INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Evaluation of the performance of the generalized estimating equations method for the analysis of crossover designs

Marie-France Valois
Department of Mathematics and Statistics
McGill University, Montreal
May, 1997

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science.

Abstract

Crossover designs are widely used in clinical trials. The main advantage of this type of design is that the treatments are compared within subjects. That is, every subject provides a direct comparison of the treatments he or she has received. In general, a smaller number of subjects is needed to obtain the same precision than with a cross-sectional design. However, because of the correlations within subjects arising from the repeated measurements, the usual analysis of variance based on ordinary least squares (OLS) may be inappropriate to analyze crossover designs. Some approximate likelihood based tests that take into account the structure of the covariance matrix have recently been proposed in the literature.

The aim of this thesis is to compare the performance of the OLS method and two of the approximate likelihood based tests to a non-likelihood based method, the generalized estimating equations, for testing the treatment and carryover effects, in crossover designs, under the assumption of multivariate normality.

Résumé

Les plans croisés sont souvent utilisés dans les essais cliniques. Le principal avantage de ce type de plans est de comparer les effets de traitement avec la variabilité intra-sujet. En d'autres termes, chaque sujet fournit une comparaison directe entre les traitements qu'il reçoit. Donc, en général, un nombre moindre de sujets est nécessaire pour obtenir la même précision qu'avec un plan d'analyse de variance classique. Cependant, à cause des corrélations entre les mesures répétées d'un même sujet, l'analyse de variance basée sur la méthode des moindres carrés ordinaires peut être inappropriée pour l'analyse des plans croisés. Quelques tests approximatifs basés sur la vraisemblance et qui tiennent compte de la structure de la matrice de covariance ont été proposés récemment dans la littérature.

L'objectif de ce mémoire est de comparer la performance de la méthode des moindres carrés ordinaires et de deux tests approximatifs basés sur la méthode du maximum de vraisemblance à une méthode qui n'est pas basée sur la vraisemblance, soit la méthode des équations généralisées d'estimation, pour confronter les hypothèses d'absence d'effets de traitement et d'effets rémanents dans les plans croisés sous le présupposé de multi-normalité.

Acknowledgements

I would like to thank Assistant Professor François Bellavance, whose constant guidance and encouragement made this thesis a pleasure to write. I would also like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial help and Marc De Sales Laterrière for his moral and technical support.

Contents

Chapter 1

Introduction

Nowadays, longitudinal data are frequently found in research studies, especially in epidemiology and in clinical trials. In a cross-sectional study, only one observation of the response variable is taken for each subject. On the other hand, in longitudinal studies, a sequence of repeated measurements of the response variable is observed over time. Longitudinal studies can be viewed as a large number of short time series, one for each subject. An important advantage of longitudinal studies is that it can discriminate changes over time within individuals from between individuals differences (i.e. differences among subjects in their baseline levels). Although, in general, a smaller number of subjects is needed in longitudinal studies to obtain the same precision than in cross-sectional studies, the multiple observations per subject in the former design may involve higher costs. The choice, if the researcher has any, between using a longitudinal design or a cross-sectional design will depend, among other things, on the relative cost of recruiting sub-

jects and the cost of taking repeated measurements on the subjects. Another drawback of longitudinal studies is the special and more complex statistical methods required because of the correlations between the set of observations from each subject.

In longitudinal studies, a subject may receive only a single treatment that is evaluated at different time points (repeated measures studies) or he may receive a sequence of several treatments with the order of presentation of the treatments varying between subjects. The latter type of longitudinal study is commonly called a crossover design. An overview of the analyses of repeated measurements studies can be found in the books by Diggle, Liang and Zeger (1994) and Crowder and Hand (1990). In this thesis we will consider the analysis of a very important particular class of longitudinal studies : crossover designs.

The classical analysis of data from a crossover design, assuming that the vector of observations for a subject follows a multivariate normal distribution, is the analysis of variance based on ordinary least squares (OLS) (Jones and Kenward, 1989). However, for this analysis to be valid, the covariance matrix must have a sphericity structure (Bellavance, 1994). Since this latter assumption is most of the time violated in practice, Bellavance, Tardif and Stephens (1996) proposed and studied the performance of some approximate likelihood based tests for the analysis of crossover designs that take into account the covariance structure. In particular, they considered two methods that use an estimate of the covariance matrix. Another method exists that allows the use of different structures of the covariance matrix in the case where the vector of observations has a multivariate normal distribution; this

is the generalized estimating equations (GEE) method introduced by Liang and Zeger (1986). The method is useful because of the lack of tests available for the case where the distribution of the data is not multivariate normal. Thus, the GEE method can also be used with different distributions like the Poisson, binomial or gamma. However, the purpose of this thesis is to compare the performance of the GEE method, the analysis of variance based on the ordinary least squares and two of the approximate tests studied by Bellavance, Tardif and Stephens (1996) in the case of small and medium sample sizes and for different structures of the covariance matrix with observations coming from a multivariate normal distribution. In Chapter 2, the OLS and two approximate likelihood based tests are presented. In Chapter 3, the GEE method is described. A numerical example is presented in Chapter 4 and the Monte-Carlo simulations and their results are given in Chapter 5. Finally, some conclusions are drawn in Chapter 6.

Chapter 2

Crossover designs and likelihood based methods

2.1 Crossover designs

Two sources of variation in data from experimental designs with repeated measures are the within-subject and between-subjects variations. However, most of the information for treatment comparisons is contained in the within-subject variation. Hence, to achieve sufficient precision from small trials for treatment comparisons, it is desirable, when possible, to reduce or eliminate the between-subject variation and to maximize the information obtained from each subject. This is the main advantage of repeated measurement designs in general and of crossover designs in particular.

In crossover designs, each subject receives a sequence of treatments over different periods of time. Although the main aim of crossover trial is to

compare the effects of two or more treatments, there are some nuisance parameters that need to be considered in the model. Indeed, even if two treatments have identical effects, a large difference between two measurements on a subject might be obtained if, for some reason, the measurements in one treatment period were significantly lower or higher than those in the other treatment period. To avoid confounding period and treatment effects, more than one sequence must be used. Hence, the subjects are randomly assigned to prespecified sequences of treatments, and it is therefore possible to account for the presence of a period effect in the statistical model and obtain unbiased estimates of the treatment effects.

The use of repeated measurements on the same subject brings with it great advantages, but it also brings a potential disadvantage. This disadvantage can be largely overcome if three or more treatment periods are used, and is only serious in the simplest crossover design known as the 2×2 design. The disadvantage to which we refer to is the possibility in drug trials that the effect of a drug given in one period might still be present at the start of the following treatment period. The effect of a treatment that persists after the end of the treatment period is called the carryover effect. In the standard 2 treatment - 2 period crossover design where each subject receives both treatments, which are conventionally labelled as A and B, the test for carryover effects lacks power because it is based on between-subject variation. Moreover, in the presence of a carryover effect, it is not possible to get an unbiased estimate of the treatment effect using the within-subject variation; using a "wash-out" period between the two treatment periods should lessen the chances of a significant carryover effect. However, the use of a "wash-

out" period increases the length of the study. Furthermore, it is often not ethical to include a wash-out period when a standard and effective treatment exists. The use of higher-order designs, designs including either more than two sequences or more than two treatment periods or both, would yield unbiased within-subject estimates of the treatment and carryover effects. For a thorough discussion of the advantages and disadvantages of using crossover designs, see Jones and Kenward (1989).

Here are some examples of different crossover designs (note : the different letters represent different treatments).

