0.029 (0.000)	0.705 (0.828)	0.148 (0.100)	0.112 (0.072)	0.006 (0.000)
Compound Symmetry	1	0.980	0.944	13	0.114	0.033	0.044 (0.005)	0.006 (0.008)	0.873 (0.909)	0.075 (0.078)	0.002 (0.000)
AR1 (ρ=0.2)	1	0.309	0.139	13	0.125	0.046	0.046 (0.003)	0.403 (0.520)	0.135 (0.116)	0.406 (0.359)	0.010 (0.002)
AR1 (ρ=0.8)	1	1.000	1.000	13	0.114	0.039	0.044 (0.002)	0.000 (0.000)	0.038 (0.041)	0.907 (0.956)	0.011 (0.001)
Antedependence	10	1.000	1.000	4	0.291	0.113	0.408 (0.195)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.592 (0.805)
Unstructured	14	1.000	1.000	—	—	—	1.000 (1.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Unstr (QRE)	14	1.000	1.000	—	—	—	1.000 (1.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

Table 2.2.1: *Results from 1000 simulations of the Pilot Study: a simple Repeated Measures experiment with 16 subjects. (UN=Unstructured, ID = Identity, CS=Compound Symmetry, ANTE=Antedependence).*

of compound symmetry, AR1 and antedependence forms. These were found to be close to the nominal levels of both 1% and 5% in each case for the two low parameter models (compound symmetry and AR1), but slightly inflated for the antedependence form. (Details omitted).

Once the nominal properties of a test have been established it is appropriate to consider power. The proportion of significant test results from the estimated 'true' structure against the identity in Table 2.2.1 gives the power of the likelihood ratio test to detect departures from independence. The results presented show that the reduced likelihood ratio test is very powerful in detecting such departures, and hence the need for modelling of the covariance structure. The exception to this is for data generated from an underlying AR1 covariance structure with low correlation, where the null hypothesis of independence is rejected on only 30% of occasions at the 5% level of significance. This is however unsurprising as this model may be considered close to an identity structure.

Likelihood Ratio Test: Structured Covariance Model v's Unstructured - 'Goodness of Fit'

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_S \quad \text{v's} \quad H_1 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_U$$

As the unstructured covariance form will always give the best fit to the data, the proportion of significant tests from reduced likelihood ratio tests of the estimated 'true' covariance structure against an unstructured form gives the actual size of the test for detecting a more parsimonious model where an appropriate one exists. Table 2.2.1 shows that in this respect the reduced likelihood ratio test does not achieve its nominal size, with actual sizes in excess of 3% and 12% for nominal levels of 1% and 5% respectively. This means that the use of such a test to assess the fit of a structured model against that of an unstructured one will result in the unstructured form being chosen more often than would be expected from the nominal level. These sizes are further inflated for the test of an antedependence form against the unstructured. However, whilst these structures are not close they are closer than the unstructured is to say an AR1 or compound symmetry form.

Underlying 'True' Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (H_0: No Treatment/Time Interaction)			
		Size			Power
		χ^2	F	KR	KR
Identity	Unstr	0.178	0.161	0.044	0.747
	Identity	0.062	0.045	0.045	0.903
Compound Symmetry	Unstr	0.193	0.175	0.047	0.754
	Comp Sym	0.072	0.053	0.048	0.897
AR1 ($\rho = 0.2$)	Unstr	0.200	0.174	0.051	0.765
	AR1	0.065	0.055	0.047	0.901
AR1 ($\rho = 0.8$)	Unstr	0.179	0.160	0.039	0.744
	AR1	0.049	0.044	0.031	0.879
Antedependence	Unstr	0.172	0.149	0.036	0.764
	Ante	0.108	0.096	0.051	0.859
Unstructured	Unstr	0.176	0.158	0.047	0.767
Unstr (QRE)	Unstr	0.225	0.198	0.057	0.782

Table 2.2.2: *Results from 1000 simulations of the Pilot Study: a simple Repeated Measures experiment with 16 subjects. Size, χ^2, F and KR, gives the proportion of Type 1 errors obtained using the χ^2 and F-Wald statistics with residual and Kenward Roger adjusted denominator degrees of freedom.*

AICR and BICR

The measures AICR and BICR have no nominal properties against which to assess their behaviour, so need to be assessed in an absolute sense. Table 2.2.1 appears to show that they generally have a high proportion of success in identifying the 'true' covariance structure. Where there is a spread in these values, for example for data arising from the AR1 (low correlation) or antedependence models, it is because these models are not far removed from their alternatives.

Testing of Fixed Effects

Table 2.2.2 shows the actual sizes, testing under the null hypothesis, of the various Wald statistics for no treatment/time interaction. The results from the simulation show that both the asymptotic chi-squared and F-statistics (using residual degrees of freedom) have greatly inflated type 1 error rates, even for this moderately small sample size. If the 'true' structure is estimated, then for low dimension models (identity, compound symmetry and AR1) the error rates although inflated are closer

to the nominal level of 5%. For all underlying covariance structures, the adjusted F-statistic using the Kenward Roger adjustment gives an actual size close to the nominal level when using either the unstructured or 'true' model. It has been shown in a further study that where the estimated covariance structure is not 'true', use of the Kenward Roger adjustment can lead to inflated test sizes. (Details omitted).

It should be noted that for this model, using a saturated means model with complete and balanced data results in mean estimates which are equivalent to ordinary least squares estimates which do not depend on the covariance structure. Also, use of the unstructured covariance matrix in testing the interaction is equivalent to using Hotelling's T^2 test. That is, this is an exact small sample test, as is the split-plot ANOVA statistic which results when a compound symmetry structure is chosen, so that the nominal properties of these tests is known. When the identity model is adopted, the Kenward Roger adjustment is equivalent to the usual Wald F-test using the residual degrees of freedom.

Again, once the size of the test has been fixed at the nominal level, it is appropriate to consider power. Here this is done by adding terms which are linearly increasing in time to the responses of one of the treatment groups. These terms were chosen to set the power using the unstructured model at around 75%. Table 2.2.2 shows that if the 'true' covariance structure can be determined, then the resulting Wald tests will be more powerful than if the unstructured model is adopted. This is expected, but the reduction in power of up to 15% from adopting the unstructured model is non-negligible.

2.3 Further Simulations

The pilot study above involved only a moderately small sample size. Also, the design of the study meant that the covariance structure played no part in the estimation of the mean parameters being tested, and allowed Wald tests based on the Kenward Roger adjustment to be exact with nominal properties.

Three further designs are therefore considered to investigate the effects of reduced sample size, missing values and model design.

(**A**) A simple repeated measures experiment, with 10 subjects randomly allocated to two treatment groups (of equal size), and a response recorded for each subject at each of five time points.

(**B**) As design (A), but with missing values. One subject in each treatment group drops out at some random time following the first observation.

(**C**) A five treatment-five period crossover trial, with 10 subjects allocated randomly to treatments according to Table 2.3.1, using a pair of Williams' squares.

Subject	Period				
	1	2	3	4	5
1	A	B	C	D	E
2	B	D	A	E	C
3	D	E	B	C	A
4	E	C	D	A	B
5	C	A	E	B	D
6	E	D	C	B	A
7	C	E	A	D	B
8	A	C	B	E	D
9	B	A	D	C	E
10	D	B	E	A	C

Table 2.3.1: *A crossover design for 5 treatments (A, B, C, D, E).*

2.3.1 Methodology

Each design is treated as before, with 1000 data sets independently generated for each underlying covariance structure. In designs (A) and (B) a saturated means model is fitted, but for the crossover design (C) the mean model comprises an intercept, treatment and period effects (9 parameters). Again REML estimates of the mean models and covariance structures are adopted. Interest in designs (A) and (B) remains on the treatment/time interaction, but in design (C) the appropriate null hypothesis is that of no treatment effect.

2.3.2 Results

The results of these simulations are shown in Tables 2.3.2 - 2.3.7.

Choice of Covariance Model

Likelihood Ratio Tests

Table 2.3.2 for design (A) shows results similar to those of the pilot study. That is, the reduced likelihood ratio tests are powerful in detecting departures from independence for data arising from underlying covariance structures which are far from an identity form. This is repeated in Tables 2.3.3 and 2.3.4 for designs (B) and (C) respectively. Again a check has been undertaken to confirm the nominal properties of this test, with results similar to those noted for the pilot study.

For these designs based on a smaller sample size, the reduced likelihood ratio tests of the 'true' structure against the unstructured show that this test does not have nominal size. That is, where the data are generated from an underlying 'true' structure of few parameters, the likelihood ratio test is less effective at discriminating between this and an unstructured model in these settings. The actual sizes from the simulations are around 25% for design (A) and in excess of 40% for designs (B) and (C) compared to a nominal size of 5%. Again, the sizes are inflated further in the case of the estimated antedependence structure, which shares variances and lag 1 correlations with the unstructured model.

Little has been published on the small sample behaviour of the likelihood ratio test (LRT) in the context of choosing a covariance structure. Zucker *et al.* (2000) considers the use of a Bartlett correction (Bartlett (1937)) when making inferences about the mean parameters from a mixed linear model. For very small samples this correction still leads to an inflated test size, but can result in a test with nominal properties for tests involving single parameters only, when used in conjunction with an adjusted likelihood expression.

Underlying 'True' Covariance Structure	Reduced Likelihood Ratio Tests Proportion of Significant Tests (Out of 1000)						Proportion of Structures chosen by AICR (BICR) (No. of Covariance parameters)				
	'True' Model v's Identity			'True' Model v's Unstr			UN (15)	ID (1)	CS (2)	AR1 (2)	ANTE (11)
	df	at 5%	at 1%	df	at 5%	at 1%					
Identity	—	—	—	14	0.240	0.099	0.109 (0.052)	0.672 (0.756)	0.117 (0.105)	0.097 (0.086)	0.005 (0.001)
Compound Symmetry	1	0.874	0.769	13	0.229	0.097	0.118 (0.044)	0.049 (0.062)	0.709 (0.756)	0.123 (0.136)	0.001 (0.002)
AR1 (ρ=0.2)	1	0.210	0.077	13	0.254	0.108	0.116 (0.048)	0.456 (0.553)	0.144 (0.133)	0.277 (0.264)	0.007 (0.002)
AR1 (ρ=0.8)	1	0.999	0.995	13	0.249	0.122	0.145 (0.072)	0.000 (0.000)	0.071 (0.078)	0.771 (0.843)	0.013 (0.007)
Ante	10	1.000	0.994	4	0.474	0.247	0.568 (0.485)	0.000 (0.003)	0.000 (0.003)	0.004 (0.013)	0.428 (0.496)
Unstr	14	0.999	0.998	—	—	—	0.988 (0.966)	0.000 (0.003)	0.006 (0.020)	0.003 (0.009)	0.003 (0.002)
Unstr. (QRE)	14	1.000	1.000	—	—	—	0.778 (0.706)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.222 (0.294)

Table 2.3.2: *Results from 1000 simulations of Design (A): A simple Repeated Measures experiment with 10 subjects.*

Underlying 'True' Covariance Structure	Reduced Likelihood Ratio Tests Proportion of Significant Tests (Out of 1000)						Proportion of Structures chosen by AICR (BICR) (No. of Covariance parameters)				
	'True' Model v's Identity			'True' Model v's Unstr			UN (15)	ID (1)	CS (2)	AR1 (2)	ANTE (11)
	df	at 5%	at 1%	df	at 5%	at 1%					
Identity	—	—	—	14	0.393	0.202	0.205 (0.121)	0.581 (0.683)	0.105 (0.093)	0.098 (0.099)	0.011 (0.004)
Compound Symmetry	1	0.813	0.670	13	0.403	0.211	0.246 (0.144)	0.057 (0.084)	0.567 (0.634)	0.127 (0.136)	0.003 (0.002)
AR1 (ρ=0.2)	1	0.166	0.062	13	0.398	0.204	0.214 (0.118)	0.445 (0.539)	0.127 (0.128)	0.206 (0.213)	0.008 (0.002)
AR1 (ρ=0.8)	1	0.998	0.988	13	0.379	0.186	0.218 (0.130)	0.000 (0.001)	0.084 (0.094)	0.684 (0.764)	0.017 (0.011)
Ante	10	0.992	0.961	4	0.639	0.445	0.721 (0.644)	0.005 (0.011)	0.001 (0.004)	0.004 (0.016)	0.269 (0.325)
Unstr	14	0.999	0.992	—	—	—	0.978 (0.935)	0.003 (0.010)	0.014 (0.042)	0.003 (0.012)	0.002 (0.001)
Unstr. (QRE)	14	1.000	1.000	—	—	—	0.829 (0.768)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.171 (0.232)

Table 2.3.3: *Results from 1000 simulations of Design (B): A simple Repeated Measures experiment with 10 subjects, and missing values.*

Underlying 'True' Covariance Structure	Reduced Likelihood Ratio Tests Proportion of Significant Tests (Out of 1000)						Proportion of Structures chosen by AICR (BICR) (No. of Covariance parameters)				
	'True' Model v's Identity			'True' Model v's Unstr			UN (15)	ID (1)	CS (2)	AR1 (2)	ANTE (11)
	df	at 5%	at 1%	df	at 5%	at 1%					
Identity	—	—	—	14	0.410	0.262	0.265 (0.185)	0.543 (0.642)	0.081 (0.075)	0.102 (0.094)	0.009 (0.004)
Compound Symmetry	1	0.912	0.796	13	0.468	0.287	0.323 (0.233)	0.015 (0.022)	0.548 (0.624)	0.112 (0.121)	0.002 (0.000)
AR1 (ρ=0.2)	1	0.204	0.073	13	0.468	0.289	0.312 (0.206)	0.394 (0.486)	0.092 (0.092)	0.196 (0.213)	0.006 (0.003)
AR1 (ρ=0.8)	1	0.998	0.997	13	0.485	0.289	0.326 (0.231)	0.001 (0.001)	0.055 (0.062)	0.607 (0.703)	0.011 (0.003)
Ante	10	0.995	0.952	4	0.688	0.499	0.768 (0.691)	0.000 (0.008)	0.005 (0.007)	0.009 (0.020)	0.218 (0.274)
Unstr	14	0.999	0.999	—	—	—	0.988 (0.972)	0.000 (0.001)	0.009 (0.021)	0.003 (0.006)	0.000 (0.000)
Unstr. (QRE)	14	0.980	0.980	—	—	—	0.830 (0.799)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.170 (0.201)

Table 2.3.4: *Results from 1000 simulations of Design (C): A five treatment-five period Crossover Study with 10 subjects.*

Lawley (1956) demonstrates that the Bartlett correction generally improves the small sample behaviour of the LRT statistic, and gives an expression for the correction factor in terms of the cumulants of the log-likelihood derivatives. The correction adjusts the LRT statistic to give the same moments as the chi-squared distribution, ignoring quantities of order $1/m^2$, where m is the sample size, and its behaviour is generally improved as the number of estimated parameters is small in comparison to the sample size. A separate correction would be needed for each pair of structures to be compared, and this is not pursued here.

AICR and BICR

For design (A), the measures AICR and BICR still pick out the 'true' structure on the majority of occasions, but with a greater spread towards structures which are close to the 'true' structure. For designs (B) and (C), the spread is increased, indicating the reduced effectiveness of these measures in discriminating between structures.

Testing of Fixed Effects

Table 2.3.5 shows greatly inflated type 1 error rates for the χ^2 and F Wald statistics in design (A), but the Kenward Roger adjustment is seen to fix the size of the test at the nominal level of 5% when either the unstructured or 'true' covariance estimate is adopted. Again, if the 'true' structure with few parameters is estimated, then it is more powerful in detecting a significant difference in the interaction. Table 2.3.6 for design (B) shows that when data are missing, the Kenward Roger adjustment leads to inflated type 1 error rates of around 7.5% when the unstructured or ('true') antedependence structure are used. When the estimate of the 'true' structure is based on few parameters however, the adjustment still generally fixes the test size to the nominal value.

The loss of nominal properties in these tests when using an unstructured covariance model for data with missing values is marked, although when only two subjects drop

Underlying 'True' Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (H_0: No Treatment/Time Interaction)			
		Size			Power
		χ^2	F	KR	KR
Identity	Unstr	0.329	0.290	0.053	0.735
	Identity	0.064	0.048	0.048	0.980
Compound Symmetry	Unstr	0.350	0.306	0.045	0.747
	Comp Sym	0.069	0.051	0.046	0.985
AR1 ($\rho = 0.2$)	Unstr	0.352	0.319	0.050	0.735
	AR1	0.095	0.070	0.052	0.978
AR1 ($\rho = 0.8$)	Unstr	0.344	0.318	0.056	0.775
	AR1	0.074	0.058	0.033	0.984
Antedependence	Unstr	0.332	0.300	0.058	0.760
	Ante	0.208	0.179	0.069	0.945
Unstructured	Unstr	0.348	0.316	0.045	0.768
Unstr (QRE)	Unstr	0.333	0.300	0.048	0.740

Table 2.3.5: *Results from 1000 simulations of Design (A): A simple Repeated Measures experiment with 10 subjects.*

out, the loss in data is at most 8 observations out of a possible 50. Allowing a third subject to drop out (from either treatment group) is found to inflate this test size further, to around 15%, although as the amount of missing data increases so do the problems of convergence in finding the REML estimates. With the introduction of missing data, these small sample adjustments no longer lead to exact tests. That is, the Kenward Roger adjustment for all models is now based on an approximate Taylor series expansion, which may not be appropriate for large dimension problems with such small data sets. Also, the covariance structure of the data now directly effects the estimates of the mean parameters as well as their standard errors. Table 2.3.6 also shows a loss of power as we would expect with less data, but there is a greater loss when using the unstructured model even accounting for the inflated test size.

This inflation of the type 1 error rates using the Kenward Roger adjustment is even more pronounced in design (C), Table 2.3.7, where the actual test sizes found when using the unstructured covariance matrix are very far from the nominal level for the test of no treatment effect. However, it is noted again how the tests which rely on an estimate of the 'true' structure (where this has few parameters) appear to be

Underlying 'True' Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (H_0: No Treatment/Time Interaction)			
		Size			Power
		χ^2	F	KR	KR
Identity	Unstr	0.468	0.432	0.073	0.633
	Identity	0.062	0.044	0.044	0.958
Compound Symmetry	Unstr	0.437	0.411	0.074	0.632
	Comp Sym	0.080	0.050	0.047	0.950
AR1 ($\rho = 0.2$)	Unstr	0.461	0.419	0.089	0.637
	AR1	0.094	0.063	0.050	0.976
AR1 ($\rho = 0.8$)	Unstr	0.463	0.412	0.078	0.637
	AR1	0.075	0.051	0.031	0.966
Antedependence	Unstr	0.487	0.454	0.084	0.645
	Ante	0.261	0.220	0.079	0.858
Unstructured	Unstr	0.432	0.399	0.077	0.640
Unstr (QRE)	Unstr	0.360	0.338	0.046	0.516

Table 2.3.6: *Results from 1000 simulations of Design (B): A simple Repeated Measures experiment with 10 subjects, and missing values.*

reasonably robust, with actual sizes close to the nominal level. It is not appropriate to compare power here, as the actual test sizes have not all attained the nominal level.

2.4 Discussion

The simulation studies of the previous section show that, in general, as the sample size decreases or the model becomes complex, either by design or due to missing values:

(1) Reduced likelihood ratio tests of goodness of fit (and AICR/BICR) increasingly favour the unstructured covariance model in preference to the 'true' structure, so are not necessarily appropriate means of determining fit.

(2) The power of (small sample adjusted) tests of fixed effects using the unstructured covariance estimate is diminished in detecting departures from the null hypothesis.

(3) Where an exact small sample test of the fixed effects using the unstructured

Underlying 'True' Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (H_0: No Treatment Effect)		
		Size		
		χ^2	F	KR
Identity	Unstr	0.913	0.900	0.654
	Identity	0.065	0.048	0.048
Compound Symmetry	Unstr	0.921	0.912	0.690
	Comp Sym	0.077	0.058	0.054
AR1 ($\rho = 0.2$)	Unstr	0.910	0.900	0.660
	AR1	0.085	0.061	0.041
AR1 ($\rho = 0.8$)	Unstr	0.914	0.901	0.672
	AR1	0.062	0.042	0.024
Antedependence	Unstr	0.943	0.938	0.712
	Ante	0.410	0.382	0.137
Unstructured	Unstr	0.919	0.910	0.677
Unstr (QRE)	Unstr	0.926	0.913	0.698

Table 2.3.7: *Results from 1000 simulations of Design (C): A five treatment-five period Crossover Study with 10 subjects.*

model is not possible, the approximate test is unlikely to have nominal properties. That is the actual size of the test may not be fixed at the nominal level, making inferences uncertain.

These simulations show the importance in small sample studies, where the covariance structure plays a part in the estimation of the mean parameters, of finding an appropriately fitting model with few parameters, where one exists, for validity of inference. This is because models with fewer parameters are likely to be better estimated when there is less data available, than their large dimension counterparts such as the unstructured model. Where a covariance structure is badly estimated, it affects not only the estimates of the mean parameters, $\boldsymbol{\beta}$, recall (1.2.3), but also the standard errors which are used to make inferences about them using (1.3.1).

This raises questions about the role of the covariance structure in making inferences in a repeated measures context where the sample size is very small and the model or the covariance structure are complex. Such data are often highly non-stationary, with variances and covariances which change over time, so that low-dimensional (stationary) models such as compound symmetry or AR1 are unlikely to be adequate.

However, adopting the additional parameters of an antedependence or unstructured model may lead to a dramatic loss of power, and inferences which are invalid. This suggests that the usual methods of analyses for such data are unreliable in these settings.

2.5 Some Alternative Approaches

There is clearly a need for methods which accommodate the structure of the data, but do not lead to inferences which are invalid in very small sample settings. Two general approaches are suggested:

(1) To consider a covariance structure which offers a compromise between unstructured and structured forms. A structure which retains the relevant features of the data in a model which has a low number of parameters.

(2) To drop the covariance structure from estimation of the mean parameters, basing inference on their ordinary least squares estimates and using some consistent estimate of $\mathrm{Var}(\mathbf{y})$ to derive the standard errors.

Option (1) suggests adoption of a 'smoothed' covariance estimate, whereas option (2) suggests the use of a robust (sandwich) estimator or Box's modified F-test based on an ANOVA statistic may be appropriate. These approaches are considered in turn in Parts II and III of this thesis. Our interest is in finding appropriate methods for the analysis of very small samples of repeated measurements that are widely applicable across a range of settings, which lead to tests with nominal properties for general hypotheses involving the fixed effects parameters, and have acceptable power to detect departures from those hypotheses.

Part II

Smoothing the Covariance Structure

Chapter 3

Background: Smoothing Covariance Structures

3.1 Introduction

Part 1 demonstrated the need for a generalised approach to covariance structure modelling in repeated measures analysis; where often the primary interest in an analysis is the mean response but the efficient estimation of a covariance structure is necessary for inference. There is enough common structure in the variances and covariances of repeated measurements to suggest that it should be possible to formulate models that can adapt to a variety of settings. Common features are variances and covariances that change smoothly with changes in time and time lags, and correlations that tend to be smaller with larger time lags. The underlying smoothness that is commonly seen suggests smoothing more general structures in an adaptive way.

Diggle and Verbyla (1998) suggest smoothing the components of the variogram to provide an estimated covariance matrix. The variogram of the residual process $Z(t) = Y(t) - \mu(t)$

$$\gamma(s,t) = \frac{1}{2}\mathrm{E}\left[\{Z(s) - Z(t)\}^2\right], s \neq t \tag{3.1.1}$$

relates to the covariance function $G(s,t)$ of the data via

$$\gamma(s,t) = \frac{1}{2}\Big\{G(s,s) + G(t,t) - 2G(s,t)\Big\} \tag{3.1.2}$$

This implies $\gamma(t,t) = 0$ so that the variance function $\sigma^2(t)$ must be considered additionally to account for the measurement error in the data. For a set of longitudinal data (y_{ij}, t_{ij}): $i = 1, ..., m$; $j = 1, ..., n_i$, where y_{ij} is the jth measurement on subject i and t_{ij} is the time at which this measurement is made, with mean value function $\mu_i(t)$, the variogram cloud is defined as the set of points $(t_{ij}, t_{ik}, v_{ijk})$: $i = 1, ..., m$; $j = 1, ..., n_i$; $k > j$, where

$$v_{ijk} = \frac{1}{2}\Big[\{y_{ij} - \mu_i(t_{ij})\} - \{y_{ik} - \mu_i(t_{ik})\}\Big]^2 \tag{3.1.3}$$

Diggle and Verbyla use the variogram cloud as the input data for a two-dimensional non-parametric estimator for the variogram, and the squared residuals $z_{ij}^2 = \{y_{ij} - \mu(t_{ij})\}^2$ as the input data for an estimator for the variance function. These functions are separately smoothed using a kernel weighted local linear regression with bandwidths chosen to minimise a cross-validation criterion, and combined to give a smoothed covariance estimate

$$G(s,t) = \begin{cases} \frac{1}{2}\{\sigma^2(s) + \sigma^2(t)\} - \gamma(s,t), & s \neq t \\ \sigma^2(t), & s = t \end{cases} \tag{3.1.4}$$

This approach is difficult to implement and does not guarantee an estimate which is postive-definite, so has limited use in inference. However, it is suggested as a diagnostic tool for helping to specify a plausible parametric structure. It could also be useful in providing an estimated covariance structure for data which is highly unbalanced or in which observation times are not common to all subjects, and alternative approaches are not available.

3.2 A Bayesian Approach

Shrinkage estimation in which parameter estimates are pushed towards pre-determined or believable values has a natural setting in Bayesian statistics where such shrinking is a natural consequence of reliance on informative prior distributions. Such estimation is common in the context of the multivariate linear model, where differing mean estimates can be moved closer to an overall mean. Often this involves use of an informative prior distribution for the covariance matrix which influences the mean estimates, although in many applications direct shrinking of the covariance parameters is undertaken.

In small samples it is well understood that covariance matrices provided by maximum likelihood methods are unstable with the estimated eigenvalues of the underlying population covariance matrix being biased. Typically, the largest eigenvalues are overestimated whilst the smallest are underestimated. Many authors following Stein (1975) have sought to shrink the eigenvalues of the sample covariance matrix towards more plausible values (often a common value) resulting in a more robust estimate. These are orthogonally invariant estimates of the form $\hat{\mathbf{\Sigma}} = \mathbf{O}\mathbf{\Lambda}^*(\hat{\lambda})\mathbf{O}^T$, where $\mathbf{O}$ is the matrix of normalised eigenvectors, $\hat{\lambda}$ is the vector of sample eigenvectors and $\mathbf{\Lambda}^*(\hat{\lambda}) = \text{diag}(\lambda_1^*(\hat{\lambda}), ..., \lambda_p^*(\hat{\lambda}))$, where each λ_j^* is a real-valued nonnegative function. Stein's characteristic roots method sets $\lambda_j^*(\hat{\lambda}) = n\hat{\lambda}_j/\alpha_j$, where

$$\alpha_j = n - p + 1 + 2\hat{\lambda}_j \sum_{i \neq j} \frac{1}{\hat{\lambda}_j - \hat{\lambda}_i} \tag{3.2.1}$$

Others have sought to achieve such stability through a decomposition of the covariance matrix and placing prior distributions on the separate elements. Lin and Perlman (1985) and Barnard *et al.* (2000) model the covariance matrix in terms of standard deviations and correlations. (See Daniels and Kass (1999, 2001)).

Chen (1979) and others have proposed shrinking the unstructured sample covariance matrix towards a structured form, by using the structured form as the scale matrix

in the inverse Wishart prior. Increasing shrinkage towards the structured form can be achieved by adopting a large value for the associated degrees of freedom. Chen develops his approach for a multivariate linear model where he shrinks the sample covariance matrix towards a factor analysis form.

Chen's approach can be described as follows for data $\mathbf{Y}_i \sim \mathrm{N}(0, \boldsymbol{\Sigma})$, where $\mathbf{A} = \sum_i \mathbf{Y}_i^T \mathbf{Y}_i$, and $\hat{\boldsymbol{\Sigma}} = \mathbf{A}/n$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$. The unstructured covariance matrix $\boldsymbol{\Sigma}$ is modelled as an inverse Wishart distribution $\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1} \sim \mathbf{W}_p\{(\nu\boldsymbol{\Omega})^{-1}, \nu\}$, centred at $\boldsymbol{\Omega}$, a $p \times p$ structured covariance matrix. Ignorance (flat) priors are then adopted for the unknown hyperparameters in $\boldsymbol{\Omega}$ and for the degrees of freedom parameter ν.

Assuming $\boldsymbol{\Omega}$ and ν are known initially, we have the posterior distributions,

$$\begin{aligned} (\boldsymbol{\Lambda}|\text{data}) &\sim \mathbf{W}_p\{(\mathbf{A} + \nu\boldsymbol{\Omega})^{-1}, n + \nu\}, \text{and} \\ (\boldsymbol{\Sigma}|\text{data}) &\sim \mathbf{W}_p^{-1}\{(\mathbf{A} + \nu^*\boldsymbol{\Omega}^*), n + \nu^* + p + 1\} \end{aligned} \tag{3.2.2}$$

and it follows that the Bayes estimate of $\boldsymbol{\Sigma}$, the mode of the posterior density, is given by

$$\hat{\boldsymbol{\Sigma}}^* = \frac{n}{n + \nu^*}\hat{\boldsymbol{\Sigma}} + \frac{\nu^*}{n + \nu^*}\boldsymbol{\Omega}^* \tag{3.2.3}$$

which is a weighted average of the unstructured estimated matrix and the given structured matrix. The degrees of freedom parameter ν^* weights ones prior belief about the specified structure against the sample evidence.

ν^* and $\boldsymbol{\Omega}^*$ are calculated by reference to the joint density of the data $\mathbf{A}$ and the inverse covariance matrix $\boldsymbol{\Lambda}$

$$f(\mathbf{A}, \boldsymbol{\Lambda}|\boldsymbol{\Omega}, \nu) = f_1(\mathbf{A}|\boldsymbol{\Lambda}) f_2(\boldsymbol{\Lambda}|\boldsymbol{\Omega}, \nu) \tag{3.2.4}$$

from which we obtain the marginal density of $\mathbf{A}$

$$\mathrm{g}(\mathbf{A}|\mathbf{\Omega},\nu) = \int_{\{\mathbf{\Lambda}>0\}} f(\mathbf{A},\mathbf{\Lambda}|\mathbf{\Omega},\nu)d\mathbf{\Lambda} \tag{3.2.5}$$

Chen uses the EM algorithm for the hyperparameter estimation, where $(\mathbf{\Omega}^*,\nu^*)$ are chosen to maximise this marginal likelihood. By considering $(\mathbf{A},\mathbf{\Lambda})$ as the complete data with $\mathbf{\Lambda}$ unobservable and $\mathbf{A}$ alone incomplete . Given $(\mathbf{\Omega}^{(k)},\nu^{(k)})$ the current values of $(\mathbf{\Omega},\nu)$ after k cycles of the algorithm, we have

E step:

$$\mathbf{\Lambda}^{(k+1)} = E(\mathbf{\Lambda}|\mathbf{A},\mathbf{\Omega}^{(k)},\nu^{(k)}) = (\nu^{(k)}+n)(\mathbf{A}+\nu^{(k)}\mathbf{\Omega}^{(k)})^{-1} \tag{3.2.6}$$

M step:

Find $\mathbf{\Omega}^{(k+1)}$which maximises $H(\mathbf{\Omega}) = \log|\mathbf{\Omega}| - \mathrm{tr}(\mathbf{\Omega}\mathbf{\Lambda}^{(k+1)})$

Then, solve for ν (3.2.7)

$$g(\nu) - g(\nu^{(k)}+n) = -\log|\mathbf{\Omega}^{(k+1)}\mathbf{\Lambda}^{(k+1)}|$$

where, $\mathrm{g}(\nu) = p\log(\nu/2) - \sum_{j=1}^{p}\Psi\{(\nu+1-j)/2\}$ and $\Psi(t) = (d/dt)\log\Gamma(t)$ is the digamma function.

Chen shows (under the usual 'regularity conditions') that $\hat{\mathbf{\Sigma}}^*$ is a consistent estimate of $\mathbf{\Sigma}$, and that $\hat{\mathbf{\Sigma}}^*$ is asymptotically at least as efficient as the maximum likelihood estimate $\hat{\mathbf{\Sigma}}$ for all $\mathbf{\Sigma}$.

This approach is attractive in principle with one shrinkage parameter (degrees of freedom) for all elements of the covariance matrix being determined by the data. However, there are necessary restrictions in practice as the degrees of freedom must be greater than $p-1$ in order to maintain a proper prior, forcing the resulting estimate some way towards the structured form, even when the data suggest otherwise.

Chen's approach has been adopted by several authors, including Brown *et al.* (1994)

in the context of spatial interpolation. Also, Friedman (1989) takes an analogous approach in choosing between two covariance estimates for problems in discriminant analysis which are ill- or poorly-posed (i.e. have sample sizes below the number of parameters to be estimated). Friedman proposes a 'regularisation' between using the separate group covariance structure (quadratic discriminant analysis, QDA) and adopting a single pooled covariance structure for all groups (linear discriminant analysis, LDA). The regularised covariance structure reduces the high variance associated with small sample based estimation at the expense of bias. Two smoothing parameters are introduced, the first of which determines the amount of regularisation between QDA and LDA, whilst the second reduces the inherent bias in the process. These parameters are chosen to jointly minimise an estimate of the future misclassification risk.

Daniels and Kass (1999) also propose using an estimated covariance matrix which is a compromise between an unstructured and a parametric (diagonal) form. They achieve this through the introduction of two hierarchical priors (HPs) for the covariance matrix based on two different matrix decompositions. These are compared with the standard Bayesian approach of using a hierarchical Wishart prior, where a Wishart prior is introduced for $\mathbf{\Sigma}^{-1}$ which has an unknown scale matrix with a diagonal form and unknown degrees of freedom, which allows the data to determine the extent to which diagonality is supported.

Their first approach is to place a hierarchical prior on the correlations using the variance/correlation decomposition of the covariance matrix $\mathbf{\Sigma} = \text{diag}(\mathbf{S})\mathbf{R}\text{diag}(\mathbf{S})$, where $\mathbf{S}$ is a vector of standard deviations and $\mathbf{R}$ is the matrix of correlations. Daniels and Kass place a normal distribution on Fisher's z-transform of the correlations $0.5 \log\{(1-\rho)/(1+\rho)\} \sim \text{N}(0, \tau^2)$ which will shrink the correlations towards zero. These normal distributions are truncated to ensure that the resulting estimator is positive definite. Also, flat priors are placed on the logarithms of the diagonal elements of $\mathbf{\Sigma}$ and a 'uniform shrinkage' prior of the form $\pi(\tau^2) \propto (c+\tau^2)^{-2}$ is used for the unknown variance τ^2, where c is chosen to be the asymptotic variance of the

z-transform of the correlations, $1/(k-3)$, where k is the number of subjects.

The second approach is based on the spectral decomposition $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix of ordered eigenvalues and $\mathbf{P}$ is the orthogonal matrix which is the product of the unique rotation matrices defined by $p(p-1)/2$ Givens angles. Here, a normal distribution centered at zero is placed on the logit transformation of the Givens angles,

$$\log\left\{\frac{\pi/(2+\theta)}{\pi/(2-\theta)}\right\} \sim \mathrm{N}(0, \tau^2)$$

As with the correlation shrinkage method, flat priors are placed on the logarithms of eigenvalues of $\mathbf{\Sigma}$ and a prior of the 'uniform shrinkage' form is placed on the unknown variance τ^2. Here, however, the choice of the constant c is more difficult, because there is no expression for the asymptotic variance of the logit of θ. Daniels and Kass set $c = 1$.

The risk of the three estimators is compared to that given by standard priors (e.g. Jeffreys and Wishart) in a simulation study, using data $\mathbf{Y} \sim \mathrm{N}(0, \mathbf{\Sigma})$ for a variety of *true* covariance matrices ranging from the identity to somewhat ill-conditioned matrices. using the risk function $\mathbf{R}(\tilde{\mathbf{\Sigma}}, \mathbf{\Sigma}) = E_{\mathbf{\Sigma}}\{L(\tilde{\mathbf{\Sigma}}, \mathbf{\Sigma})\}$ associated with Stein's loss $L(\tilde{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \mathrm{tr}(\tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log \mid \tilde{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} \mid -p$, where $\tilde{\mathbf{\Sigma}}$ is the Bayes estimator.

The hierarchical Wishart prior performs well for matrices which are close to diagonal but struggles otherwise as the degrees of freedom parameter pushes against the boundary $p-1$, a property shared with the non-hierarchical Wishart prior. This gives the edge to the correlation and Givens-angle HPs where the parameter τ^2, which works analogously to the degrees of freedom parameter in determining the amount of shrinkage, may become arbitrarily small.

One drawback of these estimators however is the computational problems involved in sampling from such non-conjugate priors. Using Gibbs sampling when the full conditional of $\mathbf{\Sigma}$ is not inverse Wishart leads to a univariate approach generating

the full conditional of $\boldsymbol{\Sigma}$ componentwise requiring $p(p+1)/2$ evaluations. Approximations to the full conditional of $\boldsymbol{\Sigma}$ which allow it to be generated at once often require an expensive maximisation at each iteration and can also lead to problems because of the high correlation between estimated parameters. To avoid these problems, when modelling the covariance matrix in the second stage of a hierarchical model, Daniels and Kass propose a combination of approximation and importance sampling.

This approach is generalised in Daniels and Kass (2001), where the correlation shrinkage and rotation shrinkage estimators are developed to provide for shrinkage towards any structured form, and asymptotic distributions are adopted to simplify the computational aspects of their use. They consider specifically the use of these estimators in a fixed effects regression model with correlated errors, useful in repeated measures analysis.

$$\mathbf{Y}_i \sim \mathrm{N}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}) \tag{3.2.8}$$

Their approach is to first fit a model by maximum likelihood with an unstructured form for $\boldsymbol{\Sigma}$, and conditional on $\hat{\boldsymbol{\beta}}$ compute the observed information matrix based on one of the two parameterizations. They then propose a two level Normal-Normal model where the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is firstly approximated by a Normal distribution with variance the inverse of the observed information matrix, which is taken to be block diagonal. Then, a normal prior is used to shrink the correlations and eigenvalues(or variances and angles) towards a structured form estimated from the data. This requires the estimation of only two variance components (shrinkage parameters) from the data in addition to the parameters of the structured model.

For the correlation shrinkage estimator, the normal prior on the correlations used in Daniels and Kass (1999) is replaced by the more general $z(\rho) \sim \mathrm{N}(z(\rho_s), \tau_\rho^2)$, where ρ_s are the correlations under the assumed covariance structure. Using an asymptotic

approximation first proposed by Lin and Perlman (1985), the effect of this approach is to simply replace the sample correlations, given in the vector $\hat{\boldsymbol{\rho}}$, with

$$\tilde{\boldsymbol{\rho}} = z^{-1}\left([\mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}^{-1}+\hat{\tau}^2\mathbf{I}]^{-1}\hat{\tau}^2 z(\hat{\boldsymbol{\rho}})+[\mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}^{-1}+\hat{\tau}^2\mathbf{I}]^{-1}[\mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}^{-1}z(\hat{\boldsymbol{\rho}}_s)]\right) \quad (3.2.9)$$

where $\mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}$ is the observed information matrix for the z-transform of the correlations, and

$$\hat{\tau}^2_\rho = \{(z(\hat{\boldsymbol{\rho}}) - z(\boldsymbol{\rho}_s)\}^T \mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}\{z(\hat{\boldsymbol{\rho}}) - z(\boldsymbol{\rho}_s)\} - \frac{p(p-1)/2}{\mathrm{tr}[\mathrm{I}\{z(\hat{\boldsymbol{\rho}})\}]} \quad (3.2.10)$$

If it is required to shrink the covariance matrix rather than the correlation matrix, then the variances must also be shrunk using $\log(\hat{\sigma}^2) \sim \mathrm{N}(\log(\sigma^2_s), \tau^2_\sigma)$, which replaces the sample variances, given in the vector $\hat{\boldsymbol{\sigma}}^2$, with

$$\tilde{\boldsymbol{\sigma}}^2 = \log^{-1}\Big([\mathrm{I}\{\log(\hat{\boldsymbol{\sigma}}^2)\}^{-1} + \hat{\tau}^2\mathbf{I}]^{-1}\hat{\tau}^2\log(\hat{\boldsymbol{\sigma}}^2) + [\mathrm{I}\{\log(\hat{\boldsymbol{\sigma}}^2)\}^{-1} + \hat{\tau}^2\mathbf{I}]^{-1}\mathrm{I}\{\log(\hat{\boldsymbol{\sigma}}^2)^{-1}\log(\hat{\boldsymbol{\sigma}}^2_s)\}\Big) \quad (3.2.11)$$

where $\mathrm{I}\{\log(\hat{\boldsymbol{\sigma}}^2)\}$ is the observed information matrix for the logarithm of the variance. This assumes that the variances and correlations are asymptotically independent which is unlikely to be the case, although Daniels and Kass find that this is not a problem in practice. Also, this estimator is not guaranteed to be positive definite even where the unstructured and structured estimates from which they are derived are.

For the rotation shrinkage estimator, the prior distribution for the logit of the Givens angles becomes

$$\mathrm{logit}(\theta) = \log\left(\frac{\pi/2+\theta}{\pi/2-\theta}\right) \sim \mathrm{N}(\mathrm{logit}(\theta_s), \tau^2_\theta)$$

where θ_s are the Givens angles under the assumed covariance structure. Again, assuming an asymptotic distribution for the Givens angles, a simple shrinkage estimator is found by replacing the sample Givens angles, given in the vector $\hat{\boldsymbol{\theta}}$, with

$$\tilde{\boldsymbol{\theta}} = \text{logit}^{-1}\Big([\text{I}\{\text{logit}(\hat{\boldsymbol{\theta}})\}^{-1} + \hat{\tau}^2\mathbf{I}]^{-1}\hat{\tau}^2\text{logit}(\hat{\boldsymbol{\theta}}) + [\text{I}\{\text{logit}(\hat{\boldsymbol{\theta}})\}^{-1} + \hat{\tau}^2\mathbf{I}]^{-1}\text{I}\{\text{logit}(\hat{\boldsymbol{\theta}})^{-1}\text{logit}(\hat{\boldsymbol{\theta}}_s)\}\Big) \quad (3.2.12)$$

where $I\{\text{logit}(\hat{\boldsymbol{\theta}})\}$ is the observed information matrix for the logit of the Givens angles, and

$$\hat{\tau}^2_\theta = \{z(\hat{\boldsymbol{\theta}}) - z(\boldsymbol{\theta}_\text{s})\}^T I\{z(\hat{\boldsymbol{\theta}})\}\{z(\hat{\boldsymbol{\theta}}) - z(\boldsymbol{\theta}_\text{s})\} - \frac{p(p-1)/2}{\text{tr}[\text{I}\{z(\hat{\boldsymbol{\theta}})\}]} \quad (3.2.13)$$

To shrink the eigenvalues they use

$$\tilde{\lambda}_i = \exp\left\{ \frac{2/n}{2/n + \hat{\tau}^2}\log(\hat{\lambda}_s) + \frac{\hat{\tau}^2}{2/n + \hat{\tau}^2}\log(\hat{\lambda}_i) \right\} \quad (3.2.14)$$

Daniels and Kass refer to shrinkage of the eigenvalues alone, using (3.2.14) as the structured log eigenvalue estimator. This is considered to be more reliable in practice than the full rotation shrinkage estimator as the Givens angles in the rotation method are slow to converge towards their asymptotic distribution, making these estimates unreliable in small samples.

These two 'shrinkage to structure' approaches are compared with conventional eigenvalue shrinkage approaches through a simulation study, which calculates the percentage reduction in average loss for each estimator using the loss function of Daniels and Kass (1999). This is the difference between the risk of the sample covariance matrix and the risk of the estimator divided by the risk of the sample covariance matrix. It is concluded that both estimators have smaller risk than the unstructured maximum likelihood estimator in small to medium sized samples and better than

the structured estimator where the hypothesised structure is incorrect. They also consider the effect of such estimates on the estimated regression parameters, and report gains in the mean squared error of these estimates. The correlation shrinkage method is preferred, and results from this estimator are good particularly when the chosen structure is close to the true one. In large samples, both estimators do well as the data dominates the specified structure, whether or not it is correctly chosen. In this case the shrinkage estimates and unstructured REML estimates become indistinguishable.

The estimators are described as optimally efficient, in the sense that the covariance estimates are consistent and the resulting regression coefficients are consistent and asymptotically efficient.

Daniels and Kass recommend that the eigenvalues of the unstructured covariance matrix are first shrunk using the Stein estimator to provide a more robust estimate of the sample covariance matrix, before shrinking towards some parametric structure. They suggest the following approach to shrinking the estimated covariance matrix.

(1) Fit the model using an unstructured covariance matrix, $\hat{\mathbf{\Sigma}}$

(2) Shrink the eigenvalues of the unstructured estimator to obtain a more stable estimate, $\hat{\mathbf{\Sigma}}_{st}$

(3) Fit the model assuming some covariance structure, $\hat{\mathbf{\Sigma}}_{s}$.

(4) Compute estimates of the parameters that determine the amount of shrinkage using $\hat{\mathbf{\Sigma}}_{st}$ and $\hat{\mathbf{\Sigma}}_{s}$

(5) Compute a shrinkage estimator of the covariance matrix, $\hat{\mathbf{\Sigma}}_{sh}$ using the estimates in steps 2-4.

(6) Refit the model conditional on $\hat{\mathbf{\Sigma}}_{sh}$

The advantage of this approach is that it is easily computed using existing statistical software such as SAS proc mixed (SAS Institute, 1999) with steps (4) and (5)

requiring a simple macro. The structured covariance matrix which the unstructured matrix is smoothed towards may be chosen according to the usual criteria (AIC, BIC or likelihood ratio tests). However, the approximations and asymptotic results needed to simplify the computations means that these estimates are less reliable in the small sample situations where they are most likely to be useful. Also, it is not clear if independent modelling of the variances and correlations in the sample covariance matrix is appropriate in a repeated measures or longitudinal data setting.

Daniels and Kass suggest alternative parameterizations should be pursued for shrinking the covariance matrix towards a structure which do not involve asymptotic approximations and are computationally efficient. They suggest also that the approach be extended to include linear mixed effects models. Daniels and Cressie (2001) and Daniels and Pourahmadi (2002) provide further examples of this approach in applications involving time series and longitudinal data respectively.

Chapter 4

A Direct Smoothing Approach

4.1 Introduction

Although there have been recent advances in Bayesian techniques for smoothing between unstructured and structured covariance estimates where the data are not sufficient for a well-fitting unstructured model, such methods are not accessible to those involved in everyday analysis. Where straight-forward 'plug-in' estimates are provided by way of approximations, these are not reliable in the small sample setting where they are most likely to be adopted. The aim of this research is to consider the development of adaptive-estimation techniques in a frequentist setting that attempt to retain the degree of structure appropriate for the analysis of small samples.

We approach the problem of choosing between unstructured and structured covariance estimators by attempting to smooth directly between them, with the amount of smoothing being determined by the data.

An intuitive first approach, analogous to that of Chen (1979), is to consider the linear combination of the unstructured and structured covariance estimators $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{S}}$, given by

$$\hat{\boldsymbol{\Sigma}}_{\lambda} = (1-\lambda)\hat{\boldsymbol{\Sigma}} + \lambda\hat{\boldsymbol{\Sigma}}_{\mathrm{S}}, \qquad 0 \leq \lambda \leq 1 \tag{4.1.1}$$

where the smoothing parameter, λ, determines the degree of smoothing. That is, for $\lambda = 0$ we have the unstructured estimate, whilst at the other extreme $\lambda = 1$ results in the structured estimate being adopted.

Initially, simple forms of $\hat{\mathbf{\Sigma}}_S$ such as identity and compound symmetry are considered. The effects of smoothing towards such structures will be illustrated here and in subsequent chapters by reference to the following data set.

4.1.1 Cardiac Enzyme Data

Data were collected to compare the effects of four preserving liquids on the enzyme content of 23 preserved dog hearts. The four test liquids were defined by the presence and absence of two components A and B, and repeated measurements of the adenosine triphosphate (ATP) level as a percentage of total enzyme (%ATP) were made on 9 occasions at one and two hourly intervals.

The sample variance-correlation matrix for these data is shown in Table 4.1.1. (This shows variances on the diagonal, with covariances above and correlations below). Subject and mean profiles for the four preserving liquids are shown in Figures 4.1.1 and 4.2.2 respectively.

29.01	12.26	10.41	21.39	8.76	27.68	28.62	16.08	8.40
0.36	**39.86**	10.25	27.70	7.72	30.22	45.62	15.20	1.38
0.31	0.26	**38.71**	-1.20	7.97	9.84	30.14	-5.48	0.76
0.39	0.43	-0.02	**105.12**	-1.95	41.60	66.27	34.87	-2.33
0.24	0.18	0.19	-0.03	**45.31**	39.36	37.15	13.65	42.26
0.43	0.40	0.13	0.34	0.49	**141.36**	99.53	72.19	105.61
0.42	0.57	0.38	0.51	0.44	0.66	**159.70**	73.00	59.21
0.27	0.21	-0.08	0.30	0.18	0.54	0.51	**126.22**	79.08
0.12	0.02	0.01	-0.02	0.50	0.71	0.37	0.56	**158.01**

Table 4.1.1: *Sample Variance-Correlation Matrix for the Cardiac Enzyme Data*

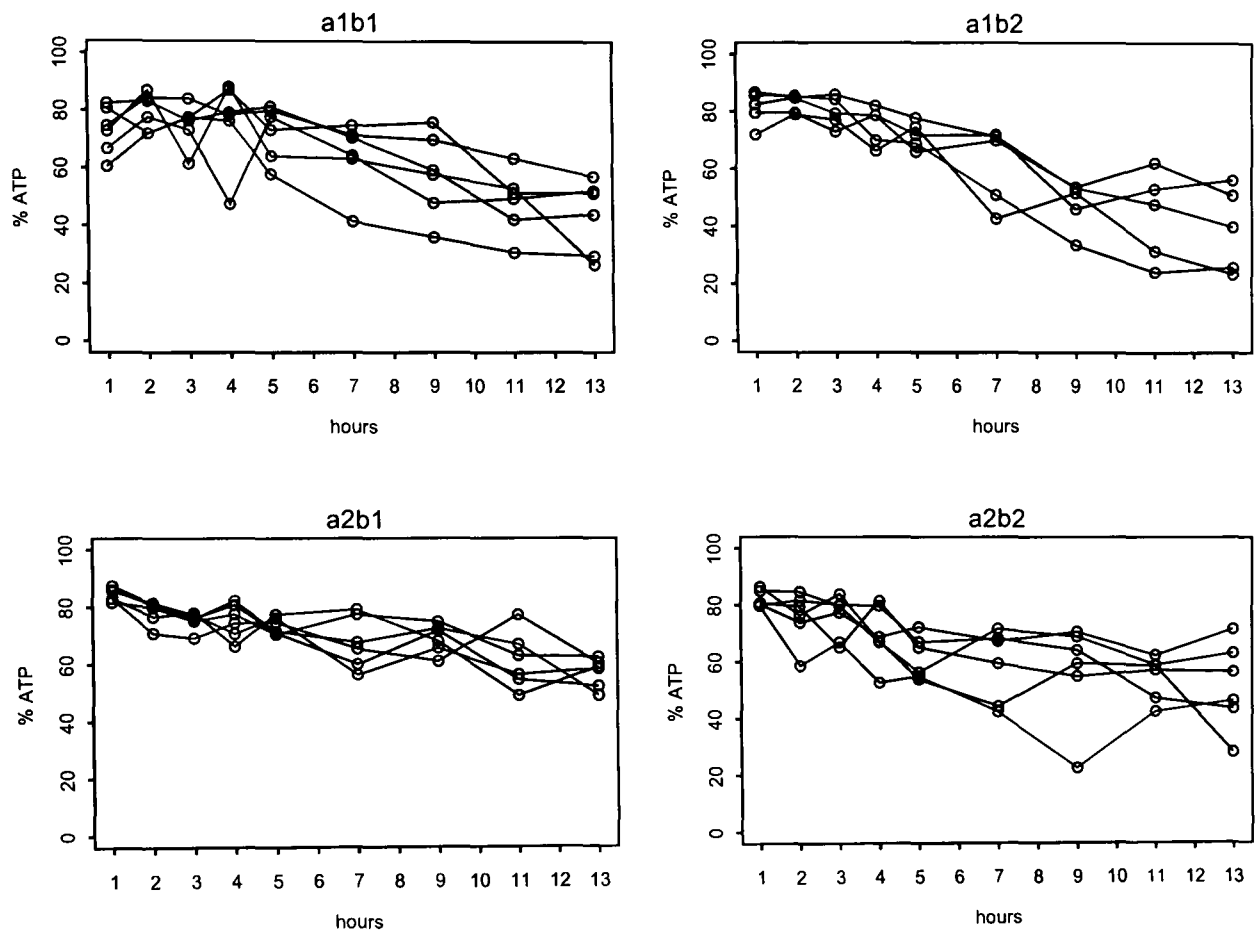


Figure 4.1.1: *Cardiac Enzyme data: Subject Profiles for the Four preserving liquids*

4.2 Smoothing Towards an Identity Structure

We begin by smoothing the unstructured estimated covariance structure $\hat{\mathbf{\Sigma}}$ towards an identity structure. To motivate development of this technique, attention is restricted in the first instance to balanced and complete data, for which the REML log-likelihood of $\mathbf{\Sigma}$ collapses to the log-likelihood of the Wishart distribution. That is $\hat{\mathbf{\Sigma}} = \mathbf{S}$, the sample covariance estimate, follows a Wishart distribution, $m\hat{\mathbf{\Sigma}} \sim \mathbf{W}_p(\mathbf{\Sigma}, m)$. It follows that $\hat{\mathbf{\Sigma}}_{\mathrm{I}} = (1/p)\mathrm{tr}(\mathbf{S})\mathbf{I}$, the identity (structured) maximum likelihood estimator of $\mathbf{\Sigma}$ given the sample covariance matrix $\mathbf{S}$.

Thus, identity smoothing is defined as

$$\hat{\mathbf{\Sigma}}_\lambda = (1-\lambda)\hat{\mathbf{\Sigma}} + \lambda\frac{\mathrm{tr}(\hat{\mathbf{\Sigma}})}{p}\mathbf{I} \qquad 0 \leq \lambda \leq 1 \tag{4.2.1}$$

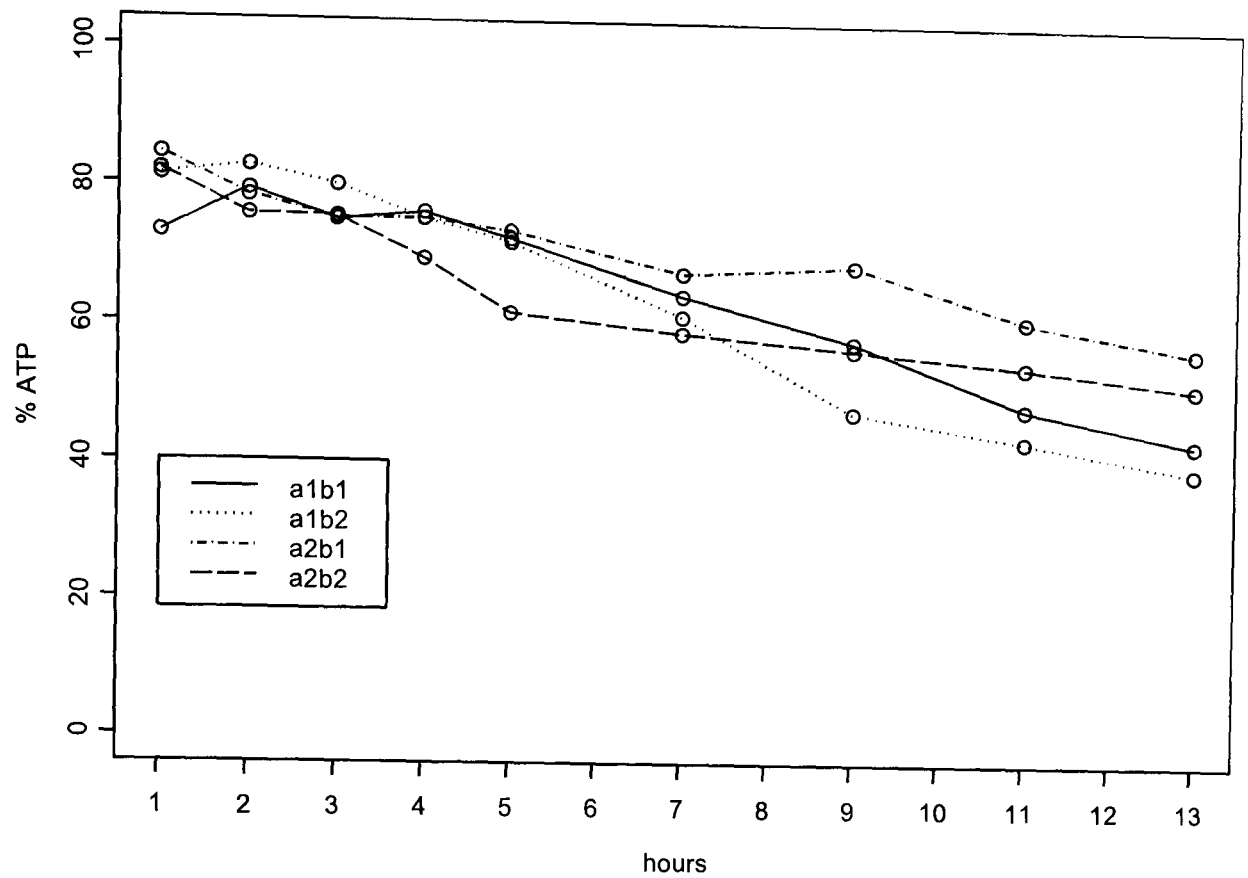


Figure 4.1.2: *Cardiac Enzyme data: Mean Profiles*

That is, given a suitable value of the parameter λ, the unstructured matrix is smoothed towards an identity structure with a constant variance across time (that is the average of the unstructured variances). i.e. $\text{tr}(\hat{\boldsymbol{\Sigma}}_\lambda) = \text{tr}(\hat{\boldsymbol{\Sigma}}) = p$, for all $0 \leq \lambda \leq 1$, so that the average variance is maintained.

Figure 4.2.1 shows the effect of smoothing the variances for various values of λ for the Cardiac Enzyme data.

Two questions are immediately apparent about such a smoothing approach:

(1) How is the fit of the smoothed matrix to be assessed? For example, can the (reduced) log-likelihoods of both $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_\lambda$ be compared?

(2) What is the effective number of parameters in the smoothed matrix? Is there an intermediate number of parameters between the $p(p+1)/2$ of the unstructured model and the single parameter of the identity structure for $0 \leq \lambda \leq 1$?

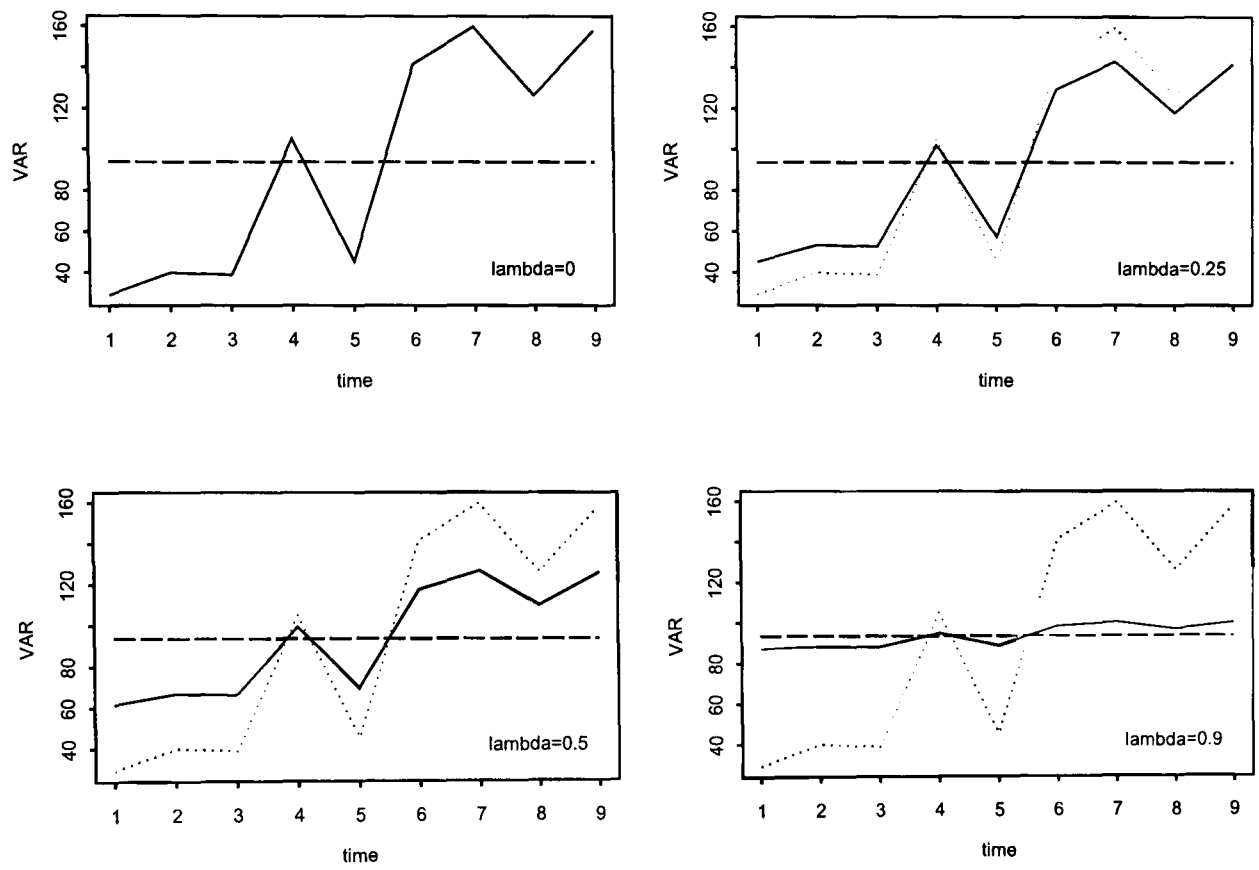


Figure 4.2.1: *Variances by Time for the Cardiac Enzyme data. Direct Smoothing to an Identity form. Legend: Solid Line, smoothed estimate; Dashed Line, identity form; Dotted Line, unstructured form.*

Appropriate answers to these questions may lead to an AIC type penalised likelihood measure, which can be used to determine an appropriate value of λ.

An alternative procedure would be to assess the variabilty and inherent bias in the smoothed estimator, and to choose λ to minimise the mean squared error of a suitable function of $\hat{\mathbf{\Sigma}}_\lambda$.

$$\begin{aligned} \text{M.S.E.}\{f(\hat{\mathbf{\Sigma}}_\lambda)\} &= \text{Var}\{f(\hat{\mathbf{\Sigma}}_\lambda)\} + [\text{bias}\{f(\hat{\mathbf{\Sigma}}_\lambda)\}]^2 \\ &= \text{Var}\{f(\hat{\mathbf{\Sigma}}_\lambda)\} + \{f(\hat{\mathbf{\Sigma}}_\lambda) - f(\mathbf{\Sigma}_\lambda)\}^2 \end{aligned} \tag{4.2.2}$$

4.2.1 A Likelihood Based Approach to finding λ

We begin by considering a likelihood based approach. Note that $\hat{\boldsymbol{\Sigma}}$ maximises the Wishart log-likelihood

$$\ell(\boldsymbol{\Sigma};\mathbf{S}) = \text{const.} - \frac{m}{2}\log|\boldsymbol{\Sigma}| + \frac{1}{2}(m-p-1)\log|\mathbf{S}| - \frac{m}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) \tag{4.2.3}$$

so that equivalently, ignoring constants, $\hat{\boldsymbol{\Sigma}}$ $(= \mathbf{S})$ minimises (minus the log-likelihood)

$$-\ell(\boldsymbol{\Sigma};\mathbf{S}) \equiv \log|\boldsymbol{\Sigma}| + \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) \tag{4.2.4}$$

Writing ϕ_i, $(i = 1,\ldots,p)$, to be the eigenvalues of $\hat{\boldsymbol{\Sigma}}$, we have that the minimised value of (4.2.4) is given by

$$\log|\hat{\boldsymbol{\Sigma}}| + \text{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}) = \log(\prod_{i=1}^{p}\phi_i) + \text{tr}(\mathbf{I}) = \sum_{i=1}^{p}\log(\phi_i) + p \tag{4.2.5}$$

The smoothed estimator $\hat{\boldsymbol{\Sigma}}_\lambda$ does not follow a Wishart distribution, but it is informative to consider the value of the (maximised) log-likelihood of $\boldsymbol{\Sigma}$ when $\boldsymbol{\Sigma}_\lambda$ is substituted via the 'smaller is better' formulation

$$-\ell_{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Sigma}}_\lambda;\mathbf{S}) = \log|\hat{\boldsymbol{\Sigma}}_\lambda| + \text{tr}(\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\hat{\boldsymbol{\Sigma}}) = \log\left[\prod_{i=1}^{p}\{(1-\lambda)\phi_i + \lambda\overline{\phi}\}\right] + \sum_{i=1}^{p}\left[\frac{\phi_i}{(1-\lambda)\phi_i + \lambda\overline{\phi}}\right] \tag{4.2.6}$$

where $\overline{\phi} = \sum_{i=1}^{p}\phi_i/p$ is the mean eigenvalue of $\hat{\boldsymbol{\Sigma}}$.

Consider now the difference in the (maximised) log-likelihoods of the unstructured and smoothed estimates, $\ell(\hat{\boldsymbol{\Sigma}}_\lambda;\mathbf{S}) - \ell(\hat{\boldsymbol{\Sigma}};\mathbf{S})$, for $0 \leq \lambda \leq 1$.

$$\begin{aligned}
\ell(\hat{\boldsymbol{\Sigma}}_\lambda;\mathbf{S}) - \ell(\hat{\boldsymbol{\Sigma}};\mathbf{S}) &= \log|\hat{\boldsymbol{\Sigma}}_\lambda| + \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\hat{\boldsymbol{\Sigma}}) - \{\log|\hat{\boldsymbol{\Sigma}}| + \mathrm{tr}(\mathbf{I})\} \\
&= \log\left[\prod_{i=1}^{p}\{(1-\lambda)\phi_i + \lambda\overline{\phi}\}\right] + \sum_{i=1}^{p}\left[\frac{\phi_i}{(1-\lambda)\phi_i + \lambda\overline{\phi}}\right] - \log(\prod_{i=1}^{p}\phi_i) - p \\
&= \sum_{i=1}^{p}\left\{\frac{(1-\lambda)\phi_i + \lambda\overline{\phi}}{\phi_i}\right\} - \sum_{i=1}^{p}\log\left\{\frac{\phi_i}{(1-\lambda)\phi_i + \lambda\overline{\phi}}\right\} - p \\
&= \sum_{i=1}^{p}\{\Phi_i - \log(\Phi_i)\} - p
\end{aligned}$$

where $\Phi_i = \dfrac{\phi_i}{(1-\lambda)\phi_i + \lambda\overline{\phi_i}}$ (are the eigenvalues of $\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\hat{\boldsymbol{\Sigma}}$).

That is, the difference in the maximised log-likelihoods of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_\lambda$ may be written

$$\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\hat{\boldsymbol{\Sigma}}) - \log|\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\hat{\boldsymbol{\Sigma}}| - p \tag{4.2.7}$$

Also, it is easily seen that for $0 \leq \lambda \leq 1$

$$\begin{aligned}
\ell_{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Sigma}}_\lambda;\mathbf{S}) - \ell(\hat{\boldsymbol{\Sigma}};\mathbf{S}) &= \sum_{i=1}^{p}\{\Phi_i - \log(\Phi_i)\} - p \\
&\geq \sum_{i=1}^{p} 1 - p = p - p = 0
\end{aligned} \tag{4.2.8}$$

(since $\log(x)$ is concave and below all of its tangents, so that $x - 1 \geq \log(x)$. i.e. $x - \log(x) \geq 1$, for all $x > 0$). That is, as we might expect, the fit of the smoothed estimate to the data reduces as the degree of smoothing, determined by λ, increases. However, this reduction in fit may be offset by penalising the likelihood measure by the effective number of parameters in the smoothed estimate $\hat{\boldsymbol{\Sigma}}_\lambda$, leading to a measure which can be used to determine a suitable value of λ.

An intuitive estimate of the degrees of freedom associated with $\hat{\boldsymbol{\Sigma}}_\lambda$ is given by

$$(1-\lambda)\left\{\frac{p(p+1)}{2}\right\} + \lambda \tag{4.2.9}$$

which is a simple weighted combination of the $\frac{1}{2}p(p+1)$ parameters of the unstructured estimate and the single parameter of the identity form.

This suggests a penalised measure of the form

$$\ell_{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Sigma}}_\lambda; \mathbf{S}) - g \tag{4.2.10}$$

where g is a suitably chosen function of the effective number of parameters in $\hat{\boldsymbol{\Sigma}}_\lambda$.

Unfortunately this method proves to be too heavily dependent on a suitable choice of function g. Also, it is not obvious that the effective number of parameters in $\hat{\boldsymbol{\Sigma}}_\lambda$ will change linearly in λ as (4.2.9) suggests. Indeed, work presented in the next chapter suggests that the smoothed estimator differs more greatly for low values of λ than for large.

Such a penalised approach is not pursued further in this context. A penalised likelihood approach to the problem of smoothing between unstructured and structured covariance estimators is presented in Chapter 5.

4.2.2 Finding λ via the M.S.E. of a function of $\hat{\boldsymbol{\Sigma}}_\lambda$

An alternative approach to the direct smoothing problem is to consider a method for choosing λ in $\hat{\boldsymbol{\Sigma}}_\lambda$ via a suitable measure of the mean squared error (M.S.E.). To do this, however, we need to define a meaningful measure of $\mathrm{Var}(\hat{\boldsymbol{\Sigma}}_\lambda)$ which leads to a scalar quantity which can be minimised in the objective function. Two measures of variation suggested from multivariate statistics are $\mathrm{tr}(\boldsymbol{\Sigma})$, the total variance, and $|\boldsymbol{\Sigma}|$, the generalised variance.

As we have seen, when smoothing towards an identity structure, $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_\lambda) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}})$, for all $0 \leq \lambda \leq 1$, so that there is clearly no advantage in considering the former. However it may be possible to find λ through mimimising the mean squared error of the generalised variance via M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$.

M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$

The mean squared error of the quantity $\ln|\hat{\boldsymbol{\Sigma}}_\lambda|$ is defined

$$\text{M.S.E.}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|) = \text{Var}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|) + \{\text{bias}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)\}^2 \tag{4.2.11}$$

Here, approximate expressions can be obtained for $\text{E}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$, using $\text{E}(\hat{\boldsymbol{\Sigma}})=\boldsymbol{\Sigma}$ so that $\text{E}(\hat{\sigma}_{ij}) = \sigma_{ij}$, i.e. using the rule that for a general variable X, $\text{E}\{f(X)\} = f\{\text{E}(X)\}$. Also, $\text{Var}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$ can be evaluated by considering the Taylor expansion of $\ln|\hat{\boldsymbol{\Sigma}}_\lambda|$ as a function of $\text{vec}(\hat{\boldsymbol{\Sigma}}) = (\hat{\sigma}_{11}, \hat{\sigma}_{21}, \ldots, \hat{\sigma}_{pp})$, via the 'delta' method. (See, for example, Stuart and Ord (1994), Chapter 10). That is, we have

$$\begin{aligned}\text{M.S.E.}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|) &= \text{Var}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|) + \{\text{bias}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)\}^2 \\ &\approx \Delta^{\text{T}}\mathbf{W}\Delta + (\ln|\boldsymbol{\Sigma}_\lambda| - \ln|\boldsymbol{\Sigma}|)^2\end{aligned} \tag{4.2.12}$$

where $\mathbf{W} = \text{Cov}\{\text{vec}(\hat{\boldsymbol{\Sigma}}), \text{vec}(\hat{\boldsymbol{\Sigma}})\}$, and

$$\Delta^{\text{T}} = \left(\frac{\partial \ln|\hat{\boldsymbol{\Sigma}}_\lambda|}{\partial \hat{\sigma}_{11}}\bigg|_{\hat{\sigma}_{11}=\sigma_{11}}, \frac{\partial \ln|\hat{\boldsymbol{\Sigma}}_\lambda|}{\partial \hat{\sigma}_{21}}\bigg|_{\hat{\sigma}_{21}=\sigma_{21}}, \ldots, \frac{\partial \ln|\hat{\boldsymbol{\Sigma}}_\lambda|}{\partial \hat{\sigma}_{pp}}\bigg|_{\hat{\sigma}_{pp}=\sigma_{pp}} \right) \tag{4.2.13}$$

Also, following Searle (1982), defining $\mathbf{K}$ as the 'commutative' matrix, $\mathbf{K} = \sum_{i,j=1}^{p}(\mathbf{H}_{ij} \otimes \mathbf{H}_{ij}^T)$, (where $\mathbf{H}_{ij}$ has i,jth element 1 and all other elements zero), we have

$$\mathbf{W} = \text{Cov}\{\text{vec}(\hat{\boldsymbol{\Sigma}}), \text{vec}(\hat{\boldsymbol{\Sigma}})\} = \frac{1}{m}(\mathbf{I}_{p^2} + \mathbf{K})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \tag{4.2.14}$$

Recall for an $(m \times n)$ matrix $\mathbf{A}$, with i,jth element a_{ij}, and a $(p \times q)$ matrix $\mathbf{B}$, the kronecker product $\mathbf{A} \otimes \mathbf{B}$ is given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \ldots & a_{mn}\mathbf{B} \end{bmatrix}$$

Since,

$$\frac{\partial(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)}{\partial\hat{\sigma}_{ij}} = \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}_\lambda}{\partial\hat{\sigma}_{ij}}\right) \tag{4.2.15}$$

and

$$\begin{aligned}\frac{\partial\hat{\boldsymbol{\Sigma}}_\lambda}{\partial\hat{\sigma}_{ij}} &= \frac{\partial}{\partial\hat{\sigma}_{ij}}\left\{(1-\lambda)\hat{\boldsymbol{\Sigma}} + \frac{\lambda}{p}\mathrm{tr}(\hat{\boldsymbol{\Sigma}})\mathbf{I}\right\} \\ &= (1-\lambda)(\mathbf{H}_{ij} + \mathbf{H}_{ji} - \delta_{ij}\mathbf{H}_{ij}) + \delta_{ij}\frac{\lambda}{p}\mathbf{I}\end{aligned} \tag{4.2.16}$$

where, as before, $\mathbf{H}_{ij}$ is a matrix of 0's with 1 in the i,jth position and $\delta_{ij} = 1$ when $i = j$, else $\delta_{ij} = 0$. We have the following expression (elementwise) for Δ.

$$\left.\frac{\partial\ln|\hat{\boldsymbol{\Sigma}}_\lambda|}{\partial\hat{\sigma}_{11}}\right|_{\hat{\sigma}_{ij}=\sigma_{ij}} = \mathrm{tr}\left[\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\left\{(1-\lambda)(\mathbf{H}_{ij} + \mathbf{H}_{ji} - \delta_{ij}\mathbf{H}_{ij}) + \delta_{ij}\frac{\lambda}{p}\mathbf{I}\right\}\right] \tag{4.2.17}$$

This allows us to construct M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$ to plot, or we can find λ by minimising (4.2.12) using a generic optimization routine in *SAS* or *S-Plus*.

Another approach is to consider M.S.E.$(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$ where, for a non-null vector $\mathbf{a}$ (with $\mathbf{a}^T\mathbf{a} = 1$), $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$ is a linear combination of the elements of $\hat{\boldsymbol{\Sigma}}_\lambda$. This is developed below.

M.S.E.$(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$

λ is chosen to minimise the mean squared error of the quantity $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$, given by

$$\text{M.S.E.}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) = \mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) + \{\mathrm{bias}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})\}^2 \tag{4.2.18}$$

Now,

$$\begin{aligned}\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) &= \mathrm{Var}\{(1-\lambda)\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a} + \lambda\frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}})}{p}\} \\ &= (1-\lambda)^2\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}) + \frac{\lambda^2}{p^2}\mathrm{Var}\{\mathrm{tr}(\hat{\boldsymbol{\Sigma}})\} + 2(1-\lambda)\frac{\lambda}{p}\mathrm{Cov}\{\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}, \mathrm{tr}(\hat{\boldsymbol{\Sigma}})\}\end{aligned}$$

so that, noting that $\hat{\boldsymbol{\Sigma}}$ follows a Wishart distribution, $m\hat{\boldsymbol{\Sigma}} \sim W_p(\Sigma, m)$, $\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$ can be evaluated by reference to the standard results

$$\frac{\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}}{\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}} \sim \chi^2_m \quad \text{and} \quad \mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = \sum_{i=1}^{p}\phi_i\chi^2_m$$

where ϕ_i are the eigenvalues of $\boldsymbol{\Sigma}$, (see, for example, Anderson (1958) or Muirhead (1982)). However, we have the complicated covariance term $\mathrm{Cov}\{\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}, \mathrm{tr}(\hat{\boldsymbol{\Sigma}})\}$ to deal with. A simpler approach is to consider the alternative writing of $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$ as

$$\begin{aligned}\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a} = (1-\lambda)\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a} + \lambda\frac{tr(\boldsymbol{\Sigma})}{p} &= \mathrm{tr}[\hat{\boldsymbol{\Sigma}}\{(1-\lambda)\mathbf{a}\mathbf{a}^T + \frac{\lambda}{p}\mathbf{I}\}] \\ &= \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{A})\end{aligned} \tag{4.2.19}$$

where

$$\mathbf{A} = (1-\lambda)\mathbf{a}\mathbf{a}^T + \frac{\lambda}{p}\mathbf{I} \tag{4.2.20}$$

Using this form, we can determine the following expressions for the mean and variance of $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$

$$\begin{aligned}\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) &= \mathrm{Var}\{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{A})\} \\ &= \frac{2}{m}\mathrm{tr}(\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A})\end{aligned} \tag{4.2.21}$$

so that,

$$\begin{aligned}\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) &= \frac{2}{m}\mathrm{tr}\left[\boldsymbol{\Sigma}\{(1-\lambda)\mathbf{a}\mathbf{a}^T+\frac{\lambda}{p}\mathbf{I}\}\boldsymbol{\Sigma}\{(1-\lambda)\mathbf{a}\mathbf{a}^T+\frac{\lambda}{p}\mathbf{I}\}\right]\\ &= \frac{2}{m}\mathrm{tr}\left[(1-\lambda)^2(\boldsymbol{\Sigma}\mathbf{a}\mathbf{a}^T)^2+\frac{\lambda}{p}(1-\lambda)\boldsymbol{\Sigma}\mathbf{a}\mathbf{a}^T\boldsymbol{\Sigma}+\frac{\lambda}{p}(1-\lambda)\boldsymbol{\Sigma}^2\mathbf{a}\mathbf{a}^T+\frac{\lambda^2}{p^2}\boldsymbol{\Sigma}^2\right]\\ &= \frac{2}{m}\left[(1-\lambda)^2(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2+2\frac{\lambda}{p}(1-\lambda)\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}+\frac{\lambda^2}{p^2}\mathrm{tr}(\boldsymbol{\Sigma}^2)\right]\end{aligned} \tag{4.2.22}$$

Also,

$$\begin{aligned}\mathrm{bias}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}) &= \mathrm{E}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a})-\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}\\ &= \frac{\lambda}{p}\mathrm{tr}(\boldsymbol{\Sigma})-\lambda\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}\end{aligned} \tag{4.2.23}$$

This leads to the following (quadratic) expression in λ for the mean squared error.

$$\begin{aligned}\mathrm{M.S.E.}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}) = \frac{2}{m}\Bigg\{&(1-\lambda)^2(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2+2\frac{\lambda}{p}(1-\lambda)\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}+\frac{\lambda^2}{p^2}\mathrm{tr}(\boldsymbol{\Sigma}^2)\Bigg\}\\ &+\frac{\lambda^2}{p^2}\{\mathrm{tr}(\boldsymbol{\Sigma})\}^2-2\frac{\lambda^2}{p}\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}\{\mathrm{tr}(\boldsymbol{\Sigma})\}+\lambda^2(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2\end{aligned} \tag{4.2.24}$$

Recalling that $\mathrm{tr}(\boldsymbol{\Sigma})=\sum_{i=1}^{p}\phi_i$ and $\mathrm{tr}(\boldsymbol{\Sigma}^2)=\sum_{i=1}^{p}\phi_i^2$, the R.H.S. can be written,

$$\begin{aligned}&\frac{1}{mp}\{2p(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2-4\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}+2\mathrm{Var}\{\phi_i\}+2(\mathrm{E}\{\phi_i\})^2+mp(\mathrm{E}\{\phi_i\}-\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2\}\lambda^2\\ &\qquad-\frac{4}{mp}\{p(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2-\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}\}\lambda+\frac{2}{m}(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2\end{aligned} \tag{4.2.25}$$

where $\{\phi_i\}$ are the eigenvalues of $\boldsymbol{\Sigma}$. It follows that, for a suitable choice of $\mathbf{a}$, $\mathrm{M.S.E.}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a})$ is minimised by

$$\lambda=\frac{2p(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2-2\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}}{2p(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2-4\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a}+2\mathrm{Var}\{\phi_i\}+2(\mathrm{E}\{\phi_i\})^2+mp(\mathrm{E}\{\phi_i\}-\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2} \tag{4.2.26}$$

But how should **a** be chosen?

Noting that $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}$ is bounded within the interval $[\phi_{\min}, \phi_{\max}]$, with **a** the corresponding eigenvector, (c/f Principal Components Analysis), it seems sensible to take $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} = \phi_*$, a (particular) eigenvalue of $\boldsymbol{\Sigma}$. That is, if **a** is an eigenvector of $\boldsymbol{\Sigma}$, we have $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} = \phi_*$, the corresponding eigenvalue, and we have

$$\lambda = \frac{2(p-1)\phi_*^2}{2(p-2)\phi_*^2 + 2\text{Var}\{\phi_i\} + 2(\text{E}\{\phi_i\})^2 + mp(\text{E}\{\phi_i\} - \phi_*)^2} \tag{4.2.27}$$

Writing $A = \dfrac{\text{Var}\{\phi_i\}}{\phi_*^2}$ and $B = \dfrac{\text{E}\{\phi_i\}}{\phi_*}$, we obtain

$$\lambda = \frac{2(1 - \frac{1}{p})}{2(1 - \frac{2}{p}) + \frac{2}{p}(A + B^2) + m(B-1)^2}$$

so that it would seem sensible to choose ϕ_* to be the largest eigenvalue of $\boldsymbol{\Sigma}$. As $\phi_* \to \text{E}\{\phi_i\}$, so that $\text{Var}\{\phi_i\} \to 0$, we have $\lambda \to 1$, since the underlying structure will be close to an identity form. It is also seen that $\lambda \to 0$ as $m \to \infty$ as we would intuitively expect.

The above approaches to determining the value of the smoothing parameter λ for directly smoothing towards an identity structure will be tested in Section 4.4 through a simulation study.

4.3 Smoothing Towards a Compound Symmetry Structure

In the context of repeated measures, where data often display high dependencies, an identity (independence) form may be considered too restrictive as a suitable choice of 'smoothing structure'. Hence, consideration is also given to smoothing towards a compound symmetry structure, which allows for a uniform correlation between

observations on a subject.

Again, for data which is balanced and complete, we have

$$\hat{\boldsymbol{\Sigma}}_{\text{CS}} = \frac{1}{p(p-1)}\left[\{\text{tr}(\mathbf{S}\mathbf{1}\mathbf{1}^T) - \text{tr}(\mathbf{S})\}\mathbf{1}\mathbf{1}^T + \{p\text{tr}(\mathbf{S}) - \text{tr}(\mathbf{S}\mathbf{1}\mathbf{1}^T)\}\mathbf{I}\right] \tag{4.3.1}$$

the compound symmetry (structured) maximum likelihood estimator of $\boldsymbol{\Sigma}$ given the sample covariance matrix $\mathbf{S}$. In this case the smoothed estimate $\hat{\boldsymbol{\Sigma}}_\lambda$ preserves the total of the diagonal entries, that is $\text{tr}(\hat{\boldsymbol{\Sigma}}_\lambda) = \text{tr}(\hat{\boldsymbol{\Sigma}})$ for all $0 \leq \lambda \leq 1$, and also the total of the off-diagonal entries, $\text{tr}(\hat{\boldsymbol{\Sigma}}_\lambda \mathbf{1}\mathbf{1}^T) = \text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T)$.

Then, for $0 \leq \lambda \leq 1$, compound symmetry smoothing is defined

$$\hat{\boldsymbol{\Sigma}}_\lambda = (1-\lambda)\hat{\boldsymbol{\Sigma}} + \frac{\lambda}{p(p-1)}\left[\{\text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T) - \text{tr}(\hat{\boldsymbol{\Sigma}})\}\mathbf{1}\mathbf{1}^T + \{p\text{tr}(\hat{\boldsymbol{\Sigma}}) - \text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T)\}\mathbf{I}\right] \tag{4.3.2}$$

As with the identity smoothing of the previous section, expressions can be obtained for λ via the minimisation of the mean squared error of the quatntities $\ln|\hat{\boldsymbol{\Sigma}}_\lambda|$ and $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$. Attention here is restricted to a statement of comparable results to those given for identity smoothing in subsection 4.2.2.

M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$

To obtain the mean squared error of $\ln|\hat{\boldsymbol{\Sigma}}_\lambda|$, for smoothing towards a compound symmetry structure, we proceed as in the previous section, with

$$\begin{aligned}\frac{\partial\hat{\boldsymbol{\Sigma}}_\lambda}{\partial\hat{\sigma}_{ij}} &= \frac{\partial}{\partial\hat{\sigma}_{ij}}\left((1-\lambda)\hat{\boldsymbol{\Sigma}} + \frac{\lambda}{p(p-1)}\left[\{\text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T) - \text{tr}(\hat{\boldsymbol{\Sigma}})\}\mathbf{1}\mathbf{1}^T + \{p\text{tr}(\hat{\boldsymbol{\Sigma}}) - \text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T)\}\mathbf{I}\right]\right) \\ &= (1-\lambda)(\mathbf{H}_{ij} + \mathbf{H}_{ji} - \delta_{ij}\mathbf{H}_{ij}) + \frac{\lambda}{p(p-1)}\left[\{(2-\delta_{ij}) - \delta_{ij}\}\mathbf{1}\mathbf{1}^T + \{p\delta_{ij} - (2-\delta_{ij})\}\mathbf{I}\right]\end{aligned} \tag{4.3.3}$$

so that,

$$\left.\frac{\partial \ln|\hat{\boldsymbol{\Sigma}}_\lambda|}{\partial \hat{\sigma}_{11}}\right|_{\hat{\sigma}_{ij}=\sigma_{ij}} = \mathrm{tr}\left\{\hat{\boldsymbol{\Sigma}}_\lambda^{-1}\left((1-\lambda)(\mathbf{H}_{ij}+\mathbf{H}_{ji}-\delta_{ij}\mathbf{H}_{ij}) + \frac{\lambda}{p(p-1)}\left[2(1-\delta_{ij})\mathbf{1}\mathbf{1}^T + \{\delta_{ij}(p+1)-2\}\mathbf{I}\right]\right)\right\} \quad (4.3.4)$$

M.S.E.$(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$

Again, the value of λ is chosen to minimise the mean squared error of the quantity $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$. Proceeding as before, but for compound symmetry smoothing, we can write

$$\begin{aligned}\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a} &= (1-\lambda)\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a} + \frac{\lambda}{p(p-1)}\left[\{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T)-\mathrm{tr}(\hat{\boldsymbol{\Sigma}})\}\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a} - \{p\,\mathrm{tr}(\hat{\boldsymbol{\Sigma}})-\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T)\}\right] \\ &= (1-\lambda)\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a} + \frac{\lambda}{p(p-1)}\left[(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T) - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\mathrm{tr}(\hat{\boldsymbol{\Sigma}})\right] \\ &= \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{A})\end{aligned} \quad (4.3.5)$$

where $\mathbf{A}$ is now the matrix

$$\mathbf{A} = (1-\lambda)\mathbf{a}\mathbf{a}^T + \frac{\lambda}{p(p-1)}\left\{(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\mathbf{1}\mathbf{1}^T - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\mathbf{I}\right\} \quad (4.3.6)$$

Then, the variance and bias of $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$ are given as

$$\begin{aligned}\mathrm{Var}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) &= \frac{2}{m}\mathrm{tr}(\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}) \\ &= \frac{2}{m}\mathrm{tr}\left(\left[(1-\lambda)\boldsymbol{\Sigma}\mathbf{a}\mathbf{a}^T + \frac{\lambda}{p(p-1)}\left\{(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\boldsymbol{\Sigma}\mathbf{1}\mathbf{1}^T - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\boldsymbol{\Sigma}\right\}\right]^2\right)\end{aligned} \quad (4.3.7)$$

and,

$$\begin{aligned}\text{bias}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a}) &= \mathrm{E}\left[(1-\lambda)\mathbf{a}^T\hat{\boldsymbol{\Sigma}}\mathbf{a} + \frac{\lambda}{p(p-1)}\left\{(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{1}\mathbf{1}^T) - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\text{tr}(\hat{\boldsymbol{\Sigma}})\right\}\right] \\ &\quad - \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} \\ &= -\lambda\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} + \frac{\lambda}{p(p-1)}\left\{(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\text{tr}(\boldsymbol{\Sigma}\mathbf{1}\mathbf{1}^T) - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\text{tr}(\boldsymbol{\Sigma})\right\}\end{aligned} \tag{4.3.8}$$

which again leads to a (quadratic) expression in λ for the mean squared error. This is given by

$$\begin{aligned}\text{M.S.E.}(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}) = \frac{1}{mp^2(p-1)^2}&\left\{2p^2(p-1)^2A^2 - 4p(p-1)B + 2C + mD^2\right\}\lambda^2 \\ &- \frac{4}{mp(p-1)}\left\{p(p-1)A^2 - B\right\}\lambda + \frac{2}{m}A^2\end{aligned} \tag{4.3.9}$$

where,

$$\begin{aligned}A &= \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} \\ B &= (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2 - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a} \\ C &= (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)^2(\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})^2 - 2(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\mathbf{a}^T\boldsymbol{\Sigma}^2\mathbf{a} + (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)^2\text{tr}(\boldsymbol{\Sigma}^2) \\ D &= p(p-1)\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} - \{(\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-1)\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} - (\mathbf{a}^T\mathbf{1}\mathbf{1}^T\mathbf{a}-p)\text{tr}(\boldsymbol{\Sigma})\}\end{aligned} \tag{4.3.10}$$

It is easily shown that the M.S.E. is minimised by choosing

$$\lambda = \frac{2p^2(p-1)^2A^2 - 2p(p-1)B}{2p^2(p-1)^2A^2 - 4p(p-1)B + 2C + mD^2} \tag{4.3.11}$$

4.4 A Simulation Study

Here we attempt to compare the estimates of the smoothing parameter λ, arising from the various direct smoothing techniques discussed, using simulated data. The datasets arising from the Pilot Study of Chapter 2, Section 2.2, are used initially since they are complete and balanced, as our methods based on the Wishart likelihood necessitate. That is, we have 1000 datasets, each comprising two (identical) treatment groups of size 8 (m=16) measured at p=5 time points arising from seven underlying covariance structures including identity, compound symmetry, AR1 (low and high correlation), antedependence and two 'unstructured' forms.

Since the methods developed are based on minimisation of the mean squared error of a function of the smoothed estimate $\hat{\mathbf{\Sigma}}_\lambda$, and for simulated data the underlying ('true') covariance structure is necessarily known, it is possible to check the performance of the mean squared error approach in determining an appropriate value of the smoothing parameter λ, against an empirical measure, *viz.*

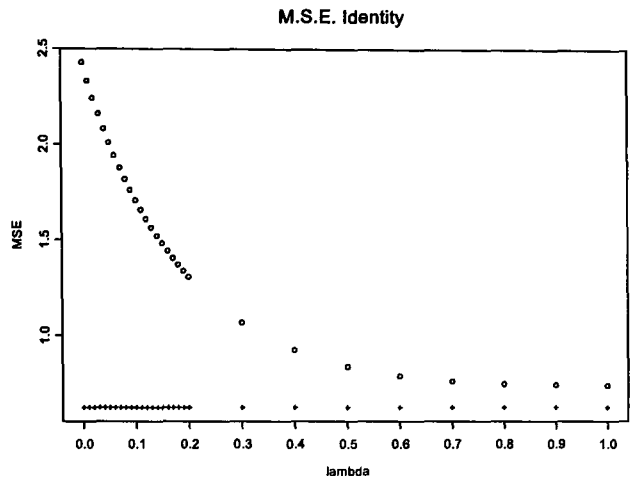
$$\text{Empirical M.S.E.}\{f(\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a})\} = \frac{1}{N}\sum_{i=1}^{N}\{f(\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a}) - f(\mathbf{a}^T\mathbf{\Sigma}\mathbf{a})\}^2 \qquad (4.4.1)$$

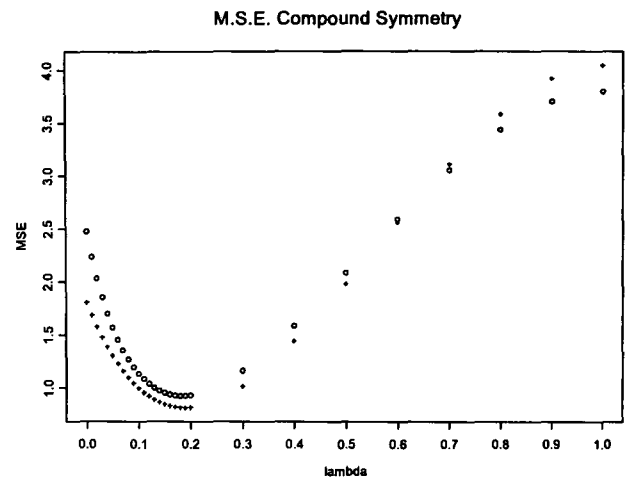
where N (=1000) is the number of simulations. This can be compared with the theoretically derived M.S.E. substituting the known underlying covariance structures in the approaches developed in sections 4.2 and 4.3.

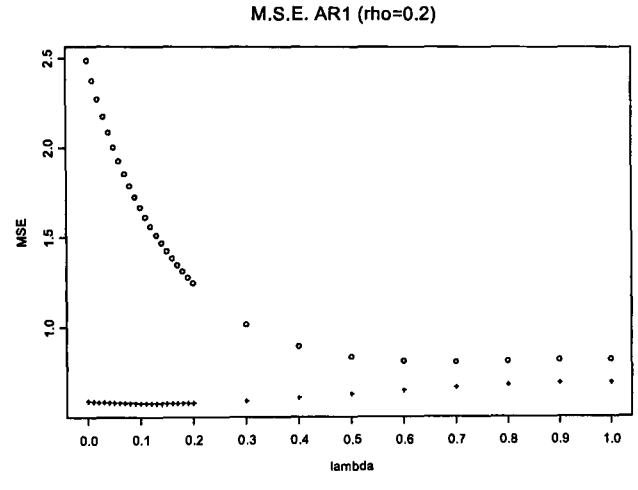
We begin by assessing the use of M.S.E.($\ln|\hat{\mathbf{\Sigma}}_\lambda|$) in providing an estimate of the smoothing parameter λ.

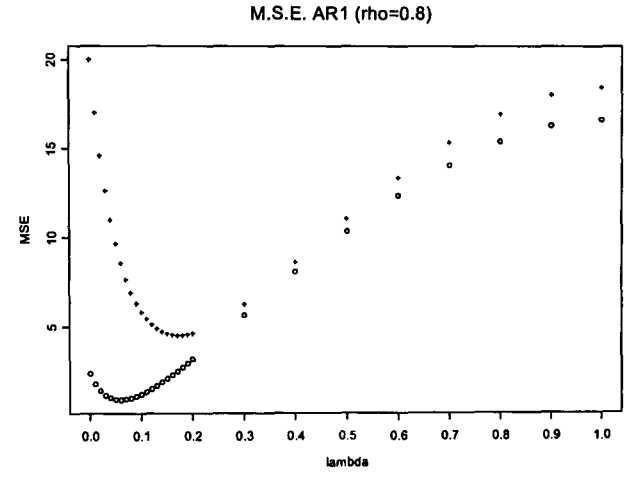
M.S.E.($\ln|\hat{\mathbf{\Sigma}}_\lambda|$)

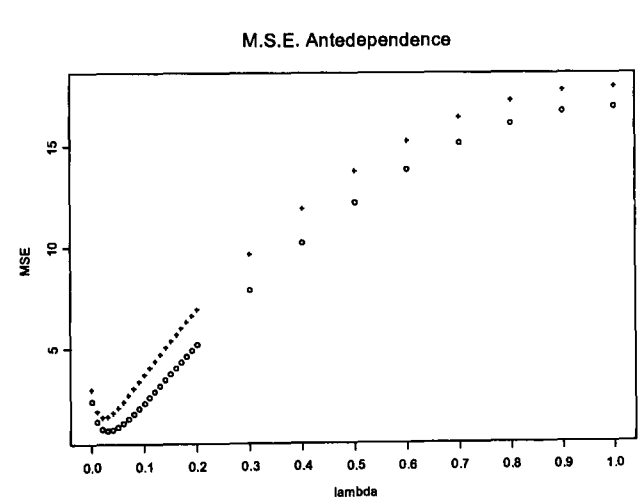
Figures 4.4.1 and 4.4.2 show plots of the empirically and theoretically calculated M.S.E. for data arising from the underlying covariance structures for smoothing towards an identity and compound symmetry form respectively.

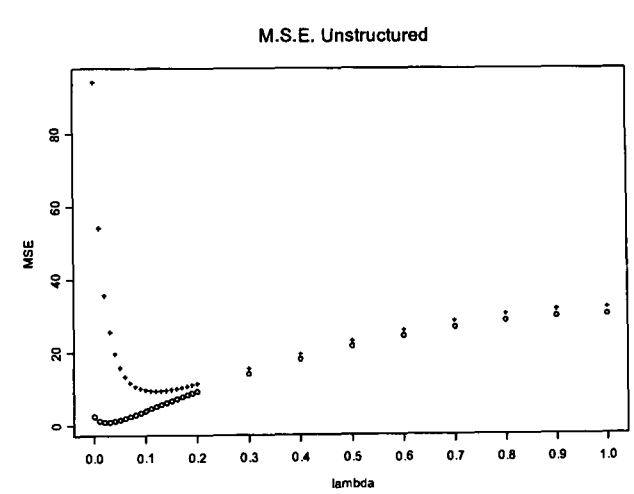












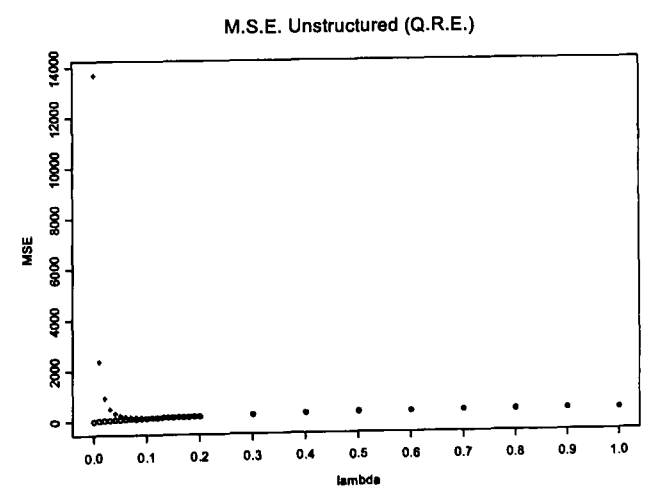


Figure 4.4.1: *Plots of both Empirical and Theoretical M.S.E. functions for given* λ *in* $\text{M.S.E.}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$, *Identity Smoothing. 1000 simulations of the Pilot Study. Legend:* o *Empirical,* + *Theoretical*

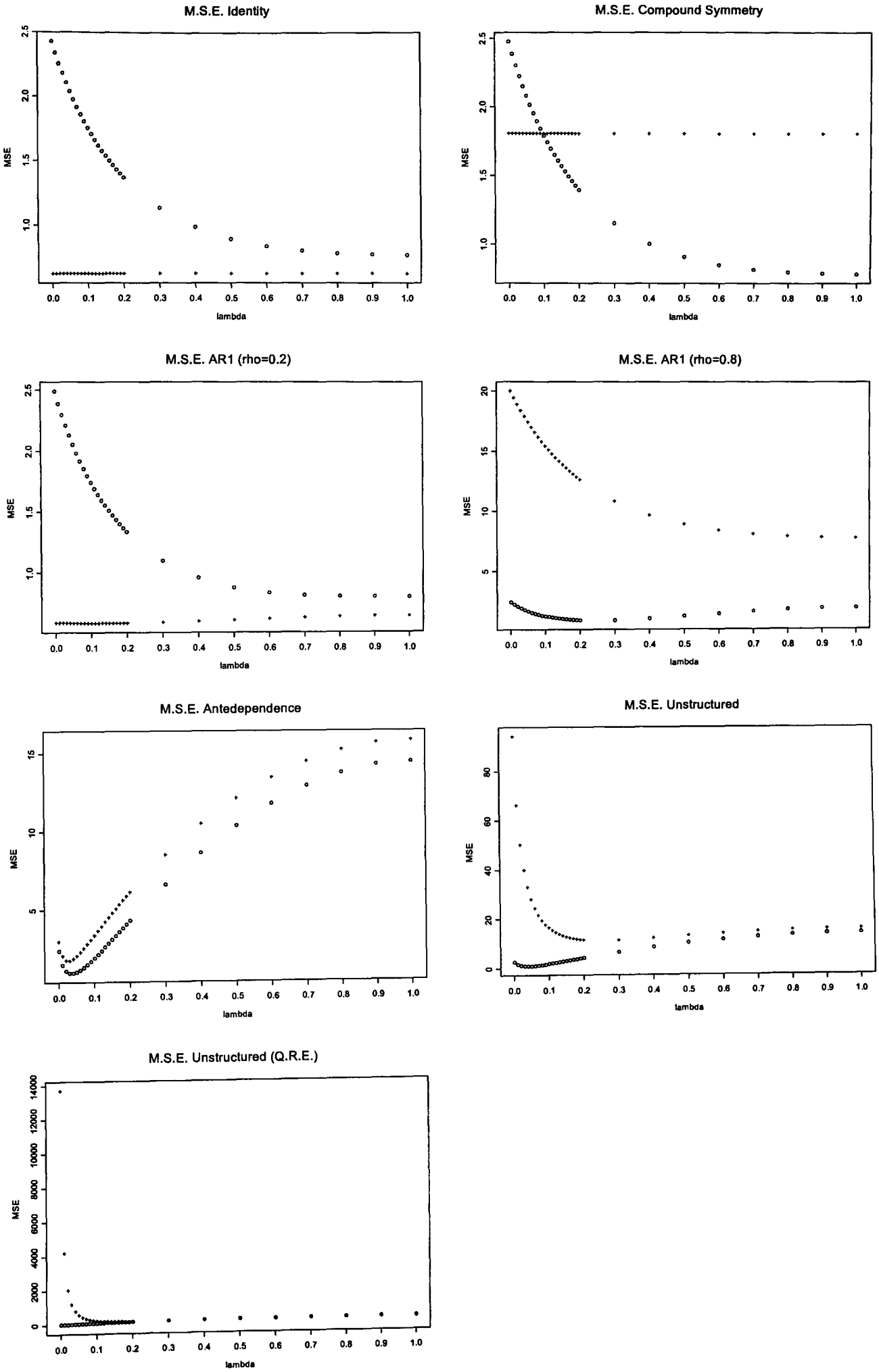


Figure 4.4.2: *Plots of both Empirical and Theoretical M.S.E. functions for given* λ *in* M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$, *Compound Symmetry Smoothing. 1000 simulations of the Pilot Study. Legend:* o *Empirical,* + *Theoretical*

Underlying ('True') Covariance Structure	λ Known$\boldsymbol{\Sigma}$	λ Estimated from Data			
		Mean	S.D.	Min	Max
Identity Smoothing					
Identity	-	0.086	0.024	0.016	0.157
Compound Symmetry	0.189	0.147	0.024	0.037	0.190
AR1 ($\rho = 0.2$)	0.133	0.092	0.026	0.025	0.163
AR1 ($\rho = 0.8$)	0.172	0.159	0.015	0.088	0.192
Antedependence	0.023	0.031	0.015	0.003	0.097
Unstructured	0.251	0.117	0.018	0.061	0.162
Unstruct. (Q.R.E.)	0.104	0.101	0.005	0.089	0.128
Compound Symmetry Smoothing					
Identity	-	0.092	0.031	0.015	0.269
Compound Symmetry	-	0.536	0.320	0.034	1.000
AR1 ($\rho = 0.2$)	0.125	0.102	0.038	0.021	0.327
AR1 ($\rho = 0.8$)	1.000	0.748	0.285	0.113	1.000
Antedependence	0.026	0.037	0.022	0.003	0.228
Unstructured	0.119	0.258	0.126	0.065	1.000
Unstruct. (Q.R.E.)	0.190	0.185	0.011	0.148	0.258

Table 4.4.1: *Minimising values of* λ *in* M.S.E.$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$, *Identity and Compound Symmetry Smoothing. 1000 simulations of the Pilot Study.*

Two things are apparent from these plots. Firstly, that the theoretically calculated M.S.E. gives a flat profile when smoothing towards the actual 'true' underlying structure. Secondly, that there is some difference between the theoretically and empirically calculated mean squared errors. These features are repeated across the range of data offered by the Pilot study.

Further investigations show that the latter point is explained by the Taylor Series approximation for $\text{Var}(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$ which is not close for small sample sizes rendering such theoretically determined estimates unreliable. That is increasing m, the number of subjects, results in a closer match between the theoretically and empirically calculated M.S.E.s.

Using this method to calculate λ from each of the 1000 datasets arising from the seven underlying covariance structures gives the results in Table 4.4.1.

The table shows that little smoothing is suggested for direct smoothing towards an identity structure, with a maximum value of λ suggested from 1000 simulations of each of seven underlying covariance structures of 0.192. Smoothing towards a

compound symmetry structure results in a greater degree of smoothing as we might expect. The amount of smoothing being generally greater for data arising from structures close to compound symmetry, such as the AR1 structure with a high correlation ($\rho = 0.8$), with a mean value of λ of 0.748 from 1000 such datasets.

M.S.E.$(\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a})$

Figures 4.4.3 and 4.4.4 show plots of the theoretically and empirically calculated M.S.E. for data arising from the underlying covariance structures for smoothing towards an identity and compound symmetry form respectively, using $\phi_* = \phi_{\max}$, the largest eigenvalue of $\mathbf{\Sigma}$.

It is seen that there is close convergence between the theoretical and empirical M.S.E.s with minimising values of the smoothing parameter λ close to that we would intuitively expect. That is, for smoothing towards an identity structure we obtain $\lambda = 1$ for data arising from a 'known' underlying identity structure and significantly lower smoothing parameter estimates from structures which are known to be far from independence.

A similar pattern is noted for smoothing towards a compound symmetry form, although the flat profiles for data arising from underlying structures close to compound symmetry should be noted. This is because the eigenvector corresponding to the largest eigenvalue $\phi_{\max}$ is proportional to $\mathbf{1}$, the unity vector, for which $\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a}$ is constant for all $0 \leq \lambda \leq 1$.

Estimates of the smoothing parameter λ for 1000 simulations from each of the underlying structures are given in Table 4.4.2.

These values match those of the plots. For data arising from underlying covariance structures which are close to independence, such as the Identity or an AR1 form with a low correlation, a higher value of the smoothing parameter is suggested when smoothing towards an identity structure, as we might expect.