Ex. 1 : The standard 2 treatments × 2 periods × 2 sequences crossover design

		Period	
Sequence		1	2
	1	*A*	*B*
	2	*B*	*A*

Ex. 2 : Three possible 2 treatments × 3 periods × 2 sequences crossover designs

		Period		
Seq.		1	2	3
	1	*A*	*A*	*B*
	2	*B*	*B*	*A*

		Period		
Seq.		1	2	3
	1	*A*	*B*	*B*
	2	*B*	*A*	*A*

		Period		
Seq.		1	2	3
	1	A	B	A
	2	B	A	B

Ex. 3 : Two possible 4 treatments × 4 periods × 4 sequences crossover designs

		Period			
Seq.		1	2	3	4
	1	A	D	B	C
	2	B	A	C	D
	3	C	B	D	A
	4	D	C	A	B

		Period			
Seq.		1	2	3	4
	1	A	B	C	D
	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C

Clearly, it is possible to choose among a large number of designs to compare a specific number of treatments. The problem of deciding which design to use to estimate the treatment effects has been considered by a number of researchers (see Jones and Kenward, 1989, for a review). The "optimal" designs chosen provide minimum-variance unbiased estimates of the effects of interest.

2.2 Model

In general, consider a p -period crossover design comparing t -treatments with n subjects. If the responses are continuous and are put in a single np

dimensional vector Y, then the usual linear model for analyzing these data is

$$Y = X\beta + \varepsilon$$

$$= 1_{np}\mu + (I_n \otimes 1_p)\alpha + (1_n \otimes I_p)\pi + X_\tau\tau + X_\lambda\lambda + \varepsilon, \qquad (2.1)$$

where $Y = (y_{11}, y_{12}, \ldots, y_{1p}, y_{21}, y_{22}, \ldots, y_{2p}, \ldots, y_{n1}, y_{n2}, \ldots, y_{np})'$, y_{jk} being the response of subject j in period k, $j = 1, 2, \ldots, n$, $k = 1, 2, \ldots, p$, $X = (1_{np} \mid (I_n \otimes 1_p) \mid (1_n \otimes I_p) \mid X_\tau \mid X_\lambda)$, $\beta' = (\mu\ \alpha'\ \pi'\ \tau'\ \lambda')$, I_m is the identity matrix of order m and 1_m is an m-dimensional vector of ones. The symbol $\otimes$ denotes the Kronecker product. In model (2.1) $\varepsilon = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{np})'$ is a vector of normally distributed errors with zero mean. Also, μ, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$, $\pi = (\pi_1, \pi_2, \ldots, \pi_p)'$, $\tau = (\tau_1, \tau_2, \ldots, \tau_t)'$, and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)'$ represent the overall mean, fixed subject, period, direct treatment and carryover effects respectively. The $np \times t$ matrices X_τ and X_λ are the design matrices associated with τ, and λ respectively. For example, in the case of the crossover design with the two sequences ABB and BAA and two subjects per sequence, the above design matrices are the following :

$$Y = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix} \qquad 1_{12} \cdot \mu = \begin{pmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \end{pmatrix} \qquad I_4 \otimes 1_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}$$

$$1_4 \otimes I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} \quad X_\tau = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \tau = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$$

(Note : here, τ_1 and τ_2 represent the effects of A and B respectively.)

$$X_\lambda = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \end{pmatrix}$$

2.3 Likelihood based methods

2.3.1 The ordinary least squares analysis

The hypotheses of interest are

$$H_{0\tau} : \tau = 0, \text{ i.e. no treatment effects or } \tau_1 = \tau_2 = \ldots = \tau_t = 0$$
$$\text{and } H_{0\lambda} : \lambda = 0, \text{ i.e. no carryover effects or } \lambda_1 = \lambda_2 = \ldots = \lambda_t = 0.$$

In order to present the statistics used to test these two hypotheses, it is useful to first expose some important concepts of linear algebra. Let B be a matrix of dimension $p \times p$:

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pp} \end{bmatrix}.$$

1. The trace of the matrix B is the sum of the diagonal elements of B :

$$tr(B) = \sum_{i=1}^{p} b_{ii}.$$

2. The matrix B is idempotent if $B \cdot B = B^2 = B$.

3. A generalized inverse of a matrix B is defined as any matrix B^- that satisfies the equation $BB^-B = B$.

4. The matrix B is nonnegative definite if $X'BX \geq 0$ for all vector X of dimension p.

5. A covariance matrix Σ of dimension $p \times p$ with a compound symmetry structure has the form

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}.$$

6. A covariance matrix $\Sigma = (\sigma_{kk'})$, $k, k' = 1, 2, \dots, p$ has a sphericity structure if Σ has elements of the form :

$\sigma_{kk'} = \alpha_k + \alpha_{k'} + \lambda\delta_{kk'}$ where α_k are constants, $k = 1, 2, \dots, p$, $\lambda > 0$ and $\delta_{kk'}$ is equal to 1 if $k = k'$ and 0 otherwise.

Note that the compound symmetry structure is a particular case of the sphericity structure.

Now, define the following matrices :

$$X = (X_1 \mid X_\tau \mid X_\lambda),$$

$$M_\tau = (X_1 \mid X_\lambda),$$

$$M_\lambda = (X_1 \mid X_\tau),$$

where

$$X_1 = (1_{np} \mid I_n \otimes 1_p \mid 1_n \otimes I_p).$$

Also, define :

$$E = [I - X(X'X)^- X'],$$

$$A_\tau = [X(X'X)^- X' - M_\tau(M_\tau' M_\tau)^- M_\tau'],$$

$$A_\lambda = [X(X'X)^- X' - M_\lambda(M_\lambda' M_\lambda)^- M_\lambda'],$$

where B^- is a generalized inverse of the matrix B. Note that there exists a multitude of generalized inverse B^- for a matrix B (Searle, 1987). However, the operations performed in the statistical methods presented in this thesis are invariant to the choice of the generalized inverse.

In the ordinary least squares analysis (OLS), the following F-ratio tests are used for testing the treatment effects adjusted to carryover effects (i.e. testing for the presence of treatment effects considering that the carryover effects are already included in the model) and for testing carryover effects adjusted for treatment effects :

$$F_\tau = \frac{r(E)}{r(A_\tau)} \frac{Y'A_\tau Y}{Y'EY} \text{ for } H_{0\tau}$$

and ·

$$F_\lambda = \frac{r(E)}{r(A_\lambda)} \frac{Y'A_\lambda Y}{Y'EY} \text{ for } H_{0\lambda}$$

where $r(B)$ is the rank of B.

For both F -ratios, F_τ and F_λ, in the case where Y has a multivariate normal distribution, the covariance structure of the vector of errors, ϵ, has to have a sphericity structure for the quadratic forms of the numerator and denominator to be independent and χ^2 -distributed (Bellavance, 1994). Under these assumptions, the F -ratios F_τ and F_λ have an exact F -distribution. Hence, we will reject the null hypotheses $H_{0\tau}$ and $H_{0\lambda}$ for large values of F_τ and F_λ respectively. Unfortunately, in practice the assumption of sphericity structure for the covariance matrix is rarely met. This is why Bellavance, Tardif and Stephens (1996) proposed and compared three alternative likelihood based tests that take into account the covariance structure. Two of these tests will be used in this study and will be presented in the next two sections.