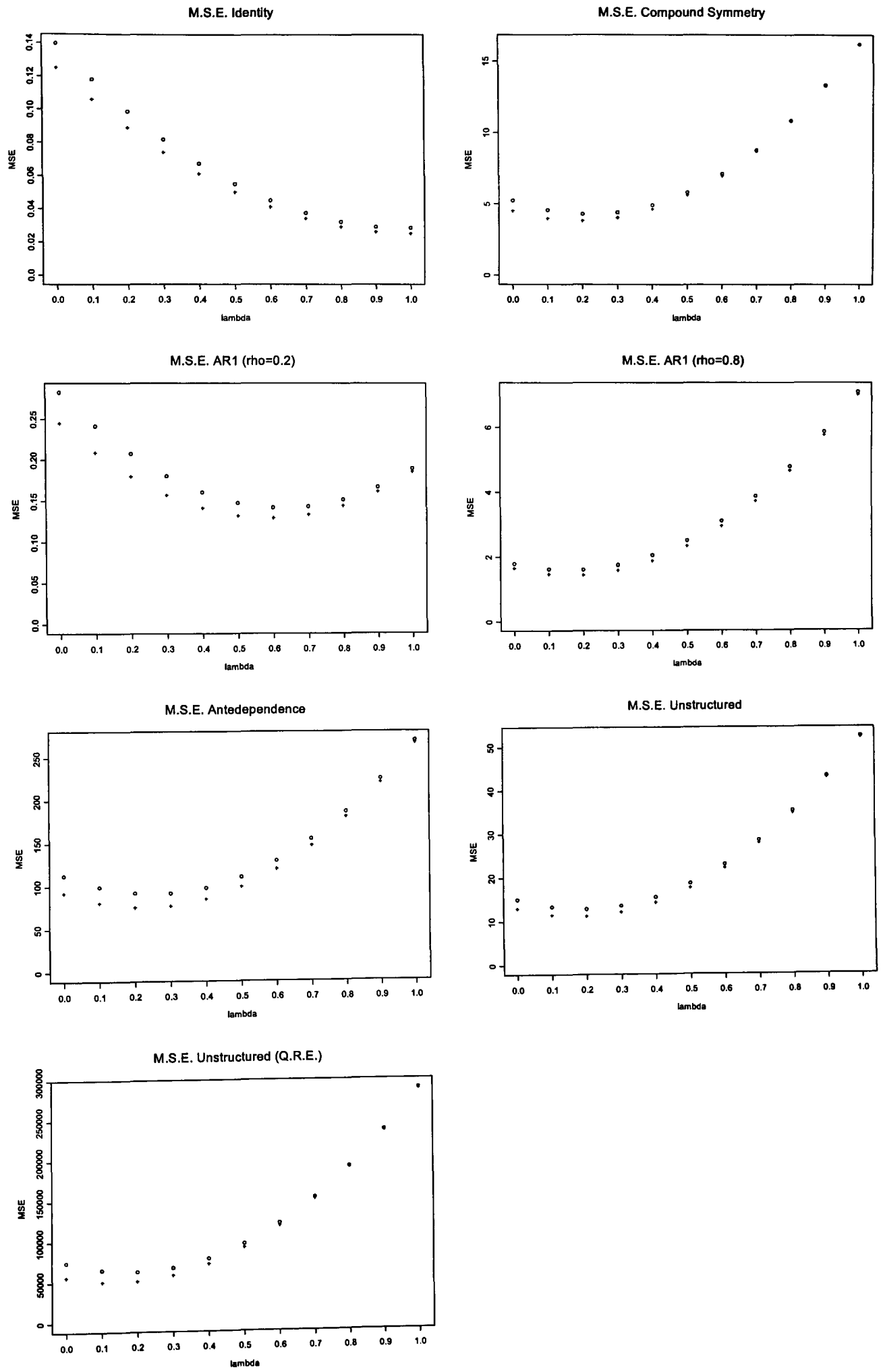


Figure 4.4.3: *Plots of both Empirical and Theoretical M.S.E. functions for given* λ *in* M.S.E.$(\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a})$, *Identity Smoothing. 1000 simulations of the Pilot Study. Legend:* o *Empirical,* + *Theoretical*

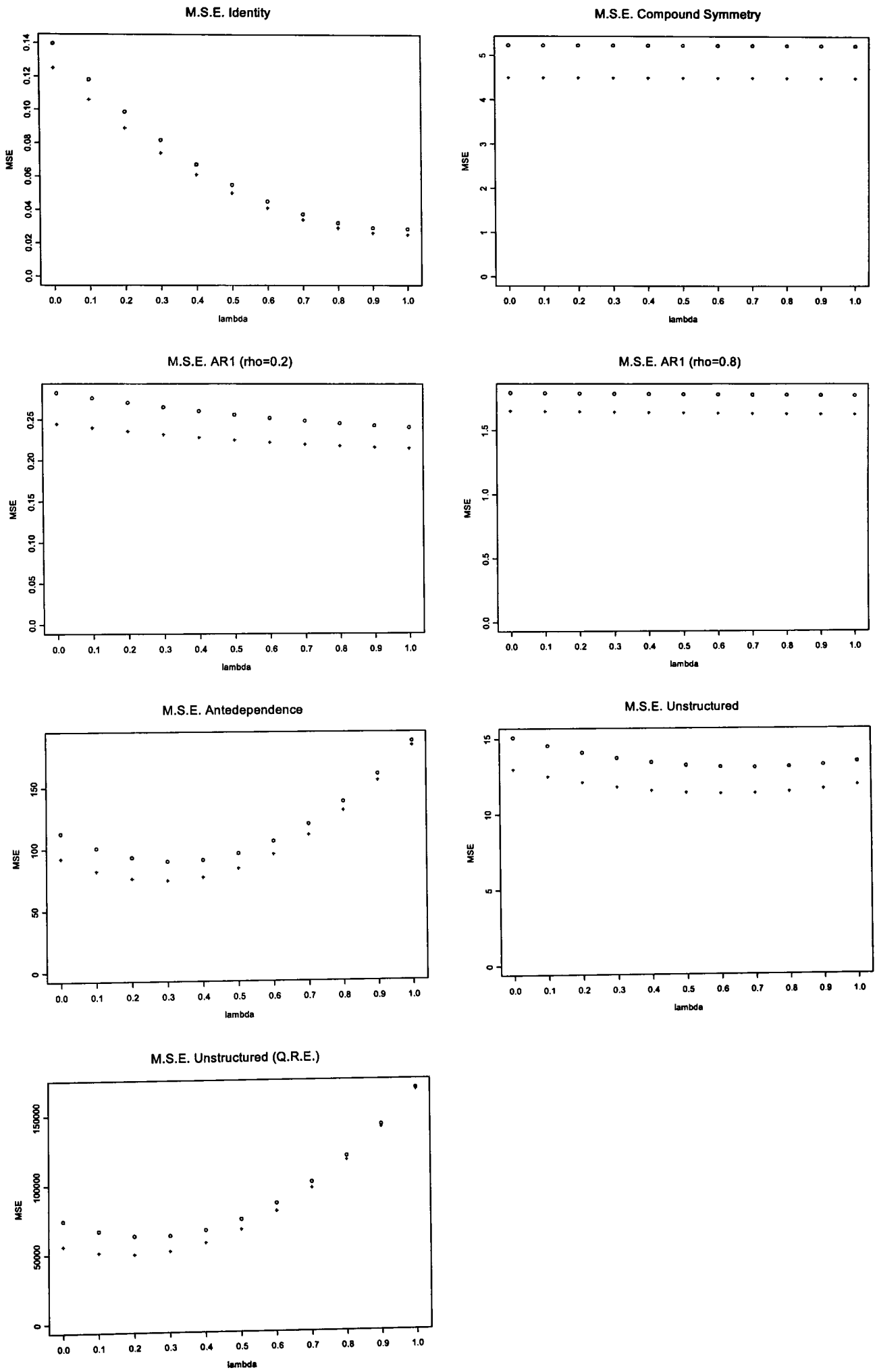


Figure 4.4.4: *Plots of both Empirical and Theoretical M.S.E. functions for given* λ *in* M.S.E.$(\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a})$, *Compound Symmetry Smoothing. 1000 simulations of the Pilot Study. Legend:* $\circ$ *Empirical,* $+$ *Theoretical*

Underlying ('True') Covariance Structure	λ Known$\boldsymbol{\Sigma}$	λ Estimated from Data			
		Mean	S.D.	Min	Max
Identity Smoothing					
Identity	1.000	0.323	0.067	0.189	0.615
Compound Symmetry	0.190	0.191	0.031	0.151	0.476
AR1 ($\rho = 0.2$)	0.589	0.299	0.061	0.183	0.538
AR1 ($\rho = 0.8$)	0.165	0.167	0.015	0.143	0.244
Antedependence	0.229	0.206	0.029	0.154	0.336
Unstructured	0.173	0.174	0.018	0.145	0.269
Unstruct. (Q.R.E.)	0.139	0.139	0.0002	0.139	0.140
Compound Symmetry Smoothing					
Identity	1.000	0.361	0.114	0.200	1.119
Compound Symmetry	-	1.509	1.246	0.242	12.858
AR1 ($\rho = 0.2$)	1.273	0.403	0.235	0.193	2.987
AR1 ($\rho = 0.8$)	6.997	2.422	2.343	0.203	23.920
Antedependence	0.288	0.262	0.085	0.158	0.716
Unstructured	0.628	0.494	0.130	0.175	1.322
Unstruct. (Q.R.E.)	0.187	0.187	0.004	0.175	0.202

Table 4.4.2: *Minimising values of* λ *in* M.S.E.$(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$, *Identity and Compound Symmetry Smoothing. 1000 simulations of the Pilot Study.*

In practice, however, the underlying covariance structure is unknown, and calculating λ using the eigenvalues of $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$, the estimated sample covariance matrix, appears problematic. For badly estimated covariance structures, the eigenvalues are also badly estimated, and the simulation show that the estimates of λ obtained from the data are not close to those found theoretically. Since such estimates always result in covariance structures with greater heterogeneity, the values of λ obtained are always smaller than would be the case if the real underlying structure were known.

For direct smoothing towards a compound symmetry form, estimates are more problematic, since taking $\phi_* = \phi_{\max}$ can lead to estimates of $\lambda > 1$. This is more likely to occur for data arising from underlying covariance structures close to compound symmetry which are likely to have flat profiles (e.g. for compound symmetry or AR1 (high correlation)), but not exclusively so. In such cases there is little to be lost in simply adopting $\lambda = 1$ as the estimate of the smoothing parameter, but this does point to problems with this approach.

A further possibility to improve estimation in the case of direct smoothing towards an

Underlying ('True') Covariance Structure	λ Known$\boldsymbol{\Sigma}$	λ Estimated from Data			
		Mean	S.D.	Min	Max
Identity Smoothing					
Identity	1.000	0.726	0.302	0.219	1.000
Compound Symmetry	0.190	0.231	0.090	0.163	1.000
AR1 ($\rho = 0.2$)	0.589	0.618	0.308	0.210	1.002
AR1 ($\rho = 0.8$)	0.165	0.186	0.023	0.153	0.322
Antedependence	0.229	0.257	0.078	0.168	1.000
Unstructured	0.173	0.197	0.029	0.155	0.390
Unstruct. (Q.R.E.)	0.139	0.148	0.0003	0.148	0.150

Table 4.4.3: *Minimising values of* λ *in* M.S.E.$(\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a})$, *Identity Smoothing (using the smoothed eigenstructure of* $\hat{\boldsymbol{\Sigma}}$*). 1000 simulations of the Pilot Study.*

identity structure, is to take $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} = (1-\lambda)\hat{\phi}_{\max} + \lambda\overline{\phi}$, the corresponding (largest) eigenvalue of the smoothed matrix Σ_λ and determining λ iteratively to improve estimation when the underlying structure of $\boldsymbol{\Sigma}$ is unknown. Results from 1000 simulations of the Pilot Study are shown in Table 4.4.3.

This somewhat ad-hoc approach does show some improvement in the estimates of λ, and is probably the best approach in this badly estimated case. It is worthy of note however that this iterative approach:

1. increases the estimate of λ in each case (over that obtained using $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} = \phi_{\max}$), and

2. estimates of λ outside the range [0,1] are possible. (Also, care must be taken since $\lambda = 1$ is also a solution of the iterative procedure in each case).

4.5 Discussion

Methods have been presented which attempt to find a suitable value for the smoothing parameter λ in the estimator $\hat{\boldsymbol{\Sigma}}_\lambda$, based on minimising the mean squared error of a function of $\hat{\boldsymbol{\Sigma}}_\lambda$. Two such functions were considered, namely $\ln|\hat{\boldsymbol{\Sigma}}_\lambda|$ and $\mathbf{a}^T\hat{\boldsymbol{\Sigma}}_\lambda\mathbf{a}$ (for a non-null vector $\mathbf{a}$). In the first case M.S.E.$(|\hat{\boldsymbol{\Sigma}}_\lambda|)$ was calculated dependent on a Taylor series approximation of Var$(\ln|\hat{\boldsymbol{\Sigma}}_\lambda|)$ which has been shown to be inac-

curate in a small sample setting. The alternative formulation, M.S.E.($\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a}$) gave differing results in the cases of smoothing towards identity and compound symmetry forms respectively.

There was found to be close convergence between the theoretical and empirical forms with minimising values of the smoothing parameter λ close to that which we would intuitively expect in each case. However, in practice where the underlying structure is unknown there were found to be problems with substituting the eigenvalues from the sample covariance matrix, since such estimates were necessarily more heterogenous than those of the true underlying structure so that little smoothing was suggested. There are also doubts about the validity of the procedure in introducing an unknown vector $\mathbf{a}$ whose choice is subjective and clearly affects the chosen value of λ. This is more evident in the case of smoothing towards a compound symmetry structure where there is less justification for taking $\mathbf{a}^T\hat{\mathbf{\Sigma}}_\lambda\mathbf{a} = \phi_{\max}$, the largest eigenvalue of $\hat{\mathbf{\Sigma}}_\lambda$.

Neither of the chosen methods for determining λ are successful in practice. Estimates are improved when m, the number of subjects is greatly increased but in this case little smoothing is warranted when the covariance estimates are well-estimated. Also, it is not clear how, so dependent on the distributional properties of the Wishart distribution, they could be easily adapted to more complex situations. The approach of directly combining the two forms of covariance structure is appealing however and in the absence of alternative methods for determining an appropriate value for λ cross-validation may be used. Such an approach will be investigated in Chapter 5.

Chapter 5

A Penalised Likelihood Approach

5.1 Introduction

Maximum penalised likelihood methods are used widely in non-parametric and semi-parametric regression. Such applications are reviewed by Green (1998). Here, an alternative approach to finding a smoothed covariance estimate is considered through the maximisation of the penalised (log-) likelihood

$$l_{\mathrm{p}}(\boldsymbol{\Sigma}; \mathbf{S}) = l(\boldsymbol{\Sigma}; \mathbf{S}) + \alpha l^*(\mathbf{M}; \boldsymbol{\Sigma}), \qquad \alpha \geq 0 \tag{5.1.1}$$

where the penalty $l^*(\mathbf{M}, \boldsymbol{\Sigma})$ acts as a likelihood ratio, measuring the fit of the unstructured covariance matrix $\boldsymbol{\Sigma}$ to some structured model $\mathbf{M}$. The higher the weight given to the penalty, determined by the smoothing parameter α, the more the penalised estimate $\tilde{\boldsymbol{\Sigma}}$ is forced away from the unstructured form $\boldsymbol{\Sigma}$ and towards the structured form $\mathbf{M}$. We have

$$l_{\mathrm{p}}(\boldsymbol{\Sigma}; \mathbf{S}) = -\ln|\boldsymbol{\Sigma}| - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + \alpha\big\{-\ln|\mathbf{M}| + \ln|\boldsymbol{\Sigma}| - \mathrm{tr}(\mathbf{M}^{-1}\boldsymbol{\Sigma})\big\} \tag{5.1.2}$$

Again the data are initially assumed to be balanced and complete so that the simpli-

fied log-likelihood of $\mathbf{\Sigma}$ based on the Wishart distribution has been used to motivate development, although in this approach an extension to cover data with missing values by using a penalised REML likelihood is immediately apparent.

It can be shown that maximising the penalised likelihood with respect to the parameters of both $\mathbf{\Sigma}$ and $\mathbf{M}$, we obtain $\mathbf{M}$ to be the maximum likelihood estimator of the structured form given $\mathbf{\Sigma}$. That is, to find jointly the m.l.e.s of $\mathbf{\Sigma}$ and $\mathbf{M}$ in $l_{\mathrm{p}}(\mathbf{\Sigma};\mathbf{S})$, is to solve simultaneously

$$\begin{aligned} &\frac{\partial l_p}{\partial \sigma_{(\mathbf{\Sigma})i}} = 0 \qquad \text{for } i = 1,\ldots,p(p+1)/2 \qquad (1) \\ &\frac{\partial l_p}{\partial \sigma_{(\mathbf{M})i}} = 0 \qquad \text{for } j = 1,\ldots,r \qquad (2) \end{aligned} \tag{5.1.3}$$

for each of the $p(p+1)/2$ and r parameters in the unstructured and structured estimates of $\mathbf{\Sigma}$ and $\mathbf{M}$ respectively. However, (2) implies that

$$\begin{aligned} &\alpha \frac{\partial}{\partial \sigma_{(\mathbf{M})i}} \left\{ -\ln|\mathbf{M}| + \ln|\mathbf{\Sigma}| - \mathrm{tr}(\mathbf{M}^{-1}\mathbf{\Sigma}) \right\} = 0 \\ \Rightarrow &\frac{\partial}{\partial \sigma_{(\mathbf{M})i}} \left\{ -\ln|\mathbf{M}| - \mathrm{tr}(\mathbf{M}^{-1}\mathbf{\Sigma}) \right\} = 0 \end{aligned}$$

so that $(\partial/\partial\sigma_{(\mathbf{M})i})l(\mathbf{M};\mathbf{\Sigma}) = 0$, and $\hat{\mathbf{M}}$ is the m.l.e. of $\mathbf{M}$ given $\mathbf{\Sigma}$. Then, $\mathrm{tr}(\hat{\mathbf{M}}^{-1}\mathbf{\Sigma}) = p$ and the penalised likelihood may be written

$$\begin{aligned} l_{\mathrm{p}}(\mathbf{\Sigma};\mathbf{S}) &= -\ln|\mathbf{\Sigma}| - \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) + \alpha(-\ln|\mathbf{M}^*| + \ln|\mathbf{\Sigma}|) \\ &= -(1-\alpha)\ln|\mathbf{\Sigma}| - \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) + \alpha \ln|\mathbf{M}^*| \end{aligned} \tag{5.1.4}$$

where $\mathbf{M}^*$ is the m.l.e. of $\mathbf{M}$ given $\mathbf{\Sigma}$. This simplifies the calculations where 'plug-in' estimates exist for the structured covariance forms. It can be shown further that the penalised estimate $\tilde{\mathbf{\Sigma}}$ has the property $\mathrm{tr}(\tilde{\mathbf{\Sigma}}^{-1}\mathbf{S}) = p$ which it shares with maximum likelihood estimators. This is shown below.

Writing the penalised likelihood, from (5.1.2), as

$$l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = (1-\alpha)\ln|\boldsymbol{\Sigma}^{-1}| - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \alpha\mathrm{tr}(\mathbf{M}^{-1}\boldsymbol{\Sigma}) - \alpha\ln|\mathbf{M}| \tag{5.1.5}$$

we have, differentiating with respect to an element of $\boldsymbol{\Sigma}$, σ_i say, for a given $\mathbf{M}$

$$\frac{\partial l_p}{\partial \sigma_i} = (1-\alpha)\mathrm{tr}\left(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\right) - \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\mathbf{S}\right) - \alpha\mathrm{tr}\left(\mathbf{M}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\right) \tag{5.1.6}$$

Now,

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} = -\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\boldsymbol{\Sigma}^{-1}$$

so that,

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} = -\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}$$

and hence

$$\begin{aligned}\frac{\partial l_p}{\partial \sigma_i} &= (1-\alpha)\mathrm{tr}\left(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\right) - \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\mathbf{S}\right) + \alpha\mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}\mathbf{M}^{-1}\boldsymbol{\Sigma}\right) \\ &= \mathrm{tr}\left[\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\left\{(1-\alpha)\boldsymbol{\Sigma} - (\mathbf{S} - \alpha\boldsymbol{\Sigma}\mathbf{M}^{-1}\boldsymbol{\Sigma})\right\}\right]\end{aligned} \tag{5.1.7}$$

so that, equating to zero, we find $\tilde{\boldsymbol{\Sigma}}$ is a solution of

$$\frac{\partial l_p}{\partial \sigma_i} = \mathrm{tr}\left\{\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}(\boldsymbol{\Sigma} - \mathbf{S}^*)\right\} = 0 \qquad \text{for } i = 1,\ldots,p(p+1)/2 \tag{5.1.8}$$

where,

$$\mathbf{S}^* = \frac{1}{1-\alpha}(\mathbf{S} - \alpha\boldsymbol{\Sigma}\mathbf{M}^{-1}\boldsymbol{\Sigma})$$

Now, considering the unstructured (penalised) covariance matrix as $\boldsymbol{\Sigma} = \sum_{j=1}^{r} \sigma_j \mathbf{Z}_j$ for fixed matrices $\mathbf{Z}_j$, we have further that

$$\boldsymbol{\Sigma} = \sum_{j=1}^{r} \eta_j \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j} \quad \text{and so} \quad \boldsymbol{\Sigma}^{-1} = \sum_{j=1}^{r} \eta_j^* \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j}$$

It follows that $\tilde{\boldsymbol{\Sigma}}$ satisfies

$$\sum_{j=1}^{r} \eta_j^* \frac{\partial l_p}{\partial \sigma_j} = 0$$

That is,

$$\sum_{j=1}^{r} \eta_j^* \mathrm{tr}\left\{ \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j} (\tilde{\boldsymbol{\Sigma}} - \mathbf{S}^*) \right\} = 0 \tag{5.1.9}$$

which implies that,

$$\begin{aligned} & \mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\Sigma}} - \mathbf{S}^*)\} = 0 \\ \Rightarrow & \mathrm{tr}\left\{ \mathbf{I} - \frac{1}{1-\alpha}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \alpha \mathbf{M}^{-1}\tilde{\boldsymbol{\Sigma}} \right) \right\} = 0 \\ \Rightarrow & p - \frac{1}{1-\alpha}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) - \alpha p = 0 \\ \Rightarrow & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) = p \end{aligned} \tag{5.1.10}$$

Also, (5.1.8) suggests the following iterative approach to finding $\tilde{\boldsymbol{\Sigma}}$ for a given value of smoothing parameter α,

$$\tilde{\boldsymbol{\Sigma}}_{r+1} = \mathbf{S}^* = \frac{1}{1-\alpha}\left(\mathbf{S} - \alpha \tilde{\boldsymbol{\Sigma}}_r \hat{\mathbf{M}}_r^{-1} \tilde{\boldsymbol{\Sigma}}_r \right) \tag{5.1.11}$$

where $\hat{\mathbf{M}}_r$ is the m.l.e. of $\mathbf{M}$ given $\tilde{\boldsymbol{\Sigma}}_r$. Unfortunately, this method converges on the correct solution only for small values of α ($<< 0.2$), so is not reliable in general. However, for a given $\alpha \geq 0$, the penalised estimator $\tilde{\boldsymbol{\Sigma}}$ can be computed

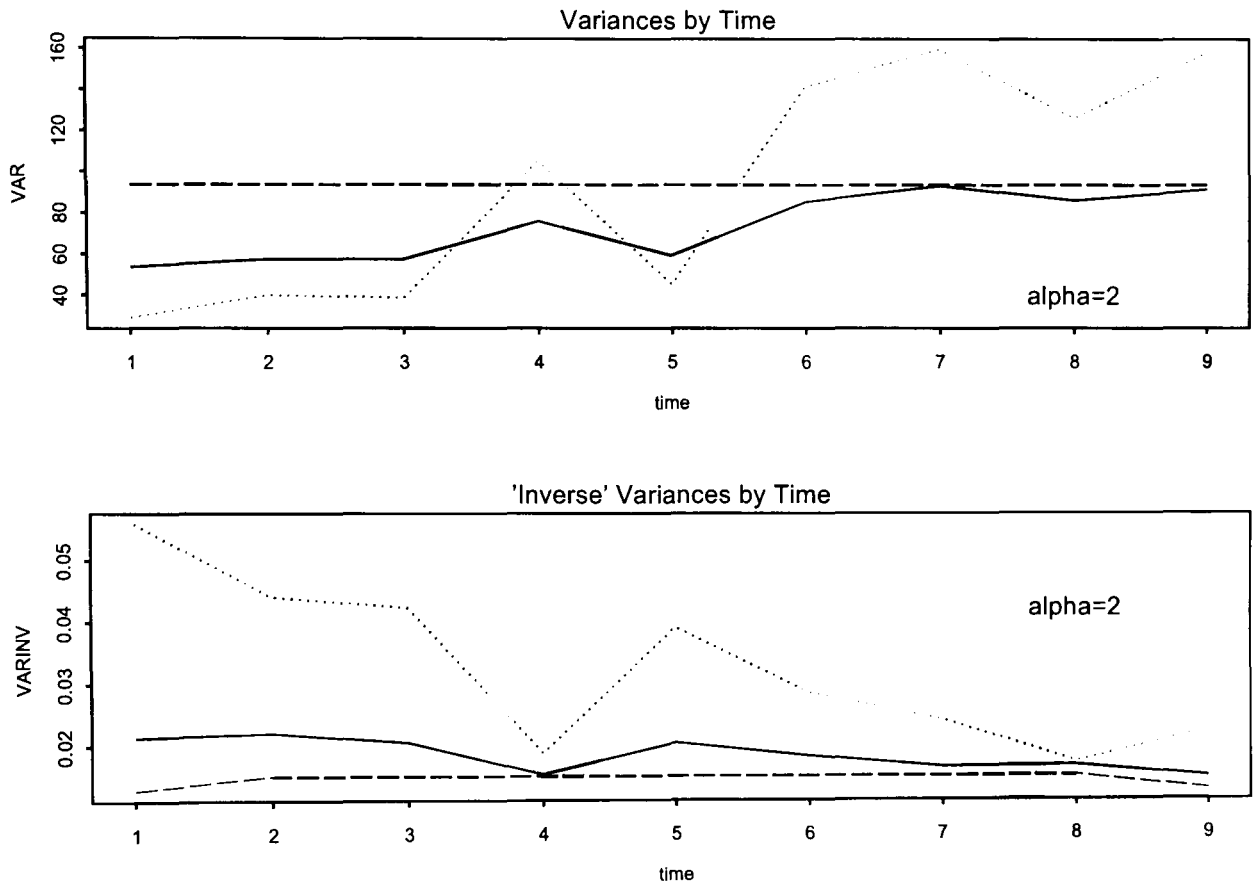


Figure 5.1.1: *Variances and 'Inverse' Variances by Time for the Cardiac Enzyme data. Penalised Likelihood Smoothing to an AR1 form. Legend: Solid Line, smoothed estimate; Dashed Line, AR1 form; Dotted Line, unstructured form.*

by maximising the penalised likelihood given in (5.1.4) using a numerical procedure (such as Newton Raphson, 'B.F.G.S.') in *S-Plus* or *PROC IML* in *SAS*.

The direct smoothing approach of the previous Chapter has shown that little smoothing results where the identity form is adopted as the smoothing structure, the form that the estimated covariance structure is directed towards. In this Chapter, we consider smoothing towards both compound symmetry and AR1 structured forms.

Figure 5.1.1 shows the effect of smoothing the variances and diagonal elements of the inverse covariance structure ('inverse variances'), when $\alpha = 2$ for smoothing the Cardiac Enzyme data towards an AR1 form. It can be seen that the effect of the penalised likelihood approach is actually to smooth between the inverses of the unstructured and structured covariance matrices. This may be considered advantageous, as our interest in the covariance structure is in its use in inference

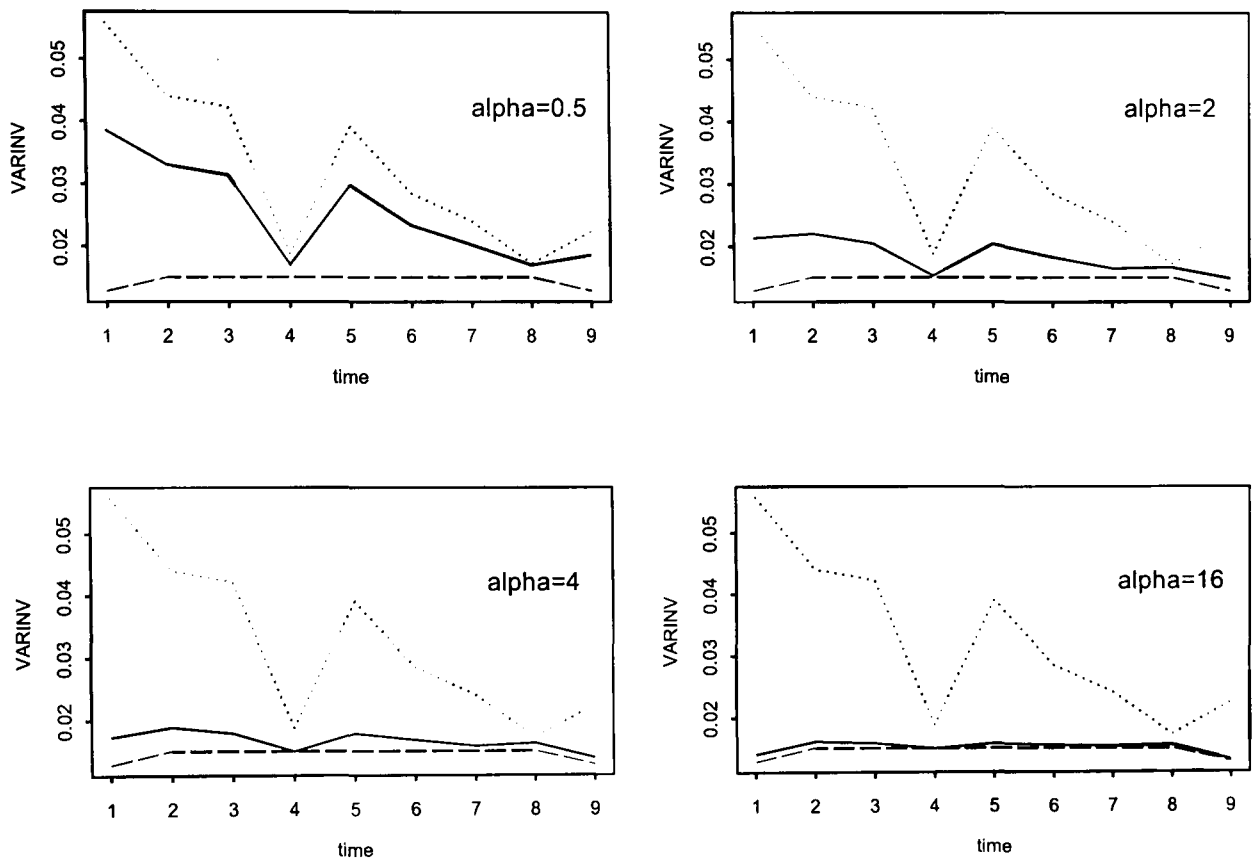


Figure 5.1.2: *'Inverse' Variances by Time for the Cardiac Enzyme data. Penalised Likelihood Smoothing to an AR1 form. Legend: Solid Line, smoothed estimate; Dashed Line, identity form; Dotted Line, unstructured form.*

where the inverse matrix is used. It is also a major difference between this method and the direct smoothing approach where the smoothing is directly on the matrix itself. Figure 5.1.2 shows the effect of smoothing the diagonal elements of the inverse covariance matrix similarly, for various values of α.

5.1.1 A Cross-validation algorithm for finding α

The value of the smoothing parameter can be chosen by cross-validation, (see, for example, Efron and Tibshirani (1993), Chapter 17). Using a saturated means model the approach is to drop subjects in turn and to use the smoothed covariance estimates based on the remaining data (and the means from the complete data) to predict each of the dropped observations. The smoothing parameter is chosen to minimise the total predictive squared error across all subjects. This is a computationally intensive

method as it requires the maximisation of the penalised likelihood for each value of the smoothing parameter considered and observation dropped, but it can be carried out reasonably efficiently in *SAS* (*PROC IML*) for small to medium sized data sets.

Given an $m \times p$ data matrix $\mathbf{Y}$, the cross-validation algorithm is as follows:

(**A**) For a given choice of α:

(1). Remove subject i from the data, so we have the $(m-1) \times p$ data matrix $\mathbf{Y}_{|i}$.

(2). Calculate the sample covariance matrix, $\mathbf{S}_{|i}$, for the reduced data matrix $\mathbf{Y}_{|i}$.

(3). Find the penalised likelihood estimate $\tilde{\mathbf{\Sigma}}$ of the covariance structure of the reduced data matrix, which minimises the penalised (log-) likelihood
$-\ln|\mathbf{\Sigma}| - \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}_{/i}) + \alpha\{-\ln|M^*\mathbf{\Sigma}| + \ln|\mathbf{\Sigma}|\}$

(4). Use $\tilde{\mathbf{\Sigma}}$ and the means $\hat{\mu}_j$ (calculated from the complete data), to predict the values of y_{ij}, $j = 2, \ldots, p$, given the previous observations on that subject, using the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$, i.e. using
$\mathrm{E}(y_2|\mathbf{y}_1) = \mu_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, where $y_2 = y_{ij}$ and $\mathbf{y}_1^T = (y_{i1}, y_{i2}, \ldots, y_{i(j-1)})$, $j = 2, \ldots, p$, so that

$$\hat{y}_{i2} = \hat{\mu}_2 + \frac{\tilde{\sigma}_{21}}{\tilde{\sigma}_{11}}(y_{i1} - \hat{\mu}_1),\ \hat{y}_{i3} = \hat{\mu}_3 + \begin{pmatrix}\tilde{\sigma}_{31} & \tilde{\sigma}_{32}\end{pmatrix}\begin{pmatrix}\tilde{\sigma}_{11} & \tilde{\sigma}_{21}\\ \tilde{\sigma}_{21} & \tilde{\sigma}_{22}\end{pmatrix}\begin{pmatrix}y_{i1} - \hat{\mu}_1\\ y_{i2} - \hat{\mu}_2\end{pmatrix}, \text{ etc.}$$

(5). Calculate the (total) predictive squared error $\mathrm{PSE}_i = \sum_{j=2}^{p}(\hat{y}_{ij} - \hat{\mu}_j)^2$.

(6). Repeat steps (1)-(5) removing each subject, $i = 1, \ldots, m$ in turn.

(7). Calculate the mean (total) predictive squared error, $\frac{1}{m}\sum_{i=1}^{m}\mathrm{PSE}_i$.

(**B**) Repeat steps (1)-(7) from (**A**) for various choices of α, taking as the value of α that which minimises the mean (total) predictive squared error.

The procedure is easily adapted to grouped data by smoothing the pooled within-groups (sample) covariance structure. In this case the treatment group of a 'dropped'

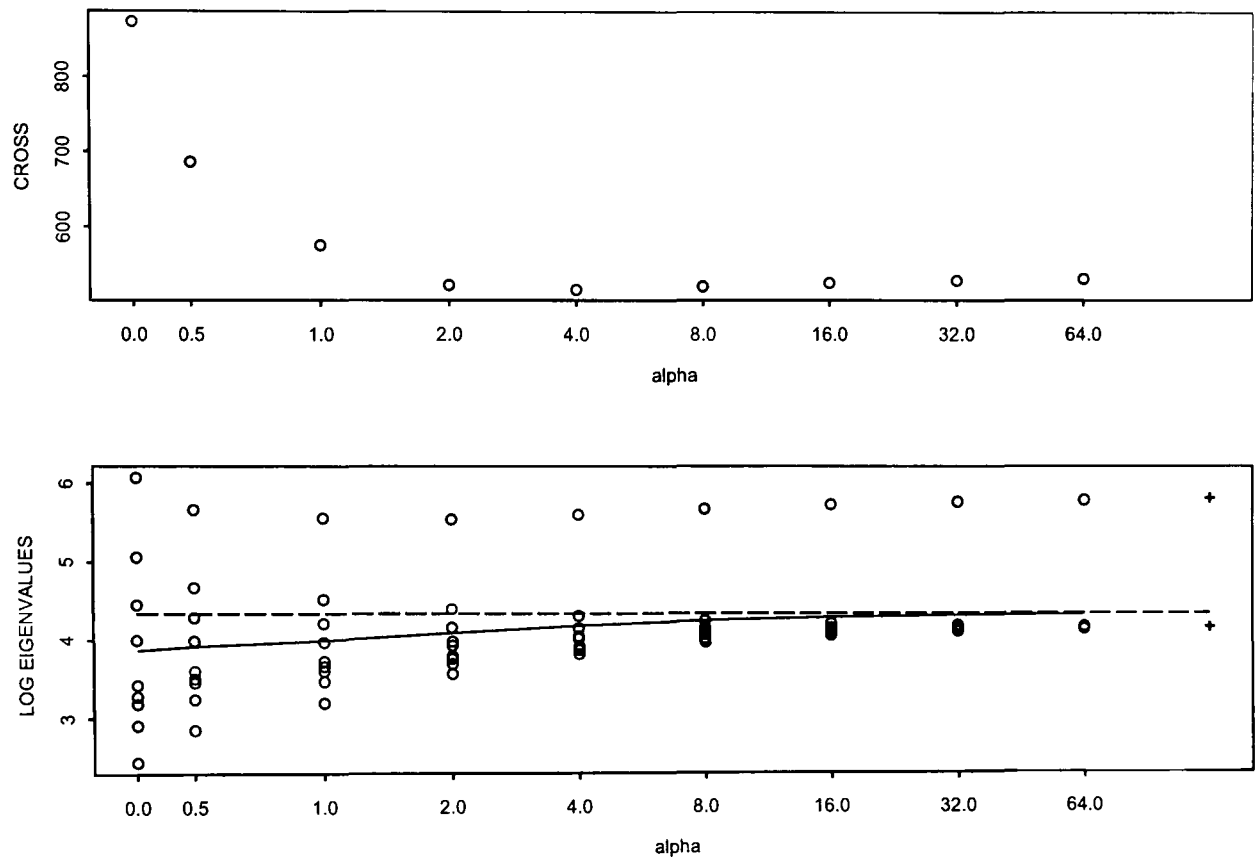


Figure 5.1.3: *Choice of α by cross-validation for the Cardiac Enzyme data. Penalised Likelihood Smoothing to a Compound Symmetry form.*

subject is noted before smoothing the revised within-groups covariance structure and predicting the dropped observations using the means from that complete treatment group.

The approach can be speeded up by noticing that in many cases there is a flattening out of the cross-validatory function above a certain value of α, which suggests that further smoothing is of little consequence. One way to determine whether this is the case is to consider at each value of α the difference in the log-determinants between the current smoothed covariance structure and the last. Then, the optimal value of the smoothing parameter is reached when either there is a rise in the cross-validatory function so that a 'proper' minimum has been attained or the difference in the log-determinants falls below a prescribed level. Adopting such a procedure allows the automation of the cross-validatory approach to determining the smoothing parameter. An example of choosing α by cross-validation for the Cardiac Enzyme

data, smoothing towards a compound symmetry structure, is shown in Figure 5.1.3. Values of α were restricted to the set $\{0, 0.5, 1, 2, 4, 8, 16, 32, 64\}$ and the automated procedure chose $\alpha = 4$ in this instance.

5.2 A Comparison with Direct Smoothing

We have seen that the two approaches differ in that the penalised likelihood method acts upon the inverse covariance matrix whilst the direct approach smooths the elements of the covariance matrix itself. Also, the penalised likelihood method is grounded in the well understood principles of maximum likelihood, but is computationally intensive. The direct smoothing approach is simplistic but is easily computed (λ can be found by substituting $\hat{\mathbf{\Sigma}}_\lambda = (1-\lambda)\hat{\mathbf{\Sigma}} + \lambda\hat{\mathbf{\Sigma}}_S$ in step $(\mathbf{A})(4)$ of the cross-validation procedure described in the preceding section). It is appropriate however to determine whether there is any real difference between the approaches in terms of inference.

The pilot study of Chapter 2 is again repeated to compare the efficiency of each of the two approaches to smoothing, in making inferences. The two smoothing approaches were compared on a number of simulated data sets with $p = 5$ time points and $m = 16$ subjects, arbitrarily split into two (identical) treatment groups of 8. In each case a saturated means model was fitted and the smoothed covariance matrices for a suitable value of the relevant smoothing parameter chosen by cross-validation were calculated. Values of α and λ considered in the cross-validation procedure were restricted to the sets $\alpha \in \{0, 0.5, 1, 2, 4, 8, 16, 32, 64\}$ and $\lambda \in \{0, 0.1, 0.2, \ldots, 1.0\}$ respectively.

The methods were compared for smoothing towards both compound symmetry and AR1 forms and for a number of underlying covariance structures including identity, compound symmetry, AR1 (low correlation), AR1 (high correlation) and two 'unstructured' forms. Using 1000 simulated data sets from each of the various un-

derlying covariance structures gives values of both α and λ as we would intuitively expect. That is, a low degree of smoothing is suggested for data which arises from an underlying covariance structure which is far from the smoothed form, with greater amounts of smoothing suggested as the underlying form approaches the form of the structured matrix. The distributions of these parameters can be seen in Figure 5.2.1. Note that for data sets arising from the 'badly behaved' Q.R.E. structure no smoothing at all was suggested towards either of the stationary Compound Symmetry or AR1 structures in the majority of cases. (i.e. $\alpha = 0$ and $\lambda = 0$).

The smoothed covariance estimates were then used to test an overall interaction and a set of polynomial contrasts for each data set. For each smoothing method and underlying covariance structure, the number of significant test results from 1000 such data sets given by each smoothing approach was noted. These were compared with the number of significant test results which would have been obtained if the unstructured covariance estimate had been used.

5.2.1 A 'Scaled' Kenward Roger Adjustment

In comparing the results using the smoothed estimates with those achieved using the unstructured and structured estimates themselves, it is necessary to consider small sample adjustments. It was earlier shown that small sample Wald tests have nominal properties for complete data problems using the Kenward Roger (KR) adjustment with an unstructured covariance estimator or an estimator with a low number of parameters, where this is appropriate for the data. That is, for testing the general linear hypothesis

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$$

for $\mathbf{L}$ $(l \times r)$, we use the usual Wald test statistic, (1.3.3)

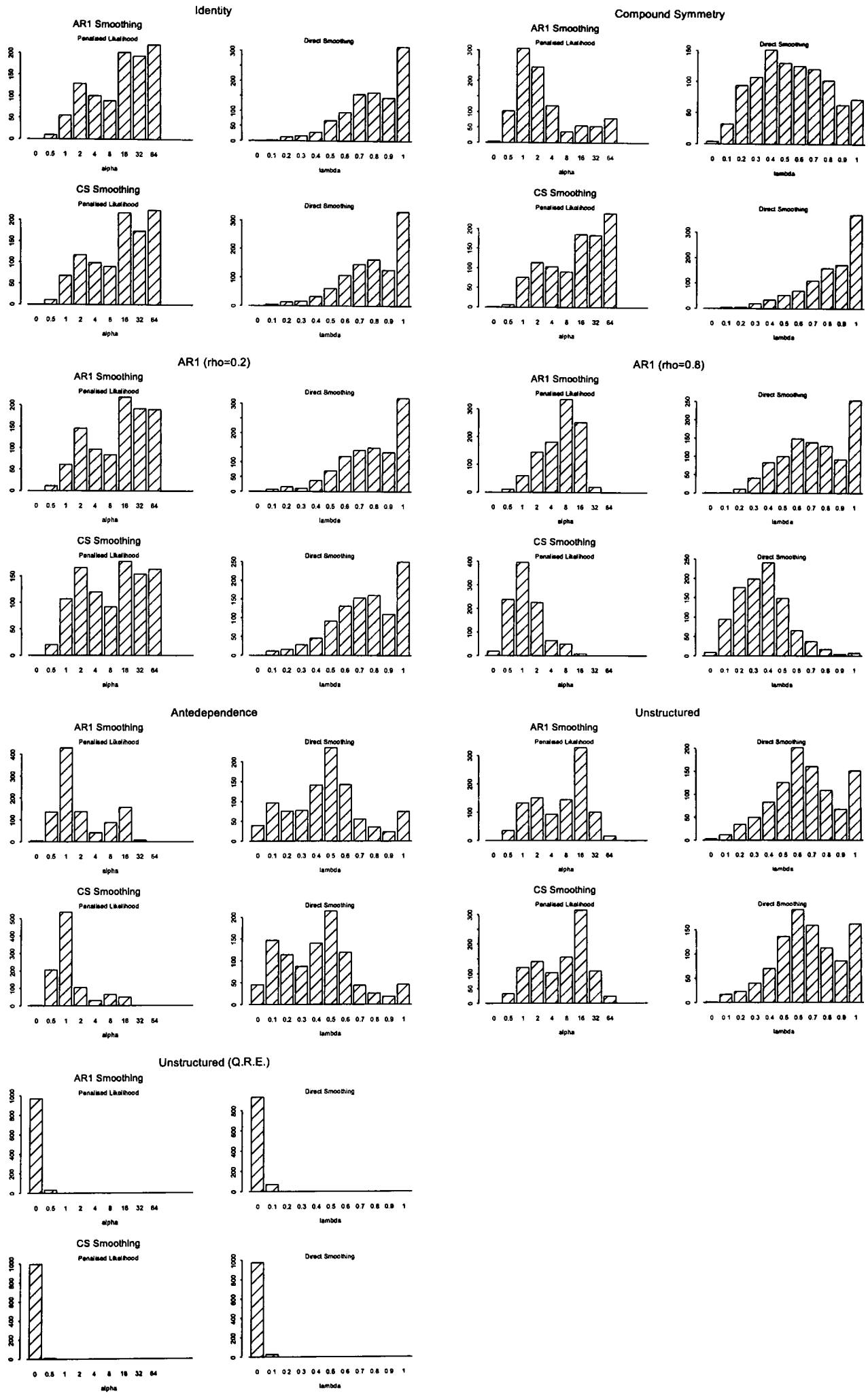


Figure 5.2.1: *Distributions of Smoothing Parameter Estimates α and λ obtained from 1000 simulations of the Pilot Study. Penalised Likelihood and Direct Smoothing towards both AR1 and Compond Symmetry forms*

$$\mathrm{F} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T (\mathbf{L}\boldsymbol{\Phi}\mathbf{L}^T)^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{l}$$

where $\boldsymbol{\Phi} = \mathrm{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X})^{-1}$ is replaced by $\boldsymbol{\Phi}_A = \boldsymbol{\Phi} + 2\boldsymbol{\Lambda}$ in the test statistic F^* and tests $\gamma \mathrm{F}^* \sim \mathrm{F}(l, m)$, where m is the adjusted denominator degrees of freedom parameter.

An intuitive if somewhat *ad-hoc* approach therefore to small sample adjustments using smoothed covariance estimates is to adopt a 'scaled' KR adjustment. That is, in the case of penalised smoothing towards a compound symmetry structure, for example, we use the KR adjustment with

$$\begin{aligned} \gamma &= \gamma_{\mathrm{UN}} + \lambda^*(\gamma_{\mathrm{CS}} - \gamma_{\mathrm{UN}}) \\ m &= m_{\mathrm{UN}} + \lambda^*(m_{\mathrm{CS}} - m_{\mathrm{UN}}) \\ \boldsymbol{\Phi}_A &= \boldsymbol{\Phi} + 2\boldsymbol{\Lambda}_{\mathrm{CS}} \end{aligned} \tag{5.2.1}$$

where the subscript terms UN and CS refer to the KR parameters under the unstructured and compound symmetry forms respectively. λ^* is chosen by a suitable transformation of the smoothing parameter $\alpha \in [0, \infty)$ onto the interval $[0, 1]$ to give a precise measure of the degree of smoothing towards the structured estimate.

A family of transformations $\lambda^* = 1 - \exp(-k\alpha)$ was investigated for differing values of k, $0 < k \leq 1$ but the preferred transformation for the penalised smoothing parameter is

$$\lambda^* = \log\left\{1 + \frac{\alpha}{1+\alpha}(\mathrm{e} - 1)\right\} \tag{5.2.2}$$

which gave the closest match to nominal test sizes across the range of tests. The effect of this transformation for values of $\alpha \in \{0, 0.5, 1, 2, 4, 8, 16, 32, 64\}$ is shown in Figure 5.2.2. (This function behaves similarly to the earlier transformation taking $k = 1$).

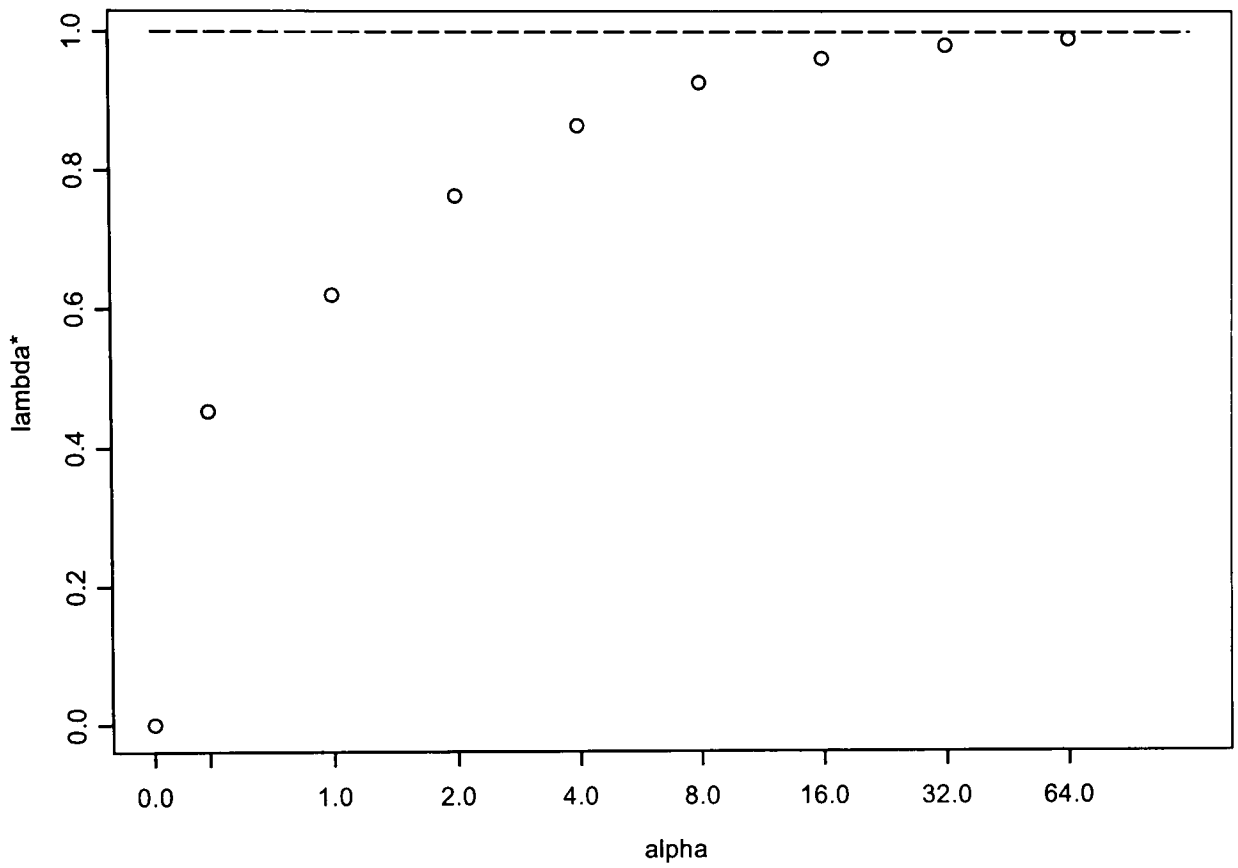


Figure 5.2.2: *Transforming* $\alpha \in [0, \text{inf})$ *to* $\lambda^* \in [0, 1]$.

Surprisingly, in providing a 'scaled' KR adjustment for inference using a directly smoothed covariance estimator, taking $\lambda^* = \lambda$ did not achieve nominal test sizes for the corresponding adjusted Wald statistics. Instead the non-linear transformation of the smoothing parameter λ, given by

$$\lambda^* = \log\left\{1 + \frac{\lambda}{1+\lambda} 2(\mathrm{e} - 1)\right\} \tag{5.2.3}$$

is adopted to better represent the degree of smoothing.

5.2.2 Results

Tables 5.2.1 and 5.2.2 show the test sizes for Wald tests for an overall treatment time interaction and a set of orthogonal polynomial contrasts (C1, linear; C2, quadratic; C3, cubic; and C4, quartic) for data arising from the various stationary and non-

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Stationary Structures						
Identity	Unstr	0.044	0.055	0.037	0.046	0.049
	AR1	0.051	0.058	0.037	0.056	0.057
	Penal(AR1)	0.058	0.060	0.043	0.058	0.060
	Direct(AR1)	0.049	0.059	0.038	0.052	0.054
	Comp Sym	0.046	0.053	0.040	0.055	0.059
	Penal(CS)	0.054	0.056	0.044	0.059	0.061
	Direct(CS)	0.050	0.050	0.041	0.052	0.059
Compound Symmetry	Unstr	0.047	0.059	0.063	0.059	0.056
	AR1*	0.045	0.004	0.016	0.060	0.082
	Penal(AR1)	0.053	0.022	0.036	0.073	0.080
	Direct(AR1)	0.039	0.014	0.028	0.060	0.066
	Comp Sym	0.048	0.050	0.054	0.056	0.055
	Penal(CS)	0.053	0.055	0.058	0.062	0.058
	Direct(CS)	0.048	0.048	0.054	0.054	0.051
AR1 ($\rho = 0.2$)	Unstr	0.051	0.056	0.048	0.054	0.065
	AR1	0.047	0.054	0.048	0.048	0.060
	Penal(AR1)	0.050	0.057	0.054	0.052	0.065
	Direct(AR1)	0.044	0.052	0.048	0.045	0.063
	Comp S*	0.050	0.080	0.052	0.039	0.036
	Penal(CS)	0.050	0.083	0.059	0.048	0.041
	Direct(CS)	0.044	0.070	0.050	0.041	0.042
AR1 ($\rho = 0.8$)	Unstr	0.039	0.062	0.044	0.041	0.033
	AR1	0.031	0.051	0.038	0.032	0.020
	Penal(AR1)	0.034	0.059	0.041	0.040	0.025
	Direct(AR1)	0.027	0.053	0.038	0.034	0.024
	Comp S*	0.082	0.195	0.034	0.003	0.002
	Penal(CS)	0.062	0.137	0.052	0.033	0.016
	Direct(CS)	0.042	0.098	0.041	0.024	0.010

Table 5.2.1: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of type 1 errors (Size), using a Wald statistic with a 'KR' adjustment.*

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Non-Stationary Structures						
Ante-dependence	Unstr	0.036	0.051	0.049	0.043	0.051
	AR1*	0.075	0.052	0.053	0.049	0.027
	Penal(AR1)	0.069	0.087	0.086	0.077	0.059
	Direct(AR1)	0.037	0.054	0.055	0.045	0.041
	Comp S*	0.072	0.116	0.062	0.036	0.010
	Penal(CS)	0.074	0.131	0.101	0.074	0.045
	Direct(CS)	0.029	0.075	0.058	0.040	0.023
Unstr.	Unstr	0.044	0.047	0.041	0.052	0.052
	AR1*	0.057	0.033	0.026	0.051	0.059
	Penal(AR1)	0.057	0.044	0.039	0.053	0.066
	Direct(AR1)	0.047	0.038	0.028	0.052	0.054
	Comp S*	0.061	0.065	0.040	0.050	0.044
	Penal(CS)	0.059	0.072	0.047	0.055	0.055
	Direct(CS)	0.045	0.064	0.041	0.052	0.049
Unstr. (Q.R.E.)	Unstr	0.057	0.053	0.056	0.056	0.063
	AR1*	0.137	0.278	0.017	0.000	0.000
	Penal(AR1)	0.053	0.057	0.058	0.054	0.062
	Direct(AR1)	0.051	0.054	0.056	0.053	0.060
	Comp S*	0.142	0.321	0.002	0.000	0.000
	Penal(CS)	0.056	0.053	0.056	0.055	0.063
	Direct(CS)	0.049	0.053	0.056	0.053	0.059

Table 5.2.2: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of type 1 errors (Size), using a Wald statistic with a 'KR' adjustment.*

stationary covariance structures which should be compared with a nominal level of 5%.