2.3.2 Modified F-test approximation (MFA)

Suppose that in model (2.1), the vector ε has a multivariate normal distribution with mean zero and positive definite covariance matrix Σ ($\Sigma > 0$). Thus, $Y \sim N(X\beta, \Sigma)$. In general, the quadratic forms of the numerator and denominator of both F -ratios in section 2.3.1 are dependent. To present the alternative test, the following results on quadratic forms are needed. Let

$Q = Y'DY$ for a matrix D symmetric and nonnegative definite. Therefore

$$Q \sim \sum_{i=1}^{r(D)} \theta_i \chi_1^2 \tag{2.2}$$

where the θ_is are the $r(D)$ nonzero eigenvalues of the matrix $D\Sigma$ and χ_h^2 represents a random variable having a chi-squared distribution with h degrees of freedom. In equation (2.2) the χ_1^2 are independent. In the case where the covariance structure of the vector of errors, ϵ, is spheric, all eigenvalues of $D\Sigma$ are equal, $\theta_i = \theta$, say, $i = 1, 2, \ldots, r(D)$. Therefore

$$Q \sim \theta \sum_{i=1}^{r(D)} \chi_1^2 = \theta \chi_{r(D)}^2,$$

that is, Q has a χ^2 -distribution with r(D) degrees of freedom multiplied by a scalar θ. In general, the expectation and variance of Q are

$$E[Q] = E[\sum_{i=1}^{r(D)} \theta_i \chi_1^2] = \sum_{i=1}^{r(D)} \theta_i E[\chi_1^2] = \sum_{i=1}^{r(D)} \theta_i \cdot 1 = tr(D\Sigma)$$

and

$$Var[Q] = Var[\sum_{i=1}^{r(D)} \theta_i \chi_1^2] = \sum_{i=1}^{r(D)} \theta_i^2 Var[\chi_1^2] = \sum_{i=1}^{r(D)} \theta_i^2 \cdot 2 = 2tr(D\Sigma)^2,$$

where $tr(B)$ is the trace of the matrix B.

Box (1954 a, b) proposed the following approximation for the distribution of Q :

$$Q \approx c\chi_h^2$$

where c and h are such that Q and $c\chi_h^2$ have the same first two moments, that is

$$E(Q) = tr(D\Sigma) = E(c\chi_h^2) = ch$$

$$\text{and } Var(Q) = 2tr(D\Sigma)^2 = Var(c\chi_h^2) = 2c^2h.$$

Therefore,

$$c = \frac{tr(D\Sigma)^2}{tr(D\Sigma)} \text{ and } h = \frac{[tr(D\Sigma)]^2}{tr(D\Sigma)^2}.$$

Now, consider two nonnegative quadratic forms $Q_1 = Y'D_1Y$ and $Q_2 = Y'D_2Y$ approximated by $c_1\chi_{h_1}^2$ and $c_2\chi_{h_2}^2$ respectively. Q_1 and Q_2 need not be independent. A simple approximation to the distribution of the ratio $\frac{r(D_2)Q_1}{r(D_1)Q_2}$ is the distribution of a constant b times an F-distributed random variable with (h_1, h_2) degrees of freedom where

$$b = \frac{r(D_2)c_1h_1}{r(D_1)c_2h_2} = \frac{r(D_2)tr(D_1\Sigma)}{r(D_1)tr(D_2\Sigma)}.$$

Back to the hypotheses testing of $H_{0\tau}$ and $H_{0\lambda}$ in the analysis of crossover trials, the following approximations can be used :

$$F_\tau = \frac{r(E)Y'A_\tau Y}{r(A_\tau)Y'EY} \approx b_\tau F(h_{1\tau}, h_2)$$

where

$$b_\tau = \frac{[r(E)tr(A_\tau\Sigma)]}{[r(A_\tau)tr(E\Sigma)]},$$

$$h_{1\tau} = \frac{[tr(A_\tau\Sigma)]^2}{tr(A_\tau\Sigma)^2},$$

$$h_2 = \frac{[tr(E\Sigma)]^2}{tr(E\Sigma)^2},$$

and

$$F_\lambda = \frac{r(E)Y'A_\lambda Y}{r(A_\lambda)Y'EY} \approx b_\lambda F(h_{1\lambda}, h_2)$$

where

$$b_\lambda = \frac{[r(E)tr(A_\lambda\Sigma)]}{[r(A_\lambda)tr(E\Sigma)]},$$

$$h_{1\lambda} = \frac{[tr(A_\lambda \Sigma)]^2}{tr(A_\lambda \Sigma)^2}.$$

Unfortunately, in practice Σ is unknown and an alternative is to estimate Σ from the data and use this in place of the true value. In section 2.3, an estimate of Σ will be presented.

Note that when Σ has a sphericity structure, the above approximate tests are then exactly equivalent to the OLS method. To prove this, we first note that the matrices A_τ, A_λ and E are all idempotent and furthermore, under the assumption of sphericity for the structure of Σ, the following identities hold (Bellavance, 1994) :

$A_\tau \Sigma A_\tau = \theta A_\tau$, $A_\lambda \Sigma A_\lambda = \theta A_\lambda$, $E\Sigma E = \theta E$, $A_\tau \Sigma E = 0$ and $A_\lambda \Sigma E = 0$.

Thus, the last two identities indicate that the quadratic forms $Y'A_\tau Y$ and $Y'A_\lambda Y$ are both independent of the quadratic form $Y'EY$. Moreover, we have for $D = A_\tau$, A_λ and E

$$\begin{aligned} tr(D\Sigma) &= tr(DD\Sigma) \quad \text{(because } D \text{ is idempotent)} \\ &= tr(D\Sigma D) \\ &= tr(\theta D) \quad \text{(first three identities above)} \\ &= \theta\, tr(D) \\ &= \theta\, r(D) \quad \text{(because D is idempotent),} \end{aligned}$$

and

$$\begin{aligned} tr(D\Sigma)^2 &= tr(D\Sigma D\Sigma) \\ &= tr(\theta D\Sigma) \end{aligned}$$

$$= \theta\, tr(D\Sigma)$$
$$= \theta^2\, r(D).$$

Hence,

$$b_\tau = \frac{r(E)\theta\, r(A_\tau)}{r(A_\tau)\theta\, r(E)} = 1,\ b_\lambda = \frac{r(E)\theta\, r(A_\lambda)}{r(A_\lambda)\theta\, r(E)} = 1, h_{1\tau} = \frac{[\theta\, r(A_\tau)]^2}{\theta^2\, r(A_\tau)} = r(A_\tau),$$

$$h_{1\lambda} = \frac{[\theta\, r(A_\lambda)]^2}{\theta^2\, r(A_\lambda)} = r(A_\lambda),\ \ h_2 = \frac{[\theta\, r(E)]^2}{\theta^2\, r(E)} = r(E).$$

2.3.3 Empirical generalized least squares (EGLS)

The empirical generalized least squares technique has been suggested by Jones and Kenward (1989) for the analysis of crossover designs. Basically, it consists of a transformation of the vector of observations Y and of the design matrix X. Since Σ is positive definite, there exists a non singular matrix K such that $\Sigma = KK'$. The following transformations are performed :

$$Z = K^{-1}Y,\ W = K^{-1}X \text{ and } \eta = K^{-1}\varepsilon.$$

The following transformed model is then obtained

$$Z = W\beta + \eta, \text{ where } \eta \sim N(0, I).$$

Then the OLS method can be conducted on the transformed vector of observations Z and the transformed design matrix W to make inferences about β. Here again, Σ is needed to be able to find K, but in practice it is replaced by an estimate.

2.4 Estimation of Σ

If we assume that the errors on each subject are independent and have the same dispersion matrix V of order p, then

$$\Sigma = I_n \otimes V.$$

The experimental design has s sequences and we suppose that $n_i \geq 2$ subjects per sequence, $i = 1, 2, \ldots, s$. Then an unbiased estimate of V is given by the within-sequence sample dispersion matrix

$$S = \frac{1}{n-s} \sum_{i=1}^{s} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)',$$

where

$$y_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijp})' \text{ and } \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij},\ i = 1, 2, \ldots, s.$$

Chapter 3

Generalized estimating equations (GEE) method

3.1 Introduction

As mentioned in the last chapter, models for longitudinal data with multivariate Gaussian outcomes already exist (Laird and Ware, 1982; Ware, 1985). For the particular case of crossover trials, Jones and Kenward (1989) present the ordinary least squares and the empirical generalized least squares methods. Bellavance, Tardif and Stephens (1996) proposed the modified F-test and a Pearson curve approximation. However, all these methods assume that the observations are multivariate Gaussian. In the case of binary outcomes, likelihood based analysis is possible but computation is difficult (Stiratelli et al., 1984). With other types of outcomes, multivariate distributions for y_{jk} ($k = 1, \ldots, p_j$) similar to the multi-normal distribution are not available.