For each underlying ('true') structure the Kenward Roger adjusted sizes are given for the appropriate Wald tests based on an unstructured, AR1 or compound symmetry covariance structure, which can be compared with the test sizes resulting from the smoothed estimate and adopting the scaled adjustment. Test sizes for AR1 and compound symmetry structures are marked with an asterisk (∗) where they are not appropriate for the underlying data.

Note that under the null hypothesis of no difference between treatment groups a 95% probability interval for the proportion of tests out of 1000 leading to a rejection

of the test is (0.036, 0.064). This indicates that the test size for the interaction using an AR1 structure for data which actually does arise from an AR1 structure ($\rho = 0.8$) is a little low, falling below the lower bound of this interval and may impact upon the scaled adjustments of the smoothed estimates. As discussed previously the adjustments for the AR1 structure are not exact and may not perform quite so well in such small samples.

A further ten simulations of 1000 data sets showed that whilst there is a tendancy for adjustments based on a 'true' AR1 structure to be low (below the nominal level of 0.05) they are largely within the 95% probability interval suggesting that the extremely low size of 0.031 was obtained by chance.

Test sizes for the smoothed estimates are close to nominal levels using the scaled Kenward Roger adjustment for all underlying covariance structures, even where smoothing towards a structure that is inappropriate for the data.

Having fixed the size of these tests at the nominal level, power can be meaningfully considered. Power levels for the Wald tests are shown in Tables 5.2.3 and 5.2.4.

Power was calculated by superimposing a (linear) difference between the means of the two treatment groups in each simulated data set, and retesting using the small sample adjustments as previously described. As with the Pilot Study in Chapter 2, for each underlying structure the mean difference between groups was fixed to achieve a power of around 75% using an unstructured matrix.

For each of the underlying stationary covariance structures the tests for no treatment/time interaction using the smoothed estimates show a greater power than that obtained using the unstructured matrix, and close to that obtained using an AR1 or compound symmetry matrix where these are the true underlying types. A similar pattern of results is displayed for the linear contrast C1. The power calculations suggest no real advantage of one method over another.

For the data arising from the non-stationary covariance structures, the pattern is

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Stationary Structures						
Identity	Unstr	0.747	0.974	0.037	0.046	0.049
	AR1	0.913	0.988	0.037	0.056	0.057
	Penal(AR1)	0.906	0.982	0.043	0.058	0.060
	Direct(AR1)	0.897	0.980	0.038	0.052	0.054
	Comp Sym	0.896	0.980	0.040	0.050	0.059
	Penal(CS)	0.910	0.979	0.044	0.059	0.061
	Direct(CS)	0.903	0.978	0.042	0.052	0.063
Compound Symmetry	Unstr	0.754	0.965	0.063	0.059	0.056
	AR1*	0.523	0.819	0.016	0.060	0.082
	Penal(AR1)	0.741	0.919	0.036	0.073	0.080
	Direct(AR1)	0.686	0.904	0.028	0.060	0.066
	Comp Sym	0.897	0.979	0.057	0.058	0.056
	Penal(CS)	0.901	0.980	0.058	0.062	0.058
	Direct(CS)	0.893	0.979	0.054	0.054	0.051
AR1 ($\rho = 0.2$)	Unstr	0.765	0.978	0.048	0.054	0.065
	AR1	0.901	0.988	0.048	0.048	0.060
	Penal(AR1)	0.907	0.991	0.054	0.052	0.052
	Direct(AR1)	0.900	0.988	0.048	0.045	0.063
	Comp S*	0.939	0.995	0.052	0.039	0.036
	Penal(CS)	0.938	0.996	0.059	0.048	0.041
	Direct(CS)	0.927	0.991	0.050	0.041	0.042
AR1 ($\rho = 0.8$)	Unstr	0.744	0.967	0.044	0.041	0.033
	AR1	0.879	0.977	0.038	0.032	0.020
	Penal(AR1)	0.895	0.976	0.041	0.040	0.025
	Direct(AR1)	0.880	0.975	0.038	0.034	0.024
	Comp S*	0.984	0.994	0.036	0.003	0.002
	Penal(CS)	0.956	0.991	0.052	0.033	0.016
	Direct(CS)	0.890	0.983	0.041	0.024	0.010

Table 5.2.3: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of significant results (Power), using a Wald statistic with a 'KR' adjustment.*

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Non-Stationary Structures						
Ante-dependence	Unstr	0.764	0.748	0.049	0.043	0.051
	AR1*	0.543	0.789	0.053	0.049	0.027
	Penal(AR1)	0.774	0.848	0.086	0.077	0.059
	Direct(AR1)	0.598	0.787	0.055	0.045	0.041
	Comp S*	0.688	0.876	0.062	0.036	0.010
	Penal(CS)	0.864	0.882	0.101	0.074	0.045
	Direct(CS)	0.682	0.818	0.058	0.040	0.023
Unstr.	Unstr	0.777	0.948	0.041	0.052	0.052
	AR1*	0.824	0.957	0.026	0.051	0.059
	Penal(AR1)	0.857	0.962	0.039	0.053	0.066
	Direct(AR1)	0.853	0.957	0.028	0.052	0.054
	Comp S*	0.902	0.980	0.040	0.050	0.044
	Penal(CS)	0.911	0.979	0.047	0.055	0.055
	Direct(CS)	0.898	0.973	0.041	0.052	0.049
Unstr. (Q.R.E.)	Unstr	0.782	0.177	0.056	0.056	0.063
	AR1*	0.287	0.511	0.017	0.000	0.000
	Penal(AR1)	0.778	0.189	0.058	0.054	0.062
	Direct(AR1)	0.744	0.179	0.056	0.053	0.060
	Comp S*	0.332	0.556	0.002	0.000	0.000
	Penal(CS)	0.782	0.181	0.056	0.055	0.063
	Direct(CS)	0.770	0.178	0.056	0.053	0.059

Table 5.2.4: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of significant results (Power), using a Wald statistic with a 'KR' adjustment.*

a little different. For the highly non-stationary antedependence structure, the penalised likelihood estimates give greater power than the directly smoothed estimates. Also, more power is gained when smoothing towards a compound symmetry structure rather than an AR1 structure, and the power achieved using a structure smoothed towards AR1 falls short of that obtained from simply adopting an unstructured estimate.

This latter point is perhaps surprising since the AR1 structure is nested within the antedependence structure and would have appeared to be the more reasonable choice of smoothing structure in this case.

The results from the data arising from the underlying unstructured covariance forms

are of further interest. The smoothed estimates for the Q.R.E. structure appear to behave no differently in general from the unstructured estimates, but as we have seen there was no smoothing in the majority of cases so that the smoothed and unstructured estimates were in fact the same. For the data arising from the remaining unstructured form, the results follow a similar pattern as for the underlying stationary structures.

5.3 Discussion

The results reported above may be summarised as follows.

- Both the direct and penalised likelihood smoothing approaches appear to be reasonable, giving scaled Wald tests which attain nominal properties.
- For data arising from covariance structures which are stationary, adopting a smoothed estimate for the covariance structure results in greater power over adopting an unstructured estimate.
- For data arising from covariance structures which are highly non-stationary, the choice of smoothing approach and smoothing structure appears to be important.
- For very highly non-stationary data, both smoothing approaches lead to little or no smoothing, as we might expect.

To investigate further the properties of the smoothed estimates for data which are non-stationary, the simulations were repeated using the correlation structures from the antedependence and unstructured forms, but allowing the variances to vary. These further results showed that little or no smoothing was obtained from the Q.R.E structure with equal variances, and from the remaining structures as the difference in variances (by time) became large. This confirms that these methods are not suitable for data which are highly non-stationary, in correlations or variances.

since either little smoothing will be necessary or the resulting estimates may not be more powerful than simply adopting an unstructured estimate. This is important as it is the case in many repeated measures problems that data will be non-stationary.

A further modification to the penalised likelihood smoothed estimate will be presented in Chapter 6, which attempts to improve the performance of the smoothed covariance estimate, for data which are non-stationary, by adopting a local-stationarity penalised likelihood approach.

However, the restriction of the smoothing techniques to data which is balanced and complete is clearly too great, making them of little benefit in practice. To be of practical use it will be necessary to extend both the direct smoothing and penalised likelihood methods to include data sets which are unbalanced or have missing values. In the case of the penalised likelihood approach this will mean defining a penalised REML likelihood. Recalling the penalised likelihood form, $l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = l(\boldsymbol{\Sigma};\mathbf{S}) + \alpha l^{*}(\mathbf{M};\boldsymbol{\Sigma})$, we could use the REML log-likelihood

$$l(\boldsymbol{\Sigma};\mathbf{S}) = -\ln|\boldsymbol{\Sigma}| - \ln|\mathbf{X}^{T}\boldsymbol{\Sigma}^{-1}\mathbf{X}| - \mathbf{y}^{T}\mathbf{H}_{\boldsymbol{\Sigma}}\mathbf{y} \tag{5.3.1}$$

where, $\mathbf{H}_{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^{T}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{T}\boldsymbol{\Sigma}^{-1}$, and the penalty term becomes

$$l^{*}(\mathbf{M};\boldsymbol{\Sigma}) = -\ln|\mathbf{M}| - \ln|\mathbf{X}^{T}\mathbf{M}^{-1}\mathbf{X}| - \mathbf{y}^{T}\mathbf{H}_{\mathbf{M}}\mathbf{y} \tag{5.3.2}$$

where $\boldsymbol{\Sigma}$ and $\mathbf{M}$ now have block diagonal structures. The direct smoothing approach can be updated by simply replacing the maximum likelihood estimates of the unstructured and structured forms, $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{S}$ in $\hat{\boldsymbol{\Sigma}}_{\lambda}$ with the corresponding block diagonal REML estimates.

Further, the extension of such procedures to use block diagonal forms will prove computationally expensive in the case of the penalised likelihood approach where numerical algorithms are required to solve the likelihood. This may make the use of

cross-validation impractical as a method for determining the value of the smoothing parameter α in anything other than small data sets with few covariance parameters. Given the numerical problems encountered in the restrictive setting considered (it was increasingly infeasible to carry out the penalised likelihood procedure as the number of time points increased), this is unlikely to prove successful. It is also unlikely to be of any practical benefit given the results obtained from the simple setting.

Another solution to this problem would be to determine the effective degrees of freedom (number of parameters) of the smoothed covariance estimate, which would allow the formulation of a measure such as Akaike's information criterion for determining the value of the smoothing parameter. Little has been published in this area in the context of covariance structures, although Schwarz (1978) and Spiegelhalter *et al.* (2002) use Bayesian formulations to determine the appropriate dimension of a polynomial regression model and a hierarchical model respectively, where the effective number of parameters is not clearly defined.

It should also be noted that the *ad-hoc* approach to scaling the Kenward Roger small sample adjustments, although useful in providing a simple comparison of the techniques, lacks rigour and a formal methodology for testing using smoothed estimates would need to be developed.

It increasingly appears that, in small sample settings, there is a need for methods which are less dependent on the estimated covariance structure. This will be the focus in parts III and IV of this thesis.

Chapter 6

Local-Stationarity Penalised Likelihood

6.1 Introduction

It is clear from the examples presented in the previous chapter that very small values were achieved for the smoothing parameters where the underlying covariance structure of the data was highly non-stationary, such as the Q.R.E. structure, even where the small number of observations means that such structures are poorly estimated. A consideration therefore is the generalisation of the technique in smoothing towards additional structured, particularly non-stationary covariance matrices. The success of a general smoothing approach in practice will ultimately be dependent on a suitable choice of 'smoothing' structure. Another approach to this would be to consider localised smoothing. Such an approach was suggested by Kenward (1997), who proposed the penalised likelihood function,

$$l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = l(\boldsymbol{\Sigma};\mathbf{S}) + \alpha \sum_{t=1}^{p-q+1} l^{*}(\mathbf{M};\boldsymbol{\Sigma}_t) \tag{6.1.1}$$

where $\boldsymbol{\Sigma}$ has a non-stationary antependence(1) structure and $\mathbf{M}$ has a stationary AR(1) form estimated from successive submatrices $\boldsymbol{\Sigma}_t$ from the diagonal of $\boldsymbol{\Sigma}$. i.e. $\boldsymbol{\Sigma}_t$ is the submatrix of $\boldsymbol{\Sigma}$ at the q successive times $t, \ldots, t+q-1$. This penalises

the fit of the covariance structure $\boldsymbol{\Sigma}$ by its departure from $\mathbf{M}$ along its diagonal.

By taking $\boldsymbol{\Sigma}$ to be an antedependence(1) structure and $\mathbf{M}$ to be an AR(1), initially, we are penalising departure from local stationarity. The choice of these structures is intuitive as both depend only on the elements of the diagonal and subdiagonal of $\mathbf{S}$ (the sample covariance matrix), so that there is no loss of information in either case by considering only submatrices $\boldsymbol{\Sigma}_t$ from the diagonal. As $\alpha \to \infty$ we move towards an AR(1) estimate, but for $\alpha = 0$ we are effectively smoothing the off-diagonal elements of the covariance structure by adopting an antedependence structure as a starting point rather than the unstructured. The generalisation to higher order structures is obvious.

Additionally to the choice of smoothing parameter α we need to consider an appropriate number of time points q to be included in the submatrix $\boldsymbol{\Sigma}_t$.

6.2 Properties of the Estimator

As in the previous chapter, development is motivated by restricting ourselves initially to data which are balanced and complete so that the simplified log likelihood expressions for $\boldsymbol{\Sigma}$ based on the Wishart distribution are appropriate.

The local stationarity penalised log likelihood may be written

$$l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = -\ln|\boldsymbol{\Sigma}| - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + \alpha \sum_{t=1}^{p-q+1} \{-\ln|\mathbf{M}| + \ln|\boldsymbol{\Sigma}_t| - \mathrm{tr}(\mathbf{M}^{-1}\boldsymbol{\Sigma}_t)\} \quad (6.2.1)$$

or equivalently,

$$l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = \ln|\boldsymbol{\Sigma}^{-1}| - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \alpha \sum_{t=1}^{p-q+1} (\ln|\mathbf{M}^*| + \ln|\boldsymbol{\Sigma}_t|) \quad (6.2.2)$$

where, as before, $\mathbf{M}^*$ is the maximum likelihood estimator of the structured form

given $\boldsymbol{\Sigma}_t$. Again, the penalised estimator $\tilde{\boldsymbol{\Sigma}}$ has the desirable property $\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) = p$.

To show this, we write from (6.2.1)

$$l_{\mathrm{p}}(\boldsymbol{\Sigma};\mathbf{S}) = \ln|\boldsymbol{\Sigma}^{-1}| - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + \alpha \sum_{t=1}^{p-q+1} \{\ln|\mathbf{M}^{-1}| - \ln|\boldsymbol{\Sigma}_t^{-1}| - \mathrm{tr}(\mathbf{M}^{-1}\boldsymbol{\Sigma}_t)\} \quad (6.2.3)$$

so that for a given $\mathbf{M}$, differentiating with respect to an element of $\boldsymbol{\Sigma}$, σ_i say,

$$\begin{aligned}\frac{\partial l_p}{\partial \sigma_i} &= \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}\right) - \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\mathbf{S}\right) + \alpha \sum_{t=1}^{p-q+1}\left\{-\mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}\right) + \mathrm{tr}\left(\mathbf{M}^{-1}\frac{\partial \boldsymbol{\Sigma}_t}{\partial \sigma_i}\right)\right\} \\ &= \mathrm{tr}\left\{\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\left(\boldsymbol{\Sigma} - \mathbf{S}\right)\right\} - \alpha \sum_{t=1}^{p-q+1} \mathrm{tr}\left\{\frac{\partial \boldsymbol{\Sigma}_t^{-1}}{\partial \sigma_i}\left(\boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_t\mathbf{M}^{-1}\boldsymbol{\Sigma}_t\right)\right\}\end{aligned} \quad (6.2.4)$$

And, following the same argument of Chapter 5, we have, writing

$$\boldsymbol{\Sigma}^{-1} = \sum_{j=1}^{r} \eta_j^* \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j}$$

so that $\tilde{\boldsymbol{\Sigma}}$ satisfies

$$\sum_{j=1}^{r} \eta_j^* \frac{\partial l_p}{\partial \sigma_j} = 0$$

$$\begin{aligned}&\sum_{j=1}^{r} \eta_j^* \left[\mathrm{tr}\left\{\frac{\partial \tilde{\boldsymbol{\Sigma}}^{-1}}{\partial \sigma_i}\left(\tilde{\boldsymbol{\Sigma}} - \mathbf{S}\right)\right\} - \alpha \sum_{t=1}^{p-q+1} \mathrm{tr}\left\{\frac{\partial \tilde{\boldsymbol{\Sigma}}_t^{-1}}{\partial \sigma_i}\left(\tilde{\boldsymbol{\Sigma}}_t - \tilde{\boldsymbol{\Sigma}}_t\mathbf{M}^{-1}\tilde{\boldsymbol{\Sigma}}_t\right)\right\}\right] = 0 \\ \Rightarrow &\mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\Sigma}} - \mathbf{S})\} - \alpha \sum_{t=1}^{p-q+1} \mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}_t^{-1}(\tilde{\boldsymbol{\Sigma}}_t - \tilde{\boldsymbol{\Sigma}}_t\mathbf{M}^{-1}\tilde{\boldsymbol{\Sigma}}_t)\} = 0 \\ \Rightarrow &\mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) - \alpha \sum_{t=1}^{p-q+1} \{\mathrm{tr}(\mathbf{I}_t) - \mathrm{tr}(\mathbf{M}^{-1}\tilde{\boldsymbol{\Sigma}}_t)\} = 0 \\ \Rightarrow &\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) = p\end{aligned} \quad (6.2.5)$$

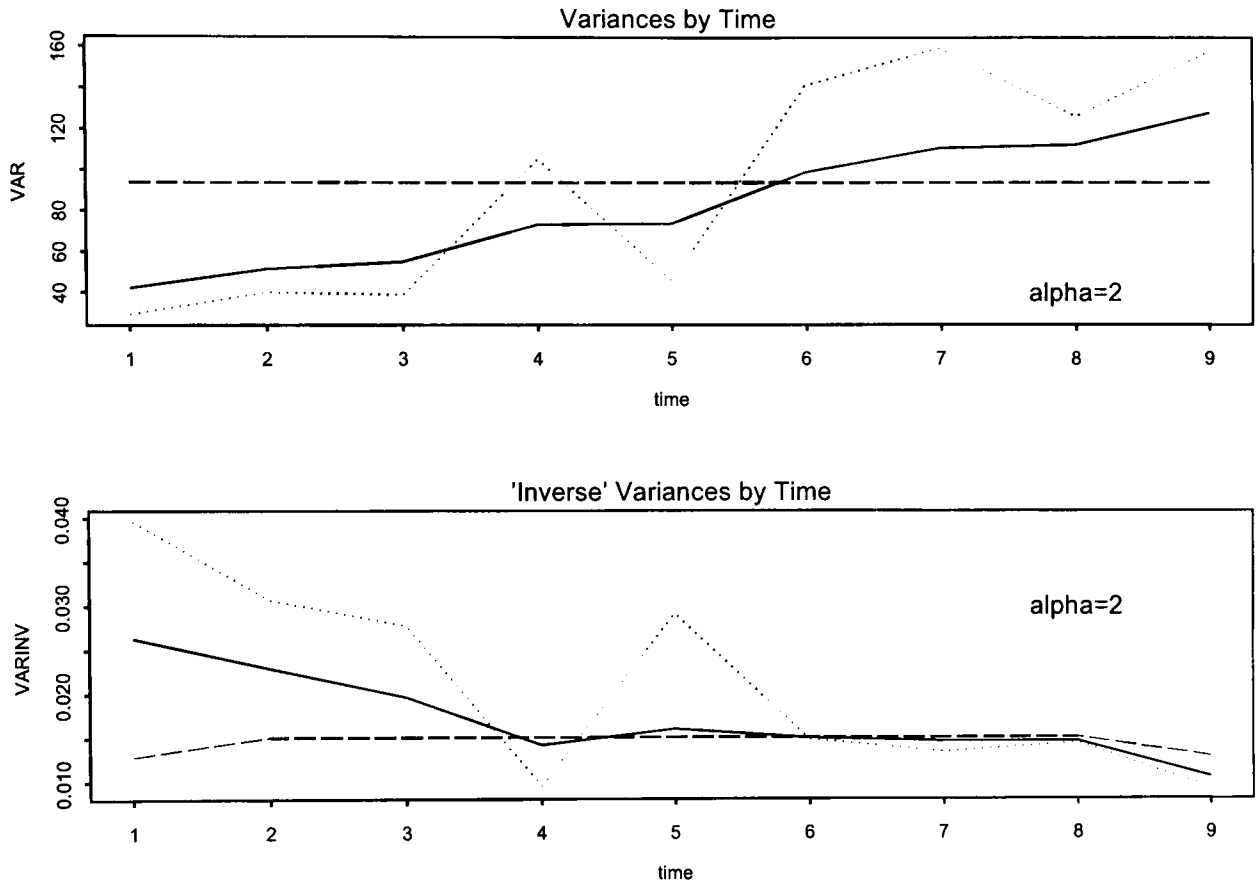


Figure 6.2.1: *Variances and 'Inverse' Variances by Time for the Cardiac Enzyme data. Local-Stationarity Penalised Likelihood Smoothing. Legend: Solid Line, smoothed estimate; Dashed Line, AR1 form; Dotted Line, antedependence form.*

The effect of this form of smoothing can be seen by looking at plots of variances and 'inverse variances' by time, (see Figures 6.2.1 and 6.2.2). As with the penalised likelihood smoothing of the previous chapter it can be seen that we are smoothing the elements of inverse covariance matrix. The size of the submatrix $\mathbf{\Sigma}_t$ is taken as $q = 3$, the smallest possible value, which ensures a fine degree of smoothing allowing for a maximum account to be taken of local stationarity.

Adopting an autoregressive parametrisation of $\mathbf{\Sigma}$ allows us to consider the effects of smoothing both the autoregressive and variance parameters separately. That is, for an antedependence(1) structure on p time measurements we may write

$$y_1 = \epsilon_1$$

$$y_t = \phi_t y_{t-1} + \epsilon_t, \quad t = 2, \ldots, p$$

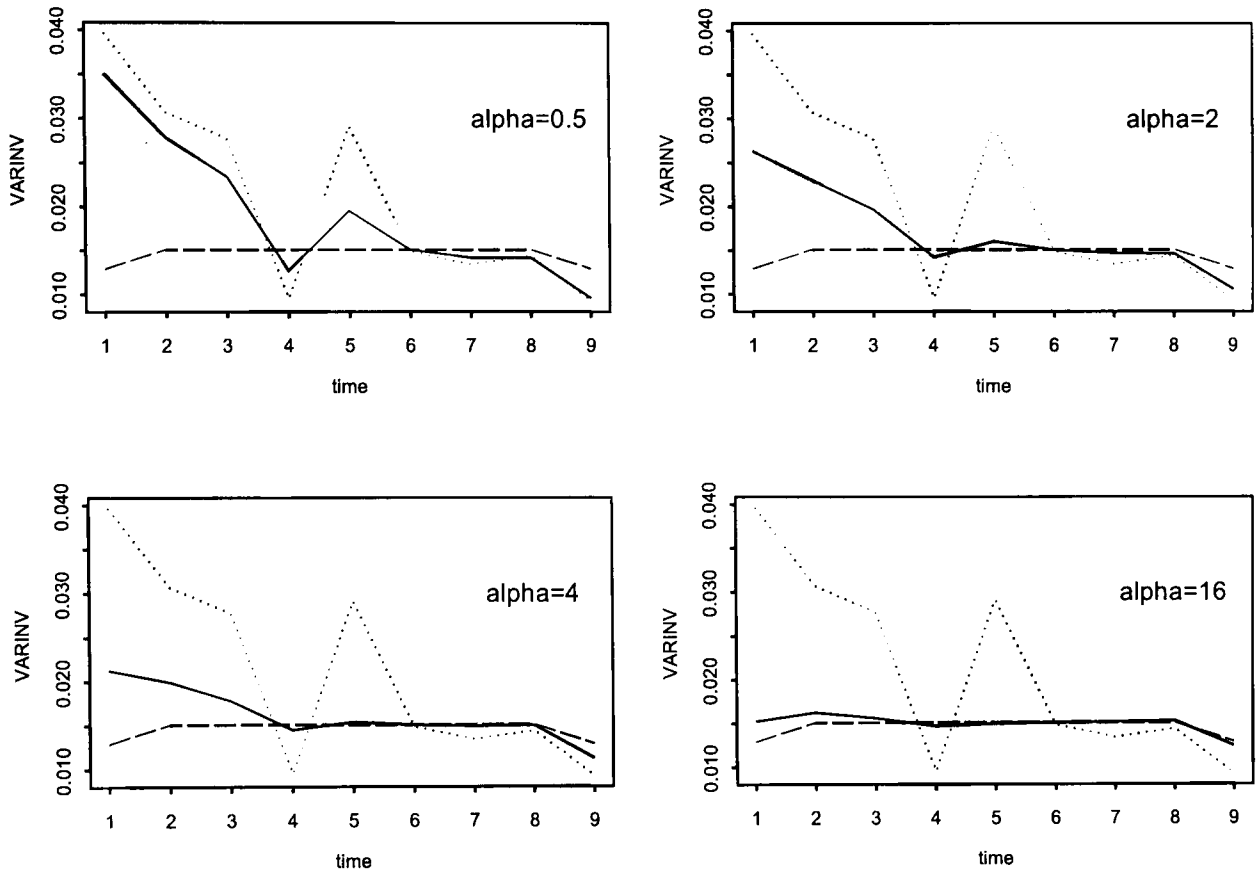


Figure 6.2.2: *'Inverse' Variances by Time for the Cardiac Enzyme data. Local-Stationarity Penalised Likelihood Smoothing. Legend: Solid Line, smoothed estimate; Dashed Line, AR1 form; Dotted Line, antedependence form.*

or equivalently,

$$\begin{aligned} y_1 &= \epsilon_1 \\ y_t - \phi_t y_{t-1} &= \epsilon_t, \quad t = 2, \ldots, p \end{aligned} \tag{6.2.6}$$

where $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{Var}(\epsilon_t) = \sigma_t^2$, $t = 1, \ldots, p$. In vector-matrix form this may be written as $\mathbf{Uy} = \mathbf{e}$, where

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ -\phi & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \ldots & -\phi & 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{p-1} \\ y_p \end{pmatrix}, \mathbf{e} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{p-1} \\ \epsilon_p \end{pmatrix}$$

and $\mathrm{Var}(\mathbf{Uy}) = \mathrm{Var}(\mathbf{e}) = \Lambda = \mathrm{diag}(\phi_i)$.

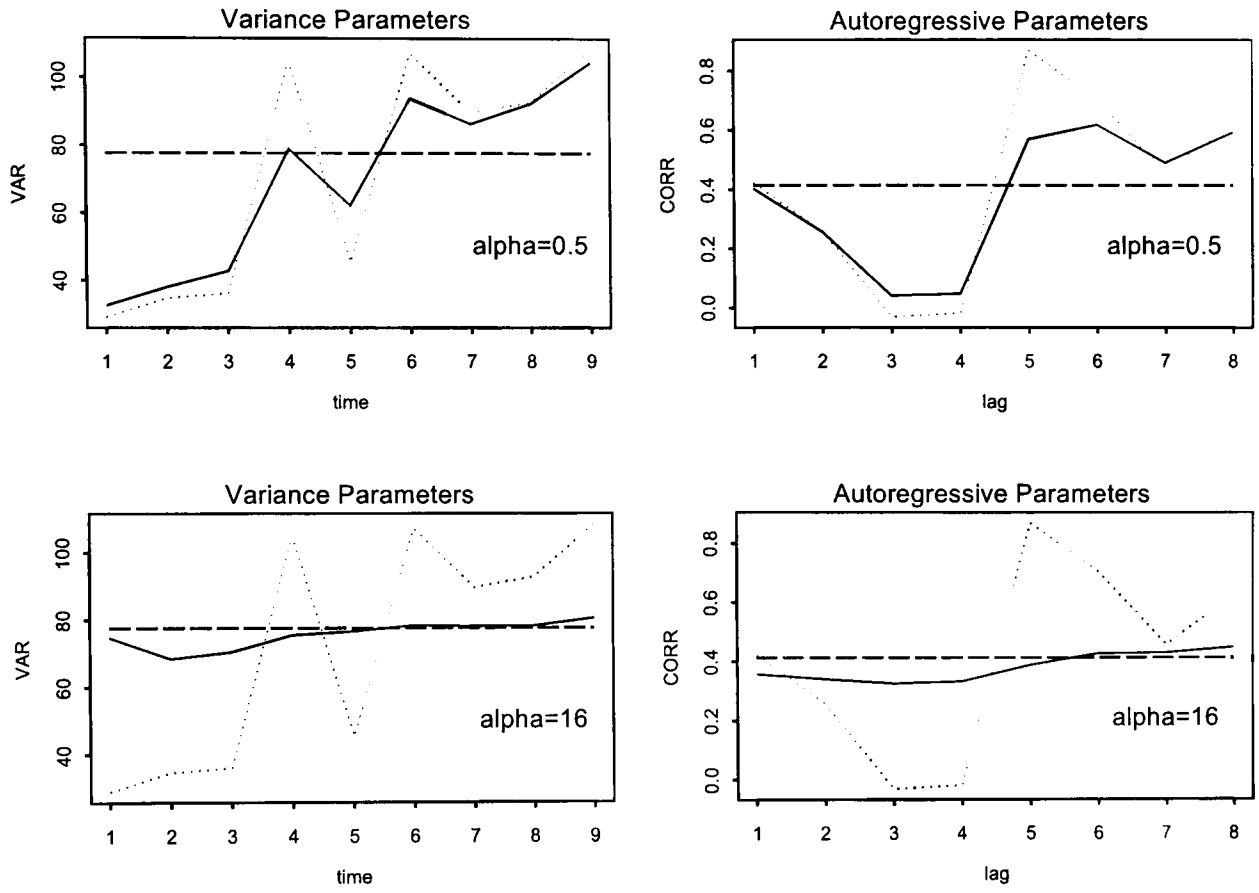


Figure 6.2.3: *Variance and Autoregressive Parameters by Time for the Cardiac Enzyme data. Local-Stationarity Penalised Likelihood Smoothing. Legend: Solid Line, smoothed estimate; Dashed Line, AR1 form; Dotted Line, antedependence form.*

Then $\Lambda = \text{Var}(\mathbf{Uy}) = \mathbf{U}\text{Var}(\mathbf{y})\mathbf{U}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T$ so that $\boldsymbol{\Sigma} = \mathbf{U}\Lambda(\mathbf{U}^T)^{-1}$ and $\boldsymbol{\Sigma}^{-1} = \mathbf{U}^T\Lambda^{-1}\mathbf{U}$.

Hence, the antedependence $\boldsymbol{\Sigma}$ can be partitioned in terms of its autoregressive parameters (ϕ_i) and variance parameters (σ_t^2). Figure 6.2.3 shows the effect of smoothing both the autoregressive and variance parameters of $\boldsymbol{\Sigma}$ for various values of α.

6.3 A comparison with other forms of smoothing

The value of the smoothing parameter α for a particular data set can be determined by cross validation using the algorithm of the previous chapter, and inferences made using a 'scaled' Kenward Roger (KR) adjustment. For the cardiac enzyme data we

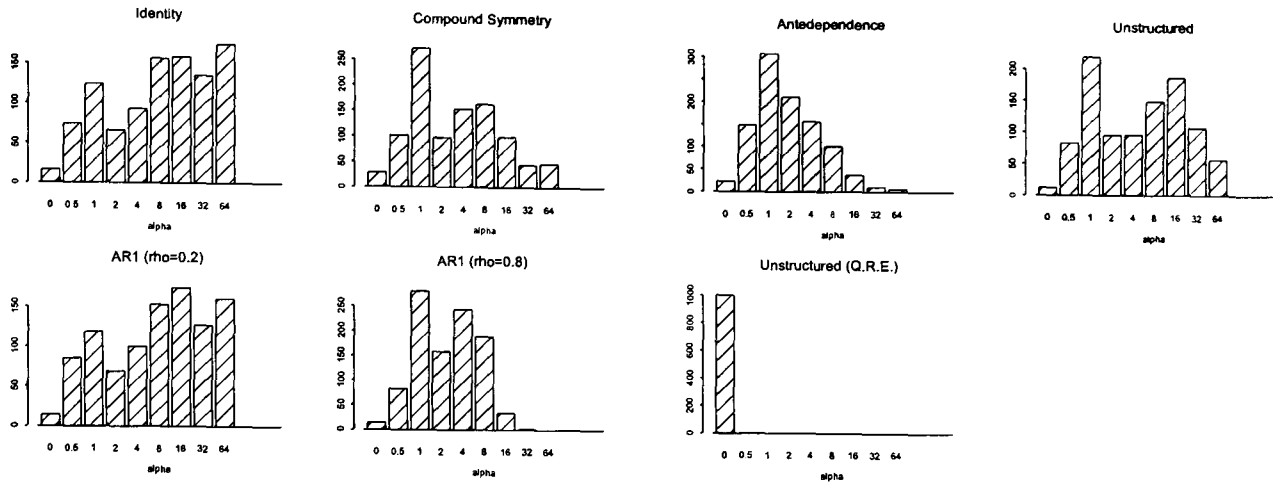


Figure 6.3.1: *Distributions of Smoothing Parameter Estimates α obtained from 1000 simulations for Local-Stationarity Penalised Likelihood Smoothing.*

obtain a value of $\alpha = 1$ which indicates only moderate smoothing. This is perhaps unsurprising as likelihood ratio tests indicate the lack of fit of an AR1 structure to this data.

We repeat the simulations of the pilot study of Chapters 2 and 5 to assess the efficiency of this method of smoothing, with α chosen by cross validation from the set $\alpha \in \{0, 0.5, 1, 2, 4, 8, 16, 32, 64\}$ for data arising from a number of underlying covariance structures including identity, compound symmetry, AR1 (low and high correlations), antedependence and two 'unstructured' forms.

The distributions of α from 1000 simulations of each of these underlying structures can be seen in Figure 6.3.1. Again, we see what we might intuitively expect, although some differences with penalised smoothing which are worthy of note. In particular, the distributions show that this method suggests a higher proportion of smaller values of α for data arising from underlying stationary covariance structures (identity, compound symmetry etc) where departures from constant variance in the poorly estimated covariance structures are given greater weight. The distributions are more uniformly distributed than for the other methods in these cases. There is still a preference for small values of α for data arising from non-stationary structures. For the Q.R.E. structure, $\alpha = 0$ in all 1000 datasets, indicating no smoothing at all.

6.3.1 Results

Table 6.3.1 shows the test sizes for Wald tests for Wald tests for an overall treatment-time interaction and a set of orthogonal polynomial contrasts (C1-C4) for the simulated data described above.

Once again, for each underlying ('true') structure Kenward Roger adjusted sizes are given for the appropriate Wald tests based on an unstructured, antedependence or AR1 covariance structure, which can be compared with the test sizes resulting from the smoothed estimate adopting the scaled adjustment. Test sizes resulting from AR1 and antedependence estimates are marked with an asterisk (*) where they are not appropriate for the underlying data. Recall that under the null hypothesis of no difference between treatment groups, a 95% probability interval for the proportion of tests out of 1000 leading to a rejection of the test is (0.036, 0.064).

Test sizes for the smoothed estimates are close to nominal levels using the scaled KR adjustment for all underlying covariance structures (except Q.R.E.), even where smoothing towards an AR1 structure that appears inappropriate for the data.

For the data arising from the Q.R.E. structure, the scaled test sizes are necessarily the same as for the antedependence structure. The nominal test size is not achieved in this instance since the antedependence structure is not appropriate for these data.

Power levels for the Wald tests are shown in Table 6.3.2, for detecting a linear difference between the means of two treatment groups (fixed to achieve a power of around 75% using the unstructured matrix). Hence, the power levels are directly comparable with those given in Tables 5.2.3 and 5.2.4 of Chapter 5 for the direct and penalised smoothing estimates.

As before, we see a noticeable improvement in power over adopting an unstructured covariance estimate in all cases, except for data arising from an underlying compound symmetry structure. In this case both ends of the smoothing spectrum

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Stationary Structures						
Identity	Unstr	0.044	0.055	0.037	0.046	0.049
	Ante	0.061	0.062	0.044	0.059	0.060
	AR1	0.051	0.058	0.037	0.056	0.057
	Loc-Penal	0.056	0.058	0.039	0.060	0.062
Compound Symmetry	Unstr	0.047	0.059	0.063	0.059	0.056
	Ante*	0.053	0.003	0.022	0.072	0.100
	AR1*	0.045	0.004	0.016	0.060	0.082
	Loc-Penal	0.055	0.005	0.019	0.066	0.092
AR1 ($\rho = 0.2$)	Unstr	0.051	0.056	0.048	0.054	0.065
	Ante	0.060	0.058	0.059	0.055	0.066
	AR1	0.047	0.054	0.048	0.048	0.060
	Loc-Penal	0.052	0.061	0.052	0.052	0.067
AR1 ($\rho = 0.8$)	Unstr	0.039	0.062	0.044	0.041	0.033
	Ante	0.050	0.059	0.048	0.042	0.035
	AR1	0.031	0.051	0.038	0.032	0.020
	Loc-Penal	0.037	0.057	0.043	0.036	0.029
Non-Stationary Structures						
Ante-dependence	Unstr	0.036	0.051	0.049	0.043	0.051
	Ante	0.051	0.049	0.047	0.051	0.064
	AR1*	0.075	0.052	0.053	0.049	0.027
	Loc-Penal	0.065	0.076	0.072	0.079	0.067
Unstr.	Unstr	0.044	0.047	0.041	0.052	0.052
	Ante*	0.042	0.040	0.034	0.049	0.065
	AR1*	0.057	0.033	0.026	0.051	0.059
	Loc-Penal	0.060	0.037	0.032	0.055	0.067
Unstr. (Q.R.E.)	Unstr	0.057	0.053	0.056	0.056	0.063
	Ante*	0.100	0.055	0.057	0.074	0.135
	AR1*	0.137	0.278	0.017	0.000	0.000
	Loc-Penal	0.100	0.055	0.057	0.074	0.135

Table 6.3.1: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of type 1 errors (Size), using a Wald statistic with a 'KR' adjustment.*

Underlying Covariance Structure	Covariance Estimate	Proportion of Significant Test Results (out of 1000) (Null model- No treatment/time differences)				
		Inter-action	Orthogonal Polynomial Contrasts			
			C1	C2	C3	C4
Non-Stationary Structures						
Identity	Unstr	0.747	0.974	0.037	0.046	0.049
	Ante	0.842	0.981	0.044	0.059	0.060
	AR1	0.896	0.982	0.037	0.056	0.057
	Loc-Penal	0.897	0.984	0.039	0.060	0.062
Compound Symmetry	Unstr	0.754	0.965	0.063	0.059	0.056
	Ante*	0.486	0.833	0.022	0.072	0.100
	AR1*	0.523	0.819	0.016	0.060	0.082
	Loc-Penal	0.544	0.833	0.019	0.066	0.092
AR1 ($\rho = 0.2$)	Unstr	0.765	0.978	0.048	0.054	0.065
	Ante	0.864	0.988	0.059	0.055	0.066
	AR1	0.901	0.988	0.048	0.048	0.060
	Loc-Penal	0.906	0.991	0.052	0.052	0.067
AR1 ($\rho = 0.8$)	Unstr	0.744	0.967	0.044	0.041	0.033
	Ante	0.862	0.975	0.048	0.042	0.035
	AR1	0.879	0.977	0.038	0.032	0.020
	Loc-Penal	0.886	0.979	0.043	0.036	0.029
Non-Stationary Structures						
Ante-dependence	Unstr	0.764	0.748	0.049	0.043	0.051
	Ante	0.859	0.778	0.047	0.051	0.064
	AR1*	0.534	0.789	0.053	0.049	0.027
	Loc-Penal	0.822	0.838	0.072	0.079	0.067
Unstr.	Unstr	0.777	0.948	0.041	0.052	0.052
	Ante*	0.845	0.949	0.034	0.049	0.065
	AR1*	0.824	0.957	0.026	0.051	0.059
	Loc-Penal	0.852	0.960	0.032	0.055	0.067
Unstr. (Q.R.E.)	Unstr	0.782	0.177	0.056	0.056	0.063
	Ante*	0.801	0.182	0.057	0.074	0.135
	AR1*	0.287	0.511	0.017	0.000	0.000
	Loc-Penal	0.801	0.182	0.057	0.074	0.135

Table 6.3.2: *Summary of results from 1000 simulations of Pilot Study Design. Table gives the proportion of significant results (Power), using a Wald statistic with a 'KR' adjustment.*

(antedependence to AR1) are inappropriate for these data and so that the tests using the penalised estimate are less powerful, as we might expect. It is also worthy of note that for data arising from an underlying antedependence form, the local-stationarity penalised estimates are less powerful than simply adopting an antedependence structure for the data, even in such a badly estimated context.

In all cases the power levels achieved using the local-stationarity penalised estimates are below those for the corresponding tests using the direct and penalised smoothing estimates of Chapter 5.

6.4 Discussion

The results reported above are somewhat disappointing in that they do not point to any significant advantages in adopting a local-stationarity penalised likelihood approach over the methods discussed in Chapter 5 in this small sample setting. There are two possible reasons for this.

- Firstly, the starting point for the smoothing process being an antedependence(1) structure may not be entirely appropriate for data which is highly non-stationary as it matches only the first two diagonals of the unstructured matrix. (This effectively smooths the remaining covariance elements before we start). This could be remedied by considering smoothing between higher order structures, such as antedependence(2) and ar(2) etc.

- Secondly, it is also possible that with only $p = 5$ time points considered in the pilot study, the 'smoothing window' of size $q = 3$ yields only 3 submatrices from the diagonal from which to assess departures from local stationarity, thus limiting the procedure. There is no way to improve upon this in the case of only 5 time points, but it is possible that considering data sets with a greater number of time points would allow the procedure greater scope. (e.g. $p = 10$ time points yields 7 submatrices of size 3).

In both these cases numerical problems are likely to be a limiting factor. As the number of time points increases, the maximisation of the penalised likelihood expression in (6.2.2) by numerical methods becomes unstable, and increasing the available data makes cross-validation infeasible as a means of choosing the smoothing parameter. Such problems were encountered in the penalised smoothing estimator of Chapter 5, but are amplified in this approach with the additional fitting (by maximum likelihood) of the AR(1) estimates of the submatrices from the diagonal of $\boldsymbol{\Sigma}$. Also, were more data are available, there would be little reason for not simply adopting the (unstructured) sample covariance matrix without recourse to computationally expensive techniques. Such problems can only be further amplified in extending these procedures beyond the restrictive setting of balanced and complete data.

It seems, as was flagged in the discussion of the previous chapter, that for data sets arising from very small samples of repeated measurements, there is little scope for adaptive methods based on the estimation of a covariance model which improve upon existing methods.

Part III

Methods with Less Dependence on the Covariance Structure

Chapter 7

The Empirical 'Sandwich' Estimator

7.1 Introduction

Part II showed the problems involved in developing appropriate adaptive methods for the estimation of the covariance structure in repeated measures analyses with small samples. It is clearly not possible in the very small sample setting to determine an adequate model for the covariance structure of the data of low dimensionality (i.e. based on few parameters) that outperforms the unstructured estimate. This was seen to be the case in the restricted setting of complete and balanced data, where a saturated means model ensured that the estimated covariance structure did not inform the mean estimates, but was necessary only for the appropriate estimation of their standard errors for the purposes of inference. In more complicated settings, where data are unbalanced (by design or due to missing values), the covariance structure, which may be poorly determined, affects both the parameter estimates $\boldsymbol{\beta}$ and their standard errors.

In this part of the Thesis attention is focused on methods which are less dependent on the covariance structure. That is, on methods which do not rely upon the covariance structure for estimation of the mean parameters, but make inferences using their ordinary least squares estimates and use some consistent estimate of $\mathrm{Var}(\mathbf{y})$ to derive

their standard errors.

One such method is the so called robust approach of Section 1.3 which uses the empirical sandwich estimator for the covariance matrix of the fixed effects parameters $\boldsymbol{\beta}$, which are based on their least squares estimates. Use of this estimator attempts to make inferences about $\boldsymbol{\beta}$ robust to misspecification of the covariance structure. It is the favoured approach when analysing categorical data, in finding a marginal regression for the mean response through a multivariate extension to quasilikelihood methods for generalised linear models given by generalised estimating equations (GEEs).

It is well known that such a robust approach has poor small sample properties (see, for example, Gunsolley *et al.* (1995)), but recent attempts to improve the performance of the sandwich estimator are worthy of consideration in the context of small samples, to determine whether the use of a modified estimator, together with a suitable adjustment, can lead to a test with nominal properties for continuous, normally distributed data.

7.2 The GEE Approach

Let y_{ij} represent jth observation on the ith subject, $j = 1, \ldots, n_i$ and $i = 1, \ldots, , m$. That is there are n_i observations on subject i and $n = \sum_{i=1}^{m} n_i$ observations in total. Let the vector of observations on the ith subject be $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ with corresponding vector of means $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^T$. Also, let $\mathbf{x}_{ij}$ be the r-dimensional vector of explanatory variables , so that $\mathbf{X}_i = (\mathbf{x}_{i1}^T, \ldots, \mathbf{x}_{in_i}^T)^T$ is the design matrix corresponding to the ith subject.

The generalised linear model approach is to model

$$g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} \tag{7.2.1}$$

where g is a ('link') function of the expected response and $\boldsymbol{\beta}$ is the usual vector of estimated regression coefficients.

Writing,

$$\mathbf{D}_i^T = \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \dfrac{x_{i11}}{g'(\mu_{i1})} & \cdots & \dfrac{x_{in_i1}}{g'(\mu_{in_i})} \\ \vdots & & \vdots \\ \dfrac{x_{i1r}}{g'(\mu_{i1})} & \cdots & \dfrac{x_{in_ir}}{g'(\mu_{in_i})} \end{bmatrix}$$

the generalised estimating equation (GEE) approach of Liang and Zeger (1986) finds $\boldsymbol{\beta}$ by solving the following estimating equations

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \{\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = \mathbf{0} \tag{7.2.2}$$

where $\mathbf{V}_i$ is the covariance matrix of the $\mathbf{y}_i$, which is typically modelled as $\mathbf{V}_i = \phi \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{\frac{1}{2}}$, where $\mathbf{A}_i$ is an $n_i \times n_i$ diagonal matrix with $\mathrm{Var}(\mu_{ij})$ as the jth diagonal element, ϕ is a dispersion parameter, and $\mathbf{R}_i(\boldsymbol{\alpha})$ is a 'working' correlation matrix that is specified by the vector of parameters $\boldsymbol{\alpha}$. Hence, if $\mathbf{R}_i(\boldsymbol{\alpha})$ is the true correlation matrix of $\mathbf{y}_i$, then $\mathbf{V}_i$ is the true covariance matrix of $\mathbf{y}_i$.

The equations are solved iteratively and (under mild regularity conditions) the solutions in the vector $\hat{\boldsymbol{\beta}}$ are consistent and asymptotically normal. The covariance matrix of the fixed effects estimates $\hat{\boldsymbol{\beta}}$ is consistently estimated by the empirical 'sandwich' estimator

$$\begin{aligned} \mathbf{V}_S &= \left(\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \left\{ \sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathrm{Var}(\mathbf{y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \left(\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \\ &= \mathbf{V}_M \left\{ \sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathrm{Var}(\mathbf{y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \mathbf{V}_M \end{aligned} \tag{7.2.3}$$

where $\mathbf{V}_M = \left(\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1}$ is the model based estimate of $\text{Var}(\hat{\boldsymbol{\beta}})$. In practice $\text{Var}(\mathbf{y}_i)$ is replaced by $\mathbf{r}_i \mathbf{r}_i^T = \{\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})\}\{\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})\}^T$, its estimate based on the ordinary model residuals.

The sandwich estimator is a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, even if the working correlation matrix is misspecified. If $\text{Var}(\mathbf{y}_i) = \mathbf{V}_i$, then the sandwich estimator reduces to $\mathbf{V}_M$, the model-based estimate.

For (continuous) responses, with normally distributed errors, and an identity link, we find

$$\mathbf{D}_i^T = \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} = \mathbf{X}_i^T$$

so that,

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \tag{7.2.4}$$

and setting this expression equal to zero gives us the usual 'normal' equations, with solution vector

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^{m} (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i \tag{7.2.5}$$

This leads to the robust approach referred to in Chapter 1, where a 'working' covariance structure $\mathbf{W}^{-1} = \mathbf{V}$ is used for calculation of the regression coefficients, but their standard errors are given by the sandwich estimator. That is,

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \tag{7.2.6}$$

with

$$\text{Var}(\hat{\boldsymbol{\beta}}_W) = \mathbf{V}_S = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\left\{\sum_{i=1}^{m}\mathbf{X}_i^T\mathbf{W}_i\text{Var}(\mathbf{y}_i)\mathbf{W}_i\mathbf{X}_i\right\}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \qquad (7.2.7)$$

where $\mathbf{V} = \text{diag}(\mathbf{V}_i)$ has a block diagonal structure.

As before, taking $\text{Var}(\mathbf{y}_i) = \mathbf{r}_i\mathbf{r}_i^T = (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^T$ uses the observed correlations between the residuals. Also, taking $\mathbf{W}^{-1} = \mathbf{I}$ means that the regression parameter estimates will be their ordinary least squares estimates, $\hat{\boldsymbol{\beta}}_{OLS}$, but with their standard errors adjusted to account for the observed correlation structure. This will be the approach adopted throughout this Chapter.

7.3 Small Sample Properties of the Sandwich Estimator

Gunsolley *et al.* (1995) compare the performance of the sandwich estimator of the covariance matrix for the fixed effects parameters $\mathbf{V}_S$ with the model based estimate $\mathbf{V}_M$ to show that testing based on the GEE approach results in type 1 error rates that are too liberal when dealing with small samples. They showed this through a number of simulation studies based on binary responses, and suggest that 'appropriate' size is obtained only when the number of subjects exceeded the number of observations per subject.

Mancl and DeRouen (2001) consider bias in the robust estimator $\mathbf{V}_S = \text{Var}(\hat{\boldsymbol{\beta}})$ due to the use of ordinary residuals $\mathbf{r}_i = \{\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i(\boldsymbol{\beta})\}$ to estimate $\text{Var}(\mathbf{y}_i)$, where the number of subjects is small.

They show this by considering a Taylor series expansion of the residual vector $\mathbf{r}_i$ about $\boldsymbol{\beta}$,

$$\mathbf{r}_i = \mathbf{e}_i + \frac{\partial \mathbf{e}_i}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \qquad (7.3.1)$$

where $\mathbf{e}_i = \{\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}$ and $i = 1, \ldots, m$.

This leads to the following expression of the expectation of $\text{Var}(\mathbf{y}_i)$,

$$\text{E}(\mathbf{r}_i\mathbf{r}_i^T) \approx (\mathbf{I} - \mathbf{H}_{ii})\text{Var}(\mathbf{y}_i)(\mathbf{I} - \mathbf{H}_{ii})^T + \sum_{j \neq i} \mathbf{H}_{ij}\text{Var}(\mathbf{y}_i)\mathbf{H}_{ij}^T \tag{7.3.2}$$

where

$$\mathbf{H}_{ij} = \mathbf{D}_i \left(\sum_{k=1}^{m} \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathbf{D}_k \right)^{-1} \mathbf{D}_j^T \mathbf{V}_j^{-1} \tag{7.3.3}$$

Taking the summation term in (7.3.2) to be negligable, and hence zero, this leads to the bias corrected sandwich covariance estimator, $\mathbf{V}_{S(BC)}$, of $\text{Var}(\hat{\boldsymbol{\beta}})$, given by

$$\mathbf{V}_{S(BC)} = \mathbf{V}_M \left\{ \sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{I} - \mathbf{H}_{ii})^{-1} \text{Var}(\mathbf{y}_i) (\mathbf{I} - \mathbf{H}_{ii}^T)^{-1} \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \mathbf{V}_M \tag{7.3.4}$$

Mancl and DeRouen show that this bias corrected robust approach improves the type 1 error rate when the sample size (number of subjects) is low, through simulations of binary responses involving 10-40 subjects in varying cluster sizes of between 16 and 24 observations. They find in separate tests, of null hypotheses involving each of two regression parameters and their interaction (all on 1 d.f.), that Wald-type tests using the bias-corrected sandwich estimator achieved test sizes closer to the nominal level of 5% than tests involving the unadjusted sandwich estimator. Their approach also outperforms other robust approaches such as the Jackknife. They do find, however, that test sizes using a χ^2 null distribution are still somewhat inflated and it is necessary to adopt an F distribution to achieve tests with nominal properties. In this context they adopt the number of subjects minus the number of regression parameters as the 'appropriate' measure of the denominator degrees of freedom, although they do recognise the arbitrary nature of this choice.

For normal errors, the Mancl and DeRouen approach suggests a bias corrected version of the sandwich estimator, which replaces $\mathbf{r}_i$ in $\mathbf{V}_S$ with $(\mathbf{I} - \mathbf{H}_i)^{-1}\mathbf{r}_i$, where

$$\mathbf{H}_i = \mathbf{X}_i \mathbf{V}_M \mathbf{X}_i^T \mathbf{W}_i \tag{7.3.5}$$

Table 7.3.1 shows the results from 1000 simulations each of study designs (A), (B) and (C) of Chapter 2, Section 2.3, for Wald tests using both the sandwich estimator and the bias corrected sandwich estimator suggested by Mancl and DeRouen in the context of a normal response. In each case the test sizes, for a treatment/time interaction on 4 d.f. for designs (A) and (B), and a treatment effect on 4 d.f. for design (C) are given with reference to both a chi-squared distribution and an F distribution using residual degrees of freedom in the denominator (the number of observations minus the number of parameters).

It can be seen from these results that, in this small sample setting, nominal test sizes (of 5%) are not achieved by the bias corrected sandwich estimator using either a χ^2 of F distribution, although this approach does offer some improvement over the unadjusted sandwich estimator. Pan and Wall (2002) note that performance of this approach, particularly the Wald chi-squared test, is worse when testing multiple parameters rather than a single parameter. This may account for the difference between these results and those reported by Mancl and DeRouen.