Therefore, likelihood based method cannot be used. On the other hand, both discrete and continuous responses can be modeled with the generalized estimating equations (GEE) approach (Liang and Zeger, 1986). Data coming from Poisson, binomial or gamma distributions are some examples that can easily be analyzed with the GEE method. This method also takes into account the time dependance structure of the data with the introduction of a working correlation matrix which will be presented in section 3.4. Another interesting point is that different subjects can have different numbers of repeated measurements and these measurements do not need to be taken at the same time intervals for all subjects. As for the OLS and EGLS methods, GEE provides estimates of the coefficients β, which is not the case for the MFA. This is one advantage of the GEE approach over the modified F-test approximation.

The GEE method estimates model parameters by iteratively solving a system of equations based on quasi-likelihood distributional assumptions (McCullagh and Nelder, 1983). Because it is based on quasi-likelihood, this method does not need a complete specification of the joint distribution of the responses but only a pre-determined form of the marginal distribution and the form of the expectation and variance. The GEE method gives consistent estimators of the regression parameters and of their variances under weak assumptions.

The remainder of this chapter presents the details and properties of the generalized estimating equations approach and its application to crossover designs.

3.2 Data layout

3.2.1 General layout for longitudinal studies

Repeated measurements (observations) are taken for each subject in longitudinal studies; the structure of each observation contains a subject identifier (subject id), an observation number, a time identifier, a response and some covariates. The time identifier variable can take different forms: time points equally spaced or not, identical for all subjects or not, and same number of time points for each subject or not. Here are three different examples :

Ex. 1. Same number of observations for all 3 subjects at the same time points (equally spaced) :

		Time points		
Subject	1	1	2	3
	2	1	2	3
	3	1	2	3

Ex. 2. Different time points not equally spaced between subjects but same number of observations for all 3 subjects :

		Time points			
Subject	1	1	2	3	4
	2	1	3	5	10
	3	1	6	9	10

Ex. 3. Different time points between subjects and different number of observations for the 3 subjects :

		Time points						
Subject	1	1	2	4	5	6	8	9
	2	1	4	5	10			
	3	1	3	5	7	9		

In this example, only subject 3 has equally spaced time points.

The notation for the p_j time points of subject j is the vector $(t_{j1}, t_{j2}, \ldots, t_{jp_j})'$, $j = 1, 2, \ldots, n$.

The response variable can be continuous or discrete. For example, the response could be the level of blood pressure (continuous), the presence or absence of a disease (1 or 0), or the number of asthma attacks over the last time period (counts). The responses for subject j, Y_j, are written as the vector $(y_{j1}, y_{j2}, \ldots, y_{jp_j})'$, $j = 1, 2, \ldots, n$. The mean will be addressed as $E(Y_j) = \mu_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jp_j})'$. Finally, the covariates (predictors) can also be continuous or discrete and can vary with time or not. A time varying covariate can take different values over the time line; for example, the proportion of carbon monoxyde (CO) in the air at a certain location varies in time. On the other hand, in all but extraordinary cases, the sex would not vary with time. The predictors of μ_{jk} form the vector of covariates $X_{jk} = (x_{jk1}, x_{jk2}, \ldots, x_{jkm})'$, $k = 1, 2, \ldots, p_j$ and $j = 1, 2, \ldots, n$. The following layout represents the general structure of the data for longitudinal studies.

subject	*observation #*	*time*	*response*	*covariates*		
(j)	(k)	t_{jk}	y_{jk}	x_{jk1}	$\cdots$	x_{jkm}
1	1	t_{11}	y_{11}	x_{111}	$\cdots$	x_{11m}
1	2	t_{12}	y_{12}	x_{121}	$\cdots$	x_{12m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
1	p_1	t_{1p_1}	y_{1p_1}	x_{1p_11}	$\cdots$	x_{1p_1m}
2	1	t_{21}	y_{21}	x_{211}	$\cdots$	x_{21m}
2	2	t_{22}	y_{22}	x_{221}	$\cdots$	x_{22m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
2	p_2	t_{2p_2}	y_{2p_2}	x_{2p_21}	$\cdots$	x_{2p_2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
j	1	t_{j1}	y_{j1}	x_{j11}	$\cdots$	x_{j1m}
j	2	t_{j2}	y_{j2}	x_{j21}	$\cdots$	x_{j2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
j	p_j	t_{jp_j}	y_{jp_j}	x_{jp_j1}	$\cdots$	x_{jp_jm}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	1	t_{n1}	y_{n1}	x_{n11}	$\cdots$	x_{n1m}
n	2	t_{n2}	y_{n2}	x_{n21}	$\cdots$	x_{n2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	p_n	t_{np_n}	y_{np_n}	x_{np_n1}	$\cdots$	x_{np_nm}

3.2.2 Particular layout for crossover studies

In crossover designs, each subject j ($j = 1, \ldots, n$) has a vector of p responses $Y_j' = (y_{j1}, y_{j2}, \ldots, y_{jp})$ and a design matrix X_j of $m = [1 + p + t + t]$ covariates that represents the intercept, p -period, t -treatment and t -carryover effects :

$$X_j = \begin{pmatrix} 1 & x_{j11} & x_{j12} & \ldots & x_{j1m} \\ 1 & x_{j21} & x_{j22} & \ldots & x_{j2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{jp1} & x_{jp2} & \ldots & x_{jpm} \end{pmatrix}.$$

This means that at time t_{jk}, for subject j at period k, the following vector is observed,

$$(y_{jk},\ x_{jk1},\ x_{jk2},\ \ldots,\ x_{jkm})\,.$$

In this vector, y_{jk} is the response, the p covariates $x_{jk1}, x_{jk2}, \ldots, x_{jkp}$ represent the p period effects $(\pi_1, \pi_2, \ldots, \pi_p)$, the t covariates $x_{jk(p+1)}, x_{jk(p+2)}, \ldots,$ $x_{jk(p+t)}$ represent the t treatment effects $(\tau_1, \tau_2, \ldots, \tau_t)$ and the last t covariates $x_{jk(p+t+1)}, x_{jk(p+t+2)}, \ldots, x_{jkm}$ represent the t carryover effects $(\lambda_1, \lambda_2, \ldots, \lambda_t)$. All these covariates will take the value 0 or 1.

Hence we have a design matrix for analysis of variance (ANOVA) models and we can then add the usual constraints on the model parameters, without lost of generality, i.e.

$$\textstyle\sum_{k=1}^{p} \pi_k = 0,\ \ \sum_{l=1}^{t} \tau_l = 0 \text{ and } \sum_{l=1}^{t} \lambda_l = 0.$$

Due to these constraints, the p periods can be represented by $(p-1)$ covariates and the t treatments and t carryovers by $(t-1)$ covariates respectively. Also,

the values taken by the covariates will be such that the sum over periods (treatments or carryovers) equals zero. For example, let $p = 5$; then the covariates $x_{jk1}, x_{jk2}, x_{jk3}$ and x_{jk4} could take these values for the 5 different periods :

		y_{jk}	x_{jk1}	x_{jk2}	x_{jk3}	x_{jk4}
Period	1	y_{j1}	1	0	0	0
	2	y_{j2}	0	1	0	0
	3	y_{j3}	0	0	1	0
	4	y_{j4}	0	0	0	1
	5	y_{j5}	-1	-1	-1	-1

The transformed design matrix obtained with these new $(p-1)+(t-1)+(t-1)$ covariates is of full rank.

Let's consider the example of the crossover design with three treatments (A, B and C), three periods (1, 2 and 3), six sequences (ABC, ACB, BAC, BCA, CAB and CBA) and two subjects per sequence. Let subjects 1 and 2 receive the treatment sequence ABC, subjects 3 and 4 the sequence ACB, subjects 5 and 6 the sequence BAC, subjects 7 and 8 the sequence BCA, subjects 9 and 10 the sequence CAB and subjects 11 and 12 the sequence CBA. The matrices of covariates for the twelve subjects are the following, where the first column represents the intercept, the second and third the period, the fourth and fifth the treatment and the last two the carryover effects :

$$X_1 = X_2 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 \end{pmatrix}$$

$$X_3 = X_4 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 & -1 & -1 \end{pmatrix}$$

$$X_5 = X_6 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 0 \end{pmatrix}$$

$$X_7 = X_8 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & -1 & -1 \end{pmatrix}$$

$$X_9 = X_{10} = \begin{pmatrix} 1 & 1 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & -1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$X_{11} = X_{12} = \begin{pmatrix} 1 & 1 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Although not needed in crossover designs, the GEE method allows the observations to be taken at different times for different subjects, and a different number of observations taken for different subjects. If the notation t_k $(k = 1, \ldots, p)$ is used, the same number of observations will be taken for each subject, at the same time points, as in the type of crossover design looked at in this thesis.

3.3 Details and properties of the generalized estimating equations method

The generalized estimating equations method is an extension of generalized linear models to the analysis of longitudinal data. The latter are themselves an extension of classical linear models.

In classical linear models with a single observation for each subject (i.e. $p_j = 1$) we have $E(Y_j) = \mu_j$ where $\mu_j = X_j'\beta$ and $\beta' = (\beta_1, \beta_2, \ldots, \beta_m)$ is the vector of unknown parameters that we want to estimate from the data. Assuming that the responses Y_j, $j = 1, 2, \ldots, n$, are independent Normal variables with constant variance σ^2, the density function is

$$f(Y_j, \mu_j, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp^{-\frac{1}{2}(\frac{Y_j - \mu_j}{\sigma})^2}$$

and the log-likelihood function is given by

$$\log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \beta) = -\frac{n}{2} \log (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (Y_j - \mu_j)^2.$$

The maximum likelihood estimator of β is the solution of the score equations :

$$\frac{\partial \log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[-\frac{n}{2} \log (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (Y_j - \mu_j)^2 \right] = 0$$

$$\Leftrightarrow -\frac{1}{2\sigma^2} \cdot 2 \sum_{j=1}^{n} \frac{\partial \mu_j}{\partial \beta} (Y_j - \mu_j) = 0$$

$$\Leftrightarrow \sum_{j=1}^{n} \frac{\partial \mu_j}{\partial \beta} \frac{(Y_j - \mu_j)}{\sigma^2} = 0.$$

For the introduction of the generalized linear models, modifications to the specification of the model need to be presented :

1. The responses Y_j, $j = 1, 2, \ldots, n$, are independent and have a probability density function in the exponential family, taking the form

$$f(Y_j, \theta_j, \phi) = \exp \{ \frac{(Y_j \theta_j - b(\theta_j))}{a(\phi)} + c(Y_j, \phi) \}$$

for some specific functions a, b and c, and where ϕ is the dispersion parameter. When ϕ is known, this is an exponential-family model with canonical parameter θ. The expectation and variance of Y_j are given by

$$E(Y_j) = \mu_j = b'(\theta_j) \text{ and } Var(Y_j) = b''(\theta_j)\, a(\phi).$$

Thus, the variance of Y_j is the product of the function $b''(\theta_j)$ which depends on the mean μ_j only, and of $a(\phi)$ independent of θ_j. The variance function $b''(\theta_j)$ considered as a function of μ_j is referred to as $V(\mu_j)$.

2. The linear combination of the β's is equal to some function of the expected value μ_j of Y_j, that is

$$g(\mu_j) = X_j'\beta, \quad j = 1, 2, \ldots, n$$

where g is a monotone, differentiable function called the link function. In this generalized model formulation, classical linear models have a normal distribution and the identity function for the link function.

The most important distributions used with generalized linear models are presented with their canonical link and variance functions in table 3.1.

Table 3.1 Link and Variance functions

Distribution	Notation	$a(\phi)$	$g(\mu)$	$V(\mu)$
Normal	$N(\mu, \sigma^2)$	σ^2	1	1
Poisson	$P(\mu)$	1	$\log(\mu)$	μ
Binomial	$Bin(1, \mu)$	1	$\text{logit}(\mu) = \log(\frac{\mu}{1-\mu})$	$\mu(1-\mu)$
Gamma	$G(\mu, \nu)$	$\nu - 1$	$\frac{1}{\mu}$	μ^2

For the exponential family, and thus for the generalized linear models, the likelihood is

$$\mathcal{L}(Y_1, Y_2, \ldots, Y_n, \theta_j, \phi) = \exp\{\sum_{j=1}^{n} [\frac{(Y_j\theta_j - b(\theta_j))}{a(\phi)} + c(Y_j, \phi)]\}$$

$$= \exp\{\textstyle\sum_{j=1}^{n} l_j\}, \text{ say,}$$

and the log-likelihood is

$$\log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \theta_j, \phi) = \sum_{j=1}^{n} l_j.$$

Using the chain rule, the score equations are

$$\frac{\partial \log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \theta_j, \phi)}{\partial \beta} = \sum_{j=1}^{n} [\frac{\partial l_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \mu_j} \frac{\partial \mu_j}{\partial \beta}] = 0.$$

We have,

$$\frac{\partial l_j}{\partial \theta_j} = \frac{Y_j - b'(\theta_j)}{a(\phi)} = \frac{Y_j - \mu_j}{a(\phi)}$$

and

$$\frac{\partial \mu_j}{\partial \theta_j} = b''(\theta_j) = V(\mu_j).$$

Therefore, the score equations reduce to

$$\sum_{j=1}^{n} \frac{\partial \mu_j}{\partial \beta} \cdot \frac{(Y_j - \mu_j)}{a(\phi) \cdot V(\mu_j)} = 0.$$

When ϕ is a known constant, the score equations can be written as

$$\sum_{j=1}^{n} \frac{\partial \mu_j}{\partial \beta} \frac{(Y_j - \mu_j)}{V(\mu_j)} = 0,$$

and the maximum likelihood estimator of β is the solution of these score equations. It is important to mention that the score equations obtained in the case of the classical linear models have the same form, where $a(\phi) = \sigma^2$ and $V(\mu_j) = 1$.

For both classical linear models and generalized linear models, the form of the distribution function of the Y_j's is known. In practice it may be unknown, but in most cases some characteristic features of the data will be : how the mean response, μ, is affected by external stimuli or treatments (covariates X_j); how the variability of the response changes with the average response; whether the observations are statistically independent; etc. With this information we have an idea of the form of the distribution even though

a complete specification of the distribution is not possible. This is the underlying principle of the quasi-likelihood theory, where the relationship between the mean μ_j and the covariates is

$$g(\mu_j) = X_j'\beta, \quad j = 1, 2, \ldots, n$$

with g being the link function, and the variance is assumed to be a known function V of the mean, that is

$$Var(Y_j) = a(\phi)V(\mu_j)$$

where $\phi > 0$ is a dispersion parameter. Where the classical linear models and the generalized linear models require a complete specification of the distribution of the response variable to find the likelihood function, here only the form of the mean and variance are needed to find the quasi-likelihood function. The quasi-likelihood estimator of β is the solution of the score like equations

$$\sum_{j=1}^{n} \frac{\partial \mu_j}{\partial \beta} \frac{(Y_j - \mu_j)}{a(\phi)V(\mu_j)} = 0. \tag{3.1}$$

Here again, we see that the equations in (3.1) have the same form than the score equations for the classical and generalized linear models. For more details about the quasi-likelihood theory in the regression context, see Wedderburn (1974) and McCullagh (1983).