7.4 An Adjusted F-test for the Sandwich Estimator

A number of attempts have been made to give adjusted tests based on the sandwich estimator to deal with the problem of inflated test sizes for small samples. For example, Kauermann and Carroll (2001) show that the (bias adjusted) sandwich estimator $\mathbf{V}_{S(BC)}$ always results in greater variance than the model based covariance matrix for the fixed effects $\mathbf{V}_M$, where the model is correct, and suggest adjusted t-tests for individual regression parameters based on the bias adjusted sandwich estimator. Alternatively, Fay and Graubard (2001) consider a bias adjustment to $\mathbf{V}_S$ which is dependent on the assumption that the working covariance matrix is a scale factor of the true covariance matrix. In the GEE case this implies that

Underlying 'True' Covariance Structure	Method of Inference	Null Dist.	Proportion of Significant Test Results (out of 1000) ('Size')		
			Design A	Design B	Design C
Identity	Sand	χ^2	0.426	0.482	0.388
		F	0.388	0.449	0.339
	Bias Adj.	χ^2	0.248	0.293	0.225
		F	0.207	0.252	0.196
Compound Symmetry	Sand	χ^2	0.419	0.444	0.397
		F	0.394	0.396	0.364
	Bias Adj.	χ^2	0.254	0.272	0.225
		F	0.223	0.229	0.198
AR1 ($\rho = 0.2$)	Sand	χ^2	0.428	0.485	0.384
		F	0.391	0.444	0.345
	Bias Adj.	χ^2	0.275	0.290	0.214
		F	0.249	0.264	0.185
AR1 ($\rho = 0.8$)	Sand	χ^2	0.416	0.452	0.380
		F	0.383	0.399	0.349
	Bias Adj.	χ^2	0.275	0.253	0.215
		F	0.245	0.221	0.188
Antedependence	Sand	χ^2	0.422	0.494	0.372
		F	0.385	0.459	0.336
	Bias Adj.	χ^2	0.257	0.292	0.208
		F	0.225	0.259	0.178
Unstructured	Sand	χ^2	0.423	0.459	0.378
		F	0.390	0.411	0.337
	Bias Adj.	χ^2	0.269	0.247	0.209
		F	0.230	0.211	0.184
Unstr (QRE)	Sand	χ^2	0.430	0.397	0.335
		F	0.386	0.349	0.294
	Bias Adj.	χ^2	0.257	0.210	0.132
		F	0.228	0.189	0.106

Table 7.3.1: *Results from 1000 simulations of Designs (A), (B) and (C).*

the correlation matrix is correctly specified. They suggest testing single parameters based on an $F(1, d)$ distribution (or equivalently a $t(d)$ distribution) where d is a function of the bias adjusted sandwich estimator.

A more general approach is suggested by Pan and Wall (2002) who propose an F-test for the Wald statistic which takes account of the additional variability inherent in the sandwich estimator, but without restriction to assumptions about the working correlation matrix. They describe their approach as being analogous to the testing the mean of a normal distribution with unknown variance where a t-test is preferred

over a z-test.

Their approach is developed in the GEE setting in terms of the hypothesis H_0:$\boldsymbol{\beta} = 0$. although they note the possibility for its extension to the general linear hypothesis H_0:$\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, which is presented below.

Consider the Wald statistic under the null hypothesis,

$$\mathrm{W} = (\mathbf{L}\hat{\boldsymbol{\beta}})^T(\mathbf{L}\mathbf{V}_S\mathbf{L}^T)^{-1}\mathbf{L}\hat{\boldsymbol{\beta}} = v(\mathbf{L}\hat{\boldsymbol{\beta}})^T(v\mathbf{L}\mathbf{V}_S\mathbf{L}^T)^{-1}\mathbf{L}\hat{\boldsymbol{\beta}} \tag{7.4.1}$$

where $\mathbf{L}\hat{\boldsymbol{\beta}} \sim \mathrm{N}_l(\mathbf{0}, \mathbf{L}\mathbf{V}_S\mathbf{L}^T)$. If we assume further that $v\mathbf{L}\mathbf{V}_S\mathbf{L}^T$ has a scaled Wishart distribution, that is $v\mathbf{L}\mathbf{V}_S\mathbf{L}^T \sim \mathrm{W}_l(v, \mathbf{L}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{L}^T)$, then it follows that W has the same distribution as Hotelling T^2. That is,

$$\frac{v - l + 1}{vl}\mathrm{W} \sim \mathrm{F}(l, v - l + 1) \tag{7.4.2}$$

where v is chosen from the data to match $\widehat{\mathrm{Var}}\{\mathrm{vec}(v\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\} = v^2\widehat{\mathrm{Var}}\{\mathrm{vec}(\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\}$ from the Wishart distribution with an empirically based estimator.

From the Wishart distribution, we have

$$\mathrm{Var}\{\mathrm{vec}(v\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\} = v(\mathbf{I}_{l^2} + \mathbf{K})\left[\{\mathbf{L}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{L}^T\} \otimes \{\mathbf{L}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{L}^T\}\right] \tag{7.4.3}$$

where $\mathbf{K}$ is the 'commutative' matrix and $\otimes$ is the kronecker product operator, (see Chapter 4, subsection 4.2.2). Also, since $\mathbf{V}_S$ is a reasonable estimator of the covariance matrix of the fixed effects parameters, we have $v\hat{\boldsymbol{\Omega}}$ is an estimator of $\mathrm{Var}\{\mathrm{vec}(v\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\}$ where

$$\hat{\boldsymbol{\Omega}} = (\mathbf{I}_{l^2} + \mathbf{K})\{(\mathbf{L}\mathbf{V}_S\mathbf{L}^T) \otimes (\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\} \tag{7.4.4}$$

To calculate an empirically based estimator, the Pan and Wall adjusted statistic assumes in the GEE approach that, under correct modelling assumptions, the model-based covariance estimator $\mathbf{V}_M$ is an appropriate measure, and hence concentrates on the estimated variance of the middle piece of $\mathbf{V}_S$ from (7.2.3), that is $\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i$, treating $\mathbf{D}_i$, $\mathbf{V}_i$ and $\mathbf{V}_M$ as fixed.

Let $\mathbf{P}_i = \text{vec}(\mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i)$. Then, the covariance matrix of the mean vector $\mathbf{Q} = \sum_{i=1}^{m} \mathbf{P}_i/m$ is estimated empirically by

$$\begin{aligned} \hat{\mathbf{T}} &= \sum_{i=1}^{m} (\mathbf{P}_i - \bar{\mathbf{P}})(\mathbf{P}_i - \bar{\mathbf{P}})^T / m(m-1) \\ &= \frac{1}{m^2} \widehat{\text{Var}}(\mathbf{P}_i) \end{aligned} \tag{7.4.5}$$

where $\bar{\mathbf{P}} = \sum_{i=1}^{m} \mathbf{P}_i/m$. This follows since $\sum_{i=1}^{m} \mathbf{P}_i/m$ is itself an unbiased estimator of $\mathbf{Q}$. Then, the covariance matrix of $\text{vec}(\mathbf{L}\mathbf{V}_S\mathbf{L}^T)$, that is

$$\text{Var}(\mathbf{L}\mathbf{V}_S\mathbf{L}^T) = (\mathbf{L}\mathbf{V}_M \otimes \mathbf{L}\mathbf{V}_M) \sum_{i=1}^{m} \mathbf{P}_i$$

is given (empirically) by

$$\begin{aligned} \hat{\mathbf{\Psi}} = \widehat{\text{Var}}\{\text{vec}(\mathbf{L}\mathbf{V}_S\mathbf{L}^T)\} &= (\mathbf{L}\mathbf{V}_M \otimes \mathbf{L}\mathbf{V}_M) \widehat{\text{Var}}(\mathbf{P}_i) (\mathbf{L}\mathbf{V}_M \otimes \mathbf{L}\mathbf{V}_M)^T \\ &= m^2 (\mathbf{L}\mathbf{V}_M \otimes \mathbf{L}\mathbf{V}_M) \hat{\mathbf{T}} (\mathbf{L}\mathbf{V}_M \otimes \mathbf{L}\mathbf{V}_M)^T \end{aligned} \tag{7.4.6}$$

v is chosen to match as closely as possible the quantities $v\hat{\mathbf{\Omega}}$, the estimated covariance matrix of $v\mathbf{L}\mathbf{V}_S\mathbf{L}^T$ under the Wishart assumption with $v^2\hat{\mathbf{\Psi}}$, its empirically calculated estimate. The solution favoured by Pan and Wall is to find v to minimise the sum of squared errors between $v\text{vec}(\hat{\mathbf{\Psi}})$ and $\text{vec}(\hat{\mathbf{\Omega}})$.

Writing, $\mathbf{a} = \text{vec}(\hat{\mathbf{\Psi}})$ and $\mathbf{b} = \text{vec}(\hat{\mathbf{\Omega}})$, the least squares estimate of v minimises $(v\mathbf{a} - \mathbf{b})^T(v\mathbf{a} - \mathbf{b})$, so that differentiating and setting equal to zero, we find

$$2v\mathbf{a}^T\mathbf{a} - 2\mathbf{a}^T\mathbf{b} = 0$$
$$v = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

That is,

$$v = \frac{\mathrm{tr}(\hat{\mathbf{\Psi}}\hat{\mathbf{\Omega}})}{\mathrm{tr}(\hat{\mathbf{\Psi}}^2)} \tag{7.4.7}$$

Pan and Wall compare their adjusted F-statistic for a Wald test using the sandwich estimator with the usual $\chi^2(l)$ distribution, and show in the context of a logistic model that their adjusted test reduces the inflated test sizes achieved using a χ^2 test to nominal levels.

Pan and Wall note that it is possible to combine their approach with a bias correction such as that of Mancl and DeRouen. This is achieved by simply replacing $\mathbf{r_i}$ with $(\mathbf{I} - \mathbf{H}_{ii})\mathbf{r}_i$ in $\mathbf{P}_i$ and using the bias-adjusted estimator to determine v. This results in a Wald statistic that accounts for both the bias and variability of the sandwich estimator. However, they find in simulations that there is no need for such a combination as their adjusted statistic alone is somewhat conservative for simple hypotheses in the GEE setting, giving type 1 error rates below the nominal level.

As before, in the context of continuous normally distributed data, we proceed in the above approach by substituting $\mathbf{P}_i = \mathrm{vec}(\mathbf{X}_i\mathbf{W}_i\mathbf{r}_i\mathbf{r}_i^T\mathbf{W}_i\mathbf{X}_i)$ in (7.4.5). To test this approach we again consider simulations of study designs (A), (B) and (C). Results from 1000 simulations are shown in Table 7.4.1, using both the adjusted F-test of Pan and Wall and a combined adjustment using the bias adjusted sandwich estimator of Mancl and DeRouen.

Looking at the table it is clear in testing the (multiple parameter) hypotheses of no treatment/time interaction in the simple repeated measures studies of designs (A) and (B), and no treatment effect in the crossover study of design (C) that adjusted tests using the Pan and Wall approach alone are still somewhat inflated, but that

Underlying 'True' Covariance Structure	Method of Inference		Proportion of Significant Test Results (out of 1000) ('Size')		
			Design A	Design B	Design C
Identity	PW		0.115	0.135	0.094
	Comb Adj.		0.055	0.056	0.039
Compound Symmetry	PW		0.119	0.130	0.081
	Comb Adj.		0.057	0.051	0.034
AR1 ($\rho = 0.2$)	PW		0.142	0.152	0.098
	Comb Adj.		0.068	0.068	0.037
AR1 ($\rho = 0.8$)	PW		0.166	0.160	0.073
	Comb Adj.		0.083	0.064	0.034
Antedependence	PW		0.109	0.132	0.051
	Comb Adj.		0.047	0.055	0.018
Unstructured	PW		0.109	0.088	0.067
	Comb Adj.		0.051	0.042	0.025
Unstr (QRE)	PW		0.206	0.136	0.027
	Comb Adj.		0.100	0.063	0.009

Table 7.4.1: *Results from 1000 simulations of Designs (A), (B) and (C).*

the combined adjustment (accounting for both bias and variability) comes close to achieving nominal levels in each of these designs. It would appear that the Pan and Wall approach, of an adjusted F-test using the sandwich estimator (combined with a suitable bias adjustment), leads to an appropriate test based on the general linear hypothesis for small sample repeated measures which are normally distributed.

Further consideration of the performance of this combined adjustment for tests involving the sandwich estimator will be presented in Chapter 9.

Chapter 8

Box's Correction: A Modified ANOVA Statistic

8.1 Introduction

We have seen that removing the estimated covariance structure from the estimation of the regression parameters leads to an improvement in the small sample behaviour of tests involving those parameters in repeated measures designs.

We saw this in the assessment of existing methods in Chapter 2, where Wald tests using the Kenward Roger adjustment for balanced data which allowed for exact tests (such as Hotelling T^2 and split-plot ANOVA) give adequate control over the type 1 error rate in the small sample setting. In such cases, the estimated covariance structure is not used in the estimation of the mean parameters, but is necessary for estimates of their precision. This control is seen to deteriorate where the covariance structure enters the estimation step, and is worse still in situations which are unbalanced.

We have also seen in Chapter 7 how testing based on robust methods, using the empirical sandwich estimator can be modified to account for bias and variability in this estimator to give an appropriate test statistic and (adjusted) null distribution based on ordinary least squares estimates of the mean parameter estimates.

A natural extension of this approach is to consider removing the estimated covariance structure from estimates of precision also. That is, to develop a method which ignores the covariance structure completely in calculating the test statistic, using ordinary regression or ANOVA, and then corrects the null distribution for the dependence in the repeated measurements.

One such approach, outlined in Chapter 1 (Section 1.3), is offered by Box's correction, Box (1954a,b), which suggests a modification to the one-way ANOVA F-statistic to account for departures from this model.

This approach is advocated by Bellavance *et al.* (1996) in the context of crossover designs, to account for the correlations within subjects arising from repeated measurements. They show that it gives adequate control over the type 1 error rate compared with ordinary least squares (OLS) for data arising from a variety of covariance structures which do not allow exact tests. It is considered here in a wider context.

8.2 Box's Correction

Let $\mathbf{X}$ be the $(n \times r)$ design matrix with all terms included, and $\mathbf{X}_R$ $(n \times (r - c))$ have the terms to be tested removed, and define $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and $\mathbf{B} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \mathbf{X}_R(\mathbf{X}_R^T\mathbf{X}_R)^{-1}\mathbf{X}_R^T$. Then, under the ANOVA assumptions of independence between observations (on a subject) and homogeneity of variance, the appropriate test statistic is given by

$$\mathrm{F} = \frac{(n-r)}{c}\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \sim \mathrm{F}(c, n-r) \tag{8.2.1}$$

Box's correction suggests an adjustment to the (null) F distribution which attempts to make the ANOVA statistic robust to departures from the assumption of independence. It is easily derived using standard results for quadratic forms in normal variates.

Let $\mathbf{y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, and consider the quadratic form $Q = \mathbf{y}^T\mathbf{A}\mathbf{y}$, for a fixed matrix $\mathbf{A}$. Then,

$$Q = \mathbf{y}^T\mathbf{A}\mathbf{y} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2 \tag{8.2.2}$$

where λ_i are the eigenvalues of $\mathbf{A\Sigma}$, and the χ_i^2 are i.i.d. $\chi^2(1)$ variates. In particular, we have the following results.

$$\mathrm{E}(Q) = \mathrm{tr}(\mathbf{A\Sigma}) \tag{8.2.3}$$

and,

$$\mathrm{Var}(Q) = 2\mathrm{tr}(\mathbf{A\Sigma A\Sigma}) \tag{8.2.4}$$

Also, for $Q_1 = \mathbf{y}^T\mathbf{B}\mathbf{y}$ and $Q_2 = \mathbf{y}^T\mathbf{A}\mathbf{y}$, we have

$$\mathrm{Cov}(Q_1, Q_2) = 2\mathrm{tr}(\mathbf{A\Sigma B\Sigma}) \tag{8.2.5}$$

The key approximation (Satterthwaite (1941)) is to consider the quadratic forms to be distributed as a constant times a chi-squared distribution. That is,

$$Q = \mathbf{y}^T\mathbf{A}\mathbf{y} \overset{\text{approx}}{\sim} g\chi^2(h)$$

The constant and the degrees of freedom parameters, g and h are chosen by matching the first and second moments of $\mathbf{y}^T\mathbf{A}\mathbf{y}$ with those of the scaled chi-squared distribution. i.e.

$$\mathrm{E}\{g\chi^2(h)\} = gh = \mathrm{tr}(\mathbf{A\Sigma})$$

and,

$$\text{Var}\{g\chi^2(h)\} = 2g^2h = 2\text{tr}(\mathbf{A\Sigma A\Sigma})$$

from which we obtain

$$g = \frac{\text{tr}\{(\mathbf{A\Sigma})^2\}}{\text{tr}(\mathbf{A\Sigma})} \text{ and } h = \frac{\{\text{tr}(\mathbf{A\Sigma})\}^2}{\text{tr}\{(\mathbf{A\Sigma})^2\}}$$

Box thus describes the ANOVA F-statistic, the ratio of quadratic terms $Q_1 = \mathbf{y}^T\mathbf{B}\mathbf{y}$ and $Q_2 = \mathbf{y}^T\mathbf{A}\mathbf{y}$, assumed independent, to be approximated by a scaled F-distribution.

$$\frac{Q_1}{Q_2} = \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} b\text{F}(h_1, h_2) \tag{8.2.6}$$

i.e. $b = \dfrac{g_1h_1}{g_2h_2} = \dfrac{\text{tr}(\mathbf{B\Sigma})}{\text{tr}(\mathbf{A\Sigma})}$, leading to the result of Section 1.3, which is repeated below.

$$\text{F} = \frac{(n-r)}{c}\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} \psi F(v1, v2) \tag{8.2.7}$$

where,

$$\psi = \frac{(n-r)}{c}\frac{\text{tr}(\mathbf{B\Sigma})}{\text{tr}(\mathbf{A\Sigma})} \tag{8.2.8}$$

$$v1 = \frac{\{\text{tr}(\mathbf{B\Sigma})\}^2}{\text{tr}\{(\mathbf{B\Sigma})^2\}} \tag{8.2.9}$$

$$v2 = \frac{\{\text{tr}(\mathbf{A\Sigma})\}^2}{\text{tr}\{(\mathbf{A\Sigma})^2\}} \tag{8.2.10}$$

In practice, any consistent estimator of $\mathbf{\Sigma} = \text{Var}(\mathbf{y})$ may be used to compute the adjusted F-distribution parameters. Jones and Kenward (2003) suggest the use of the ordinary least squares covariance estimate is in keeping with the spirit of

Underlying 'True' Covariance Structure	Method of Inference		Proportion of Significant Test Results (out of 1000) ('Size')		
			Design A	Design B	Design C
Identity	Box F		0.026	0.024	0.038
Comp Sym	Box F		0.023	0.017	0.032
AR1 ($\rho = 0.2$)	Box F		0.039	0.030	0.028
AR1 ($\rho = 0.8$)	Box F		0.035	0.024	0.020
Antedependence	Box F		0.046	0.050	0.021
Unstructured	Box F		0.027	0.025	0.021
Unstr (QRE)	Box F		0.061	0.040	0.004

Table 8.2.1: *Results from 1000 simulations of Designs (A), (B) and (C).*

this approach. However, for data which are unbalanced or have missing values, so that the OLS and REML estimates do not coincide, it may be more practical to simply adopt the unstructured REML estimate, which is widely implemented in existing software. In cases where an unstructured REML estimate cannot be computed (as can occur, for example, where there are too many measurements on too few subjects), $\mathbf{\Sigma}$ may be taken to be the most complex covariance structure that the data will support, such as a (high order) antedependence structure. A further advantage of this method is that it does not require a non-singular estimate of the covariance structure, so that, in such cases, it is possible to proceed using an 'empirical' estimator, such as the sample covariance matrix.

Bellavance *et al.* show that this modified F-test approximation gives adequate control over the type 1 error in the context of crossover studies. Table 8.2.1 shows the results of 1000 simulations of each of designs (A), (B) and (C) of Chapter 2, which are undertaken to show how Box's correction performs across a wider variety of data problems.

Looking at the table we see that Box's correction reduces the test sizes under each of the study designs to below the nominal level of 5%, resulting in a test statistic which is largely conservative.

8.3 A Further Modification to the ANOVA Statistic

The results of the simulations of the last section show that Box's modified F-statistic is conservative, giving excessive control to the type 1 error rate and resulting in test sizes well below the nominal rate. In this section we consider a modification to Box's correction, which attempts to find a simplified (ANOVA-type) test statistic for repeated measures problems, based on ordinary least squares estimates of β, and giving test sizes close to nominal levels.

One such approach is to consider the numerator term $\mathbf{y}^T\mathbf{B}\mathbf{y}$ as a scaled non-central chi-squared distribution

$$Q_1 = \mathbf{y}^T\mathbf{B}\mathbf{y} \overset{\text{approx}}{\sim} g_1\chi^2(h_1;\partial) \tag{8.3.1}$$

which under Box's assumption of independence between the numerator and denominator quadratic terms, Q_1 and Q_2, would lead to a scaled non-central F-distribution for the ratio Q_1/Q_2. i.e.

$$\frac{Q_1}{Q_2} = \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} bF(h_1, h_2;\partial)$$

This approach is intuitively attractive as it follows closely to Box's original approach, and it is hoped that the addition of the non-centrality parameter ∂ may improve the performance of the approximation. However, to solve for g_1, h_1 and ∂ in (8.3.1), requires us to match the first three cumulants (central moments) of $Q_1 = \mathbf{y}^T\mathbf{B}\mathbf{y}$ to those of the $g_1\chi^2(h_1;\partial)$ distribution, *viz.*

$$g_1(h_1 + \partial) = \text{tr}(\mathbf{B}\boldsymbol{\Sigma}) \tag{8.3.2}$$

$$2g_1^2(h_1 + 2\partial) = \text{tr}(\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) \tag{8.3.3}$$

$$8g_1^3(h_1 + 3\partial) = \text{tr}(\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) \tag{8.3.4}$$

and the addition of the third-order term $\text{tr}(\mathbf{B\Sigma B\Sigma B\Sigma})$ leads to unstable estimates of the approximated distribution parameters.

An alternative is to consider this ratio directly as a scaled F-distribution (rather than as a ratio of scaled chi-squared distributions). That is, to consider

$$\frac{Q_1}{Q_2} = \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} \lambda \text{F}(v_1, v_2) \tag{8.3.5}$$

Approximate moments of the ratio Q_1/Q_2 can be obtained using the 'delta' method for standard errors of functions of random variables, (see Chapter 4, subsection 4.2.2). In particular, for two variables, X_1 and X_2, we have, restricting ourselves to first order deviations for the mean

$$\text{E}\left(\frac{X_1}{X_2}\right) \approx \frac{\text{E}(X_1)}{\text{E}(X_2)}$$

and,

$$\begin{aligned}\text{Var}\left(\frac{X_1}{X_2}\right) &\approx \frac{\text{Var}(X_1)}{\{\text{E}(X_2)\}^2} + \frac{\{\text{E}(X_1)\}^2\text{Var}(X_2)}{\{\text{E}(X_2)\}^4} - \frac{2\text{E}(X_1)\text{Cov}(X_1, X_2)}{\{\text{E}(X_2)\}^3} \\ &= \frac{\{\text{E}(X_1)\}^2}{\{\text{E}(X_2)\}^2}\left[\frac{\text{Var}(X_1)}{\{\text{E}(X_1)\}^2} + \frac{\text{Var}(X_2)}{\{\text{E}(X_2)\}^2} - \frac{2\text{Cov}(X_1, X_2)}{\text{E}(X_1)\text{E}(X_2)}\right]\end{aligned} \tag{8.3.6}$$

Then, substituting $X_1 = Q_1 = \mathbf{y}^T\mathbf{B}\mathbf{y}$ and $X_2 = Q_2 = \mathbf{y}^T\mathbf{A}\mathbf{y}$, we obtain

$$\text{E}\left(\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}}\right) \approx \frac{\text{tr}(\mathbf{B\Sigma})}{\text{tr}(\mathbf{A\Sigma})} \tag{8.3.7}$$

and,

$$\text{Var}\left(\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}}\right) \approx \frac{\{\text{tr}(\mathbf{B\Sigma})\}^2}{\{\text{tr}(\mathbf{A\Sigma})\}^2}\left[2\frac{\text{tr}\{(\mathbf{B\Sigma})^2\}}{\{\text{tr}(\mathbf{B\Sigma})\}^2} + 2\frac{\{\text{tr}(\mathbf{A\Sigma})^2\}}{\{\text{tr}(\mathbf{A\Sigma})\}^2} - 4\frac{\text{tr}(\mathbf{A\Sigma B\Sigma})}{\text{tr}(\mathbf{A\Sigma})\text{tr}(\mathbf{B\Sigma})}\right] \tag{8.3.8}$$

Equating these moments with those of the scaled F-distribution, $\lambda \mathrm{F}(v_1, v_2)$, we obtain

$$\frac{1}{\lambda}\frac{\mathrm{tr}(\mathbf{B\Sigma})}{\mathrm{tr}(\mathbf{A\Sigma})} = \frac{v_2}{v_2 - 2} \tag{8.3.9}$$

and,

$$\frac{1}{\lambda^2}\frac{\{\mathrm{tr}(\mathbf{B\Sigma})\}^2}{\{\mathrm{tr}(\mathbf{A\Sigma})\}^2}\left[2\frac{\mathrm{tr}\{(\mathbf{B\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{B\Sigma})\}^2}+2\frac{\mathrm{tr}\{(\mathbf{A\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{A\Sigma})\}^2}-4\frac{\mathrm{tr}(\mathbf{A\Sigma B\Sigma})}{\mathrm{tr}(\mathbf{A\Sigma})\mathrm{tr}(\mathbf{B\Sigma})}\right] = \frac{2v_2^2(v_2+v_1-2)}{v_1(v_2-2)^2(v_2-4)} \tag{8.3.10}$$

Fixing $v1 = c$, the dimensionality of the test (similarly to the Kenward Roger and Sandwich adjustments), we can use these final two equations to find the scaling factor λ and the denominator degrees of freedom v_2 for our approximating distribution.

From (8.3.9), we obtain

$$\lambda = \frac{v_2 - 2}{v_2}\frac{\mathrm{tr}(\mathbf{B\Sigma})}{\mathrm{tr}(\mathbf{A\Sigma})} \tag{8.3.11}$$

so that, substituting for λ in (8.3.10), we find

$$\frac{v_2 + c - 2}{c(v_2 - 4)} = \left[\frac{\mathrm{tr}\{(\mathbf{B\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{B\Sigma})\}^2} + \frac{\mathrm{tr}\{(\mathbf{A\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{A\Sigma})\}^2} - 2\frac{\mathrm{tr}(\mathbf{A\Sigma B\Sigma})}{\mathrm{tr}(\mathbf{A\Sigma})\mathrm{tr}(\mathbf{B\Sigma})}\right] \tag{8.3.12}$$

so that, writing

$$\mathrm{V} = \left[\frac{\mathrm{tr}\{(\mathbf{B\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{B\Sigma})\}^2} + \frac{\mathrm{tr}\{(\mathbf{A\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{A\Sigma})\}^2} - 2\frac{\mathrm{tr}(\mathbf{A\Sigma B\Sigma})}{\mathrm{tr}(\mathbf{A\Sigma})\mathrm{tr}(\mathbf{B\Sigma})}\right] \tag{8.3.13}$$

it follows that

$$v_2 = \frac{c(4\mathrm{V}+1)-2}{c\mathrm{V}-1} \tag{8.3.14}$$

Assuming that the numerator and denominator terms, $\mathbf{y}^T\mathbf{B}\mathbf{y}$ and $\mathbf{y}^T\mathbf{A}\mathbf{y}$, are independent, so that the covariance term $\mathrm{Cov}(\mathbf{y}^T\mathbf{A}\mathbf{y}, \mathbf{y}^T\mathbf{B}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) = 0$, we obtain the following improved statistic based on Box's modified F-statistic.

$$\mathrm{F} = \frac{(n-r)}{c}\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} \lambda F(c, v2) \tag{8.3.15}$$

where,

$$\lambda = \frac{(n-r)}{c}\frac{v_2-2}{v_2}\frac{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})}{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})} \tag{8.3.16}$$

$$v_2 = \frac{c(4\mathrm{V}'+1)-2}{c\mathrm{V}'-1} \tag{8.3.17}$$

and,

$$\mathrm{V}' = \left[\frac{\mathrm{tr}\{(\mathbf{B}\boldsymbol{\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})\}^2} + \frac{\mathrm{tr}\{(\mathbf{A}\boldsymbol{\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\}^2}\right] \tag{8.3.18}$$

Results from 1000 simulations of designs (A), (B) and (C) are shown in Table 8.3.1, and show that this modified Box correction gives test sizes much closer to nominal levels than Box's original statistic.

A further modification was considered using second order deviations about the mean in (8.3.7). That is, taking

$$\mathrm{E}\left(\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}}\right) \approx \frac{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})}{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})}\left[1 + 2\frac{\mathrm{tr}\{(\mathbf{A}\boldsymbol{\Sigma})^2\}}{\{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\}^2} - 2\frac{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma})}{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})}\right] \tag{8.3.19}$$

and proceeding as above (with $\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) = 0$). However, simulations show that there is little advantage in following this more complicated approach, since this results in an adjusted statistic which is more conservative, with test sizes closer to those given by the Box correction which we are attempting to inflate.

Underlying 'True' Covariance Structure	Method of Inference		Proportion of Significant Test Results (out of 1000) ('Size')		
			Design A	Design B	Design C
Identity	Box F		0.026	0.024	0.038
	Mod Box		0.048	0.042	0.056
Compound Symmetry	Box F		0.023	0.017	0.032
	Mod Box		0.042	0.040	0.046
AR1 ($\rho = 0.2$)	Box F		0.039	0.030	0.028
	Mod Box		0.061	0.053	0.048
AR1 ($\rho = 0.8$)	Box F		0.035	0.024	0.020
	Mod Box		0.068	0.061	0.043
Antedependence	Box F		0.046	0.050	0.021
	Mod Box		0.078	0.096	0.042
Unstructured	Box F		0.027	0.025	0.021
	Mod Box		0.065	0.059	0.047
Unstr (QRE)	Box F		0.061	0.040	0.004
	Mod Box		0.120	0.118	0.038

Table 8.3.1: *Results from 1000 simulations of Designs (A), (B) and (C).*

8.4 Properties of Box-type Modified F-statistics

It is of interest to compare the properties of the improved Box-type statistic derived above with that of Box's original.

Note firstly that, for matrices $\mathbf{A}$ and $\mathbf{B}$ defined in the ANOVA statistic, we have

$$\begin{aligned}\mathrm{tr}(\mathbf{A}) = \mathrm{tr}\{\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\} &= \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}\{\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\} \\ &= \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}(\mathbf{I}_r) = n - r\end{aligned}$$

and,

$$\begin{aligned}\mathrm{tr}(\mathbf{B}) = \mathrm{tr}\{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \mathbf{X}_R(\mathbf{X}_R^T\mathbf{X}_R)^{-1}\mathbf{X}_R^T\} &= \mathrm{tr}(\mathbf{I}_r) - \mathrm{tr}(\mathbf{I}_{r-c}) \\ &= r - (r - c) = c\end{aligned}$$

Further, $\mathrm{tr}(\mathbf{A}^2) = \mathrm{tr}(\mathbf{A})$ and $\mathrm{tr}(\mathbf{B}^2) = \mathrm{tr}(\mathbf{B})$, since these 'projection' matrices are (symmetric and) idempotent.

Now, under the ANOVA assumptions of independence and homogeneity of variance, the covariance estimator has the block diagonal form $\mathbf{\Sigma} = \mathbf{I}_m \otimes \sigma^2 \mathbf{I}_p = \sigma^2 \mathbf{I}_n$, so that the Box Modified F-statistic parameters, (8.2.8)-(8.2.10), become

$$\psi = \frac{(n-r)}{c} \frac{\mathrm{tr}(\mathbf{B})}{\mathrm{tr}(\mathbf{A})} = 1 \tag{8.4.1}$$

$$v1 = \frac{\{\mathrm{tr}(\mathbf{B})\}^2}{\mathrm{tr}(\mathbf{B}^2)} = \mathrm{tr}(\mathbf{B}) = c \tag{8.4.2}$$

$$v2 = \frac{\{\mathrm{tr}(\mathbf{A})\}^2}{\mathrm{tr}(\mathbf{A}^2)} = \mathrm{tr}(\mathbf{A}) = n - r \tag{8.4.3}$$

That is, under the usual ANOVA assumptions, Box's Modified F-statistic recaptures the correct degrees of freedom for the exact one-way ANOVA test statistic given in (8.2.1).

$$\mathrm{F} = \frac{(n-r)}{c} \frac{\mathbf{y}^T \mathbf{B} \mathbf{y}}{\mathbf{y}^T \mathbf{A} \mathbf{y}} \sim F(c, n-r)$$

For the modified Box correction, we have, the same assumptions, from (8.3.18)

$$\begin{aligned} \mathrm{V}' = \left[\frac{\mathrm{tr}(\mathbf{B}^2)}{\{\mathrm{tr}(\mathbf{B})\}^2} + \frac{\mathrm{tr}(\mathbf{A}^2)}{\{\mathrm{tr}(\mathbf{A})\}^2} \right] &= \frac{1}{\mathrm{tr}(\mathbf{A})} + \frac{1}{\mathrm{tr}(\mathbf{B})} \\ &= \frac{1}{n-r} + \frac{1}{c} \end{aligned} \tag{8.4.4}$$

so that,

$$\begin{aligned} c(4\mathrm{V}' + 1) - 2 &= \frac{4c^2 + 4c(n-r) + c^2(n-r) - 2c(n-r)}{c(n-r)} \\ &= \frac{4c + c(n-r) + 2(n-r)}{(n-r)} \end{aligned}$$

and

$$
\begin{aligned}
c\mathrm{V}' - 1 &= \frac{c^2 + c(n-r) - c(n-r)}{c(n-r)} \\
&= \frac{c}{(n-r)}
\end{aligned}
$$

and hence the further Modified F-statistic parameters, (8.3.16)-(8.3.17), become

$$
v_2 = \frac{c(4\mathrm{V}' + 1) - 2}{c\mathrm{V}' - 1} = \frac{4c + c(n-r) + 2(n-r)}{c} \tag{8.4.5}
$$

and so,

$$
\begin{aligned}
\lambda &= \frac{(n-r)}{c}\frac{v_2 - 2}{v_2}\frac{\mathrm{tr}(\mathbf{B})}{\mathrm{tr}(\mathbf{A})} = \frac{v_2 - 2}{v_2} \\
&= \frac{2c + c(n-r) + 2(n-r)}{4c + c(n-r) + 2(n-r)}
\end{aligned} \tag{8.4.6}
$$

Hence, if the usual ANOVA assumptions are satisfied, we have,

$$
\mathrm{F} = \frac{(n-r)}{c}\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}} \overset{\text{approx}}{\sim} \lambda \mathrm{F}(c, v2) \tag{8.4.7}
$$

where,

$$
\lambda = \frac{2c + c(n-r) + 2(n-r)}{4c + c(n-r) + 2(n-r)} \tag{8.4.8}
$$

and,

$$
v_2 = \frac{4c + c(n-r) + 2(n-r)}{c} \tag{8.4.9}
$$

That is, the modified Box correction does not recover the exact one-way ANOVA statistic, where assumptions would allow its use. However, this discontinuity is of small order.

To illustrate this difference, consider the following study based on simulation study design (A). We have $m = 10$ subjects, split equally into two treatment groups, measured at each of $p = 5$ time points, so that the number of observations is $n = 50$. If we are concerned with whether there is a significant treatment-time interaction, then we have $r = 10$ parameters (1 intercept, 4 time, 1 group, 4 interaction), so that $n - r = 40$ and $c = 4$.

We have from Box, the standard ANOVA test statistic

$$\mathrm{F} \sim \mathrm{F}(4, 40)$$

but the further Modified F-statistic gives us

$$\mathrm{F} \overset{\text{approx}}{\sim} 0.96875 \times \mathrm{F}(4, 64)$$

That is, the modified correction inflates the denominator degrees of freedom to account for the 'scaling' of the statistic.

The differences in the densities of these distributions is shown in the Figure 8.4.1. It is seen that is very little difference between them and, in particular, the difference is not large in the 'tail end', where hypotheses are decided. The 95% quantiles are given by 2.606 and 2.596 for the F(4,40) and 0.96875 times the F(4,64) distributions, respectively.

A possible solution to this problem is to consider the addition of further small order terms to the approximated moments of the quadratic ratio F in (8.3.7) and (8.3.8), taking

$$\mathrm{E}\left(\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}}\right) \approx \frac{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma})}{\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})}\Big(1 + \alpha\Big) \tag{8.4.10}$$

and,

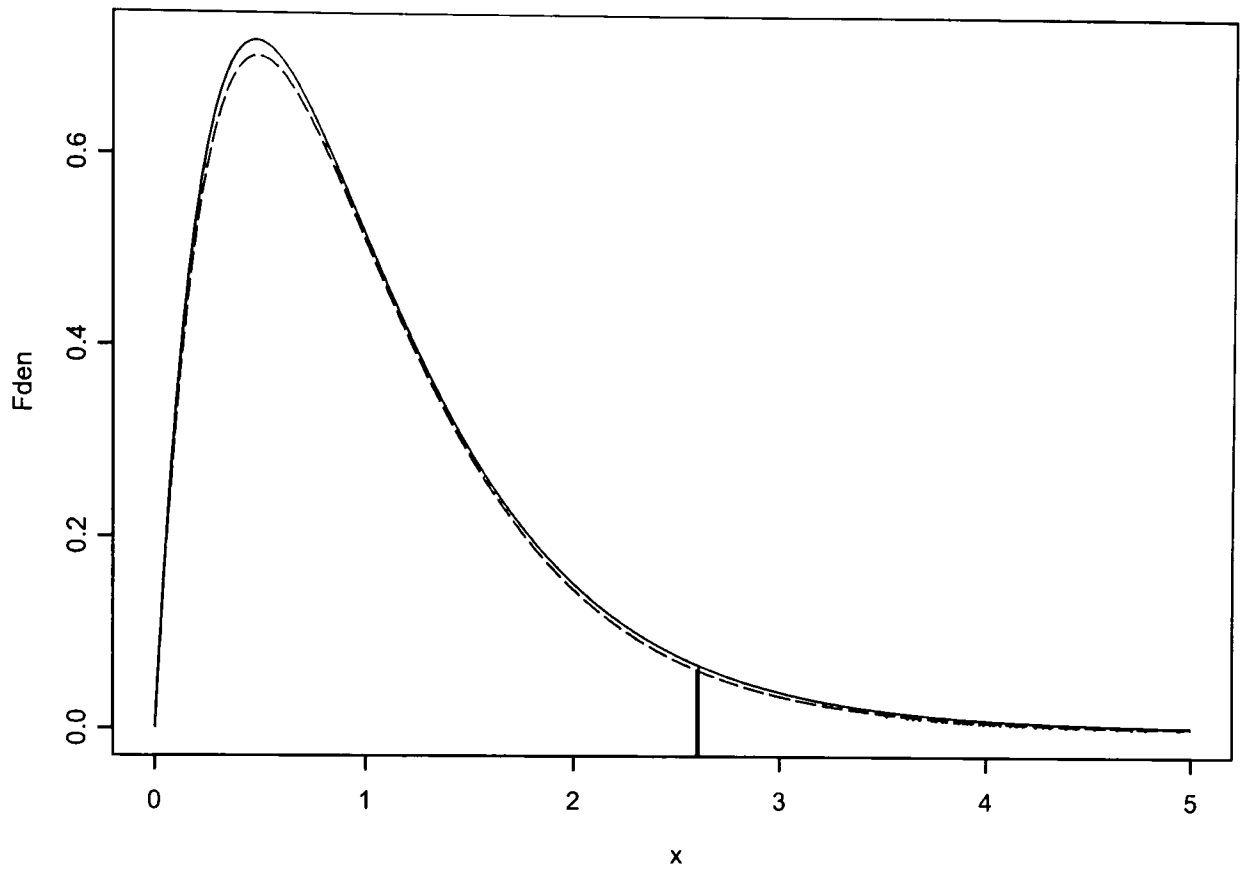


Figure 8.4.1: *F-distributions. Legend: Solid Line, the F(4,40) distribution; Dashed Line, 0.96875 times the F(4,64) distribution*

$$\text{Var}\left(\frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{A}\mathbf{y}}\right) \approx \frac{\{\text{tr}(\mathbf{B}\boldsymbol{\Sigma})\}^2}{\{\text{tr}(\mathbf{A}\boldsymbol{\Sigma})\}^2}\left[2\frac{\text{tr}\{(\mathbf{B}\boldsymbol{\Sigma})^2\}}{\{\text{tr}(\mathbf{B}\boldsymbol{\Sigma})\}^2} + 2\frac{\{\text{tr}(\mathbf{A}\boldsymbol{\Sigma})^2\}}{\{\text{tr}(\mathbf{A}\boldsymbol{\Sigma})\}^2} - 4\frac{\text{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma})}{\text{tr}(\mathbf{A}\boldsymbol{\Sigma})\text{tr}(\mathbf{B}\boldsymbol{\Sigma})} + \beta\right] \tag{8.4.11}$$

where α and β represent higher order terms in the Taylor series expansion, which are of small magnitude, but ensure the correct parameter estimates (i.e. $\lambda = 1$ and $v2 = n - r$) are obtained when the ANOVA assumptions are satisfied. This is similar to the approach taken by Kenward and Roger (1997) to ensure their small sample adjusted Wald test recovers exact tests.

In this context, the further expansion leads to complicated terms for the parameter estimates in the modified Box correction which become unstable as the estimated covariance matrix moves further away from the independence model (identity structure), and it is not pursued. Although it is desirable to have a statistic which

recovers the exact test in appropriate circumstances, it is worth noting that such circumstances are unlikely to arise in practice.

In comparing the properties of these Box-type statistics, it is useful to consider also the relationships between the parameters given by the two corrections.

Figures 8.4.1-8.4.3 show the distributions of the scaling parameter ψ given by the Box correction and its relationship with the degrees of freedom parameters, $v1$ and $v2$, from 1000 simulations of study designs (A)-(C) for data arising from the various underlying covariance structures.

The distributions of the scaling parameter λ and its relationship with $v2$, under the same scenarios, are shown similarly in Figures 8.4.4-8.4.6.

The key points to note from these plots are as follows:

- For both corrections, values greater than 1 are possible for the scaling parameters, ψ and λ.
- For data arising from an identity structure (independence), the central value of the distributions of the scaling parameters is clearly at 1.
- In general, for data arising from structures further from independence (AR1, high correlation, antedependence, etc.), the central value of the scaling parameter is below 1, and the spread of possible values is greater. The exception is for data arising from the 'unstructured' covariance matrix based on the 'badly behaved' quadratic random effects model. For these data, the estimated scaling parameters are tightly clustered around a central value at or below 0.5

The observations above are seen across the three study designs (A)-(C). However, there are some interesting differences with regards to the degrees of freedom parameters, $v1$ and $v2$. Firstly, considering the Box correction:

- In general, for increasing ψ, there is a corresponding increase in the denomina-

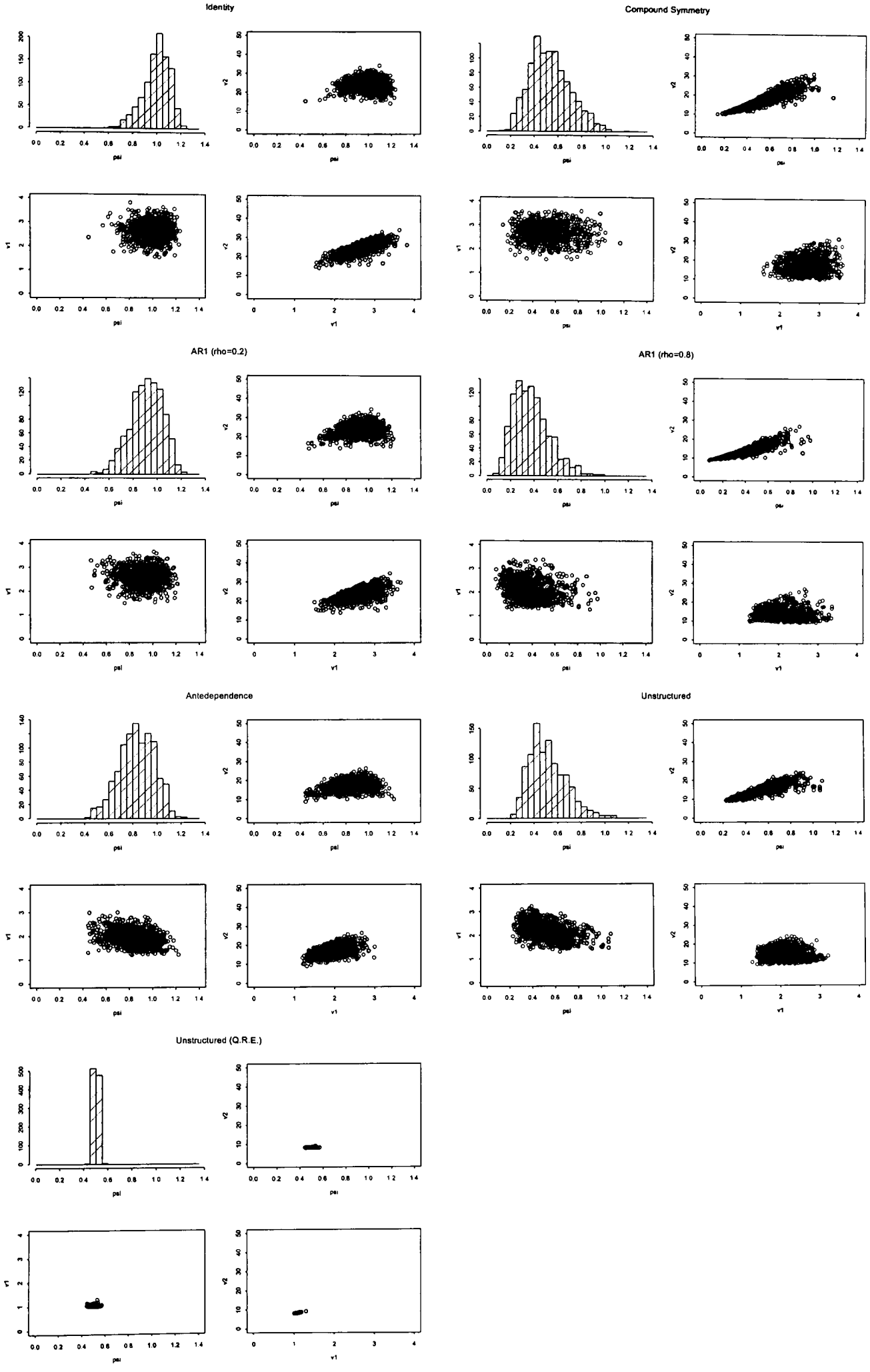


Figure 8.4.2: *Distributions of Scaling Parameter ψ from Box's correction and relationship with d.f. parameters v1 and v2 obtained from 1000 simulations of Design (A).*

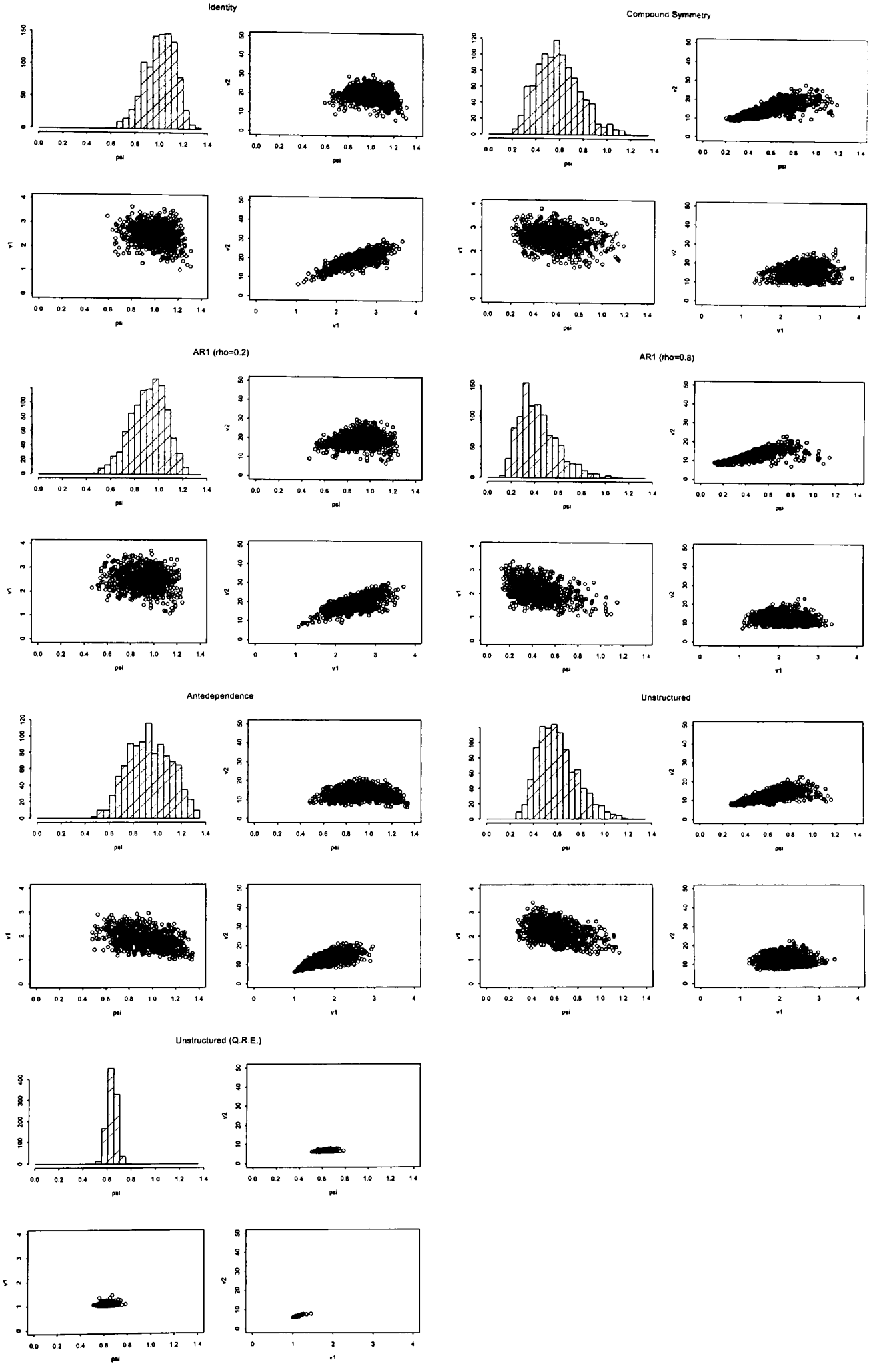


Figure 8.4.3: *Distributions of Scaling Parameter ψ from Box's correction and relationship with d.f. parameters $v1$ and $v2$ obtained from 1000 simulations of Design (B).*

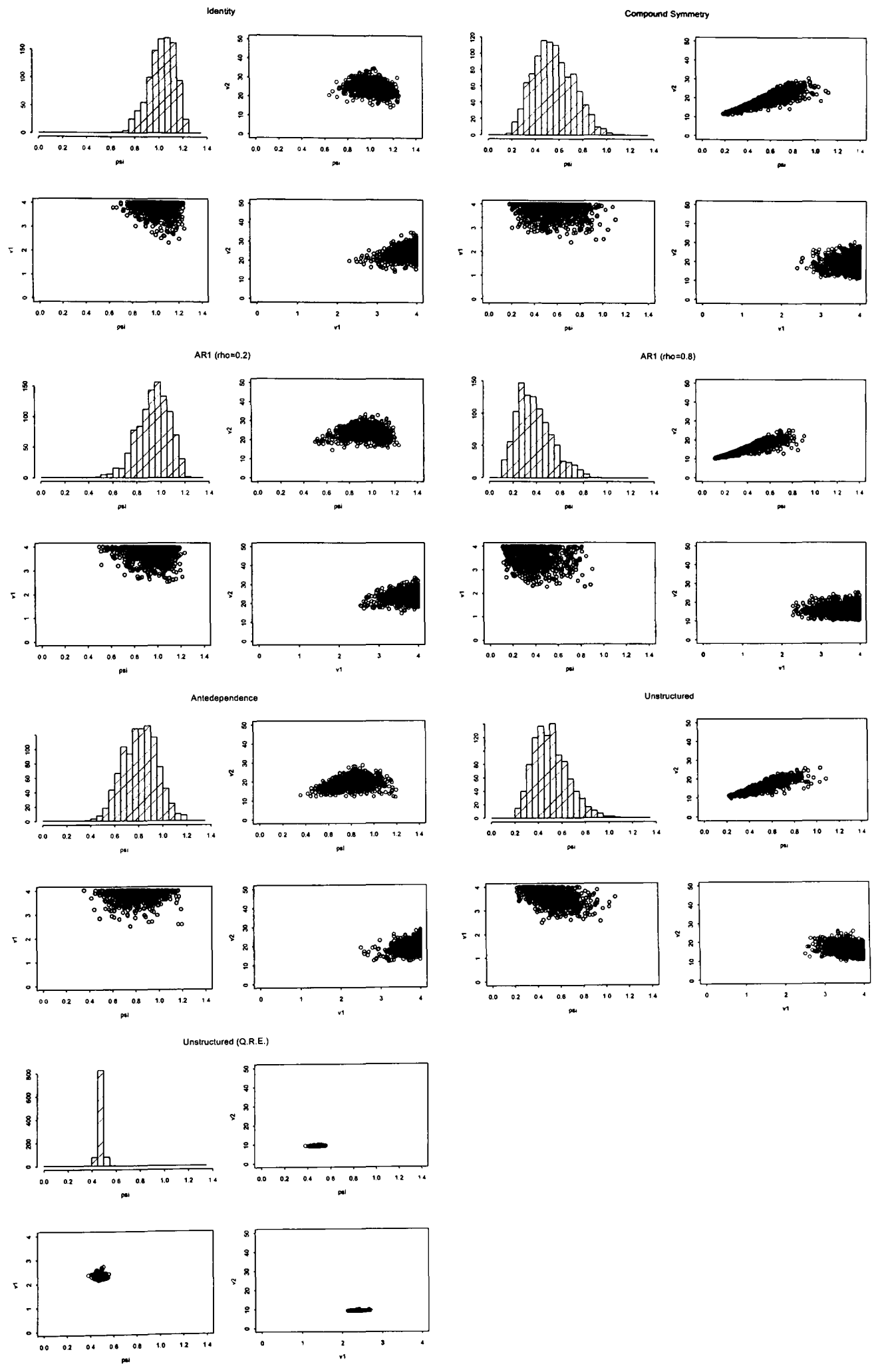


Figure 8.4.4: *Distributions of Scaling Parameter ψ from Box's correction and relationship with d.f. parameters $v1$ and $v2$ obtained from 1000 simulations of Design (C).*