The generalized estimating equations can be thought of as an extension of quasi-likelihood theory to the case where there is more than one observation per subject (i.e. $p_j > 1$). Hence we have, for subject j, the vector of observations $Y_j = (y_{j1}, y_{j2}, \ldots, y_{jp_j})'$ and its expectation and variance are

given by $E(Y_j) = \mu_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jp_j})$ and $Var(Y_j) = a(\phi)\Sigma_j$. Also, we suppose the following relationships :

$$g(\mu_{jk}) = x'_{jk}\beta \text{ and } Var(y_{jk}) = a(\phi)V(\mu_{jk}),$$

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, p_j$. The variance-covariance matrix Σ_j can be rewritten in the following form :

$$\Sigma_j = A_j^{\frac{1}{2}} R_j(\alpha) A_j^{\frac{1}{2}} \tag{3.2}$$

where $R_j(\alpha)$ is the correlation matrix for Y_j and A_j is the $(p_j \times p_j)$ diagonal matrix whose k^{th} element on the diagonal is $V(\mu_{jk})$. The $(s \times 1)$ vector α fully characterizes the structure of $R_j(\alpha)$. The dimension of the correlation matrix $R_j(\alpha)$ may vary from one subject to another depending on the number of repeated observations, but the structure is the same for all subjects. Since the nuisance parameters $\alpha_1, \alpha_2, \ldots, \alpha_s$ are usually unknown and need to be estimated, we refer to $R_j(\alpha)$ as a "working" correlation matrix. Also the name "working" correlation matrix for $R_j(\alpha)$ is used since the structure is not expected to be correctly specified. The extension of the equation system (3.1) to the repeated measurements case is

$$\sum_{j=1}^{n} \left(\frac{\partial \mu'_j}{\partial \beta}\right) [\, a(\phi)\Sigma_j \,]^{-1} (Y_j - \mu_j) = 0 \tag{3.3}$$

where

$$\frac{\partial \mu'_j}{\partial \beta} = \left(\left(\frac{\partial \mu_{jk}}{\partial \beta_l}\right)\right),\ j = 1,\ 2,\ \ldots,\ n,\ \ k = 1,\ 2,\ \ldots,\ p_j,\ \ l = 1,\ 2,\ \ldots,\ m$$

and

$$(Y_j - \mu_j) = (Y_{j1} - \mu_{j1},\ Y_{j2} - \mu_{j2},\ \ldots,\ Y_{jp_j} - \mu_{jp_j})'.$$

Σ_j could be referred to as the "working" covariance matrix. It is interesting to note that GEE reduce to score equations and maximum likelihood estimates for β when the responses are multivariate Gaussian.

Derivation

Suppose n independent multivariate responses $Y_1, Y_2, \ldots, Y_n$. Suppose also that $Y_j = (y_{j1}, y_{j2}, \ldots, y_{jp_j})$ has density functions $N(\mu_j, \tilde{\Sigma}_j)$ where $\mu_j = X_j'\beta$, β is unknown and X_j has the form presented in section 3.2.2. The multivariate normal density of Y_j is

$$f(Y_j, \mu_j, \tilde{\Sigma}_j) = (2\pi)^{-\frac{1}{2}p_j} |\tilde{\Sigma}_j|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(Y_j - \mu_j)'\tilde{\Sigma}_j^{-1}(Y_j - \mu_j)\}.$$

The likelihood of the Y_j's, $j = 1, 2, \ldots, n$ is

$$\mathcal{L}(\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_\mathrm{n}, \beta) = (2\pi)^{-\frac{1}{2}\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{p_j}} \prod_{\mathrm{j}=1}^{\mathrm{n}} [|\tilde{\Sigma}_\mathrm{j}|^{-\frac{1}{2}}] \exp\{-\frac{1}{2}\sum_{\mathrm{j}=1}^{\mathrm{n}} (\mathrm{Y_j} - \mu_\mathrm{j})'\tilde{\Sigma}_\mathrm{j}^{-1}(\mathrm{Y_j} - \mu_\mathrm{j})\}$$

and the log-likelihood is

$$\log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \beta) = -\frac{1}{2}\sum_{j=1}^{n} p_j \log(2\pi) - \frac{1}{2}\sum_{j=1}^{n} \log|\tilde{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{n} (Y_j - \mu_j)'\tilde{\Sigma}_j^{-1}(Y_j - \mu_j).$$

The score estimating equations arise by differentiating the log-likelihood with respect to β. By the chain rule we have

$$\frac{\partial \log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \beta)}{\partial \beta} = \frac{\partial \log \mathcal{L}(Y_1, Y_2, \ldots, Y_n, \beta)}{\partial \mu_j} \cdot \frac{\partial \mu_j}{\partial \beta} = 0$$

$$\Leftrightarrow \sum_{j=1}^{n} (\frac{\partial \mu_j'}{\partial \beta})\tilde{\Sigma}_j^{-1}(Y_j - \mu_j) = 0.$$

We can see that if we put $\tilde{\Sigma}_j = a(\phi)\Sigma_j$ in equation (3.3), the generalized estimating equations are identical to the score estimating equations in the multivariate Gaussian case.

Going back to the generalized estimating equations (3.3), to find the solution for the parameters estimates $\hat{\beta}$, an iteration between a modified Fisher scoring for β and a moment estimation of α and ϕ was proposed by Liang and Zeger (1986). The iterative procedure for the computation of $\hat{\beta}$ given current estimates of $\hat{\alpha}$ and $\hat{\phi}$ of the nuisance parameters is

$$\hat{\beta}_{i+1} = \hat{\beta}_i - \{\sum_{j=1}^{n} D_j'(\hat{\beta}_i)\, \tilde{V}_j^{-1}(\hat{\beta}_i)\, D_j(\hat{\beta}_i)\}^{-1}\{\sum_{j=1}^{n} D_j'(\hat{\beta}_i)\, \tilde{V}_j^{-1}(\hat{\beta}_i)\, S_j(\hat{\beta}_i)\} \quad (3.4)$$

where $D_j(\hat{\beta}_i) = \frac{\partial \mu_j}{\partial \beta}|_{\mu_j(\hat{\beta}_i)}$, $\tilde{V}_j(\hat{\beta}_i) = a(\phi)\Sigma_j\,[\hat{\beta}_i, \hat{\alpha}\{\hat{\beta}_i,\ \hat{\phi}(\hat{\beta}_i)\}]$, $S_j(\hat{\beta}_i) = (Y_j - \mu_j)|_{\mu_j(\hat{\beta}_i)}$ and $\mu_j(\hat{\beta}_i) = (\mu_{j1}(\hat{\beta}_i), \mu_{j2}(\hat{\beta}_i), \ldots, \mu_{jp_j}(\hat{\beta}_i))'$ where $\mu_{jk}(\hat{\beta}_i) = g^{-1}(x_{jk}'\hat{\beta}_i)$. The GEE design insures that the regression coefficients estimates $\hat{\beta}$ are consistent if the link function g is correctly specified (Zeger and Liang, 1986). The correlation structure $R_j(\alpha)$ does not need to be correctly specified as long as the subjects are independent.

The covariance of the estimates $\hat{\beta}$ is given in the theorem stated below. This theorem, proved by Liang and Zeger (1986), also gives the result that, under some assumptions, the estimator $\hat{\beta}$ asymptotically follows a multivariate Gaussian distribution.

Theorem

Under mild regularity conditions (see Serfling, 1980, pages 144-145) and given that :

I . $\hat{\alpha}$ is $n^{\frac{1}{2}}$-consistent estimator of α given β and ϕ;

II . $\hat{\phi}$ is $n^{\frac{1}{2}}$-consistent estimator of ϕ given β; and

III . $|\frac{\partial \hat{\alpha}(\beta,\phi)}{\partial \phi}| \leq H(Y,\beta)$ which is a function that is $O_p(1)$,

then $n^{\frac{1}{2}}(\hat{\beta}-\beta)$ is asymptotically multivariate Gaussian with zero mean and covariance matrix V_β given by

$$V_\beta = \lim_{n\to\infty} n \underbrace{(\sum_{j=1}^{n} D_j' V_j^{-1} D_j)^{-1}}_{M_0} \overbrace{\{\sum_{j=1}^{n} D_j' V_j^{-1}\, cov(Y_j)\, V_j^{-1} D_j\}}^{M_1} \underbrace{(\sum_{j=1}^{n} D_j' V_j^{-1} D_j)^{-1}}_{M_0}$$

$$= \lim_{n\to\infty} n(M_0^{-1} M_1 M_0^{-1}).$$

A consistent estimator of V_β can be obtained by replacing $cov(Y_j)$ by $(Y_j-\mu_j)(Y_j-\mu_j)'$ and α, β, ϕ by their respective estimators in V_β. Thus, the estimator of the variance of $\hat{\beta}$ is

$$\widehat{Var}(\hat{\beta}) = \hat{M}_0^{-1} \hat{M}_1 \hat{M}_0^{-1}$$

where, based on the final estimate $\hat{\beta}$ obtained from the iterative equation (3.4),

$$\hat{M}_0 = \sum_{j=1}^{n} D_j'(\hat{\beta}) V_j^{-1}(\hat{\beta}) D_j(\hat{\beta})$$

and

$$\hat{M}_1 = \sum_{j=1}^{n} D_j'(\hat{\beta})\, V_j^{-1}(\hat{\beta})\, S_j(\hat{\beta})\, S_j'(\hat{\beta})\, V_j^{-1}(\hat{\beta})\, D_j(\hat{\beta}).$$

The estimator $\widehat{Var}(\hat{\beta})$ is called the sandwich estimator because the matrix $\hat{M}_1$ is sandwiched between two instances of the matrix $\hat{M}_0^{-1}$.

It is interesting to note that the asymptotic covariance matrix estimator $\widehat{Var(\hat{\beta})}$ is robust to the choice of $\hat{\alpha}$ and $\hat{\phi}$, as long as they are $n^{\frac{1}{2}}$-consistent estimators and that the matrix M_0 in V_β converges to a fixed matrix when divided by n. Therefore it is not necessary that the observations for all subjects have the same correlation structure. One has to be careful though when there is missing data (Liang and Zeger, 1986). Since $\hat{\beta}$ and $\widehat{Var(\hat{\beta})}$ are robust to the choice of $R_j(\alpha)$, the confidence intervals and other statistical tests about β are asymptotically correct even if $R_j(\alpha)$ is misspecified, but choosing a working correlation matrix structure close to the actual one increases the efficiency of the different tests. That is the case, for example, for multivariate Gaussian outcomes (Zeger and Liang, 1986).

The matrix M_0^{-1} is a non-robust estimator of the covariance matrix of $\hat{\beta}$. This estimator is more efficient than the estimator $Var(\hat{\beta})$ only when both the working correlation structure and the mean-variance relationship for the GEE analysis are correct. Since it is impossible to know if it is really the case, this non-robust estimator of the covariance matrix is rarely used.

3.4 Working correlation matrix

As was stated before, in the case of repeated measurements data, the different observations for a subject are most often positively correlated. For each subject this dependence is represented by the correlation matrix R_j. For example, for subject j the correlation matrix would have the form

$$R_j = \begin{bmatrix} 1 & \rho_{j12} & \rho_{j13} & \cdots & \rho_{j1p_j} \\ \rho_{j12} & 1 & \rho_{j23} & \cdots & \rho_{j2p_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{j1p_j} & \rho_{j2p_j} & \cdots & \rho_{j(p_j-1)p_j} & 1 \end{bmatrix}$$

where $\rho_{jkk'} = corr(Y_{jk}, Y'_{jk})$, $k, k' = 1, 2, \ldots, p_j$ and $j = 1, 2, \ldots, n$.

As mentioned in the previous section, the dimension $(p_j \times p_j)$ of R_j can vary from subject to subject but the structure is fully specified by a $(s \times 1)$ vector of unknown parameters, α, which is the same for all subjects. Also, as can be seen in the iterative equation (3.4), the vector of parameters α will depend on the unknown scale parameter ϕ. Thus, at a given iteration i, both α and ϕ can be estimated from the current Pearson residuals defined by

$$\hat{r}_{jk} = \frac{[y_{jk} - \hat{\mu}_{jk}]}{\sqrt{V(\hat{\mu}_{jk})}}$$

and where $\hat{\mu}_{jk}$ is evaluated at the current estimated value of β, i.e. $\hat{\mu}_{jk} = g^{-1}(x'_{jk}\hat{\beta}_i)$. Then, the scale parameter ϕ can be estimated by

$$\hat{\phi}^{-1} = \sum_{j=1}^{n} \sum_{k=1}^{p_j} \frac{\hat{r}_{jk}^2}{(n-m)}.$$

I Independence structure.

The independence structure is the identity matrix of dimension $(p_j \times p_j)$. This is the simplest form and no nuisance parameter α need to be estimated.

$$R_j = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

II Exchangeable structure.

The exchangeable structure is obtained when all correlations are the same. This means $corr(Y_{jk}, Y_{jk'}) = \alpha$ for any k, k' where $k \neq k'$. In this case $R_j(\alpha)$ has the form

$$R_j(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha & \alpha & \dots & \alpha & 1 \end{bmatrix}.$$

The estimator of α is

$$\hat{\alpha} = \frac{\hat{\phi} \sum_{j=1}^{n} \sum_{k>k'} \hat{r}_{jk} \hat{r}_{jk'}}{[\sum_{j=1}^{n} \frac{1}{2} p_j (p_j - 1) - m]}.$$

III Stationary r-dependent structure.

This structure is characterized by the fact that the correlations q occasions apart are the same for $q = 1, 2, \dots, r$ and the correlations more than r occasions apart are zero, i.e.

$$
\begin{aligned}
corr(y_{jk},\, y_{j,k+1}) &= \alpha_1 \\
corr(y_{jk},\, y_{j,k+2}) &= \alpha_2 \\
\vdots \qquad\qquad & \vdots \qquad \vdots \\
corr(y_{jk},\, y_{j,k+r}) &= \alpha_r \\
corr(y_{jk},\, y_{j,k+r'}) &= 0,\ r' > r.
\end{aligned}
$$

The estimator of α_q, $1 \le q \le r$, is

$$\hat{\alpha}_q = \frac{\hat{\phi} \sum_{j=1}^{n} \sum_{k=1}^{p_j - q} \hat{r}_{jk} \hat{r}_{j,k+q}}{(p_j - q)(n - m)}.$$

IV Auto-regressive (AR-1) structure.

In the case of the auto-regressive (AR-1) structure, the correlations between two responses of the same subject are equal to a baseline correlation α to a power equal to the absolute difference between the times of the responses. This means $corr(y_{jk},\, y_{jk'}) = \alpha^{|t_{jk} - t_{jk'}|}$. Here are some examples for $R_j(\alpha)$ with an AR-1 structure.