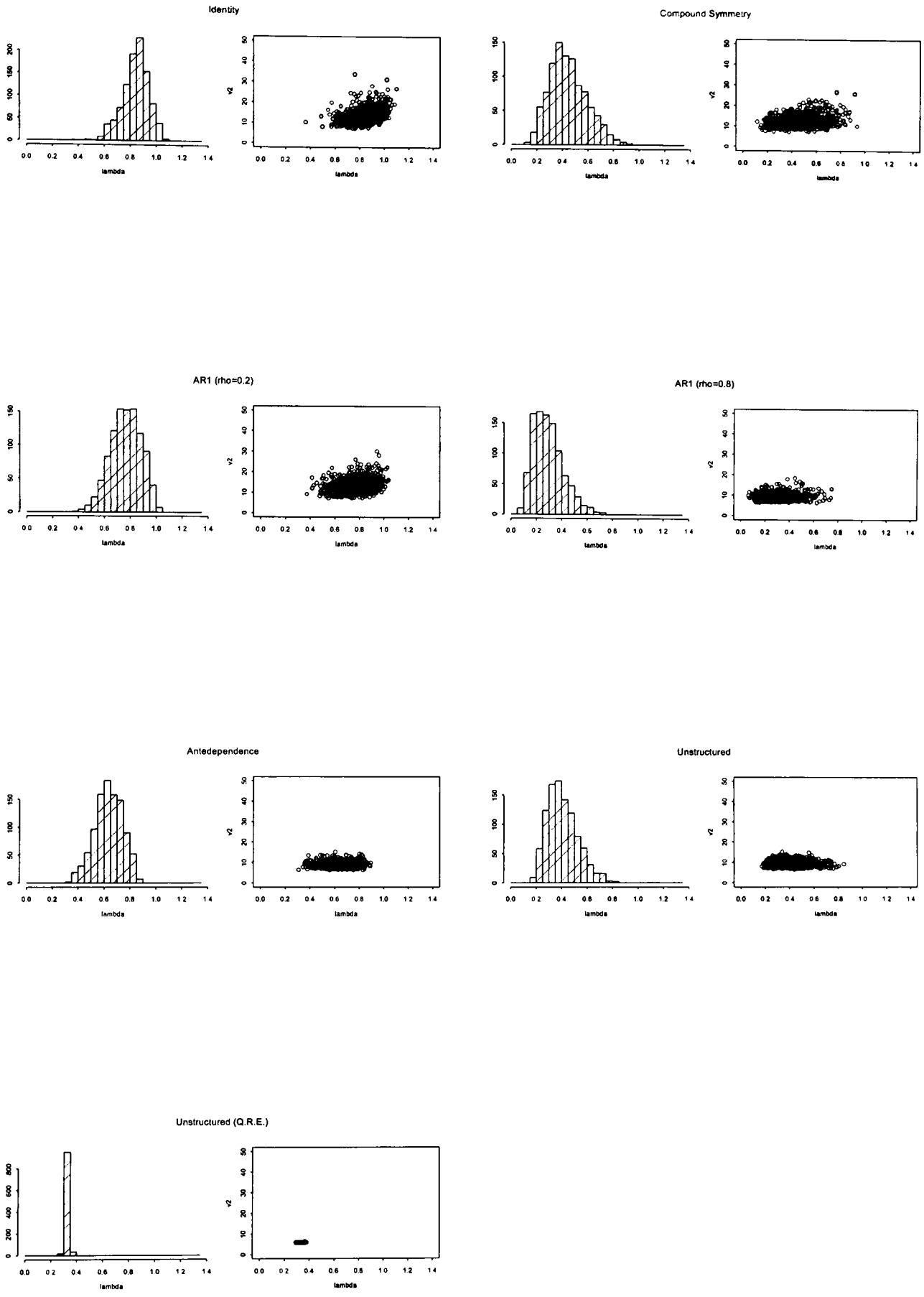


Figure 8.4.5: *Distributions of Scaling Parameter λ from the Modified Box correction and relationship with d.f. parameter v2 obtained from 1000 simulations of Design (A).*

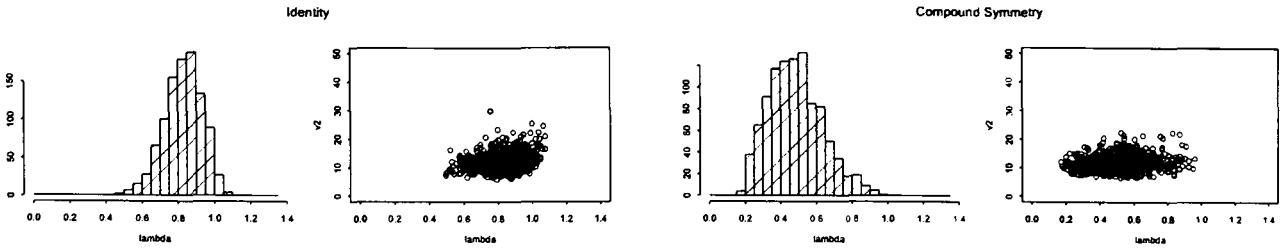


Figure 8.4.6: *Distributions of Scaling Parameter λ from the Modified Box correction and relationship with d.f. parameter v2 obtained from 1000 simulations of Design (B).*

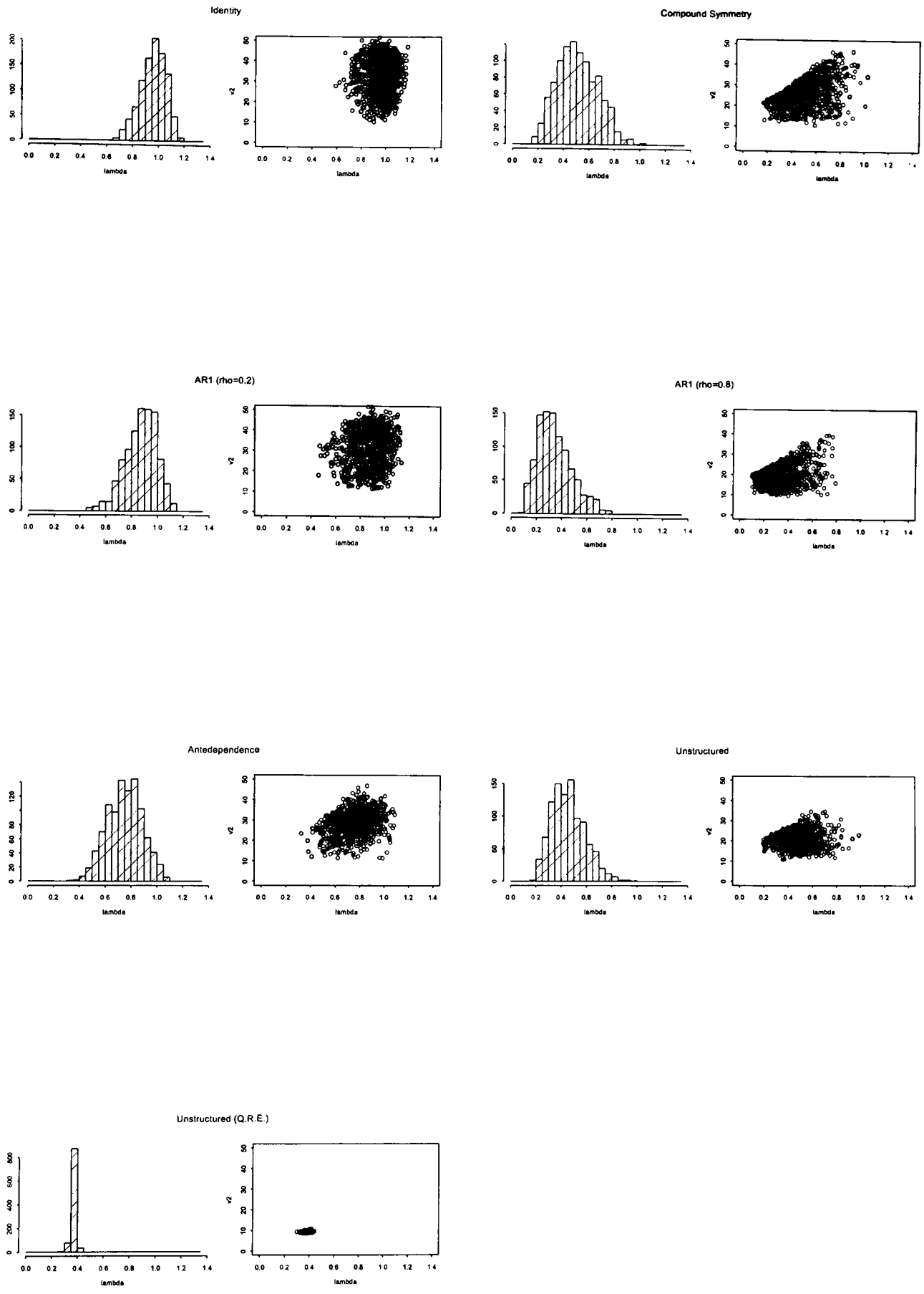


Figure 8.4.7: *Distributions of Scaling Parameter* λ *from the Modified Box correction and relationship with d.f. parameter* $v2$ *obtained from 1000 simulations of Design (C).*

tor degrees of freedom parameter $v2$, with less variation in $v2$ for data which are further from independence.

- There is a weaker relationship between ψ and the numerator degrees of freedom parameter $v1$.
- There is a moderate positive correlation between $v1$ and $v2$.
- For the crossover study of design (C), $v1$ is seen to push against its upper limit of c (=4), its value under independence. No values of $v2$ exceed $n - r$ (=40).
- There is little variation in the estimates of $v1$ and $v2$ for the data arising from the Q.R.E. structure, as we might expect from the estimates of ψ.

Secondly, for the modified Box correction, there is no clear relationship between λ and $v2$ in any of the settings, although it is interesting to note that the independence measure $n - r$ for $v2$ is exceeded in design (C).

8.5 Box's Correction and 'Compound Symmetry'

Throughout this Chapter we have considered Box's correction (and its proposed modification), which adjust for departures from the one-way ANOVA assumptions of independence between observations and homogeneity of variance. It is useful to consider, however, how such a correction behaves under the assumption of 'compound symmetry', and the relationship between this approach and repeated measures ANOVA.

Repeated measures ANOVA was often adopted for 'practical' analyses before modern computing power allowed widespread access to the general linear mixed effects model, which is more commonly used for such data. (See Fitzmaurice *et al.* (2004) for an outline of the approach, and Crowder and Hand (1990) for more detail). The general formulation, as with the one-way approach, is to treat 'time' (occasions of measurement) as an additional within-subjects factor, and to model the jth measurement on the ith subject as

$$y_{ij} = \mu_{ij} + b_i + e_{ij} \tag{8.5.1}$$

where μ_{ij} are suitably specified fixed effects, $b_i \sim \mathrm{N}(0, \sigma_b^2)$ are random (individual specific) subject effects, and $e_{ij} \sim \mathrm{N}(0, \sigma^2)$ are the usual error terms.

The two sources of variation, between-subjects and within-subjects, lead to a compound symmetry covariance structure for the repeated measurements, *viz.*

$$\mathrm{Var}(\mathbf{y}_i) = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \dots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

Figure 8.5.1: *Compound Symmetry Covariance Structure.*

This model implies that we have a constant variance, $\sigma_b^2 + \sigma^2$, at every measurement time, and also that the correlation between any pair of repeated measurements on the same subject, given by (the intraclass correlation)

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \tag{8.5.2}$$

is constant.

Such an approach is appropriate in randomised block designs, where the subjects are considered as blocks, and in extended examples involving error strata. (See Nelder (1965a,b)). However, this approach is often not appropriate for many repeated measures analyses, since measurement times are defined and cannot be allocated to the subjects at random. Also, a constant correlation between observations on the same subject is not typically seen, since correlations tend to decay in time.

Box (1954b) suggested that departures from compound symmetry could be accounted for by reducing the degree of freedom parameters for the two-way ANOVA

F ratio by a multipicative factor ϵ.

The parameter ϵ, $1/(p-1) \leq \epsilon \leq 1$, is estimated using the sample covariance matrix. This approach was extended to the split-plot design by Greenhouse and Geisser (1958) and to the (multivariate) setting of profile analysis by Greenhouse and Geisser (1959), through an adjustment to the MANOVA test statistic. Huynh and Feldt (1976) show that in general for such a model to be appropriate it is necessary only for the within-subjects covariance matrix to comply with the assumption of sphericity. This is a less restrictive condition than compound symmetry, which requires only that the variance of differences in a within-subjects design are equal across all groups.

The correction uses

$$\frac{\sigma^2}{\sigma_b^2 + \sigma^2} F \overset{\text{approx}}{\sim} F(\epsilon c, \epsilon(m-g)(p-1)) \tag{8.5.3}$$

where c is the number of terms being tested, m is the number of subjects, g is the number of treatment groups, and p is the number of repeated measurements on a subject.

The general approach to adjusting for sphericity, in tests which involve the factor 'time', is to adjust only if a significant result is found, since the corrections necessarily increases p values. If F is significant, it can be tested again using the lower bound $\epsilon = 1/(p-1)$ and only if this leads to a non-significant result will ϵ need to be estimated. Adjusted tests using ϵ defined by both Greenhouse and Geisser and Huynh and Feldt are widely implemented in software packages.

Using the Box correction of Section 8.2, under the assumption of compound symmetry, we find

$$\psi = \frac{\sigma^2}{\sigma_b^2 + \sigma^2} = 1 - \rho \tag{8.5.4}$$

That is, the Box correction adjusts the one-way ANOVA F statistic to that which would be obtained from the (more restrictive) two-way setting. This is equivalent to an (unadjusted) Wald statistic for the regression parameters in the mixed model, based on a compound symmetry covariance structure. However, although the numerator degrees of freedom are fixed as c, the denominator degrees of freedom are lower than from their two-way counterpart, since we are accounting for departures from independence. A similar relationship is found with the modified Box correction under compound symmetry.

In the context of very small samples of repeated measurements, the one-way ANOVA approach with a suitable correction is preferred, since it is more widely applicable across a range of settings. The Greenhouse and Geisser approach, based on the split-plot design, is simply too restrictive to be of use generally, since it requires complete and balanced data and would not, for example, accommodate missing data or crossover designs.

Chapter 9

A Comparative Study

9.1 Introduction

Chapters 7 and 8 considered two approaches for inference in repeated measures problems in the small sample setting, both of which are less dependent on the covariance structure.

In Chapter 7, an adjusted test based on the sandwich estimator incorporating both the bias correction of Mancl and DeRouen and the adjustment for variation proposed by Pan and Wall was seen to control the type 1 error rate, resulting in a Wald statistic with nominal properties.

In Chapter 8, Box's correction based on a modified ANOVA F-statistic was shown to also control the type 1 error rate, but was conservative, giving excessive control. A modification to this procedure was proposed which resulted in a test size closer to nominal levels.

In this chapter these approaches are compared in greater detail, using simulated data over a range of settings. In particular, having adequately controlled the type 1 error rate ('size') of tests resulting from these procedures, it is appropriate to consider their power to determine significant treatment differences or effects, where such exist. It is also informative to compare these approaches with existing methods. For small

sample repeated measures problems where a suitable low dimensional covariance structure cannot be determined, this will be a Wald test using an unstructured covariance estimate with the Kenward Roger (KR) adjustment.

We begin, once again, by revisiting the simulation study designs of Chapter 2.

9.2 An initial Simulation Study

Recall from Chapter 2, Section 2.2.2, study designs (A) and (B), each based on a simple repeated measures design with 5 time points and 10 subjects in two treatment groups. For each design, 1000 data sets are simulated from each of 7 underlying 'known' covariance structures, ranging from stationary types such as identity and compound symmetry, to highly non-stationary types such as an antedependence form (of order 1) or a 'badly behaved' structure based on a quadratic random effects model.

In design (A) the data are complete and balanced, so that a Wald test with the KR adjustment based on an unstructured covariance estimator is an exact Hotelling T^2 test. In this setting mean estimates for each treatment by time combination are equivalent to their ordinary least squares estimates, so that tests for no treatment/time interaction are directly comparable with both the adjusted sandwich and Box correction procedures. The simulation studies of Chapter 2 indicated that this is the only setting where existing methods are reliable for very small samples.

In design (B) missing values are introduced by allowing one subject in each of the two treatment groups to drop out at some random time following the first observation. In this instance mean estimates using the existing REML methods are dependent on the estimated covariance structure, and test sizes using the KR adjustment have been shown to be slightly inflated from the nominal 5% level.

The results of the simulations for study designs (A) and (B), comparing Wald tests using the unstructured covariance estimator with a KR adjustment, the adjusted

sandwich estimator, and both the Box correction and its proposed modification are shown in Tables 9.2.1 and 9.2.2.

In these tables, power levels are compared for a linear (in time) treatment difference which is fixed to achieve a power to detect a significant treatment/time interaction of around 75% using the KR adjustment in design (A). Hence, power is directly comparable between methods for each underlying covariance structure across both these tables.

Looking at the tables, we see that for design (A), each of the methods of inference considered achieves nominal test sizes close to 5%. Recall, under the null hypothesis, a 95% probability interval for the proportion of tests out of 1000 leading to a rejection of the test is (0.036, 0.064). Whilst the Box correction is conservative, the modified Box statistic achieves test sizes comfortably within these limits across the range of data arising from the different underlying covariance structures. However, the sizes using both the adjusted sandwich estimator and modified Box correction are somewhat inflated for data arising from the unstructured covariance matrix based on the quadratic random effects (Q.R.E.) model.

Considering power, we see that the power levels obtained for the two Box correction methods are consistently higher than the KR adjusted Wald test, whilst the adjusted sandwich estimator is consistently lower. The single exception to this is for the data arising from the 'badly behaved' Q.R.E. model. It is apparent however that, leaving aside this structure, the increase in power achieved by the Box corrections is less marked for data arising from the non-stationary covariance structures which are far from the independence (identity) model.

Similar results are seen for design (B), where missing data are introduced. The Box corrections appear to be robust to the missing observations, while the power is seen to fall for the Wald tests using the KR adjustment or the adjusted sandwich estimator.

Underlying Covariance Structure	Method of Inference	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment/Time Interaction)	
		Type I error rate ('Size')	Power
Stationary Structures			
Identity	Unstr (KR)	0.053	0.735
	Sand Adj	0.055	0.629
	Box F	0.026	0.947
	Mod Box	0.048	0.976
Compound Symmetry	Unstr (KR)	0.045	0.747
	Sand Adj	0.057	0.660
	Box F	0.023	0.964
	Mod Box	0.042	0.984
AR1 ($\rho = 0.2$)	Unstr (KR)	0.050	0.735
	Sand Adj	0.068	0.659
	Box F	0.039	0.980
	Mod Box	0.061	0.986
AR1 ($\rho = 0.8$)	Unstr (KR)	0.056	0.775
	Sand Adj	0.083	0.793
	Box F	0.035	0.992
	Mod Box	0.068	0.999
Non-Stationary Structures			
Ante-dependence	Unstr (KR)	0.058	0.760
	Sand Adj	0.047	0.593
	Box F	0.046	0.762
	Mod Box	0.078	0.863
Unstr.	Unstr (KR)	0.045	0.768
	Sand Adj	0.051	0.607
	Box F	0.027	0.809
	Mod Box	0.065	0.918
Unstr. (Q.R.E.)	Unstr (KR)	0.048	0.740
	Sand Adj	0.100	0.824
	Box F	0.061	0.200
	Mod Box	0.120	0.367

Table 9.2.1: *Summary of results from 1000 simulations of Design (A). Table gives the proportion of type 1 errors (Size) and Power.*

Underlying Covariance Structure	Method of Inference	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment/Time Interaction)	
		Type I error rate ('Size')	Power
Stationary Structures			
Identity	Unstr (KR)	0.073	0.633
	Sand Adj	0.056	0.569
	Box F	0.024	0.892
	Mod Box	0.042	0.949
Compound Symmetry	Unstr (KR)	0.074	0.632
	Sand Adj	0.051	0.523
	Box F	0.017	0.882
	Mod Box	0.040	0.937
AR1 ($\rho = 0.2$)	Unstr (KR)	0.089	0.637
	Sand Adj	0.068	0.592
	Box F	0.030	0.963
	Mod Box	0.053	0.984
AR1 ($\rho = 0.8$)	Unstr (KR)	0.078	0.637
	Sand Adj	0.064	0.677
	Box F	0.024	0.966
	Mod Box	0.061	0.989
Non-Stationary Structures			
Ante-dependence	Unstr (KR)	0.084	0.645
	Sand Adj	0.055	0.514
	Box F	0.050	0.619
	Mod Box	0.096	0.775
Unstr.	Unstr (KR)	0.077	0.640
	Sand Adj	0.042	0.392
	Box F	0.025	0.626
	Mod Box	0.059	0.797
Unstr. (Q.R.E.)	Unstr (KR)	0.046	0.516
	Sand Adj	0.063	0.546
	Box F	0.040	0.158
	Mod Box	0.118	0.330

Table 9.2.2: *Summary of results from 1000 simulations of Design (B). Table gives the proportion of type 1 errors (Size) and Power.*

Design (C) is a five period-five treatment crossover design with 10 subjects allocated to treatments according to a pair of Williams' designs, (see Table 2.2.3). Here KR adjusted tests for no treatment effect were shown to have type 1 error rates hugely inflated from nominal test sizes, but a Wald test using the adjusted sandwich estimator and the Box correction procedures have nominal properties. The results of these simulations are repeated in Table 9.2.3.

Since tests using the adjusted sandwich estimator and Box corrections adequately control the size, it is appropriate to consider their power, but it is not immediately apparent at what level they should be compared. However, since by the nature of these simulations the 'true' underlying covariance structure is known, it is possible to consider the true power of the Wald test to determine a treatment difference using these known structures by reference to a non-central chi-squared distribution. That is, for a Wald test of the null hypothesis $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, using the known covariance structure $\boldsymbol{\Sigma}$, the power is given by

$$1 - \mathrm{P}(W \leq \chi^2_{0.95}(l;\delta)) \tag{9.2.1}$$

where χ^2 has a (non-central) chi-squared distribution with l degrees of freedom, the dimensionality of the test, and non-centrality parameter δ, given by

$$\delta = (\mathbf{L}\hat{\boldsymbol{\beta}})^T(\mathbf{L}\boldsymbol{\Phi}\mathbf{L}^T)^{-1}\mathbf{L}\hat{\boldsymbol{\beta}} \tag{9.2.2}$$

Recall that $\hat{\boldsymbol{\beta}}$ is the vector of estimated regression parameters, with estimated standard errors $\boldsymbol{\Phi} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$.

Figure 9.2.1 shows the 'true' power of the Wald test to determine a treatment difference in design (C), for data which arise from compound symmetry and antedependence covariance structures. Superimposed are the power levels of the corresponding tests using the adjusted sandwich estimator and Box corrections from the simula-

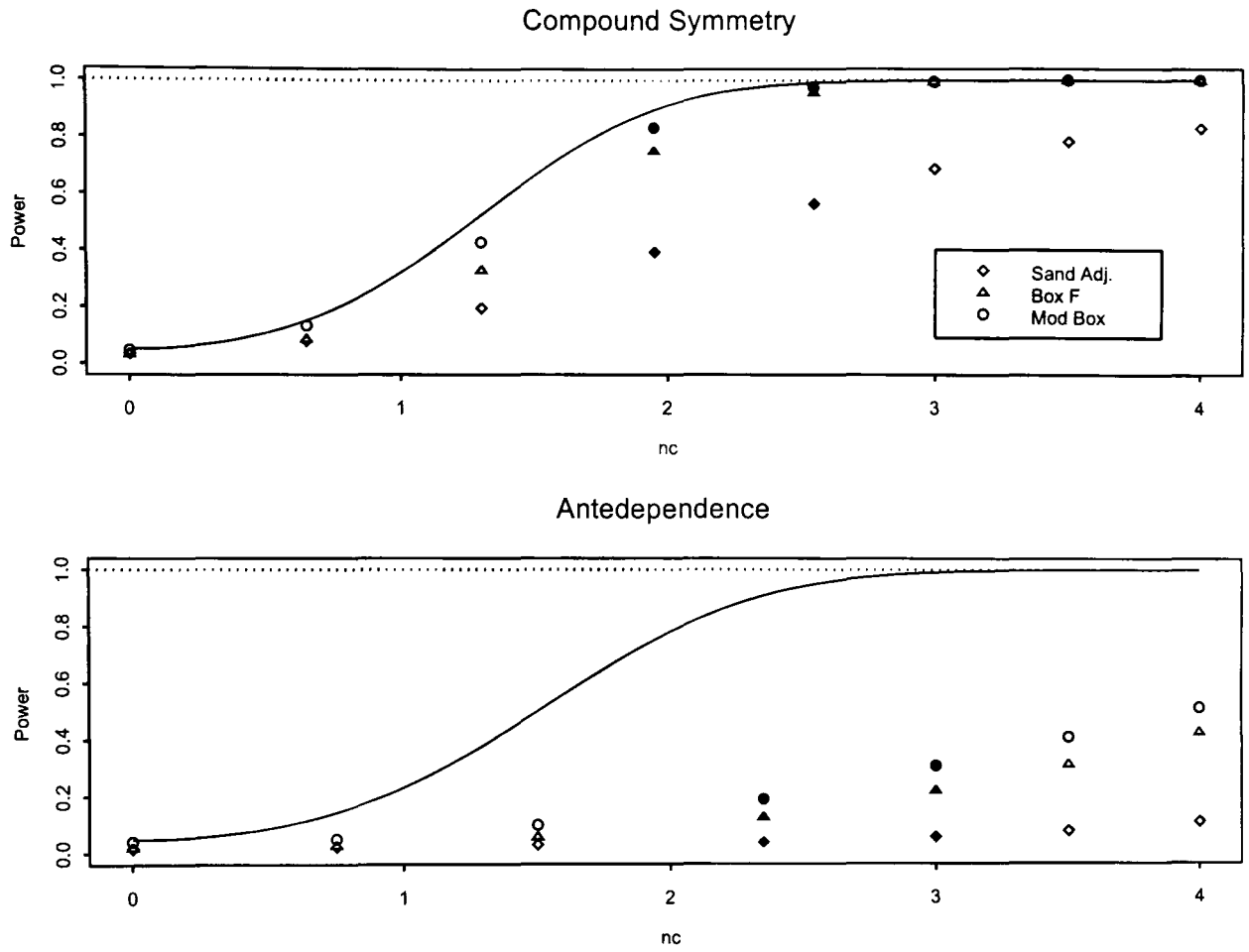


Figure 9.2.1: *Power levels of a Wald test using the 'true' compound symmetry and antedependence covariance structures.*

tions at various levels. Powers corresponding to a 'true' level using the Wald test of 90% and 100% are shown in solid.

Power levels of the tests using the adjusted sandwich estimator and the Box corrections, based on the above procedure, are compared in Table 9.2.3 at these two levels; power 1, the treatment difference at which the 'true' test achieves a 90% power level, and power 2, the level at which the 'true' test settles to 100%. Such levels are clearly unachievable in practice, where the true underlying covariance structure is unknown and must be estimated from the data. The powers noted are therefore for comparison only within the table.

Looking at the table, we see a similar pattern to that reported for designs (A) and (B). However, note that the test sizes achieved using the adjusted sandwich estimator and Box corrections are no longer inflated in the case of the Q.R.E. data. As we have previously seen, KR adjusted Wald tests using the unstructured covariance

Underlying Covariance Structure	Method of Inference	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment Effect)		
		Type I error rate ('Size')	Power 1	Power 2
Stationary Structures				
Identity	Unstr (KR)	0.654	——	——
	Sand Adj	0.039	0.385	0.568
	Box F	0.038	0.796	0.963
	Mod Box	0.056	0.837	0.977
Compound Symmetry	Unstr (KR)	0.690	——	——
	Sand Adj	0.034	0.393	0.564
	Box F	0.032	0.750	0.958
	Mod Box	0.046	0.833	0.977
AR1 ($\rho = 0.2$)	Unstr (KR)	0.660	——	——
	Sand Adj	0.037	0.391	0.568
	Box F	0.028	0.811	0.958
	Mod Box	0.048	0.848	0.969
AR1 ($\rho = 0.8$)	Unstr (KR)	0.672	——	——
	Sand Adj	0.034	0.181	0.313
	Box F	0.020	0.447	0.738
	Mod Box	0.043	0.557	0.827
Non-Stationary Structures				
Ante-dependence	Unstr (KR)	0.712	——	——
	Sand Adj	0.018	0.037	0.057
	Box F	0.021	0.126	0.221
	Mod Box	0.042	0.037	0.309
Unstr.	Unstr (KR)	0.677	——	——
	Sand Adj	0.025	0.082	0.115
	Box F	0.021	0.069	0.132
	Mod Box	0.047	0.139	0.220
Unstr. (Q.R.E.)	Unstr (KR)	0.698	——	——
	Sand Adj	0.009	0.012	0.012
	Box F	0.004	0.014	0.016
	Mod Box	0.038	0.049	0.066

Table 9.2.3: *Summary of results from 1000 simulations of Design (C). Table gives the proportion of type 1 errors (Size) and Power.*

estimate are hugely inflated, so it is not appropriate to consider their power. Again. the power obtained using the Box corrections is generally much greater than that achieved by the sandwich estimator, which gives very poor results by comparison. The power levels for the data arising from the underlying non-stationary structures is very low for these methods compared to the 'true' test.

It is perhaps worth noting that if we consider instead a null hypothesis of no period effect, all methods, including the KR adjusted Wald test give adequate control over the test size. In this case, as noted by Kenward and Roger (1997), the period effect levels are the variables defining the covariance structure and, in the absence of a treatment effect, this test takes the form an exact Hotelling T^2 test.

The results of these simulations, across study designs (A)-(C), show that the modified Box correction appears to provide a test statistic with nominal properties, being less conservative than Box's original statistic, and that this method has good power to detect significant differences from the null hypothesis across a range of settings. This power exceeds that of the exact Hotelling T^2 test, given by the KR adjusted Wald test, where the data allow. However, there is some uncertainty as to the real performance of this approach in small sample settings where the data are unbalanced, or arise from covariance structures which are non-stationary. Whilst the use of the 'badly behaved' Q.R.E. structure in simulations has been of interest, such a structure is unlikely to be found in practice, where data exhibiting large differences in variance across time would likely be stabilised prior to an analysis being undertaken.

There is a need for a wider range of simulations including consideration of changes to both the number of time points (measurement occasions) and subjects observed, in order to determine whether the modified Box correction can be widely endorsed for the analysis of very small samples of repeated measurements. This will be the focus of the next section.

9.3 Further Simulations

Consideration is given to extending the simulations based on designs (A)-(C), accommodating both a range of time points (p) and subjects (m), and including data arising from a wider variety of 'believable' non-stationary covariance structures.

The extended simulation designs are detailed below. For the simple repeated measures designs, we have

(A') A simple repeated measures experiment, with m subjects randomly allocated to two treatment groups (of equal size), and a response recorded for each subject at each of p time points.

(B') As design (A'), but with missing values. An equal number of subjects in each treatment group drop out at some random time following the first observation.

In designs (A') and (B'), we will consider $p = 5$ time points with $m = 10$ and 20 subjects, and $p = 10$ time points with $m = 20$ and 40 subjects. Additionally, in design (B'), the numbers of subjects allowed to drop out are given in Table 9.3.1, below.

Number of measurements per subject (p)	Number of subjects (m)	Number of subjects to drop out
5	10	2
	20	4
10	20	4
	40	8

Table 9.3.1: *Number of drop out subjects in extended study design (B').*

For the crossover designs, we have

(C') A five treatment-five period crossover trial, with $m = 10$ and 20 subjects allocated randomly to treatments according to Table 2.2.3, using a pair of Williams' squares.

(**D**) A nine treatment-nine period crossover trial, with $m = 18$ and 36 subjects allocated randomly to treatments according to Table 9.3.2 below.

Subject	Period								
	1	2	3	4	5	6	7	8	9
1	A	B	C	D	E	F	G	H	I
2	B	D	A	F	C	I	H	G	E
3	C	F	E	G	D	B	I	A	H
4	D	G	F	I	B	H	E	C	A
5	E	A	I	C	H	D	F	B	G
6	F	H	B	E	I	G	A	D	C
7	G	I	D	H	F	A	C	E	B
8	H	C	G	B	A	E	D	I	F
9	I	E	H	A	G	C	B	F	D
10	I	H	G	F	E	D	C	B	A
11	E	G	H	I	C	F	A	D	B
12	H	A	I	B	D	G	E	F	C
13	A	C	E	H	B	I	F	G	D
14	G	B	F	D	H	C	I	A	E
15	C	D	A	G	I	E	B	H	F
16	B	E	C	A	F	H	D	I	G
17	F	I	D	E	A	B	G	C	H
18	D	F	B	C	G	A	H	E	I

Table 9.3.2: *A crossover design for 9 treatments (A, B, C, D, E, F, G, H, I).*

For the extended simulations involving additional subjects ($m = 20$ in design (C'), and $m = 36$ in design (D)), the allocation tables are simply repeated.

As before, data samples are independently generated arising from a Gaussian distribution with zero mean for a number of underlying covariance structures, and the appropriate null hypotheses under consideration are those of no treatment/time interaction in designs (A') and (B'), and no treatment effect in designs (C) and (D). Two underlying stationary covariance structures, compound symmetry and AR1 (high correlation) are considered, together with three non-stationary structures, heterogeneous compound symmetry, heterogeneous AR1 and first order independence. For each of the non-stationary structures, variances are restricted so that they differ by no more than a factor of 10 over the range of the measurement times. These structures are shown in Figures A.2.1-A.2.5 of Appendix A.

Results from the extended simulations of study designs (A') and (B') are shown in Tables 9.3.3 and 9.3.4.

For $p = 5$ time points in design (A'), we see that as the number of subjects rises from 10 to 20, the power of the tests using the Box corrections is still above that of the KR adjusted Wald test for data arising from the stationary covariance structures. However, for data arising from the non-stationary structures, the increase in power is generally lower in comparison to the level attained by the KR adjusted test. This is particularly noticeable for the data arising from the antedependence structure, which is furthest from the ANOVA assumptions of independence and homogeneity of variance. This is as we might expect, the performance of the KR adjusted test improving as the sample size increases, and the Box corrections performing comparatively less well for large departures from independence. A similar pattern is observed in design (B'), although the loss of power relative to the KR adjusted Wald test is more apparent for increased sample sizes where we have missing values.

As the number of time points increases to $p = 10$, the loss in power of the Box corrections relative to the KR adjusted test is less marked. That is, as the number of subjects and time points is increased in the balanced and complete data setting of design (A'), the modified Box correction appears to hold its own against the KR adjusted Wald test. For design (B'), where the missing values introduce imbalance, the KR adjustments no longer give an exact Hotelling T^2 test, and must be calculated individually for each data set. For the (10×10) matrices necessitated by considering $p = 10$ time points, this is too expensive in terms of available computational time for such a practical study. In order to provide a comparison, the KR adjusted test results in Table 9.3.4 have been estimated in each case using the 'known' underlying covariance structure and an average number of observations. This gives a measure of the best that could be achieved using the KR method, that is, a 'ceiling' to its performance. (The estimated results are marked with an asterisk in the table).

Throughout Table 9.3.4, the KR adjustment is seen to give a test statistic with inflated size (for both $p = 5$ and 10 time points), although this size is seen to

Underlying Covariance Structure	Number of Times(p) & Subjects(m)	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment/Time Interaction)				
		Method of Inf.	Unstr (KR)	Sand Adj	Box F	Mod Box
Stationary Structures						
Compound Symmetry	p=5,m=10	'Size'	0.045	0.057	0.023	0.042
		Power	0.747	0.660	0.964	0.984
	p=5,m=20	'Size'	0.057	0.037	0.024	0.036
		Power	0.749	0.663	0.808	0.843
	p=10,m=20	'Size'	0.049	0.035	0.013	0.020
		Power	0.756	0.510	0.950	0.964
	p=10,m=40	'Size'	0.049	0.034	0.031	0.038
		Power	0.756	0.645	0.827	0.840
AR1 ($\rho = 0.8$)	p=5,m=10	'Size'	0.056	0.083	0.035	0.068
		Power	0.775	0.793	0.992	0.999
	p=5,m=20	'Size'	0.042	0.037	0.032	0.052
		Power	0.756	0.884	0.937	0.958
	p=10,m=20	'Size'	0.048	0.065	0.038	0.060
		Power	0.753	0.698	0.995	0.996
	p=10,m=40	'Size'	0.052	0.048	0.037	0.053
		Power	0.779	0.728	0.968	0.972
Non-Stationary Structures						
Hetero Comp Sym	p=5,m=10	'Size'	0.053	0.044	0.026	0.052
		Power	0.773	0.534	0.961	0.981
	p=5,m=20	'Size'	0.042	0.017	0.033	0.052
		Power	0.740	0.586	0.788	0.823
	p=10,m=20	'Size'	0.052	0.016	0.026	0.040
		Power	0.758	0.207	0.980	0.989
	p=10,m=40	'Size'	0.063	0.023	0.023	0.028
		Power	0.765	0.550	0.901	0.910
Hetero AR1 ($\rho = 0.8$)	p=5,m=10	'Size'	0.049	0.076	0.034	0.069
		Power	0.744	0.765	0.981	0.994
	p=5,m=20	'Size'	0.043	0.046	0.037	0.053
		Power	0.770	0.721	0.922	0.947
	p=10,m=20	'Size'	0.051	0.042	0.035	0.054
		Power	0.767	0.604	0.990	0.996
	p=10,m=40	'Size'	0.053	0.044	0.031	0.047
		Power	0.755	0.687	0.912	0.932
Ante-dependence	p=5,m=10	'Size'	0.060	0.053	0.038	0.059
		Power	0.767	0.611	0.861	0.924
	p=5,m=20	'Size'	0.058	0.032	0.041	0.053
		Power	0.741	0.600	0.624	0.688
	p=10,m=20	'Size'	0.043	0.003	0.041	0.052
		Power	0.749	0.070	0.834	0.861
	p=10,m=40	'Size'	0.056	0.010	0.038	0.043
		Power	0.764	0.403	0.704	0.725

Table 9.3.3: *Summary of results from 1000 simulations of extended Design (A'). Table gives the proportion of type 1 errors (Size) and Power.*

Underlying Covariance Structure	Number of Times(p) & Subjects(m)	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment/Time Interaction)				
		Method of Inf.	Unstr (KR)	Sand Adj	Box F	Mod Box
Stationary Structures						
Compound Symmetry	p=5,m=10 (miss 2)	'Size'	0.074	0.051	0.017	0.040
		Power	0.632	0.523	0.882	0.937
	p=5,m=20 (miss 4)	'Size'	0.059	0.036	0.041	0.050
		Power	0.644	0.505	0.733	0.772
	p=10,m=20 (miss 4)	'Size'	0.053*	0.050	0.012	0.023
		Power	0.624*	0.361	0.857	0.890
	p=10,m=40 (miss 8)	'Size'	0.057*	0.038	0.046	0.052
		Power	0.624*	0.449	0.610	0.643
AR1 ($\rho = 0.8$)	p=5,m=10 (miss 2)	'Size'	0.078	0.064	0.024	0.061
		Power	0.637	0.677	0.966	0.989
	p=5,m=20 (miss 4)	'Size'	0.067	0.050	0.052	0.067
		Power	0.662	0.586	0.869	0.900
	p=10,m=20 (miss 4)	'Size'	0.068*	0.096	0.036	0.068
		Power	0.645*	0.618	0.993	0.995
	p=10,m=40 (miss 8)	'Size'	0.062*	0.040	0.054	0.070
		Power	0.674*	0.614	0.925	0.944
Non-Stationary Structures						
Hetero Comp Sym	p=5,m=10 (miss 2)	'Size'	0.084	0.056	0.019	0.042
		Power	0.609	0.412	0.843	0.927
	p=5,m=20 (miss 4)	'Size'	0.054	0.032	0.043	0.059
		Power	0.634	0.439	0.681	0.739
	p=10,m=20 (miss 4)	'Size'	0.062*	0.025	0.022	0.037
		Power	0.627*	0.162	0.910	0.939
	p=10,m=40 (miss 4)	'Size'	0.063*	0.020	0.030	0.038
		Power	0.667*	0.318	0.796	0.810
Hetero AR1 ($\rho = 0.8$)	p=5,m=10 (miss 2)	'Size'	0.083	0.066	0.031	0.072
		Power	0.631	0.578	0.914	0.975
	p=5,m=20 (miss 4)	'Size'	0.048	0.025	0.052	0.070
		Power	0.683	0.562	0.836	0.884
	p=10,m=20 (miss 4)	'Size'	0.048*	0.049	0.048	0.075
		Power	0.602*	0.413	0.948	0.974
	p=10,m=40 (miss 4)	'Size'	0.058*	0.038	0.042	0.054
		Power	0.658*	0.505	0.851	0.894
Ante-dependence	p=5,m=10 (miss 2)	'Size'	0.065	0.042	0.038	0.072
		Power	0.664	0.494	0.748	0.862
	p=5,m=20 (miss 4)	'Size'	0.054	0.028	0.040	0.057
		Power	0.658	0.472	0.558	0.634
	p=10,m=20 (miss 4)	'Size'	0.071*	0.000	0.034	0.049
		Power	0.604*	0.040	0.767	0.813
	p=10,m=40 (miss 8)	'Size'	0.048*	0.005	0.036	0.043
		Power	0.654*	0.174	0.603	0.628

Table 9.3.4: *Summary of results from 1000 simulations of extended Design (B'). Table gives the proportion of type 1 errors (Size) and Power.*

approach the nominal level of 5% as the number of subjects increases.

In the extended simulations of study designs (A') and (B') tests involving the adjusted sandwich estimator are seen to control size, but these tests appear to have little power to detect differences from the null hypothesis in comparison to the other methods. This is particularly true for data arising from underlying covariance structures which are far from the independence (identity) 'working' covariance structure. Also, for increased time points ($p = 10$), there were problems with this adjustment in the low subject setting ($m = 20$), resulting in illegal (negative) estimates for the denominator degrees of freedom given by (7.4.2), as $v \leq (l-1)$ is possible. In these instances, requiring that $v > (l-1)$ is achieved by setting $v = l$, the dimensionality of the test, which results in a single denominator degree of freedom test. These issues do not, however, reccur as the number of subjects increases, and the results for 10 time points and 20 subjects do not appear to be out of line.

Consider now the extended crossover studies of designs (C') and (D).

Table 9.3.5 shows the results from the extended study design (C') based on the five treatment-five period design. As the number of subjects increases from 10 to 20, the test sizes using the KR adjustment are closer to the nominal level of 5%, but are still too inflated for power to be considered. As with design (C), power is compared for the adjusted sandwich estimator and the Box corrections for treatment differences which lead to 'true' powers using the Wald statistic (with 'known' covariance structure) of 90% and 100%, power 1 and power 2, respectively. Again, the modified Box correction is seen to give a test with nominal properties and good power in comparison to the 'true' test.

Results from the simulations under study design (D) are shown in Table 9.3.6. The power of tests using the adjusted sandwich estimator and Box corrections are compared, as in designs (C) and (C'), against 'true' power levels for the corresponding Wald test, and the modified Box correction is seen to give the better performance. Again the adjusted sandwich estimator results in illegal parameter estimates where

Underlying Covariance Structure	Number of Times(p) & Subjects(m)	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment Effect)				
		Method of Inf.	Unstr (KR)	Sand Adj	Box F	Mod Box
Stationary Structures						
Compound Symmetry	p=5,m=10	'Size'	0.690	0.034	0.032	0.046
		Power 1	—	0.393	0.750	0.833
		Power 2	—	0.564	0.958	0.977
	p=5,m=20	'Size'	0.126	0.045	0.049	0.062
		Power 1	—	0.679	0.860	0.881
		Power 2	—	0.900	0.981	0.983
AR1 ($\rho = 0.8$)	p=5,m=10	'Size'	0.672	0.034	0.020	0.043
		Power 1	—	0.181	0.447	0.557
		Power 2	—	0.313	0.738	0.827
	p=5,m=20	'Size'	0.123	0.024	0.036	0.053
		Power 1	—	0.325	0.584	0.642
		Power 2	—	0.533	0.837	0.867
Non-Stationary Structures						
Hetero Comp Sym	p=5,m=10	'Size'	0.682	0.032	0.016	0.047
		Power 1	—	0.233	0.563	0.649
		Power 2	—	0.372	0.840	0.899
	p=5,m=20	'Size'	0.124	0.033	0.038	0.053
		Power 1	—	0.450	0.693	0.732
		Power 2	—	0.691	0.904	0.926
Hetero AR1 ($\rho = 0.8$)	p=5,m=10	'Size'	0.597	0.037	0.012	0.028
		Power 1	—	0.126	0.225	0.295
		Power 2	—	0.226	0.442	0.536
	p=5,m=20	'Size'	0.111	0.024	0.029	0.043
		Power 1	—	0.214	0.404	0.465
		Power 2	—	0.370	0.665	0.717
Ante-dependence	p=5,m=10	'Size'	0.697	0.028	0.024	0.037
		Power 1	—	0.140	0.383	0.473
		Power 2	—	0.206	0.618	0.698
	p=5,m=20	'Size'	0.122	0.028	0.048	0.054
		Power 1	—	0.226	0.445	0.479
		Power 2	—	0.398	0.688	0.712

Table 9.3.5: *Summary of results from 1000 simulations of extended Design (C'). Table gives the proportion of type 1 errors (Size) and Power.*

Underlying Covariance Structure	Number of Times(p) & Subjects(m)	Proportion of Significant Test Results (out of 1000) (Null model- No Treatment Effect)				
		Method of Inf.	Unstr (KR)	Sand Adj	Box F	Mod Box
Stationary Structures						
Compound Symmetry	p=9,m=18	'Size'	0.472*	0.037	0.033	0.040
		Power 1	—	0.338	0.838	0.872
		Power 2	—	0.531	0.991	0.994
	p=9,m=36	'Size'	0.063*	0.029	0.041	0.047
		Power 1	0.781*	0.698	0.868	0.881
		Power 2	0.977*	0.941	0.997	0.997
AR1 ($\rho = 0.8$)	p=9,m=18	'Size'	0.497*	0.020	0.020	0.033
		Power 1	—	0.074	0.352	0.406
		Power 2	—	0.139	0.656	0.706
	p=9,m=36	'Size'	0.086*	0.023	0.037	0.052
		Power 1	0.820*	0.204	0.453	0.474
		Power 2	0.985*	0.406	0.778	0.796
Non-Stationary Structures						
Hetero Comp Sym	p=9,m=18	'Size'	0.499*	0.016	0.028	0.037
		Power 1	—	0.159	0.550	0.599
		Power 2	—	0.300	0.864	0.892
	p=9,m=36	'Size'	0.065*	0.032	0.044	0.048
		Power 1	0.783*	0.352	0.643	0.670
		Power 2	0.974*	0.671	0.922	0.929
Hetero AR1 ($\rho = 0.8$)	p=9,m=18	'Size'	0.473*	0.023	0.016	0.030
		Power 1	—	0.054	0.242	0.295
		Power 2	—	0.100	0.471	0.535
	p=9,m=36	'Size'	0.069*	0.026	0.036	0.046
		Power 1	0.795*	0.091	0.293	0.324
		Power 2	0.987*	0.237	0.591	0.619
Ante-dependence	p=9,m=18	'Size'	0.483*	0.018	0.029	0.023
		Power 1	—	0.068	0.286	0.309
		Power 2	—	0.126	0.515	0.403
	p=9,m=36	'Size'	0.066*	0.016	0.044	0.051
		Power 1	0.785*	0.123	0.324	0.338
		Power 2	0.951*	0.242	0.579	0.592

Table 9.3.6: *Summary of results from 1000 simulations of Design (D). Table gives the proportion of type 1 errors (Size) and Power.*

the number of time points is increased and the number of subjects is small ($m = 18$), so that the denominator degrees of freedom of such tests are fixed at 1, close to the boundary.

In this setting, with an estimated (9×9) covariance structure, the KR adjustments are again computationally expensive in terms of 1000 individual simulations, so are estimated, for comparison purposes, using the 'known' structure to give a 'ceiling' of performance. (Again these estimated values are marked with an asterisk in the table). These show that the test size reduces towards nominal levels as the number of subjects increases, so that it may be appropriate to consider power. However, it is clear that these values cannot be achieved in practice.

9.4 Discussion

The extensive simulation studies of this Chapter show that the modified Box correction results in a test with nominal properties which is more powerful than the other methods considered across a range of small sample settings for the analysis of repeated measurements.

Tests based on the adjusted sandwich estimator are also seen to have nominal properties across the range of settings considered, but this approach lacks power. Also, estimation of the adjusted test parameters are seen to be non-robust, resulting in illegal estimates of the denominator degrees of freedom where the number of time points is large in comparison to the number of subjects.

The simulations confirm that Wald tests using an unstructured covariance matrix with the Kenward Roger adjustment give inflated type 1 error rates where the data do not allow for exact tests, although the size of such tests does approach nominal levels as the sample size increases as we might expect. Where nominal properties are achieved, so that it is appropriate to consider power, we see that where the sample size (number of subjects) is small, tests using the Box corrections give greater power

than the corresponding Wald tests. As the sample size increases the improvement in power from using the modified Box correction over the KR adjustment becomes less as the underlying covariance structure moves further from independence.

The modified Box correction developed in Chapter 8 is preferred to Box's original statistic which is conservative and hence less powerful.

Chapter 10

Examples

10.1 Introduction

Following the simulation studies of the previous Chapter, some examples are presented which illustrate the use of the modified Box correction in practical analyses. Attention is restricted to this method since the results of the simulation studies indicate that this is the most suitable method for dealing with small sample repeated measures problems, since it produces a test statistic with nominal properties and has the greatest power, of the methods considered, across a range of settings.

10.2 Cardiac Enzyme in Preserved Dog Hearts

Consider again the Cardiac Enzyme example of Section 4.1.1, where four test liquids, defined by the presence or absence of two components A and B, were compared for their preserving qualities on 23 preserved dog hearts, measured by the percentage of total enzyme (%ATP) at nine repeated intervals of one and two hours.

Here we consider the reduced data set of 12 dog hearts, restricting our interest to the presence or absence of component A in the preserving liquid (B absent). The subject and mean profiles for the two liquids are shown in Figures 10.2.1 and 10.2.2 respectively, and the sample variance-correlation matrix is shown in Table 10.2.1.

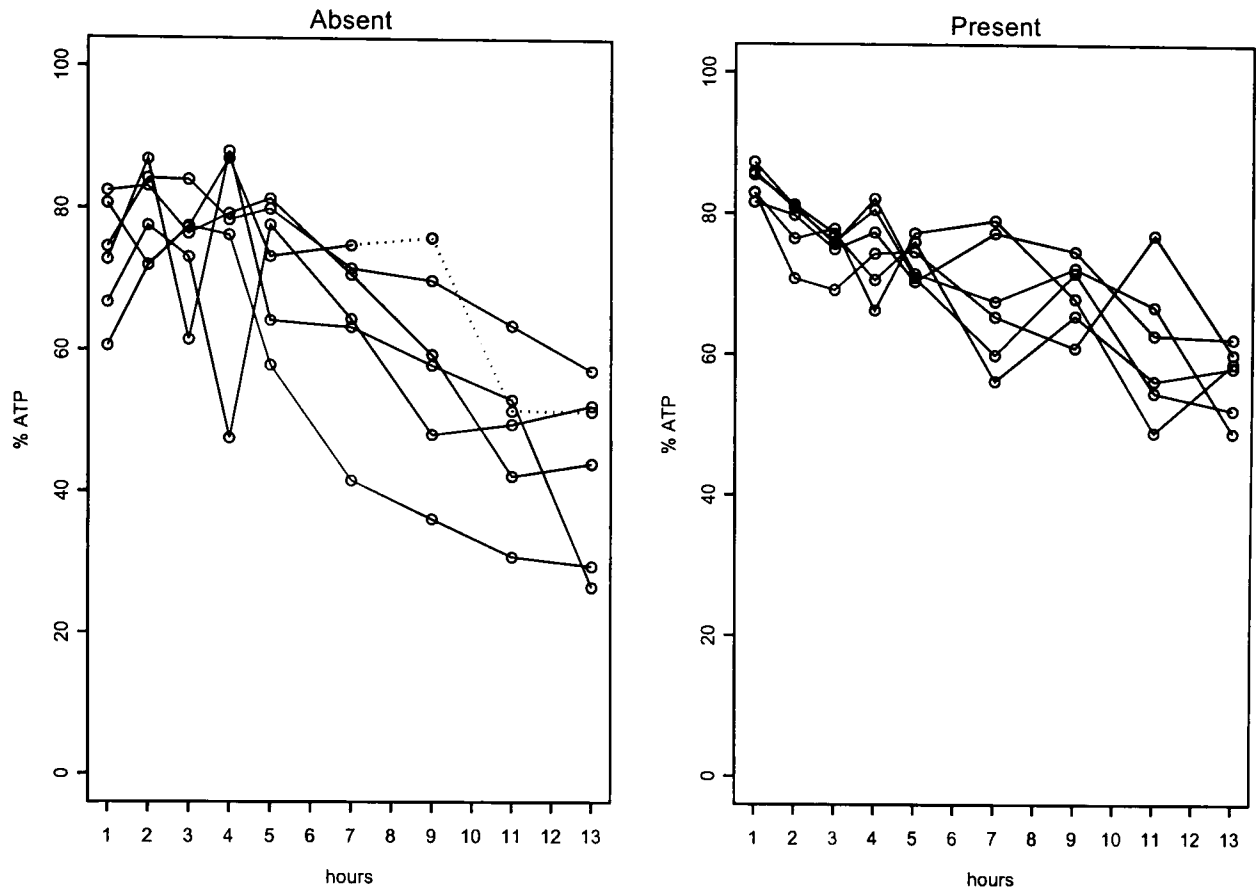


Figure 10.2.1: *Cardiac Enzyme data: Subject Profiles for the Presence/Absence of preserving liquid A*

(This shows variances on the diagonal, with covariances above and correlations below).

37.08	11.29	4.04	32.53	24.78	37.22	51.32	19.08	15.89
0.34	**29.27**	-3.52	12.80	7.64	10.02	18.66	8.14	-7.84
0.12	-0.11	**33.08**	-7.70	15.43	6.91	15.80	-11.43	30.00
0.47	0.21	-0.12	**128.08**	-27.86	6.51	58.84	19.38	-43.17
0.58	0.20	0.38	-0.35	**48.85**	46.33	33.20	24.45	53.01
0.57	0.17	0.11	0.05	0.62	**114.22**	86.48	44.59	61.27
0.78	0.32	0.25	0.48	0.44	0.75	**117.38**	51.39	48.76
0.30	0.14	-0.19	0.16	0.33	0.40	0.45	**111.24**	42.10
0.27	-0.15	0.54	-0.39	0.78	0.59	0.46	0.41	**94.24**

Table 10.2.1: *Sample Variance-Correlation Matrix for the Cardiac Enzyme Data*

Treating these data as a simple repeated measures design, a test for a treatment by time interaction using the unstructured covariance matrix with a Kenward Roger (KR) small sample adjustment is equivalent to an exact Hotelling T^2 test. Table

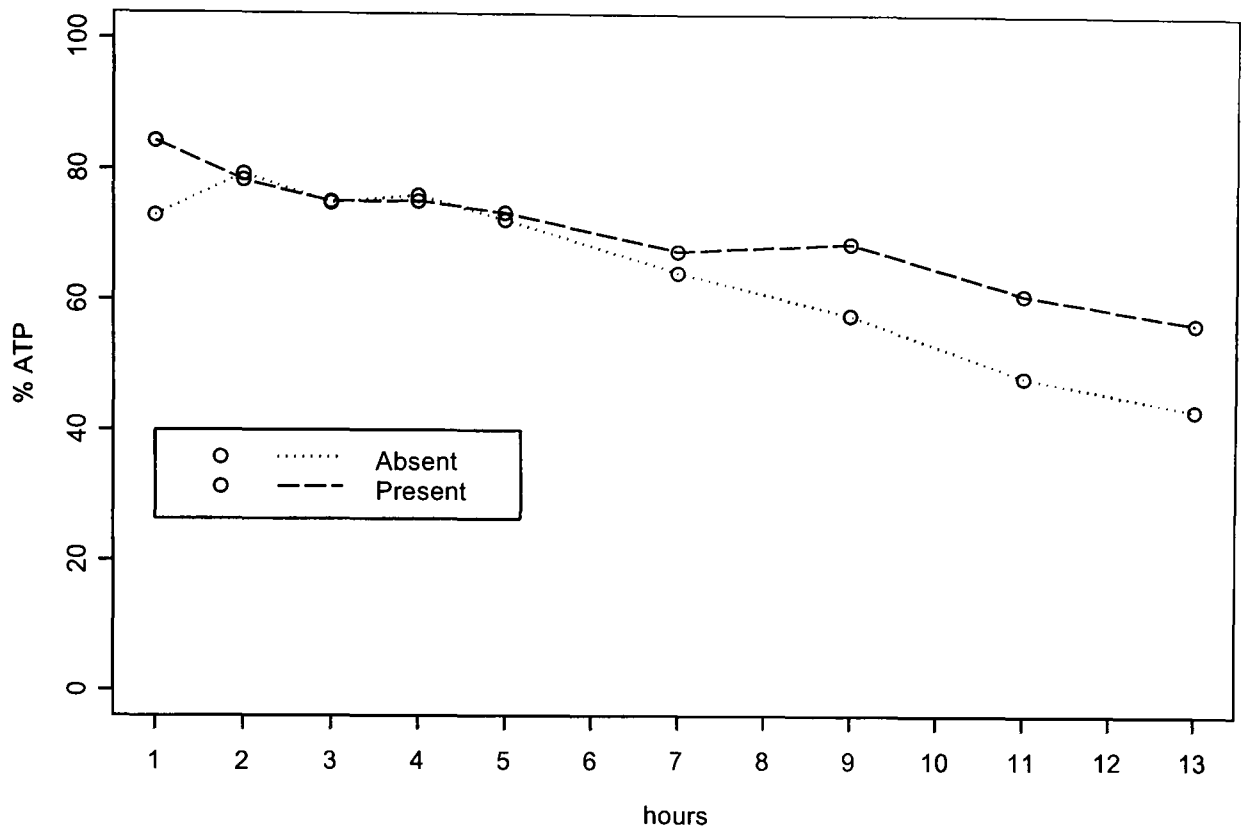


Figure 10.2.2: *Cardiac Enzyme data: Mean Profiles*

10.2.2 shows the results obtained from Wald tests using various covariance models and the modified Box correction. It can be seen that the test results differ according to the choice of structure. As the data are complete and balanced, the same mean estimates (by ordinary least squares) are obtained under each method, but they differ in their measures of the standard errors.

Covariance Structure	Num df	Den df	F	p
Identity - Independence	8	90	1.49	0.1713
Unstructured	8	3	8.73	0.0509
Compound Symmetry	8	80	2.07	0.0485
AR1	8	73.8	1.24	0.2904
Mod Box ($\lambda = 0.68$)	8	11.2	2.21	0.1109

Table 10.2.2: *Cardiac Enzyme Data: Comparison of Results, complete data*

The exact Hotelling T^2 test obtained using the unstructured covariance matrix indicates that there is insufficient evidence (at the 5% level) to reject the null hypothesis

of no treatment by time interaction. This is confirmed by the modified Box corrected statistic, although in this instance the evidence is less marginal.

To show how such results can differ where there is imbalance, and the estimates of the mean parameters as well as their standard errors are dependent on the choice of covariance structure, we introduce an artificial dropout to this reduced data set. To achieve this, consider that the three final measurements are missing from one of the dog hearts which receives the preserving liquid from which component A is absent. The profile for this heart, following dropout, is shown as a dotted line in Figure 10.2.1. Repeating the tests for a treatment/time interaction with this artificial dropout gives the results shown in Table 10.2.3.

Covariance Structure	Num df	Den df	F	p
Identity - Independence	8	87	1.87	0.0753
Unstructured	8	1.6	88.63	0.0252
Compound Symmetry	8	77.2	2.32	0.0274
AR1	8	12.2	1.51	0.1686
Mod Box ($\lambda = 0.77$)	8	2.4	10.48	0.0901

Table 10.2.3: *Cardiac Enzyme Data: Comparison of Results, with dropout*

Now the tests using the KR adjustment (with the identity, unstructured and compound symmetry structures) are no longer exact, and so these results are less believable given the evidence of our simulations. There is a higher significance of an interaction using the unstructured form once the three observations from the 'absent' group are removed, but the test using the modified Box correction remains non-significant.

10.3 Antihistamines and Mental Performance

The data in this example arise from a crossover experiment to test the effects of antihistamines on mental performance. There are 9 subjects who take part in 8 test sessions per day, with each subject taking one drug (of six) per day. Testing days

are spaced at least 7 days apart to minimise carryover effects.

The six drugs (antihistamines) are as follows:

> Placebo; Promethazine, 20 mg; Maratadine, 5 mg; Maratadine, 10 mg; Solfenadine 20 mg; Solfenadine, 40 mg.

Promethazine is a sedating antihistamine and is used as an active control, as it is known to impair performance. Maratadine and Solfenadine are thought to be non-sedating antihistamines and so performance is expected to be unaffected by these drugs.

The drugs are taken by each of the 9 subjects at 10.30 on each of six (testing) days, and are administered according to a latin square. Mental performance is tested at 8 sessions during the day:

> 9.30 (pre-dose); 11.00 (post-dose); 12.00; 13.30; 15.30; 17.30; 19.30; 21.30.

At each session a number of tests of mental performance are undertaken, but we focus here on 'choice reaction time' (CRT), which measures the speed of a correct response by subjects who are required to press one of four buttons which correspond to the position and value of single digits displayed on either the left or right side of a PC display. The response profiles of each of the 9 subjects for tests of each of the six drugs are shown in Figures 10.3.1 and 10.3.2 respectively.

Following a similar example, Example 5.6 in Jones and Kenward (2003), we analyse the data using a linear (mixed) model, with repeated measurements within periods. An appropriate model here has fixed subject effects and a factorial structure for time and treatment (drug), since no particular form for the treatment effect over time is suggested by Figure 10.3.2. Also, the use of the baseline (pre-dose) measurement as a covariate reduces the contribution of the subject effects, which account for

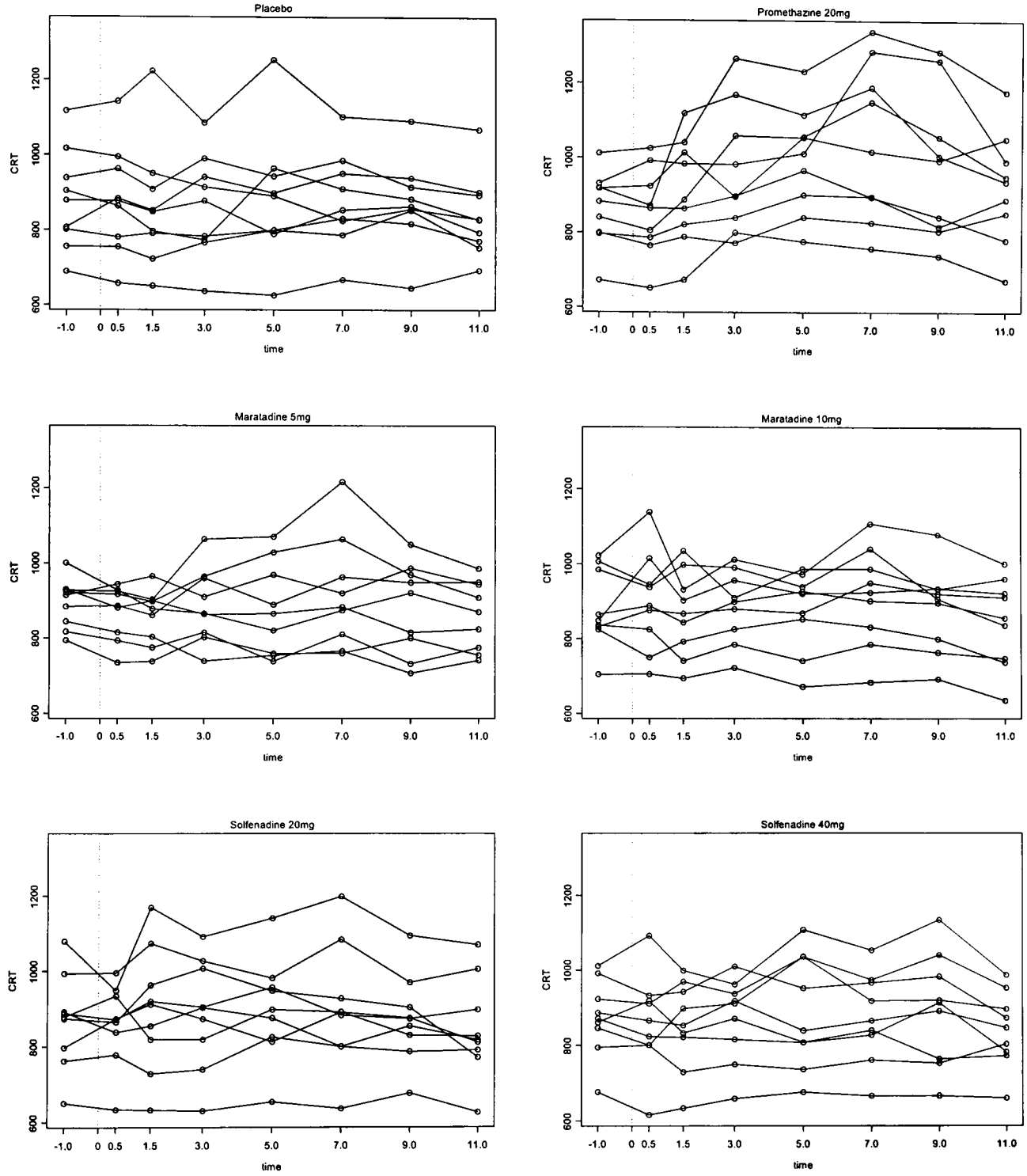


Figure 10.3.1: *CRT data: Subject Profiles for the six treatments*

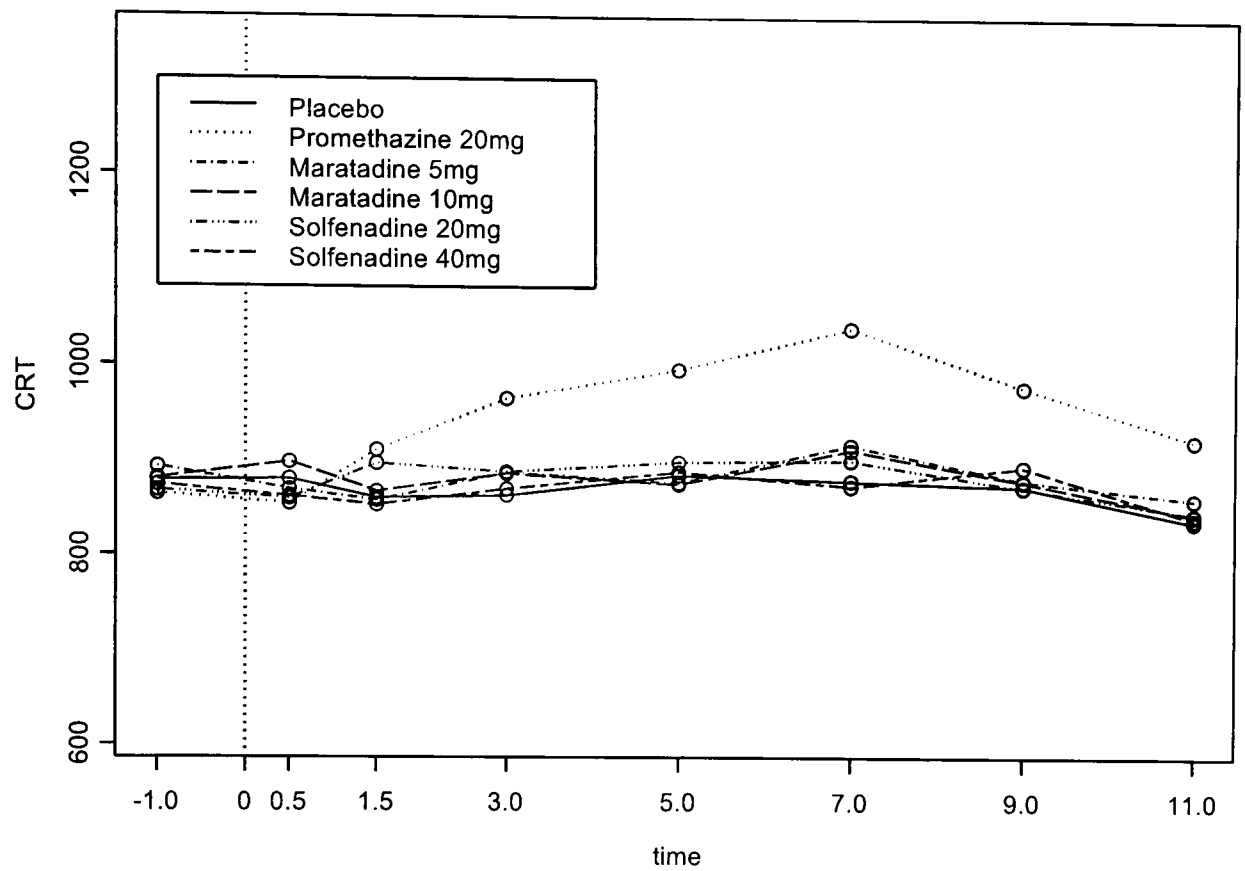


Figure 10.3.2: *CRT data: Mean Profiles*

the between-period dependencies. An unstructured model for the within-periods covariance structure is adopted, and is shown in Table 10.3.1.

2798.16	481.64	1065.70	920.05	1990.57	2092.34	1528.29
0.17	**2951.34**	2246.65	2125.46	2480.89	1747.42	1087.08
0.25	0.52	**6359.39**	3767.16	5054.03	3915.04	3257.63
0.23	0.51	0.61	**5909.75**	4870.05	4204.56	3061.88
0.40	0.49	0.67	0.67	**8840.50**	6815.84	4443.98
0.45	0.37	0.56	0.62	0.83	**7708.59**	4449.78
0.42	0.29	0.60	0.59	0.69	0.74	**4628.74**

Table 10.3.1: *Within-Period Variance-Correlation Matrix for the CRT Data*

Jones and Kenward discuss the merits of fixed versus random subject effects in this situation, and conclude that fixed effects should be used unless there is a good justification for the additional complexity inherent in modelling the random effects, such as the recovery of interblock information. In a well constructed experiment, the loss of such information should be of little consequence. In fact, due to subject-

period orthogonality in this design, tests involving time are identical under both approaches in the absence of carryover effects.

Table 10.3.2 shows the results of the analysis using Wald tests with the Kenward Roger (KR) adjustment and the modified Box correction.

Effect		Num df	Den df	F	p
Wald tests with KR adjustment					
subject		8	34	6.54	<0.0001
baseline		1	34	59.24	<0.0001
period (week)		5	27.4	0.76	0.5876
treatment (drug)		5	26.9	8.23	<0.0001
time		6	38	8.06	<0.0001
period*time		30	89	1.23	0.2284
time*treatment		30	89	1.26	0.2005
Modified Box Correction					
subject	(λ=2.76)	8	191.81	9.57	<0.0001
baseline	(λ=2.77)	1	454.75	23.70	<0.0001
period (week)	(λ=0.74)	5	208.36	2.17	0.0583
treatment (drug)	(λ=0.73)	5	214.27	4.20	0.0012
time	(λ=0.71)	6	29.04	1.20	0.3339
period*time	(λ=0.74)	30	70.78	1.11	0.3561
time*treatment	(λ=0.74)	30	70.78	2.05	0.0070

Table 10.3.2: *CRT data: Comparison of Results*

Both methods of analysis flag a significant drug effect, but the conventional (KR adjusted) Wald tests suggest the effects of drug and time are additive, whereas the modified Box correction points to a significant drug/time interaction. This suggests that the drug effect is different at different times, which would appear to be consistent with the mean profile plot in Figure 10.3.2. The modified Box correction also suggests that there is marginal evidence of a period (week) effect. That is, results on a particular day (adjusted for other variables) are low or high for no apparent reason. Where the results differ from these two methods of analysis, the simulations of the previous chapter indicate that those obtained using the modified Box correction are more believable in this setting.

A sensible route for the further analysis of these data would be to consider indi-

vidual contrasts between the drugs at particular time points to test whether the supposedly non-sedating antihistamines, maratadine and solfenadine, behave like the placebo, and that the treatment differences observed are due almost entirely to the active control. One approach to this problem using the modified Box correction would be through the appropriate specification of the design matrix. However, an alternative approach based on Scheffé's method for contrasts is presented in the next section, which accounts both for the departures from independence (measured by the modified Box correction) and for multiplicity of testing.

10.4 Electrocardiogram Abnormalities in the Guinea Pig Papillary Muscle

As a final example, recall the Guinea pig papillary muscle (GPPM) example of Brammer (2003), presented as motivation in Chapter 2, Section 2.1. These data comprise measurements taken from papillary muscles dissected from the right ventricles of each of just three guinea pigs' hearts in two experiments. The purpose of the experiments was to determine whether the compounds are likely to cause electrocardiogram abnormalities. Brammer recognises that analysis from such small samples is unlikely to be definitive, but notes that such small samples are common in isolated tissue or organ experiments.

Since the isolated tissue assays from the guinea pigs deteriorate in time, there is a limited period in which to test different concentrations of the compounds on each muscle, so a control measure is followed by six increasing concentrations of the compound. In such an ascending dose design, the carryover effect is considered to be minimal in comparison to the current dose. Concentration and time are confounded, but a separate 'control' experiment with no compound present showed that there were no important changes over time. Five variables were measured, but we focus here on AP (amplitude of action potential). Individual tissue profiles and mean profiles under each compound are shown in Figures 10.4.1 and 10.4.2 respectively.

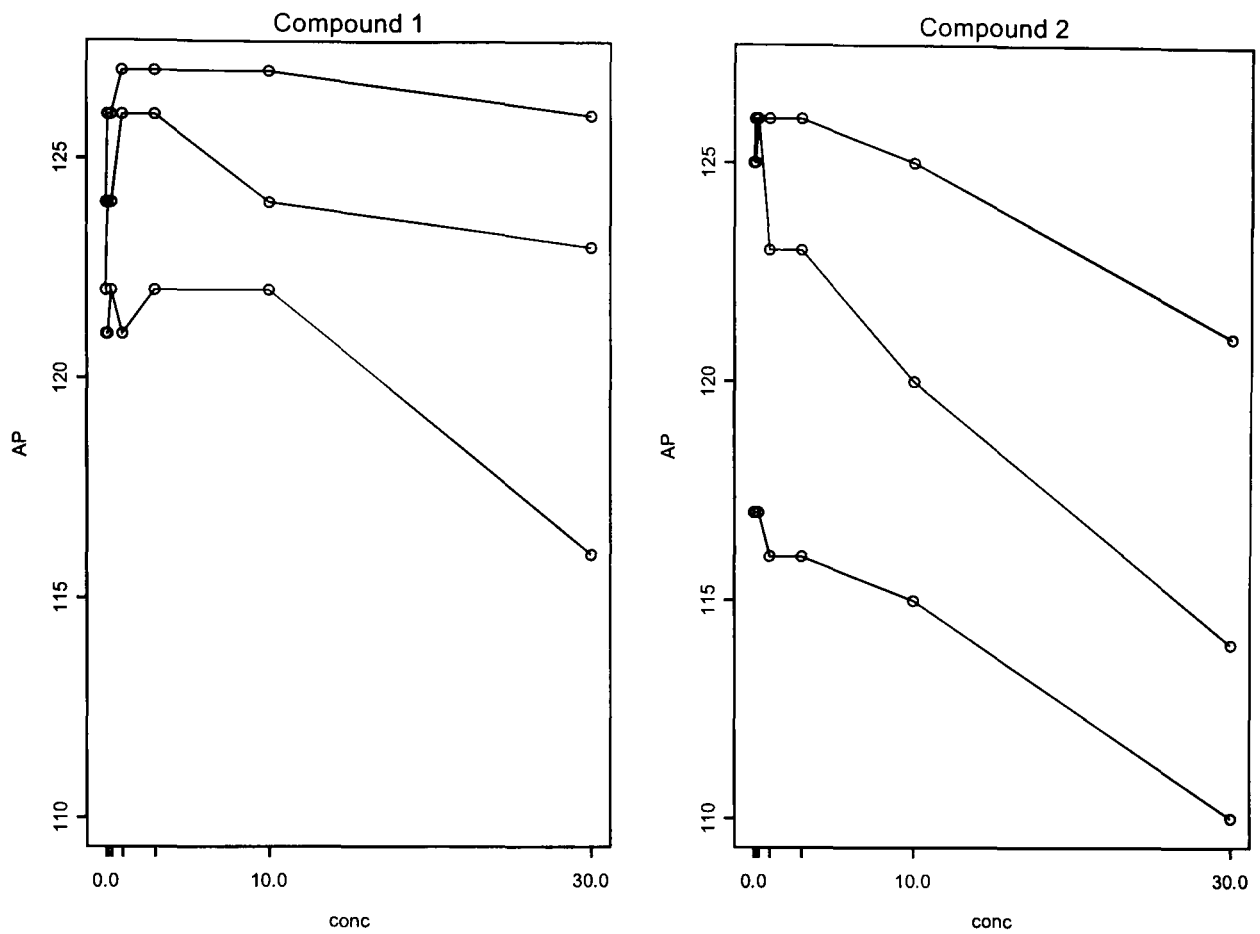


Figure 10.4.1: *GPPM Data: Subject Profiles, Compounds 1 and 2.*

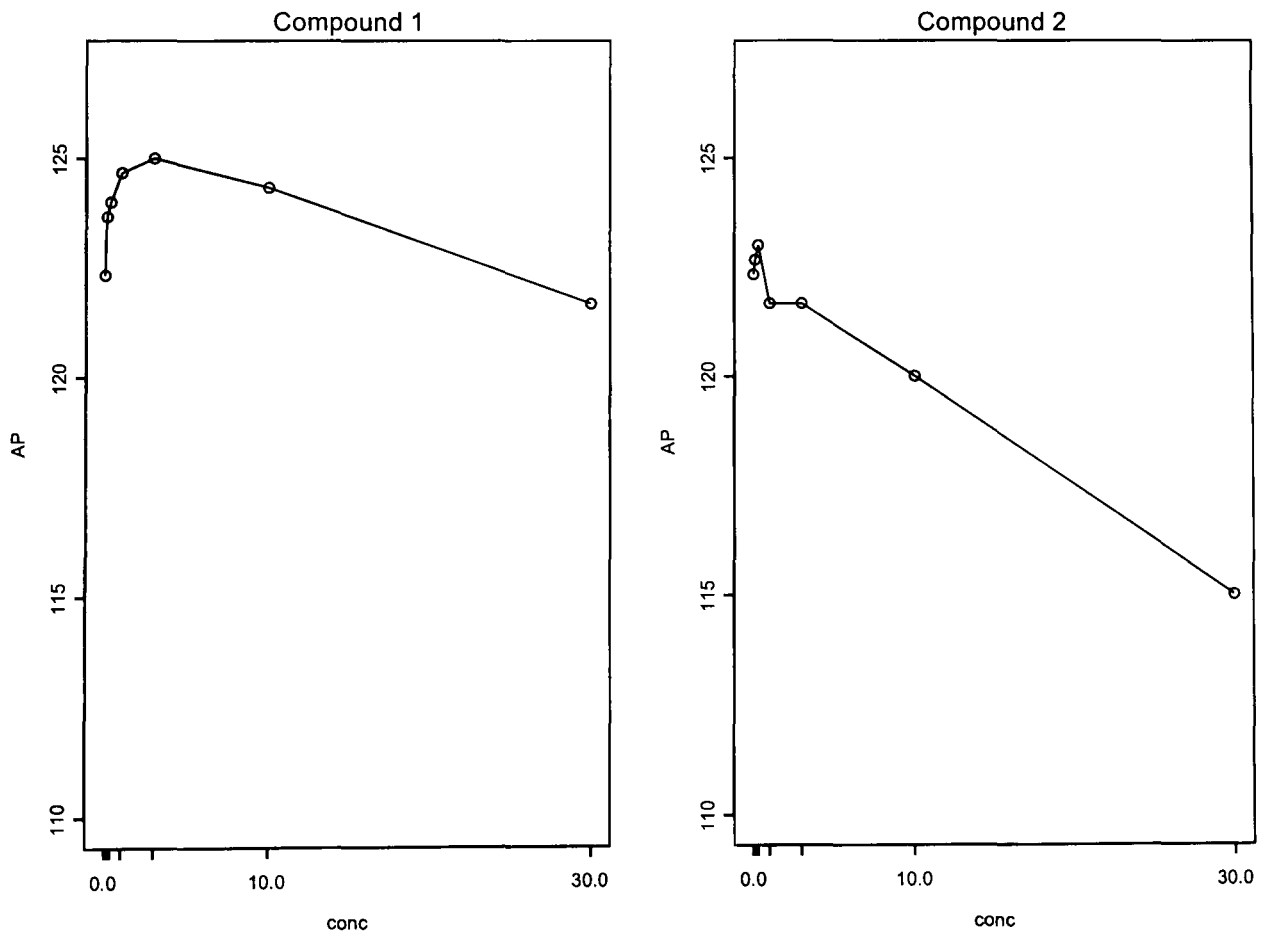


Figure 10.4.2: *GPPM Data: Mean Profiles, Compounds 1 and 2.*

These experiments can be considered as block designs, with concentration of compound as the treatment and tissue as the blocking factor, but the compound symmetry structure imposed by such a design may not be appropriate. Instead, Brammer treats the experiments as simple repeated measures designs with concentration as the time variable and tissue as the subject, and compares the resulting analyses from adopting various covariance models to account for the correlation between measurements on the same tissue (subject).

To analyse these experiments in such a way requires the fitting of a (7×7) covariance structure using just three subjects, so unsurprisingly the unstructured covariance models did not converge. Brammer compares those correlation structures that can be fitted by informal comparison of the (reduced) log-likelihoods and prefers an AR1 or heterogeneous AR1 model in each case over the more usual compound symmetry approach adopted for such experiments. However, the results of the simulations in Chapter 2 show that, for small samples, such methods are unreliable for choosing an appropriate structure.

Brammer's results for individual tests of difference from control of each concentration are reproduced in Table 10.4.1. For each compound, tests using an AR(1) and ARH(1) structure are compared with compound symmetry (2-way ANOVA) and paired t-tests of each level of concentration with the control. Additionally, Brammer fits an autoregressive moving average ARMA(1,1) model to the data for experiment 2. Also shown are results from fitting an antedependence model of order 1 to the data from compound 1, being the highest order structure which fits these data. For each method the mean levels to be compared are unchanged, so that the methods differ only in the calculation of the standard errors for the difference from the control (SEDC). Note that the results based on the paired t-tests are less powerful, being based only on 2 degrees of freedom, rather than the 11 df available for each of the other tests. Brammer makes no adjustment for small samples, or multiplicity of testing in his analysis. Table 10.4.1 also shows the effect of using the Kenward Roger (KR) adjustment to the tests for each compound.

Compound 1

Conc		AR(1)				ARH(1)				ANTE(1)				Compound Symmetry				Paired t-tests			
mM	Mean	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p
0	122.3																				
0.1	122.3	0.89	1.50	11	0.16	0.61	2.17	11	0.05	0.67	2.00	11	0.07	1.05	1.27	11	0.23	0.67	2.00	2	0.18
0.3	124.0	1.23	1.36	11	0.20	0.48	3.45	11	0.005*	0.43	3.91	11	0.002*	1.05	1.59	11	0.14	0.33	5.00	2	0.04*
1	124.7	1.47	1.59	11	0.14	1.16	2.01	11	0.07	1.15	2.03	11	0.07	1.05	2.22	11	0.05*	1.20	1.94	2	0.19
3	125.0	1.65	1.62	11	0.13	0.99	2.69	11	0.02*	0.85	3.12	11	0.01*	1.05	2.54	11	0.03*	0.88	3.02	2	0.09
10	124.3	1.80	1.11	11	0.29	1.04	1.92	11	0.08	0.92	2.17	11	0.05	1.05	1.91	11	0.08	0.58	3.46	2	0.07
30	121.7	1.93	-0.35	11	0.74	2.61	-0.26	11	0.80	2.37	-0.28	11	0.78	1.05	-0.64	11	0.54	2.19	-0.30	2	0.79
With KR small sample adjustment																					
0	122.3																				
0.1	122.3	1.16	1.15	11.6	0.27	0.64	2.07	2.31	0.16	0.71	1.89	2.00	0.20	1.05	1.27	12	0.23				
0.3	124.0	1.60	1.04	12.6	0.32	0.55	3.03	4.13	0.04*	0.48	3.47	2.16	0.07	1.05	1.59	12	0.14				
1	124.7	1.91	1.22	13.4	0.24	1.24	1.88	2.41	0.18	1.22	1.91	2.09	0.19	1.05	2.22	12	0.05*				
3	125.0	2.16	1.24	13.9	0.24	1.10	2.42	2.91	0.09	0.93	2.86	2.22	0.09	1.05	2.54	12	0.03*				
10	124.3	2.35	0.85	14.0	0.41	1.17	1.71	3.07	0.18	1.03	1.94	2.64	0.16	1.05	1.91	12	0.08				
30	121.7	2.51	-0.27	13.7	0.79	2.74	-0.24	2.20	0.83	2.48	-0.27	2.08	0.81	1.05	-0.64	12	0.54				

Compound 2

Conc		AR(1)				ARMA(1,1)				ARH(1)				Compound Symmetry				Paired t-tests			
mM	Mean	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p	SEDC	t	df	p
0	122.3																				
0.1	122.7	0.54	0.61	11	0.55	0.54	0.61	11	0.55	0.51	0.66	11	0.53	1.11	0.30	11	0.77	0.33	1.00	2	0.42
0.3	123.0	0.77	0.87	11	0.40	0.77	0.87	11	0.40	0.81	0.82	11	0.43	1.11	0.60	11	0.56	0.33	2.00	2	0.18
1	121.7	0.93	-0.71	11	0.49	0.94	-0.71	11	0.49	0.94	-0.71	11	0.49	1.11	-0.60	11	0.56	0.88	-0.76	2	0.53
3	121.7	1.07	-0.62	11	0.55	1.08	-0.62	11	0.55	1.05	-0.64	11	0.54	1.11	-0.60	11	0.56	0.88	-0.76	2	0.53
10	120.0	1.20	-1.95	11	0.07	1.20	-1.95	11	0.08	1.11	-2.10	11	0.06	1.11	-2.11	11	0.06	1.45	-1.61	2	0.25
30	115.0	1.31	-5.62	11	0.0002*	1.31	-5.61	11	0.0002*	1.39	-5.29	11	0.0003*	1.11	-6.63	11	0.0001*	2.03	-3.62	2	0.07
With KR small sample adjustment																					
0	122.3																				
0.1	122.7	0.75	0.44	12	0.66	0.75	0.44	12	0.66	0.60	0.56	5.24	0.60	1.11	0.30	12	0.77				
0.3	123.0	1.06	0.63	12.2	0.54	1.06	0.63	9.57	0.54	0.94	0.71	3.96	0.52	1.11	0.60	12	0.56				
1	121.7	0.29	-0.52	12.4	0.61	1.29	-0.52	8.33	0.62	1.11	-0.60	5.01	0.57	1.11	-0.60	12	0.56				
3	121.7	1.48	-0.45	12.6	0.66	1.48	-0.45	7.78	0.67	1.25	-0.53	5.88	0.61	1.11	-0.60	12	0.56				
10	120.0	1.65	-1.41	12.7	0.18	1.65	-1.41	7.50	0.20	1.35	-1.72	7.09	0.13	1.11	-2.11	12	0.06				
30	115.0	1.80	-4.07	12.9	0.001*	1.81	-4.06	7.36	0.004*	1.64	-4.46	4.12	0.01*	1.11	-6.63	12	0.0001*				

Table 10.4.1: *GPPM data: Comparison of results using different methods of analysis (both compounds).*

In the table results which are significant at the 5% level (or below) are flagged (*).

For compound 1, the non-stationary heterogeneous AR1 and antedependence models give a much higher SEDC for the highest concentration. Looking again at the individual profiles in Table 10.4.1 shows that this is due to the large drop in AP for one of the tissues, a possible outlier. For compound 2, such a drop is noticeable in all three tissues, and this difference from control for this concentration is seen to be significant across all methods other than the paired t-tests. It is also noticeable with compound 1 that few significant results are repeated when the KR small sample adjustment is adopted.

It is of interest to compare Brammer's approach with that offered by the modified Box correction in this extreme small sample setting. Whilst the unstructured covariance model did not converge for either of the two experiments, it is possible to construct the (singular) sample covariance matrix in each case. These are shown in Table 10.4.2 below.

Compound 1

2.33	3.67	3.00	4.17	3.50	3.83	7.17
0.95	**6.33**	5.00	7.83	6.50	6.17	12.83
0.98	0.99	**4.00**	6.00	5.00	5.00	10.00
0.85	0.97	0.93	**10.33**	8.50	7.17	16.33
0.87	0.98	0.94	1.00	**7.00**	6.00	13.50
1.00	0.97	0.99	0.89	0.90	**6.33**	12.17
0.91	0.99	0.97	0.99	0.99	0.94	**26.33**

Compound 2

21.33	22.67	24.00	22.67	22.67	20.00	20.00
0.99	**24.33**	25.50	24.83	24.83	22.50	23.00
1.00	0.99	**27.00**	25.50	25.50	22.50	22.50
0.96	0.98	0.96	**26.33**	26.33	25.00	26.50
0.96	0.98	0.96	1.00	**26.33**	25.00	26.50
0.87	0.91	0.87	0.97	0.97	**25.00**	27.50
0.78	0.84	0.78	0.93	0.93	0.99	**31.00**

Table 10.4.2: *Sample Variance-Correlation Matrix for the GPPM Data: Compounds 1 and 2.*

Whilst these matrices do not allow the construction of the usual Wald tests, which require invertible matrices, they can be used directly in the modified Box correction to allow the tests to reflect the 'observed' dependencies in the data.

Although it is possible to replicate Brammer's tests by suitable parametrization of the design matrix, with columns representing contrasts between each concentration with the control measurement, a more appropriate method is suggested as follows.

1. Use the modified Box correction to test for an overall treatment (concentration) effect.

2. If significant, use Scheffé's method, in conjunction with the adjusted F-statistic, to test individual contrasts.

This approach ensures that the type 1 error rate for individual tests is controlled for multiplicity of testing, as well as to departures from independence in the small sample setting for which the modified Box correction has been shown to be successful for the analysis of repeated measurements.

Scheffé's method (Scheffé (1953)) allows for the comparison of any or all possible contrasts between treatment means, ensuring that the type 1 error rate is at most α for any of the possible comparisons. It takes advantage of the union-intersection test properties of the ANOVA test statistic, by simultaneously considering all possible contrasts in the treatment means:

$$\Gamma_{\mathbf{a}} = a_1\mu_1 + a_2\mu_2 + \ldots + a_r\mu_r \tag{10.4.1}$$

for any $\mathbf{a}=(a_1, \ldots, a_r)$, with $\sum a_i = 0$. The corresponding contrasts in the treatment averages $\bar{y}_{i.}$, are hence

$$C_{\mathbf{a}} = a_1\bar{y}_{1.} + a_2\bar{y}_{2.} + \ldots + a_r\bar{y}_{r.} \tag{10.4.2}$$

and the standard error of this contrast is

$$S_{C_{\mathbf{a}}} = \sqrt{\hat{\sigma}^2 \sum_{i=1}^{r} (a_i^2/n_i)} \tag{10.4.3}$$

where n_i is the number of observations of the ith treatment, and $\hat{\sigma}^2$ is the mean squared error (MSE) of the data.

To use Scheffé's method with the modified Box corrected ANOVA statistic from (8.3.15), we have $\hat{\sigma}^2 = \text{MSE} = \mathbf{y}^T\mathbf{A}\mathbf{y}/(n-r)$, and the critical value to which $C_{\mathbf{a}}$ should be compared is

$$S_{\alpha,\mathbf{a}} = S_{C_{\mathbf{a}}} \sqrt{\frac{c}{\lambda} F_\alpha(c, v_2)} \tag{10.4.4}$$

so that a 100(1-α)% confidence interval for $C_{\mathbf{a}}$ is given by

$$C_{\mathbf{a}} \pm S_{\alpha,\mathbf{a}} = C_{\mathbf{a}} \pm S_{C_{\mathbf{a}}} \sqrt{\frac{c}{\lambda} F_\alpha(c, v_2)} \tag{10.4.5}$$

An advantage of Scheffé's method is that it will always agree with the ANOVA F-test in the sense that if the F-test detects differences, then at least one Scheffé test will detect a difference. Conversely, if the F-test does not detect any differences, then none of Scheffé's tests will. This is illustrated below using Brammer's data.

We begin by considering the overall tests for a concentration effect, given in Table 10.4.3.

The results show that there is only marginal evidence of a significant effect of concentration with compound 1, but the evidence of a significant effect in compound 2 is much stronger.

Turning to the Scheffé tests of individual contrasts (differences from control), we

Effect		Num df	Den df	F	p
concentration (compound 1)	(λ=0.1131)	6	5.1756	4.4977	0.0570
concentration (compound 2)	(λ=0.0429)	6	5.0861	20.8472	0.0020

Table 10.4.3: *GPPM data: Results using the modified Box correction*

have, for compound 1, $\hat{\sigma}^2$=8.9524, so that the standard error for the first, and actually all, the contrasts is given by

$$S_{C_1} = \sqrt{\hat{\sigma}^2 \sum_{i=1}^{7} (a_i^2/n_i)} = \sqrt{8.9524(1+1)/3} = 2.4430$$

and, since the 0.95-quantile of F(6,5.1756) is 4.8063, we find the 95% confidence interval for the first contrast is

$$1.333 \pm 2.4430\sqrt{6 \times 0.1131 \times 4.8063} = 1.333 \pm 4.4120 = (-3.08, 5.75)$$

Confidence intervals for the remaining contrasts are calculated similarly, (for compound 2, $\hat{\sigma}^2 = 25.9048$). The results are shown in Table 10.4.4.

As is expected, since the overall concentration effect was non-significant at the 5% level, none of the 95% confidence intervals for mean difference from control for compound 1 exclude zero. Of the contrasts with compound 2, only the final concentration is significantly different from control, 95% CI (-11.99, 2.67). (In fact this contrast is also significant at the 1% level, 99% CI (-14.14, -0.52)).

Compound 1

Conc	Mean Diff from Control	95% CI
0		
0.1	1.3333	(-3.08,5.74)
0.3	1.6667	(-2.74,6.08)
1	2.3333	(-2.07,6.75)
3	2.6667	(-1.74,7.08)
10	2.0000	(-2.71,6.41)
30	-0.6667	(-5.07,3.74)

Compound 2

Conc	Mean Diff from Control	95% CI
0		
0.1	0.3333	(-4.33,4.99)
0.3	0.6667	(-3.99,5.33)
1	-0.6667	(-5.33,3.99)
3	-0.6667	(-5.33,3.99)
10	2.3333	(-6.99,2.33)
30	-7.3333	(-11.99,-2.67)

Table 10.4.4: *GPPM data: Individual Contrasts using Scheffé's method with the modified Box correction*

Part IV

Conclusions

Chapter 11

Conclusions and Recommendations

11.1 Summary

The aim of this thesis has been to investigate methods for the analysis of repeated measurements which arise from very small samples with continuous, normally distributed responses.

We have seen in Part I, Chapters 1 and 2, that conventional methods for analysing such data, Wald tests using the small sample adjustment of Kenward and Roger (1997), are often inadequate where the sample size is very small. That is, where exact tests such as Hotelling T^2 or split-plot ANOVA are not appropriate, or where the data are unbalanced so that the covariance structure affects the mean parameter estimates as well as their standard errors. Such tests often fail to achieve their nominal properties, having largely inflated type 1 error rates.

Also, the correct choice of an appropriate covariance structure is problematic, since a low parameter model may not always be determined which provides an adequate fit to the data. That is, methods for choosing between possible covariance structures are not reliable for small samples. Often longitudinal or repeated measures data are highly non-stationary, with variances and correlations that change with time. Adoption of an unstructured form will always reflect the observed dependencies, but

this may lack power to detect differences from a general linear hypothesis involving the mean parameters.

Two approaches were suggested in the thesis for dealing with such data, both of which are primarily concerned with the role of the covariance structure in the analysis.

The first approach involved smoothing the covariance matrix between unstructured and structured forms, such as AR1 or compound symmetry. This was the focus of Part II of the thesis, Chapters 3 to 6. This intuitive approach sought to find an appropriate estimator which adequately describes the key features of the data in a low dimensional representation. A number of alternative methods were considered ranging from direct smoothing, a simple weighted average between the unstructured and structured forms, to penalised likelihood methods, which penalise the estimated covariance matrix by its lack of fit to the simpler form. Each of these alternatives was beset by computational difficulties, and development of these techniques was motivated by initially considering only complete and balanced data in order to simplify distributional assumptions. However, even in this restrictive setting, it was found that the benefits of such smoothing approaches were critically dependent on the 'correct' choice of smoothing structure, so that no general solution to the small sample problem was identified.

The second approach involved methods of analysis which are less dependent on the covariance structure. This was the focus of Part III of the thesis, Chapters 7 to 10. Two alternative methods were investigated, the first of which, the empirical sandwich estimator, borrows from the generalised estimating equations (GEE) approach. This suggests dropping the covariance structure from the estimation of the mean parameters, basing them on their ordinary least squares estimates and adjusting their standard errors in the resulting Wald statistic to reflect the observed correlations in the data. Such an approach is known to have poor small sample properties, but several recent attempts to adjust this statistic were considered. One such adjustment, by Pan and Wall (2002) was extended to deal with any general linear hypothesis,

and together with a bias adjustment (Mancl and DeRouen (2001)), was shown to result in a test statistic with nominal properties.

The second method considered in Part III was an adjusted ANOVA statistic, first suggested by Box (1954a,b), which drops the covariance structure from estimates of both the mean parameters and their precision. Instead, it suggests inferences are based on ordinary least squares (OLS) estimates under the assumption of independence, and adjusts the null distribution of the resulting test statistic to account for departures from this assumption. This method results in a type 1 error rate that is conservative, and a modified correction was developed that gives a test statistic with nominal properties.

Having identified two methods in Part III, the adjusted sandwich estimator and modified Box correction, which give adequate control over the type 1 error rates for repeated measures problems with small samples, Chapter 9 compared them extensively through a series of simulation studies. An important consideration was the power offered by these methods, and here the adjusted sandwich estimator was seen to perform poorly in comparison to the methods based on the Box correction. It was also seen to be less numerically robust in small sample settings as the number of time points (repeated measurements) increased relative to the sample size.

The modified Box correction was shown to perform adequately across the range of settings investigated, providing a test statistic which not only achieves nominal test properties (correct 'size'), but is also more powerful than exact tests such as Hotelling T^2 where they are appropriate for the data.

Chapter 10 illustrated the use of the modified Box correction through the practical analysis of three datasets.

11.2 Practical Recommendations

For repeated measures studies involving very small samples, the modified Box correction, (8.3.15) and (8.3.16) - (8.3.18), is recommended as a method of analysis which results in an adjusted test statistic with both nominal properties and acceptable power.

The extensive simulations of Chapter 9 show that where the sample size is small, the modified Box correction is more powerful than exact tests such as Hotelling T^2. As the sample size increases and the underlying covariance structure is far from independence, the advantage over such methods diminishes, but the modified correction remains an effective method for inference.

This method can, by suitable parametrization of the design matrix, be used to test any hypothesis involving fixed effects, based on their OLS estimates under the assumption of independence, and using any consistent estimator of the covariance matrix of the data, such as the REML estimator. It can be easily implemented using statistical software with minimal programming. Also, as shown in the GPPM data example of Section 10.4, it is easily combined with Scheffé's method for simultaneous contrasts to examine questions of interest arising from significant results, providing appropriate control for multiple testing.

11.3 Possible Further Work

A number of further areas for development are suggested in connection with the use of the modified Box correction in small sample analyses.

Firstly, what is an appropriate (consistent) estimator of the covariance structure for the data to be used? Jones and Kenward (2003) suggest the use of an ordinary least squares approach. Whilst this is attractive in principle, and in keeping with the spirit of the approach, it is not easily implemented. Throughout the examples and

simulations of this thesis, the unstructured REML estimator has been adopted. This is easily found, but is not always guaranteed to converge with small sample sizes, as was the case in the GPPM data example of Section 10.4. Here the singular sample covariance matrix was used in the modified Box correction, which does not require the estimator to be invertible, however, such a matrix may not be appropriate in other settings, for example where there are fixed subject effects.

Secondly, although the modified Box correction is suggested as a general solution for very small sample repeated measures problems, the question of what exactly constitutes a very small sample remains largely undefined. In particular, where exactly do the existing small sample adjustments, such as the Kenward Roger adjustment, become untenable so that the modified Box correction is preferred? Also, do any problems arise from the analysis of very small samples which are simply too small to determine anything? Certainly, the GPPM data of Brammer must be approaching such a territory. To answer such questions in a definitive way would require a very extensive series of simulations covering a great number of possible study designs.

Finally, having considered small sample issues in the context of continuous responses, can similar methods be proposed in the generalized linear modelling framework for categorical responses? That is, can such analyses be performed under the assumption of independence, and a generalized Box procedure be developed for this setting. We have seen how the adjusted sandwich estimator of Pan and Wall (2002) can be extended from hypotheses involving single parameters to the general linear hypothesis setting. Also, when combined with a bias adjustment this gives adequate control over the type 1 error rate for normal responses, but is not powerful. Guo *et al.* (2005) allude to this lack of power in turning their attention to the use of the (robust) score statistic for testing, but experience suggests that involving the likelihood is not an answer for very small samples. It may also be of interest to consider the power of tests involving the sandwich estimator more generally, across a range of sample sizes.

Bibliography

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Anderson, T. (1958) *An Introduction to Mutivariate Statistical Analysis.* Wiley.

Barnard, J., McCulloch, R. and Meng, X. (2000) Modelling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica*, **10**, 1281–1312.

Bartlett, M. S. (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **160**, 268–282.

Bellavance, F., Tardif, S. and Stephens, M. A. (1996) Tests for the analysis of variance of crossover designs with correlated errors. *Biometrics*, **52**, 608–612.

Box, G. E. P. (1954a) Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in one-way classification. *The Annals of Mathematical Statistics*, **25**, 290–302.

Box, G. E. P. (1954b) Some theorems on quadratic forms applied in the study of analysis of variance problems, ii. effects of inequality of variance and of correlation between errors in two-way classification. *The Annals of Mathematical Statistics*, **25**, 484–498.

Brammer, R. J. (2003) Modelling covariance structure in ascending dose studies of isolated tissues and organs. *Pharmaceutical Statistics*, **2**, 103–112.

Brown, H. K. and Prescott, R. J. (1999) *Applied Mixed Models in Medicine.* John Wiley and Sons Ltd.

Brown, P. J., Le, N. D. and Zidek, J. V. (1994) Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, **22**, 489–509.

Chen, C. F. (1979) Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society, Series B*, **41**, 235–248.

Crowder, M. J. and Hand, D. J. (1990) *Analysis of Repeated Measures.* Chapman and Hall.

Daniels, M. and Cressie, N. A. C. (2001) A hierarchical approach to covariance function estimation for time series. *Journal of Time Series Analysis*, **22**, 253–266.

Daniels, M. and Kass, R. (1999) Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254–1263.

Daniels, M. and Kass, R. (2001) Shrinkage estimators for covariance matrices. *Biometrics*, **57**, 1173–1184.

Daniels, M. and Pourahmadi, M. (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, **89**, 553–566.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap methods and their application.* Cambridge University Press: Cambridge.

Diggle, P. J. (1988) An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–71.

Diggle, P. J. and Verbyla, A. P. (1998) Nonparametric estimation of covariance structures in longitudinal data. *Biometrics*, **54**, 401–415.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data.* Oxford University Press.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall.

Fay, M. P. and Graubard, B. I. (2001) Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics*, **57**, 1198–1206.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004) *Applied Longitudinal Analysis*. Wiley.

Friedman, J. H. (1989) Regularised discriminant analysis. *Journal of the American Statistical Association*, **84**, 165–175.

Greenhouse, S. W. and Geisser, S. (1958) An extension of Box's results on the use of the F distribution in multivariate analysis. *The Annals of Mathematical Statistics*, **29**, 885–891.

Greenhouse, S. W. and Geisser, S. (1959) On methods in the analysis of profile data. *Psychometrika*, **24**, 95–112.

Gunsolley, J. C., Getchell, C. and Chinchilli, V. M. (1995) Small sample characteristics of generalized estimating equations. *Communications in Statistics - Simulation and Computation*, **24**, 869–878.

Guo, X., Pan, W., Connett, J. E., Hannan, P. J. and French, S. A. (2005) Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine*, **24**, 3479–3495.

Huynh, H. and Feldt, L. S. (1976) Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, **1**, 69–82.

Jones, B. and Kenward, M. G. (2003) *Design and Analysis of Cross-Over Trials*. Chapman And Hall/CRC.

Kackar, A. N. and Harville, D. A. (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of The American Statistical Association*, **79**, 853–862.

Kauermann, G. and Carroll, R. J. (2001) A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, **96**, 1387–1396.

Kenward, M. G. (1987) A method for comparing profiles of repeated measurements. *Applied Statistics*, **36**, 296–308.

Kenward, M. G. (1997) Splines and the smoothing of repeated measurements covariance structures. International Biometric Society, Australasian Region Biennial Meeting, 1997, Wirrina Cove, Australia.

Kenward, M. G. and Roger, J. H. (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997.

Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Lawley, D. N. (1956) A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.

Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lin, S. P. and Perlman, M. D. (1985) *Mulivariate Analysis*, volume 6, chapter 'A Monte Carlo comparison of four estimators for a covariance matrix', pp. 411–429. Amsterdam: North Holland. P. R. Krishnaiah (ed).

Longford, N. T. (1990) *Random Coefficient Models*. Oxford University Press.

Mancl, L. and DeRouen, T. (2001) A covariance estimator for GEE with improved small-sample properties. *Biometrics*, **57**, 26–34.

Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley.

Nelder, J. A. (1965a) The analysis of randomized experiments with orthogonal block structure. i. block structure and the null analysis of variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **283**, 147–162.

Nelder, J. A. (1965b) The analysis of randomized experiments with orthogonal block structure. ii. treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **283**, 163–178.

Pan, W. and Wall, M. M. (2002) Small-sample adjustments in using sandwich variance estimator in generalized estimating equations. *Statistics in Medicine*, **21**, 1429–1441.

Patterson, H. D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression.* Cambridge University Press.

Satterthwaite, F. E. (1941) Synthesis of variance. *Psychometrika*, **6**, 309–316.

Schaalje, G. B., McBride, J. B. and Fellingham, G. W. (2003) Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological and Environmental Services*, **7**, 512–524.

Scheffé, H. (1953) A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Searle, S. R. (1982) *Matrix Algebra Useful for Statistics.* Wiley.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

Stein, C. (1975) Estimation of a covariance matrix. Rietz Lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

Stuart, A. and Ord, J. K. (1994) *Kendall's Advanced Theory of Statistics: Vol 1. Distribution Theory.* Hodder Arnold, 6th edition.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* Springer.

Verbeke, G., Lesaffre, E. and Brant, L. (1998) The detection of serial correlation in linear mixed models. *Statistics in Medicine*, **17**, 1391–1402.

Verbyla, A. P., Cullis, B. R., Kenward, M. G. and Welham, S. J. (1999) The analysis of designed experiments and longitudinal data using smoothing splines (with discussion). *Applied Statistics*, **48**, 269–311.

Wolfinger, R. D. (1996) Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological and Environmental Statistics*, **1**, 205–230.

Zucker, D. M., Lieberman, O. and Manor, O. (2000) Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society, Series B*, **64**, 827–838.

Appendices

Appendix A

Covariance Matrices

A.1 Covariance Matrices I

The following (5×5) symmetric matrices are used as the underlying covariance structures for the data generated in the simulations of the Pilot Study and Designs (A), (B) and (C) used throughout the thesis, as outlined in Chapter 2, Section 2.2.2.

Identity

$$\sigma^2 \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

with $\sigma^2 = 1$.

Figure A.1.1: *Identity Covariance Structure.*

Compound Symmetry

$$\begin{bmatrix} \sigma_1^2+\sigma^2 & & & & \\ \sigma_1^2 & \sigma_1^2+\sigma^2 & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 \end{bmatrix} = \begin{bmatrix} 2 & & & & \\ 1 & 2 & & & \\ 1 & 1 & 2 & & \\ 1 & 1 & 1 & 2 & \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

with $\sigma_1^2 = 1$ and $\sigma^2 = 1$.

Equivalently, the Compound Symmetry form can be considered as the 'Uniform Correlation' model, with $\sigma^2 = 2$ and $\rho = 0.5$,

$$\sigma^2 \begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho & \rho & 1 & & \\ \rho & \rho & \rho & 1 & \\ \rho & \rho & \rho & \rho & 1 \end{bmatrix}$$

Figure A.1.2: *Compound Symmetry Covariance Structure.*

AR1

$$\sigma^2 \begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho^2 & \rho & 1 & & \\ \rho^3 & \rho^2 & \rho & 1 & \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Two forms of AR1 are used, both with $\sigma^2 = 1$, and with $\rho = 0.2$ and $\rho = 0.8$ respectively.

$$\begin{bmatrix} 1 & & & & \\ 0.2 & 1 & & & \\ 0.04 & 0.2 & 1 & & \\ 0.008 & 0.04 & 0.2 & 1 & \\ 0.0016 & 0.008 & 0.04 & 0.2 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & & & & \\ 0.8 & 1 & & & \\ 0.64 & 0.8 & 1 & & \\ 0.512 & 0.64 & 0.8 & 1 & \\ 0.4096 & 0.512 & 0.64 & 0.8 & 1 \end{bmatrix}$$

Figure A.1.3: *AR1 Covariance Structure.*

Antedependence

$$\begin{bmatrix} \sigma_1^2 & & & & \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & & & \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & & \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 & \\ \sigma_1\sigma_5\rho_1\rho_2\rho_3\rho_4 & \sigma_2\sigma_5\rho_2\rho_3\rho_4 & \sigma_3\sigma_5\rho_3\rho_4 & \sigma_4\sigma_5\rho_4 & \sigma_5^2 \end{bmatrix}$$

With $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\sigma_3^2 = 9$, $\sigma_4^2 = 16$, $\sigma_5^2 = 25$, $\rho_1 = 0.8$, $\rho_2 = 0.6$, $\rho_3 = 0.4$, and $\rho_4 = 0.2$, we have

$$\begin{bmatrix} 1 & & & & \\ 1.6 & 4 & & & \\ 1.44 & 3.6 & 9 & & \\ 0.768 & 1.92 & 4.8 & 16 & \\ 0.192 & 0.48 & 1.2 & 4 & 25 \end{bmatrix}$$

Figure A.1.4: *Antedependence Covariance Structure.*

Unstructured

$$\begin{bmatrix} \sigma_1^2 & & & & \\ \sigma_{21} & \sigma_2^2 & & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ 0.28 & 2 & & & \\ 1.04 & 1.96 & 3 & & \\ 0.80 & 0.56 & 2.08 & 4 & \\ 1.79 & 1.89 & 3.10 & 1.79 & 5 \end{bmatrix}$$

This has variance/correlation matrix,(variances on the diagonal, correlations below)

$$\begin{bmatrix} 1 & & & & \\ 0.2 & 2 & & & \\ 0.6 & 0.8 & 3 & & \\ 0.4 & 0.2 & 0.6 & 4 & \\ 0.8 & 0.6 & 0.8 & 0.4 & 5 \end{bmatrix}$$

Figure A.1.5: *Unstructured Covariance Structure.*

Unstructured (QRE)

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}^T + \sigma^2 \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & & & & \\ 3 & 10 & & & \\ 7 & 21 & 50 & & \\ 13 & 39 & 91 & 170 & \\ 21 & 63 & 147 & 273 & 442 \end{bmatrix}$$

i.e. with $\sigma^2 = 1$ and $\sigma_1^2 = \sigma_{21} = \sigma_2^2 = \sigma_{31} = \sigma_{32} = \sigma_3^2 = 1$. This has variance/correlation matrix

$$\begin{bmatrix} 2 & & & & \\ 0.67 & 10 & & & \\ 0.70 & 0.94 & 50 & & \\ 0.71 & 0.95 & 0.99 & 170 & \\ 0.71 & 0.95 & 0.99 & 1.00 & 442 \end{bmatrix}$$

Figure A.1.6: *Unstructured (QRE) Covariance Structure.*

A.2 Covariance Matrices II

The following (5×5) and (10×10) symmetric matrices are used as the underlying covariance structures for the data generated in the simulations of extended designs (A'), (B')and (C') used in the Further simulations of Chapter 9, as outlined in Section 9.3. The (9×9) matrices required for study design (D), are simply extracted from the first nine rows and columns of the (10×10) structures noted above.

Compound Symmetry

$$\begin{bmatrix} \sigma_1^2+\sigma^2 & & & & & & & & & \\ \sigma_1^2 & \sigma_1^2+\sigma^2 & & & & & & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & & & & & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & & & & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \cdots & \sigma_1^2+\sigma^2 & & & \end{bmatrix}$$

With $\sigma_1^2 = 1$ and $\sigma^2 = 1$, we obtain the (5×5) and (10×10) matrices

$$\begin{bmatrix} 2 & & & & \\ 1 & 2 & & & \\ 1 & 1 & 2 & & \\ 1 & 1 & 1 & 2 & \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix} \text{ and, } \begin{bmatrix} 2 & & & & & & & & & \\ 1 & 2 & & & & & & & & \\ 1 & 1 & 2 & & & & & & & \\ 1 & 1 & 1 & 2 & & & & & & \\ 1 & 1 & 1 & 1 & 2 & & & & & \\ 1 & 1 & 1 & 1 & 1 & 2 & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

Figure A.2.1: *Compound Symmetry Covariance Structures.*

AR1

$$\sigma^2 \begin{bmatrix} 1 & & & & & & \\ \rho & 1 & & & & & \\ \rho^2 & \rho & 1 & & & & \\ \rho^3 & \rho^2 & \rho & 1 & & & \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \dots & \rho \end{bmatrix}$$

With $\sigma^2 = 1$ and $\rho = 0.8$, we obtain the (5×5) and (10×10) matrices

$$\begin{bmatrix} 1 & & & & \\ 0.8 & 1 & & & \\ 0.64 & 0.8 & 1 & & \\ 0.512 & 0.64 & 0.8 & 1 & \\ 0.4096 & 0.512 & 0.64 & 0.8 & 1 \end{bmatrix} \text{ and,}$$

$$\begin{bmatrix} 1 & & & & & & & & & \\ 0.8 & 1 & & & & & & & & \\ 0.64 & 0.8 & 1 & & & & & & & \\ 0.512 & 0.64 & 0.8 & 1 & & & & & & \\ 0.4096 & 0.512 & 0.64 & 0.8 & 1 & & & & & \\ 0.3277 & 0.4096 & 0.512 & 0.64 & 0.8 & 1 & & & & \\ 0.2622 & 0.3277 & 0.4096 & 0.512 & 0.64 & 0.8 & 1 & & & \\ 0.2098 & 0.2622 & 0.3277 & 0.4096 & 0.512 & 0.64 & 0.8 & 1 & & \\ 0.1678 & 0.2098 & 0.2622 & 0.3277 & 0.4096 & 0.512 & 0.64 & 0.8 & 1 & \\ 0.1343 & 0.1678 & 0.2098 & 0.2622 & 0.3277 & 0.4096 & 0.512 & 0.64 & 0.8 & 1 \end{bmatrix}$$

Figure A.2.2: *AR1 Covariance Structures.*

Heterogeneous Compound Symmetry

Considering the Compound Symmetry form as the 'Uniform Correlation' model, from Figure A.1.2, we obtain the heterogeneous form

$$\begin{bmatrix} \sigma_1^2 & & & & & & \\ \sigma_2\sigma_1\rho & \sigma_2^2 & & & & & \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & & & & \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 & & & \\ \sigma_5\sigma_1\rho & \sigma_5\sigma_2\rho & \sigma_5\sigma_3\rho & \sigma_5\sigma_4\rho & \sigma_5^2 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\ \sigma_{10}\sigma_1\rho & \sigma_{10}\sigma_2\rho & \sigma_{10}\sigma_3\rho & \sigma_{10}\sigma_4\rho & \sigma_{10}\sigma_5\rho & \dots & \sigma_{10}^2 \end{bmatrix}$$

For the (5×5) structure, we use $\rho = 0.5$, with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, and $\sigma_5^2 = 5$, to obtain

$$\begin{bmatrix} 1 & & & & \\ 0.7071 & 2 & & & \\ 0.8660 & 1.2247 & 3 & & \\ 1.0000 & 1.4142 & 1.7321 & 4 & \\ 1.1180 & 1.5811 & 1.9365 & 2.2361 & 5 \end{bmatrix}$$

For the (10×10) structure, we use $\rho = 0.5$, with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$, $\sigma_6^2 = 6$, $\sigma_7^2 = 7$, $\sigma_8^2 = 8$, $\sigma_9^2 = 9$ and $\sigma_{10}^2 = 10$, to obtain

$$\begin{bmatrix} 1 & & & & & & & & & \\ 0.7071 & 2 & & & & & & & & \\ 0.8660 & 1.2247 & 3 & & & & & & & \\ 1.0000 & 1.4142 & 1.7321 & 4 & & & & & & \\ 1.1180 & 1.5811 & 1.9365 & 2.2361 & 5 & & & & & \\ 1.2247 & 1.7321 & 2.1213 & 2.4495 & 2.7386 & 6 & & & & \\ 1.3229 & 1.8708 & 2.2913 & 2.6458 & 2.9580 & 3.2404 & 7 & & & \\ 1.4142 & 2.0000 & 2.4495 & 2.8284 & 3.1623 & 3.4641 & 3.7417 & 8 & & \\ 1.5000 & 2.1213 & 2.5981 & 3.0000 & 3.3541 & 3.6742 & 3.9686 & 4.2426 & 9 & \\ 1.5811 & 0.2361 & 2.7386 & 3.1623 & 3.5355 & 3.8730 & 4.1833 & 4.4721 & 4.7434 & 10 \end{bmatrix}$$

Figure A.2.3: *Heterogeneous Compound Symmetry Covariance Structures.*

Heterogeneous AR1

$$\begin{bmatrix}
\sigma_1^2 & & & & & & \\
\sigma_2\sigma_1\rho & \sigma_2^2 & & & & & \\
\sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & & & & \\
\sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 & & & \\
\sigma_5\sigma_1\rho^4 & \sigma_5\sigma_2\rho^3 & \sigma_5\sigma_3\rho^2 & \sigma_5\sigma_4\rho & \sigma_5^2 & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
\sigma_{10}\sigma_1\rho^9 & \sigma_{10}\sigma_2\rho^8 & \sigma_{10}\sigma_3\rho^7 & \sigma_{10}\sigma_4\rho^6 & \sigma_{10}\sigma_5\rho^5 & \dots & \sigma_{10}^2
\end{bmatrix}$$

For the (5×5) structure, we use $\rho = 0.8$, with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, and $\sigma_5^2 = 5$, to obtain

$$\begin{bmatrix}
1 & & & & \\
0.1314 & 2 & & & \\
1.1085 & 1.9596 & 3 & & \\
1.0240 & 1.8102 & 2.7713 & 4 & \\
0.9159 & 1.6191 & 2.4787 & 3.5777 & 5
\end{bmatrix}$$

For the (10×10) structure, we use $\rho = 0.8$, with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$, $\sigma_6^2 = 6$, $\sigma_7^2 = 7$, $\sigma_8^2 = 8$, $\sigma_9^2 = 9$ and $\sigma_{10}^2 = 10$, to obtain

$$\begin{bmatrix}
1 & & & & & & & & & \\
0.1314 & 2 & & & & & & & & \\
1.1085 & 1.9596 & 3 & & & & & & & \\
1.0240 & 1.8102 & 2.7713 & 4 & & & & & & \\
0.9159 & 1.6191 & 2.4787 & 3.5777 & 5 & & & & & \\
0.8026 & 1.4189 & 2.1722 & 3.1353 & 4.3818 & 6 & & & & \\
0.6936 & 1.2261 & 1.8770 & 2.7092 & 3.7863 & 5.1846 & 7 & & & \\
0.5932 & 1.0486 & 1.6053 & 2.3170 & 3.2382 & 4.4341 & 5.9867 & 8 & & \\
0.5033 & 0.8897 & 1.3621 & 1.9661 & 2.7477 & 3.7624 & 5.0798 & 6.7882 & 9 & \\
0.4244 & 0.7503 & 1.1487 & 1.6579 & 2.3170 & 3.1727 & 4.2837 & 5.7243 & 7.5895 & 10
\end{bmatrix}$$

Figure A.2.4: *Heterogeneous AR1 Covariance Structures.*

Antedependence

$$\begin{bmatrix} \sigma_1^2 & & & & & & & & & \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & & & & & & & & \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & & & & & & & \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 & & & & & & \\ \sigma_1\sigma_5\rho_1\rho_2\rho_3\rho_4 & \sigma_2\sigma_5\rho_2\rho_3\rho_4 & \sigma_3\sigma_5\rho_3\rho_4 & \sigma_4\sigma_5\rho_4 & \sigma_5^2 & & & & & \\ \vdots & & & & & \ddots & & & & \\ \sigma_1\sigma_{10}\rho 1\rho_2\rho_3\rho_4\rho_5\rho_6\rho_7\rho_8\rho_9 & & & & & & & & & \sigma_{10}^2 \end{bmatrix}$$

With $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$, $\rho_1 = 0.8$, $\rho_2 = 0.6$, $\rho_3 = 0.4$, and $\rho_4 = 0.2$, we have the (5×5) matrix

$$\begin{bmatrix} 1 & & & & \\ 1.1314 & 2 & & & \\ 0.8314 & 1.4697 & 3 & & \\ 0.3840 & 0.6788 & 1.3856 & 4 & \\ 0.0859 & 0.1518 & 0.3098 & 0.8944 & 5 \end{bmatrix}$$

For the (10×10) structure, we use $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 3$, $\sigma_4^2 = 4$, $\sigma_5^2 = 5$, $\sigma_6^2 = 6$, $\sigma_7^2 = 7$, $\sigma_8^2 = 8$, $\sigma_9^2 = 9$ and $\sigma_{10}^2 = 10$, with $\rho_1 = 0.8$, $\rho_2 = 0.725$, $\rho_3 = 0.65$, $\rho_4 = 0.575$, $\rho_5 = 0.5$, $\rho_6 = 0.425$, $\rho_7 = 0.35$, $\rho_8 = 0.275$ and $\rho_9 = 0.2$ to obtain

$$\begin{bmatrix} 1 & & & & & & & & & \\ 0.1314 & 2 & & & & & & & & \\ 1.0046 & 1.7759 & 3 & & & & & & & \\ 0.7540 & 1.3329 & 2.2517 & 4 & & & & & & \\ 0.4847 & 1.8569 & 1.4475 & 2.5715 & 5 & & & & & \\ 0.2655 & 0.4693 & 0.7928 & 1.4085 & 2.7386 & 6 & & & & \\ 0.1219 & 0.2154 & 0.3640 & 0.6466 & 1.2572 & 2.7543 & 7 & & & \\ 0.0456 & 0.0806 & 0.1362 & 0.2419 & 0.4704 & 1.0306 & 2.6912 & 8 & & \\ 0.0133 & 0.0235 & 0.0397 & 0.0706 & 0.1372 & 0.3006 & 0.7640 & 2.3335 & 9 & \\ 0.0028 & 0.0050 & 0.0084 & 0.0149 & 0.0289 & 0.0634 & 0.1611 & 0.4919 & 1.8974 & 10 \end{bmatrix}$$

Figure A.2.5: *Antedependence Covariance Structures.*

Appendix B

Computer Code

B.1 Introduction

Examples are presented of the computer code used to implement the small sample methods investigated in this thesis. In particular, code for carrying out the Box correction methods and adjusted sandwich estimator in *SAS* is given for the reduced Cardiac Enzyme data example of Chapter 10, Section 10.2.

Subsections B.1.1-B.1.3 show how a standard linear mixed model can be fitted using *PROC MIXED*, which provides the REML estimate of the covariance structure of the data and implements the Kenward Roger small sample adjustment. The Box corrections and adjusted sandwich estimator methods require additional programming using *PROC IML*, which is illustrated in Sections B.2 and B.3.

Section B.4, gives an example of the use of cross-validation to provide a smoothed estimate of the covariance structure for the full Cardiac Enzyme data set, using the penalised likelihood method for smoothing towards a compound symmetry form, which was described in Chapter 5, Section 5.1.

We begin by inputting the Cardiac Enzyme data into *SAS* using standard *data steps*. Throughout this Appendix, code is shown as `plain text type` and the corresponding output from *SAS* is shown boxed.

B.1.1 The Cardiac Enzyme Data

The full Cardiac Enzyme data set, introduced in Chapter 4, subsection 4.1.1 is input below.

```
title1 "Cardiac Enzyme Data";
data doghearts;
   input dog a b y1-y9;
   datalines;
1  1 1 74.56 84.25 84.05 78.41 79.93 71.60 69.90 63.52 57.14
2  1 1 72.77 86.93 61.47 88.08 64.15 63.27 57.89 53.02 26.47
3  1 1 66.67 77.57 73.12 47.50 77.69 64.38 48.06 49.52 52.18
4  1 1 80.72 71.99 77.32 87.01 73.25 74.84 75.95 51.44 51.36
5  1 1 82.46 83.16 76.45 79.30 81.38 70.66 59.37 42.16 43.99
6  1 1 60.54 71.94 77.50 76.28 57.84 41.54 36.06 30.77 29.46
7  1 2 79.61 79.58 73.09 78.78 71.90 71.86 46.02 53.02 56.34
8  1 2 71.96 79.00 77.14 66.32 74.65 42.68 51.68 31.37 23.39
9  1 2 85.60 85.48 84.29 70.02 68.99 50.89 33.47 24.03 25.83
10 1 2 86.67 84.90 86.04 82.09 77.78 71.29 53.31 47.79 39.92
11 1 2 82.44 84.74 79.35 78.65 65.97 69.93 53.63 62.20 51.05
12 2 1 82.98 76.48 77.86 66.26 77.31 79.10 68.05 49.03 58.91
13 2 1 81.62 79.81 74.95 77.40 70.38 77.35 74.80 62.88 62.40
14 2 1 82.95 70.79 69.16 74.36 74.63 65.44 61.04 77.10 60.17
15 2 1 85.51 81.30 77.19 70.59 76.01 56.26 65.58 56.38 58.29
16 2 1 85.88 80.80 76.01 80.63 70.94 60.00 71.68 54.66 52.22
17 2 1 87.34 80.99 75.78 82.16 71.50 67.62 72.36 66.90 48.96
18 2 2 80.24 73.75 77.31 68.65 72.23 67.34 70.52 62.34 71.58
19 2 2 79.86 81.61 80.06 79.65 67.04 68.26 64.12 47.32 43.60
20 2 2 80.64 79.41 65.07 81.33 65.11 59.42 54.90 57.12 56.44
21 2 2 84.96 84.59 79.73 66.88 56.38 71.67 68.88 58.93 63.06
22 2 2 79.68 58.59 66.91 52.67 54.94 42.46 22.53 42.52 46.30
23 2 2 86.24 76.26 83.80 67.27 53.84 44.40 59.44 58.74 28.14
;

proc print data=doghearts;
run;
```

Cardiac Enzyme Data

Obs	dog	a	b	y1	y2	y3	y4	y5	y6	y7	y8	y9
1	1	1	1	74.56	84.25	84.05	78.41	79.93	71.60	69.90	63.52	57.14
2	2	1	1	72.77	86.93	61.47	88.08	64.15	63.27	57.89	53.02	26.47
3	3	1	1	66.67	77.57	73.12	47.50	77.69	64.38	48.06	49.52	52.18
4	4	1	1	80.72	71.99	77.32	87.01	73.25	74.84	75.95	51.44	51.36
5	5	1	1	82.46	83.16	76.45	79.30	81.38	70.66	59.37	42.16	43.99
6	6	1	1	60.54	71.94	77.50	76.28	57.84	41.54	36.06	30.77	29.46
7	7	1	2	79.61	79.58	73.09	78.78	71.90	71.86	46.02	53.02	56.34
8	8	1	2	71.96	79.00	77.14	66.32	74.65	42.68	51.68	31.37	23.39
9	9	1	2	85.60	85.48	84.29	70.02	68.99	50.89	33.47	24.03	25.83
10	10	1	2	86.67	84.90	86.04	82.09	77.78	71.29	53.31	47.79	39.92
11	11	1	2	82.44	84.74	79.35	78.65	65.97	69.93	53.63	62.20	51.05
12	12	2	1	82.98	76.48	77.86	66.26	77.31	79.10	68.05	49.03	58.91
13	13	2	1	81.62	79.81	74.95	77.40	70.38	77.35	74.80	62.88	62.40
14	14	2	1	82.95	70.79	69.16	74.36	74.63	65.44	61.04	77.10	60.17
15	15	2	1	85.51	81.30	77.19	70.59	76.01	56.26	65.58	56.38	58.29
16	16	2	1	85.88	80.80	76.01	80.63	70.94	60.00	71.68	54.66	52.22
17	17	2	1	87.34	80.99	75.78	82.16	71.50	67.62	72.36	66.90	48.96
18	18	2	2	80.24	73.75	77.31	68.65	72.23	67.34	70.52	62.34	71.58
19	19	2	2	79.86	81.61	80.06	79.65	67.04	68.26	64.12	47.32	43.60
20	20	2	2	80.64	79.41	65.07	81.33	65.11	59.42	54.90	57.12	56.44
21	21	2	2	84.96	84.59	79.73	66.88	56.38	71.67	68.88	58.93	63.06
22	22	2	2	79.68	58.59	66.91	52.67	54.94	42.46	22.53	42.52	46.30
23	23	2	2	86.24	76.26	83.80	67.27	53.84	44.40	59.44	58.74	28.14

B.1.2 The Cardiac Enzyme Data - Reduced Data set

The following data steps give the reduced Cardiac Enzyme data set (two levels of A within level 1 of B), which was analysed in Chapter 10, Section 10.2.

```
title1 "Cardiac Enzyme Data - Reduced Data";
data reddogs;
set doghearts;
where (b=1);
trt=a;
drop a b;
run;

proc print data=reddogs;
run;
```

Cardiac Enzyme Data - Reduced Data

Obs	dog	y1	y2	y3	y4	y5	y6	y7	y8	y9	trt
1	1	74.56	84.25	84.05	78.41	79.93	71.60	69.90	63.52	57.14	1
2	2	72.77	86.93	61.47	88.08	64.15	63.27	57.89	53.02	26.47	1
3	3	66.67	77.57	73.12	47.50	77.69	64.38	48.06	49.52	52.18	1
4	4	80.72	71.99	77.32	87.01	73.25	74.84	75.95	51.44	51.36	1
5	5	82.46	83.16	76.45	79.30	81.38	70.66	59.37	42.16	43.99	1
6	6	60.54	71.94	77.50	76.28	57.84	41.54	36.06	30.77	29.46	1
7	12	82.98	76.48	77.86	66.26	77.31	79.10	68.05	49.03	58.91	2
8	13	81.62	79.81	74.95	77.40	70.38	77.35	74.80	62.88	62.40	2
9	14	82.95	70.79	69.16	74.36	74.63	65.44	61.04	77.10	60.17	2
10	15	85.51	81.30	77.19	70.59	76.01	56.26	65.58	56.38	58.29	2
11	16	85.88	80.80	76.01	80.63	70.94	60.00	71.68	54.66	52.22	2
12	17	87.34	80.99	75.78	82.16	71.50	67.62	72.36	66.90	48.96	2

The data must now be formatted for analysis using *PROC MIXED* and other methods, which require each row to correspond to an observation, so that both subject (dog) and time need to be identified.

```
data d;
set reddogs;
array response y1-y9;
do time=1 to 9;
    atp=response(time);
    output;
    end;
    drop y1-y9;
run;

proc print data=d (obs=18);
run;
```

Cardiac Enzyme Data - Reduced Data

Obs	dog	trt	time	atp
1	1	1	1	74.56
2	1	1	2	84.25
3	1	1	3	84.05
4	1	1	4	78.41
5	1	1	5	79.93
6	1	1	6	71.60
7	1	1	7	69.90
8	1	1	8	63.52
9	1	1	9	57.14
10	2	1	1	72.77
11	2	1	2	86.93
12	2	1	3	61.47
13	2	1	4	88.08
14	2	1	5	64.15
15	2	1	6	63.27
16	2	1	7	57.89
17	2	1	8	53.02
18	2	1	9	26.47

B.1.3 *PROC MIXED* Analysis

The *PROC MIXED* code using the reduced data set follows. This fits a saturated means model with an unstructured covariance matrix (`type=un`) using a Kenward Roger (`ddfm=kr`) adjusted test for treatment/time interaction given by the `contrast` statement, which is equivalent to a Hotelling T^2 test. The `r` option gives the REML sample covariance matrix shown in Table 10.2.1, which is output for use in the Box correction methods. Replacing `type=un` with `type=simple`, `type=cs` and `type=ar(1)` give the test results using identity, compound symmetry and AR1 structures respectively, which were reported in Table 10.2.2.

```
ods output r=rmat;
proc mixed data=d;
class dog trt time;
model atp=trt*time/noint ddfm=kr s;
repeated time/type=un subject=dog r;
contrast "INT" trt*time 1 -1 0 0 0 0 0 0 0 -1 1 0 0 0 0 0 0 0,
               trt*time 1 0 -1 0 0 0 0 0 0 -1 0 1 0 0 0 0 0 0,
               trt*time 1 0 0 -1 0 0 0 0 0 -1 0 0 1 0 0 0 0 0,
               trt*time 1 0 0 0 -1 0 0 0 0 -1 0 0 0 1 0 0 0 0,
               trt*time 1 0 0 0 0 -1 0 0 0 -1 0 0 0 0 1 0 0 0,
               trt*time 1 0 0 0 0 0 -1 0 0 -1 0 0 0 0 0 1 0 0,
               trt*time 1 0 0 0 0 0 0 -1 0 -1 0 0 0 0 0 0 1 0,
               trt*time 1 0 0 0 0 0 0 0 -1 -1 0 0 0 0 0 0 0 1,;
run;
```

A selection of the output showing the estimated REML covariance structure and the Kenward Roger adjusted Wald test for the treatment/time interaction are included below. The unstructured covariance matrix given by REML is the sample covariance matrix reported in Table 10.2.1.

Cardiac Enzyme Data - Reduced Data

The Mixed Procedure

Estimated R Matrix for dog 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	37.0795	11.2930	4.0387	32.5322	24.7819	37.2189	51.3186	19.0824
2	11.2930	29.2673	-3.5233	12.7951	7.6419	10.0188	18.6594	8.1420
3	4.0387	-3.5233	33.0793	-7.7002	15.4256	6.9128	15.8034	-11.4270
4	32.5322	12.7951	-7.7002	128.08	-27.8605	6.5081	58.8382	19.3796
5	24.7819	7.6419	15.4256	-27.8605	48.8472	46.3273	33.2024	24.4498
6	37.2189	10.0188	6.9128	6.5081	46.3273	114.22	86.4826	44.5883
7	51.3186	18.6594	15.8034	58.8382	33.2024	86.4826	117.38	51.3943
8	19.0824	8.1420	-11.4270	19.3796	24.4498	44.5883	51.3943	111.24
9	15.8863	-7.8437	29.9951	-43.1651	53.0142	61.2687	48.7605	42.1046

Estimated R Matrix for dog 1

Row	Col9
1	15.8863
2	-7.8437
3	29.9951
4	-43.1651
5	53.0142
6	61.2687
7	48.7605
8	42.1046
9	94.2355

Solution for Fixed Effects

Effect	trt	time	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*time	1	1	72.9533	2.4859	10	29.35	<.0001
trt*time	1	2	79.3067	2.2086	10	35.91	<.0001
trt*time	1	3	74.9850	2.3480	10	31.94	<.0001
trt*time	1	4	76.0967	4.6202	10	16.47	<.0001
trt*time	1	5	72.3733	2.8533	10	25.36	<.0001
trt*time	1	6	64.3817	4.3631	10	14.76	<.0001
trt*time	1	7	57.8717	4.4231	10	13.08	<.0001
trt*time	1	8	48.4050	4.3059	10	11.24	<.0001
trt*time	1	9	43.4333	3.9631	10	10.96	<.0001
trt*time	2	1	84.3800	2.4859	10	33.94	<.0001
trt*time	2	2	78.3617	2.2086	10	35.48	<.0001
trt*time	2	3	75.1583	2.3480	10	32.01	<.0001
trt*time	2	4	75.2333	4.6202	10	16.28	<.0001
trt*time	2	5	73.4617	2.8533	10	25.75	<.0001
trt*time	2	6	67.6283	4.3631	10	15.50	<.0001
trt*time	2	7	68.9183	4.4231	10	15.58	<.0001
trt*time	2	8	61.1583	4.3059	10	14.20	<.0001
trt*time	2	9	56.8250	3.9631	10	14.34	<.0001

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
INT	8	3	8.73	0.0509

B.2 Box Correction Methods

PROC IML is used to program the Box correction methods. The data input in subsections B.1.1 and B.2.2, used in the *PROC MIXED* analysis, and the resulting REML estimate of the unstructured covariance matrix must be imported as shown below.

```
title2 "Inference Using Box Corrections";
proc iml;

m=12; /* number of subjects */
p=9;  /* number of time points */

use d var{atp};
read all;

y=atp;

use rmat var{col1 col2 col3 col4 col5 col6 col7 col8 col9};
read all;
rmat=col1||col2||col3||col4||col5||col6||col7||col8||col9;

SIGMA=rmat;
```

The following commands create the design matrices $\mathbf{X}$ and $\mathbf{X}_R$, the 'full' and reduced matrices which differ by the removal of the (interaction) terms to be tested. The columns of $\mathbf{X}$ correspond to a model with an intercept, time and treatment effects and a time by treatment interaction, given by `ic`, `time`, `trt` and `int` respectively. (Corner point constraints are adopted for the time and treatment factors, so that the first level of each is set to zero). The variables `parm` and `c` give the number of mean parameters in the model and the number of parameters in the 'dropped' terms respectively.

```
ic=j(m*p,1,1);

trt1=j(m*p/2,1,0)//j(m*p/2,1,1);
trt2=j(m*p/2,1,1)//j(m*p/2,1,0);
trt=trt2;

time=i(p);
do i=2 to m by 1;
    time=time//i(p);
    end;
time=time[,2:p];

int=j(m*p,(p-1),.);
do i=1 to (p-1);
   int[,i]=time[,i]#trt;
   end;

X=ic||trt||time||int;
Xr=ic||trt||time;

parm=ncol(X);
c=ncol(X)-ncol(Xr);
```

The next section of code is necessary to accommodate any missing values in the response. We define the variables `tot` to be the total number of observations ($m \times p$), including any missing values; `ind` to an indicator variable for whether an observation is missing (`0=missing`, `1=present`); `mobs` to be the number of observations present on each subject; and, `cnt` to be the total number of observations, excluding any missing observations.

```
tot=m*p;
ind=j(tot,1,1);
do i=1 to tot by 1;
    if  y[i]=. then ind[i]=0;
end;
nobs=sum(ind);

mobs=j(m,1,.);
do i=1 to m by 1;
    mobs[i]=sum(ind[(i-1)*p+1:(i*p)]);
end;

cnt=j(m,1,.);
do i=1 to m by 1;
    cnt[i]=sum(mobs[1:i]);
end;
```

These variables allow any missing values to be removed from the response vector and the corresponding rows in the design matrices $\mathbf{X}$ and $\mathbf{X}_R$ to be deleted. We must also create an appropriate block diagonal covariance structure.

```
yrem=y[loc(ind=1)];
Xrem=X[loc(ind=1),];
Xr_rem=Xr[loc(ind=1),];

start block_V(Mat) global(m);
    ans=i(m)@Mat;
    return(ans);
finish block_V;

start block_Vrem(Mat) global(m,p,cnt,mobs,nobs);
    ans=j(nobs,nobs,0);
    do i=1 to m by 1;
        if mobs[i]=p then Mati=Mat;
        else Mati=Mat[1:mobs[i],1:mobs[i]];
        if i=1 then ans[1:cnt[1],1:cnt[1]]=Mati;
        else ans[cnt[i-1]+1:cnt[i],cnt[i-1]+1:cnt[i]]=Mati;
    end;
    return(ans);
finish block_Vrem;

Vrem=block_Vrem(SIGMA);
```

If there are no missing values then we may proceed without the above, replacing `yrem`, `Xrem` and `Xr_rem` with `y`, `X` and `Xr` in the code that follows and using the 'full' block diagonal REML covariance structure `V` given by

```
start block_V(Mat) global(m);
    ans=i(m)@Mat;
    return(ans);
finish block_V;

V=block_Vrem(SIGMA);
```

although there is some advantage in having code that will deal appropriately with missing values where they do occur.

We are now in a position to calculate the unadjusted ANOVA statistic for the chosen interaction term, and to test using the Box and modified Box corrections, given by (8.2.7) with (8.2.8)-(8.2.10) and (8.3.15) with (8.3.16)-(8.3.18) respectively.

```
/* ANOVA F-test */

A=i(nobs)-Xrem*inv(t(Xrem)*Xrem)*t(Xrem);
B=Xrem*inv(t(Xrem)*Xrem)*t(Xrem)-Xr_rem*inv(t(Xr_rem)*Xr_rem)*t(Xr_rem);
F=(nobs-parm)#(t(yrem)*B*yrem)/(c#t(yrem)*A*yrem);

numdf=c;dendf=nobs-parm;

/* Box Correction */

psi=(nobs-parm)#trace(B*Vrem)/(c#trace(A*Vrem));
v1_BOX=((trace(B*Vrem))##2)/trace(B*Vrem*B*Vrem);
v2_BOX=((trace(A*Vrem))##2)/trace(A*Vrem*A*Vrem);
F_BOX=F/psi;

/* Modified Box Correction */

E=trace(B*Vrem)/trace(A*Vrem);
V=(trace(B*Vrem*B*Vrem)/((trace(B*Vrem))##2))
  +(trace(A*Vrem*A*Vrem)/((trace(A*Vrem))##2));
v1_MOD=c;
v2_MOD=(c#(4#V+1)-2)/(c#V-1);
lambda=((nobs-parm)/c)#((v2-2)/v2)#E;

F_MOD=F/lambda;

prob_F=1-cdf("F",F,numdf,dendf);
prob_F_BOX=1-cdf("F",F_BOX,v1_BOX,v2_BOX);
prob_F_MOD=1-cdf("F",F_MOD,v1_MOD,v2_MOD);
```

The results of the calculations performed in*SAS* can be obtained using

```
print / "ANOVA F Statistic";
print F numdf dendf prob_F;

print "Box Correction";
print psi F_BOX v1_BOX v2_BOX prob_F_BOX;

print "Modified Box Correction";
print lambda F_MOD v1_MOD v2_MOD prob_F_MOD;
```

which leads to the following output. The results concerning the modified Box correction were reported in Table 10.2.2.

Cardiac Enzyme Data - Reduced Data
Inference Using Box Corrections

ANOVA F Statistic

F	NUMDF	DENDF	PROB_F
1.4918491	8	90	0.1712727

Box Correction

PSI	F_BOX	V1_BOX	V2_BOX	PROB_F_BOX
0.7209445	2.0692983	3.7191163	32.976026	0.1114655

Modified Box Correction

LAMBDA	F_MOD	V1_MOD	V2_MOD	PROB_F_MOD
0.6772191	2.2029047	8	11.175407	0.1108818

B.3 Adjusted Sandwich Estimator

Continuing in *PROC IML*, we begin by reparametrising the design matrix based on the saturated means model.

```
title2 "Inference Using Adjusted Sandwich Estimator";

r1=j(m/2,1,1)//j(m/2,1,0);
r2=j(m/2,1,0)//j(m/2,1,1);
XX=r1||r2;

X=XX@i(p);

parm=ncol(X);

Xrem=X[loc(ind=1),];
```

We can calculate the ordinary least squares estimates of the mean parameters (one per treatment by time combination), using an identity structure for the working covariance matrix.

```
Wrem=inv(block_Vrem(i(p)));
Bsand=inv(t(Xrem)*Wrem*Xrem)*t(Xrem)*Wrem*yrem;
```

The estimates in the parameter vector `Bsand` match the fixed effects solutions reported earlier from *PROC MIXED*.

We create the bias-adjusted 'sandwich' covariance estimator `covBsand`, based on (7.3.4) and (7.3.5), for the fixed effects parameters as follows. The function `vec()` is defined to construct a vector of elements from an $h \times k$ matrix by stacking the k columns.

```
start vec(M);
    h=ncol(M);
    k=nrow(M);
    ans=j(1,(h*k),.);
    do i=1 to k by 1;
        ans[((i-1)*h+1):(i*h)]=M[i,];
    end;
    return(t(ans));
finish vec;

ressig=j(parm,parm,0);
T=j(parm#parm,parm#parm,0);
Pbar=j(parm#parm,1,0);


do i=1 to m by 1;
    if i=1 then do;
        yremi=yrem[1:mobs[i]];
        Xremi=Xrem[1:mobs[i],];
        Wremi=Wrem[1:mobs[i],1:mobs[i]];
        end;
        else do;
            yremi=yrem[cnt[i-1]+1:cnt[i]];
            Xremi=Xrem[cnt[i-1]+1:cnt[i],];
            Wremi=Wrem[cnt[i-1]+1:cnt[i],cnt[i-1]+1:cnt[i]];
            end;
    Hi=Xremi*inv(t(Xrem)*Wrem*Xrem)*t(Xremi)*Wremi;
    iparm=i(mobs[i]);
    Ei=yremi-Xremi*Bsand;
    Pi=vec(t(t(Xremi)*Wremi*inv(iparm-Hi)*Ei*t(Ei)*inv(iparm-t(Hi))*Wremi*Xremi))
    ressig=ressig+t(Xremi)*Wremi*inv(iparm-Hi)*Ei*t(Ei)*inv(iparm-t(Hi))*Wremi*Xr
    Pbar=Pbar+Pi;
end;

Pbar=Pbar/m;

covM=inv(t(Xrem)*Wrem*Xrem);
CovBsand=covM*ressig*covM;
```

We now calculate the scale parameter v, given by (7.4.7), used in the adjusted Wald statistic. We must also define the matrix $\mathbf{L}$, `LIN`, used to define the treatment-time interaction in the general linear hypothesis based on the mean parameters.

```
LIN={1 -1 0 0 0 0 0 0 0 -1 1 0 0 0 0 0 0 0,
     1 0 -1 0 0 0 0 0 0 -1 0 1 0 0 0 0 0 0,
     1 0 0 -1 0 0 0 0 0 -1 0 0 1 0 0 0 0 0,
     1 0 0 0 -1 0 0 0 0 -1 0 0 0 1 0 0 0 0,
     1 0 0 0 0 -1 0 0 0 -1 0 0 0 0 1 0 0 0,
     1 0 0 0 0 0 -1 0 0 -1 0 0 0 0 0 1 0 0,
     1 0 0 0 0 0 0 -1 0 -1 0 0 0 0 0 0 1 0,
     1 0 0 0 0 0 0 0 -1 -1 0 0 0 0 0 0 0 1};

n_LIN=nrow(LIN);
```

```
do i=1 to m by 1;
    if i=1 then do;
        yremi=yrem[1:mobs[i]];
        Xremi=Xrem[1:mobs[i],];
        Wremi=Wrem[1:mobs[i],1:mobs[i]];
        end;
        else do;
            yremi=yrem[cnt[i-1]+1:cnt[i]];
            Xremi=Xrem[cnt[i-1]+1:cnt[i],];
            Wremi=Wrem[cnt[i-1]+1:cnt[i],cnt[i-1]+1:cnt[i]];
            end;
    Hi=Xremi*inv(t(Xrem)*Wrem*Xrem)*t(Xremi)*Wremi;
    iparm=i(mobs[i]);
    Ei=yremi-Xremi*Bsand;
    Pi=vec(t(t(Xremi)*Wremi*inv(iparm-Hi)*Ei*t(Ei)*inv(iparm-t(Hi))*Wremi*Xremi))
    T=T+(Pi-Pbar)*t(Pi-Pbar);
end;

T=T/(m#(m-1));

COVSAM=(m##2)#((LIN*covM)@(LIN*covM))*T*t((LIN*covM)@(LIN*covM));

EE=j(n_LIN##2,n_LIN##2,0);

do i=1 to n_LIN;
    do j=1 to n_LIN;
        Eij=j(n_LIN,n_LIN,0);Eij[i,j]=1;
        EE[(i-1)#n_LIN+1:i#n_LIN,(j-1)#n_LIN+1:j#n_LIN]=Eij;
    end;
end;

COVWIS=(i(n_LIN##2)+EE)*((LIN*covBsand*t(LIN))@(LIN*covBsand*t(LIN)));

v=trace(COVSAM*COVWIS)/trace(COVSAM*COVSAM);
```

The following code corrects for numerical issues in the estimation, by ensuring that a negative value for the denominator degrees of freedom cannot be obtained. In such cases the scale parameter v is set to the dimension of the test (the number of rows in LIN), which results in a single denominator degree of freedom as described in Chapter 9, p173, and a warning message is given.

```
if (v<7) then do;
    print / "********************WARNING MESSAGE********************";
    print v;
    print "v error    v set to 8";
    print   "*******************************************************";
    v=8;
    end;
```

Finally, the Wald statistic can be calculated and the adjusted test undertaken as described in (7.4.2).

```
Bsand_LIN=LIN*Bsand;
F_SAND=(t(Bsand_LIN)*inv(LIN*covBsand*t(LIN))*Bsand_LIN);

ddf_ADJ=v-n_LIN+1;

F_SAND_ADJ=(ddf_PW/(v#n_LIN))#F_SAND;

prob_SAND_ADJ=1-CDF("F",F_SAND_ADJ,n_LIN,ddf_ADJ);
```

The test results for the reduced Cardiac Enzyme data set are reported below.

```
print Bsand;
print F_SAND_ADJ n_LIN ddf_ADJ prob_SAND_ADJ;
```

```
                Cardiac Enzyme Data - Reduced Data
              Inference Using Adjusted Sandwich Estimator

      **********WARNING MESSAGE**********

                                V

                           5.7220992

                    v error    v set to 8

      ************************************************

          F_SAND_ADJ     N_LIN   DDF_ADJ PROB_SAND_ADJ

           3.0309872         8         1     0.4185099
```

B.4 Penalised Likelihood Smoothing

The following code in *PROC IML* uses the full Cardiac Enzyme dataset defined in subsection B.1.1. The variable `trt` is defined to indicate the four treatment combinations of the components A and B in the preserving liquids, whilst `ms` notes the number of subjects (dogs) in each of the treatment groups.

```
proc iml;

m=23; /* no. of subjects */
p=9;  /* no. of time points */
t=4;  /* no. of treatments */

use doghearts var {dog a b y1 y2 y3 y4 y5 y6 y7 y8 y9};
read all;
yy=y1||y2||y3||y4||y5||y6||y7||y8||y9;

trt=j(6,1,1)//j(5,1,2)//j(6,1,3)//j(6,1,4);

ms={6 5 6 6};
```

We construct the sample covariance matrix for this balanced complete data set as follows. The output matrix **sigmahat** is shown in Table 4.1.1.

```
means=j(t,p,.);
do i=1 to t by 1;
    yyr=yy[loc(trt=i),];
    means[i,]=yyr[:,];
    end;

yy1=j(ms[1],p,.);
do i=1 to p by 1;
    yyr=yy[loc(trt=1),];
    yy1[,i]=yyr[,i]-means[1,i];
end;
yy2=j(ms[2],p,.);
do i=1 to p by 1;
    yyr=yy[loc(trt=2),];
    yy2[,i]=yyr[,i]-means[2,i];
end;
yy3=j(ms[3],p,.);
do i=1 to p by 1;
    yyr=yy[loc(trt=3),];
    yy3[,i]=yyr[,i]-means[3,i];
end;
yy4=j(ms[4],p,.);
do i=1 to p by 1;
    yyr=yy[loc(trt=4),];
    yy4[,i]=yyr[,i]-means[4,i];
end;

sigmahat=(1/(m-4))#(t(yy1)*yy1+t(yy2)*yy2+t(yy3)*yy3+t(yy4)*yy4);
print sigmahat;
```

Cardiac Enzyme Data

SIGMAHAT

	COL1	COL2	COL3	COL4	COL5
ROW1	29.006726	12.262463	10.409955	21.386835	8.7577418
ROW2	12.262463	39.863795	10.250122	27.704943	7.7182342
ROW3	10.409955	10.250122	38.707932	-1.203641	7.9699756
ROW4	21.386835	27.704943	-1.203641	105.1243	-1.954497
ROW5	8.7577418	7.7182342	7.9699756	-1.954497	45.313931
ROW6	27.676798	30.2239	9.8373281	41.595911	39.355078
ROW7	28.617607	45.615449	30.139332	66.27293	37.151005
ROW8	16.083665	15.197651	-5.481736	34.871456	13.65297
ROW9	8.3978712	1.3838009	0.7635205	-2.329594	42.259708

SIGMAHAT

	COL6	COL7	COL8	COL9
ROW1	27.676798	28.617607	16.083665	8.3978712
ROW2	30.2239	45.615449	15.197651	1.3838009
ROW3	9.8373281	30.139332	-5.481736	0.7635205
ROW4	41.595911	66.27293	34.871456	-2.329594
ROW5	39.355078	37.151005	13.65297	42.259708
ROW6	141.35633	99.529375	72.186262	105.61149
ROW7	99.529375	159.69981	73.00254	59.208524
ROW8	72.186262	73.00254	126.22189	79.076215
ROW9	105.61149	59.208524	79.076215	158.00595

The following code creates the penalised likelihood function `penlik()` which is *minus* that given by (5.1.4) and, for a given value of the smoothing value α, is *minimised* using `nlpqn`, a quasi-Newton optimisation technique included in *PROC IML*. This procedure is illustrated for $\alpha = 4$.

```
title2 "Penalised Likelihood Smoothing (Compound Symmetry)";
start cs(M);
    p=nrow(M);
    a=(1/(p#(p-1)))#(trace(M*j(p))-trace(M));
    b=(1/(p#(p-1)))#(p#trace(M)-trace(M*j(p)));
    ans=a#j(p)+b#I(p);
    return (ans);
finish cs;

start penlik(X1) global(sigmahat,alpha);
    ans=(1-alpha)#log(det(sqrsym(X1)))+trace(inv(sqrsym(X1))*sigmahat)
        +alpha#log(det(cs(sqrsym(X1))));
    return(ans);
finish penlik;

alpha=4;
print alpha;

    x=symsqr(sigmahat);
    call nlpqn(rc,xres,"penlik",x);

    sigmapen=sqrsym(xres);
    print sigmapen;
```

Cardiac Enzyme Data
Penalised Likelihood Smoothing (Compound Symmetry)

ALPHA
4

SIGMAPEN

	COL1	COL2	COL3	COL4	COL5
ROW1	67.536641	20.341755	20.327469	22.151662	19.808992
ROW2	20.341755	69.646537	19.974233	22.964736	19.473603
ROW3	20.327469	19.974233	69.697676	17.459412	19.463804
ROW4	22.151662	22.964736	17.459412	82.207395	17.546892
ROW5	19.808992	19.473603	19.463804	17.546892	70.763039
ROW6	23.686856	24.037177	20.16306	25.909484	25.055023
ROW7	23.47902	26.851307	24.605973	30.170389	25.29094
ROW8	21.081442	20.733047	16.950912	24.221455	19.788758
ROW9	19.417517	18.087486	18.441087	17.784237	25.339604

SIGMAPEN

	COL6	COL7	COL8	COL9
ROW1	23.686856	23.47902	21.081442	19.417517
ROW2	24.037177	26.851307	20.733047	18.087486
ROW3	20.16306	24.605973	16.950912	18.441087
ROW4	25.909484	30.170389	24.221455	17.784237
ROW5	25.055023	25.29094	19.788758	25.339604
ROW6	87.624447	35.758485	29.671442	35.947795
ROW7	35.758485	92.115266	30.801187	27.950366
ROW8	29.671442	30.801187	85.306744	30.866428
ROW9	35.947795	27.950366	30.866428	89.203153

The following code allows the appropriate value of α to be determined by the data, using the cross-validation algorithm outlined in Chapter 5, Section 5.1.1. Attention is restricted to values of α from the set $\{0, 0.5, 1, 2, 4, 8, 16, 32, 64\}$.

```
alphaset={0 0.5 1 2 4 8 16 32 64};
nal=ncol(alphaset);

start penlikmod(X1) global(sigmahatrem,alpha);
    ans=(1-alpha)#log(det(sqrsym(X1)))+trace(inv(sqrsym(X1))*sigmahatrem)
        +alpha#log(det(cs(sqrsym(X1))));
    return(ans);
finish penlikmod;

cross=j(m,nal,.);

do j=1 to nal by 1;
    alpha=alphaset[j];
    do i=1 to m by 1;
        q=trt[i];
        trtrem=trt[loc(dog^=i)];
        msrem=j(t,1,.);
        do i1=1 to t by 1;
            msrem[i1]=nrow(yy[loc(trtrem=i1),]);
        end;

        yrem=yy[i,];
        yyrem=yy[loc(dog^=i),];

        meansrem=j(t,p,.);
do i2=1 to t by 1;
    yyr=yyrem[loc(trtrem=i2),];
    meansrem[i2,]=yyr[:,];
end;

yy1=j(msrem[1],p,.);
do i2=1 to p by 1;
    yyr=yyrem[loc(trtrem=1),];
    yy1[,i2]=yyr[,i2]-meansrem[1,i2];
end;
yy2=j(msrem[2],p,.);
do i2=1 to p by 1;
    yyr=yyrem[loc(trtrem=2),];
    yy2[,i2]=yyr[,i2]-meansrem[2,i2];
end;
yy3=j(msrem[3],p,.);
do i2=1 to p by 1;
    yyr=yyrem[loc(trtrem=3),];
    yy3[,i2]=yyr[,i2]-meansrem[3,i2];
end;
yy4=j(msrem[4],p,.);
do i2=1 to p by 1;
    yyr=yyrem[loc(trtrem=4),];
    yy4[,i2]=yyr[,i2]-meansrem[4,i2];
end;
sigmahatrem=(1/(m-4-1))#(t(yy1)*yy1+t(yy2)*yy2+t(yy3)*yy3+t(yy4)*yy4);
```

```
        x=symsqr(sigmahatrem);
        call nlptr(rc,xres,"penlikmod",x);

        sigmapenrem=sqrsym(xres);

        ypred=j(1,(p-1),.);
        do k=1 to (p-1) by 1;
            srem11=sigmapenrem[1:k,1:k];
            srem21=sigmapenrem[(k+1),1:k];
            ypred[k]=means[q,k+1]+(srem21)*inv(srem11)*(yrem[1:k]-t(means[q,1:k])
        end;
        cross[i,j]=sum((yrem[2:p]-t(ypred))##2);
    end;
end;

crossval=j(1,nal,.);
do i=1 to nal by 1;
    crossval[i]=sum(cross[,i])/m;
    end;

plotmat=t(alphaset)||t(crossval);
print plotmat;
```

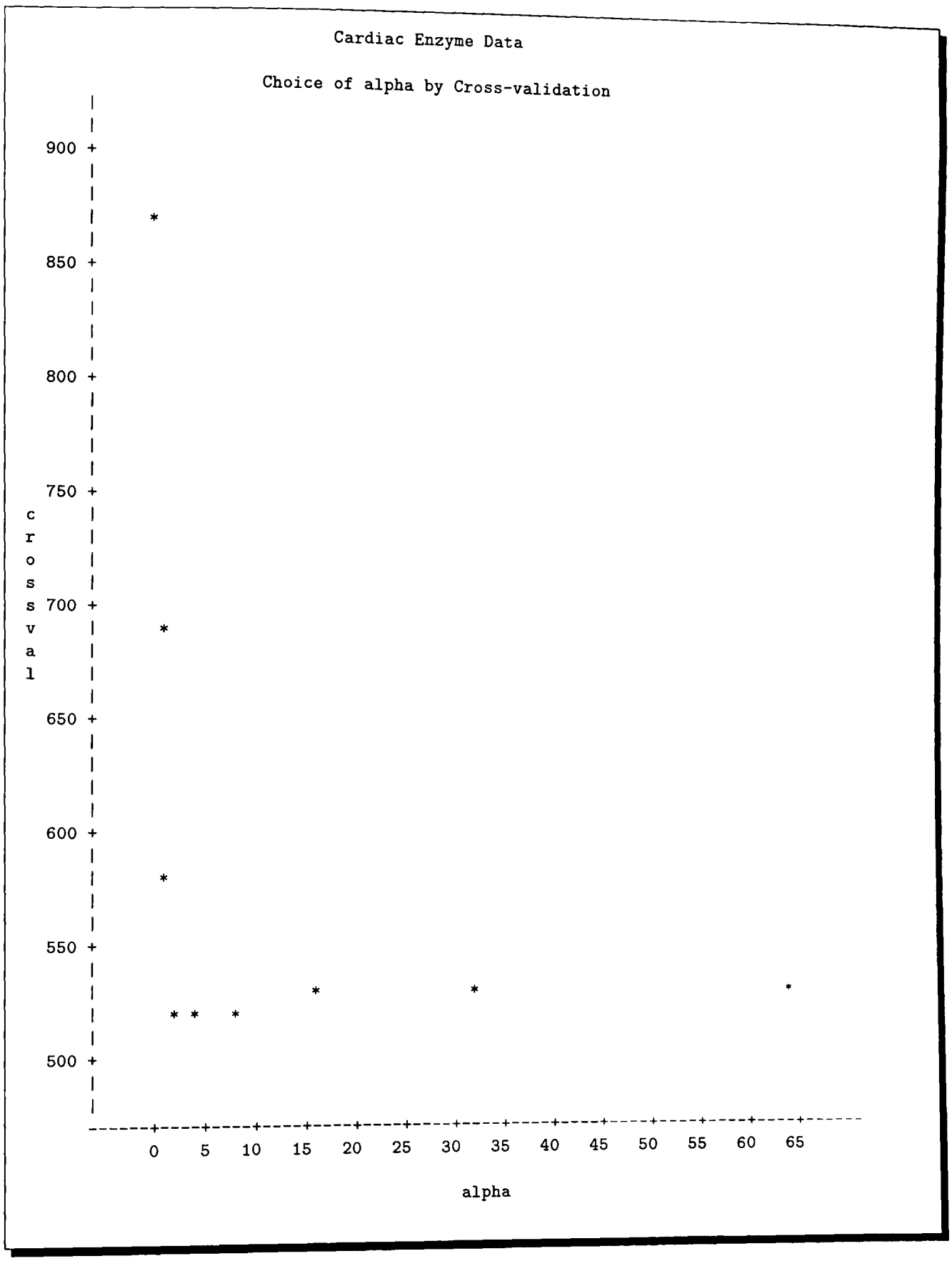
Cardiac Enzyme Data
Penalised Likelihood Smoothing (Compound Symmetry)

PLOTMAT

0	873.22578
0.5	686.13708
1	575.49412
2	522.66917
4	516.61074
8	520.75059
16	525.45114
32	528.75805
64	530.7112

The output shows that the (total) predictive squared error given by the cross-validatory function is minimised for $\alpha = 4$. this is confirmed by the plot below which was included in Figure 5.1.3.

```
call pgraf(plotmat,"*","alpha","crossval","Choice of alpha by Cross-validation")
```


Cardiac Enzyme Data
Choice of alpha by Cross-validation
cross val
900
850
800
750
700
650
600
550
500
0
5
10
15
20
25
30
35
40
45
50
55
60
65
alpha