Ex. 1 :

$$p_j = 3,\ t_{j1} = 1,\ t_{j2} = 2 \text{ and } t_{j3} = 3.$$

$$R_j(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix}.$$

Ex. 2 :

$$p_j = 5,\ t_{j1} = 1,\ t_{j2} = 2,\ t_{j3} = 3,\ t_{j4} = 4 \text{ and } t_{j5} = 5.$$

$$R_j(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 \\ \alpha^3 & \alpha^2 & \alpha & 1 & \alpha \\ \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix}.$$

Ex. 3 :

$$p_j = 4,\ t_{j1} = 1,\ t_{j2} = 3,\ t_{j3} = 4 \text{ and } t_{j4} = 4.5.$$

$$R_j(\alpha) = \begin{bmatrix} 1 & \alpha^2 & \alpha^3 & \alpha^{3.5} \\ \alpha^2 & 1 & \alpha & \alpha^{1.5} \\ \alpha^3 & \alpha & 1 & \alpha^{0.5} \\ \alpha^{3.5} & \alpha^{1.5} & \alpha^{0.5} & 1 \end{bmatrix}.$$

The estimator of α is given by the slope from the regression of the $\log(\hat{r}_{jk}, \hat{r}_{jk'})$ on $\log(|t_{jk} - t_{jk'}|)$.

V Unspecified correlation structure.

For this structure the same number of observations for all the subjects is needed, i.e. $p_j = p$ for all j. Here, $R_j(\alpha)$ has no constrained so that the vector α is of dimension $(\frac{p(p-1)}{2} \times 1)$. In the next example, the explicit form of $R_j(\alpha)$ can be seen.

Example :

$$p = 4, \ \frac{p(p-1)}{2} = 6.$$

Then $\alpha = (\alpha_{12}, \alpha_{13}, \alpha_{14}, \alpha_{23}, \alpha_{24}, \alpha_{34})'$ and

$$R_j(\alpha) = R(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 \end{bmatrix}.$$

The estimator of $R(\alpha)$ is

$$R(\hat{\alpha}) = \frac{\hat{\phi}}{n} \sum_{j=1}^{n} [A_j(\hat{\beta})]^{-\frac{1}{2}} \, S_j(\hat{\beta}) \, S_j'(\hat{\beta}) \, [A_j(\hat{\beta})]^{-\frac{1}{2}}.$$

Unfortunately, in some specific cases, the solution to the estimators of $R(\alpha)$ given for each type of structures may not exist. Crowder (1995) gave a counterexample for which there is no real solution for $\hat{\alpha}$ in the case of the auto-regressive correlation structure and he suggested some other ways to estimate $R(\alpha)$.

Chapter 4

Numerical example

4.1 Description of the experiment

An illustration of how the different methods presented in chapters 2 and 3 can be used in practice is performed using the data from example 6.2 in Jones and Kenward (1989). In this example, a three treatment three period crossover design was considered. The effects of the three treatments on blood pressure were to be compared. Treatments A and B consisted of the trial drug at 20 mg and 40 mg respectively and treatment C was a placebo. For each of the six possible treatment sequences, ABC, ACB, BAC, BCA, CAB and CBA, there was two replicates for a total of twelve subjects. The response was the level of systolic blood pressure (in mm Hg) taken under each treatment at ten successive times : 30 and 15 minutes before treatment and 15, 30, 45, 60, 75, 90, 120 and 240 minutes after treatment. For this particular illustration, only the response at 60 minutes after treatment will be considered.

4.2 Results from the different statistical methods

The model (2.1) with same dispersion matrix for all sequences was assumed. The OLS method gave $F_\tau = 5.57$ with 2 and 18 degrees of freedom and the p-value for the test on treatment effects was 0.0131. For the carryover effects we obtained $F_\lambda = 0.39$ with 2 and 18 degrees of freedom, p-value = 0.6799. For the MFA and EGLS methods an estimate of Σ was first calculated using the estimator of Σ presented in section 2.4. In the matrix S below, the variances are on the diagonal, the covariances are above the diagonal and the correlations below the diagonal.

$$S = \begin{pmatrix} 111.75 & 99.00 & 91.42 \\ 0.92 & 103.67 & 118.75 \\ 0.57 & 0.77 & 228.50 \end{pmatrix}$$

The MFA method produced the following estimates and level of significance for treatment and carryover effects respectively : $h_2 = 10.276$, $h_{1\tau} = 2$, $b_\tau = 1.457$ and p-value = 0.0573; $h_{1\lambda} = 2$, $b_\lambda = 1.502$ and p-value = 0.7742. The EGLS method gave $F_\tau = 20.84$ with p-value < 0.0001 and $F_\lambda = 6.62$ with p-value = 0.0070.

With the GEE method, the five different working correlation matrices presented in section 3.4 were considered, namely, the identity, exchangeable, 2-dependent, AR-1 and unspecified structures. A summary of all the p-values obtained for both the treatment and carryover effects are given in Table 4.1.

Table 4.1 P – values of the different statistical methods to test the absence of treatment and carryover effects

	Statistical methods							
				GEE with working correlation matrix structure				
Effects	*OLS*	*MFA*	*EGLS*	*Identity*	*Exch.*	*2 – Dept.*	*AR – 1*	*Unsp.*
treatment	0.0131	0.0573	< 0.0001	0.0008	0.0006	< 0.0001	< 0.0001	0.0002
carryover	0.6799	0.7742	0.0070	0.8020	0.7405	0.3023	0.3376	0.2572

At the 5% level of significance not all methods of analysis lead to the same conclusions. For the treatment effects, only the MFA method arrives to the conclusion of no treatment effects but the p-value is close to 5% (5.73%). For the carryover effects, the significance level of the EGLS method is very different than all the other statistical tests. The GEE method leads to the same conclusions regardless of the working correlation matrix used. Note however that the p-values can be quite different from one another (range from 0.2572 to 0.8020 for the carryover effects).

The estimates of the covariates obtained with OLS and the GEE method are given in Table 4.2. The corresponding standard errors and p-values are also presented. As was mentioned before, one advantage of the GEE method is that estimates of the covariates can be calculated. This is not possible with the MFA method.

In light of the differences observed among the tests performed, the need for an investigation on the performance of the different methods is justified. This is the subject of the next chapter.

Table 4.2 Coefficients estimates of the covariates and their standard errors and $p-values$

		Statistical methods					
			GEE with with correlation matrix structure				
Covariates		*OLS*	*Identity*	*Exch.*	$2-Dept.$	$AR-1$	*Unsp.*
treat. A	*estimate*	−0.1250	0.5667	−0.0381	−0.5716	−0.5091	−1.4576
	s.e.	1.6867	1.7280	1.7438	1.4424	1.4646	1.5232
	$p-value$	0.9417	0.7430	0.9826	0.6919	0.7281	0.3386
treat. B	*estimate*	4.9375	4.8167	4.9223	5.8963	5.8027	6.5092
	s.e.	1.6867	1.5093	1.4980	1.4204	1.4243	1.6254
	$p-value$	0.0090	0.0014	0.0010	< 0.0001	< 0.0001	< 0.0001
carry. A	*estimate*	−1.8750	0.2000	−1.6143	−3.1144	−2.9969	−3.4011
	s.e.	2.2629	3.6637	2.5522	2.2178	2.2510	2.2218
	$p-value$	0.4182	0.9565	0.5271	0.1602	0.1831	0.1258
carry. B	*estimate*	1.5625	1.2000	1.5170	2.6213	2.5274	2.5846
	s.e.	2.2629	2.8456	2.0403	1.8759	1.8846	1.8139
	$p-value$	0.4987	0.6732	0.4572	0.1623	0.1799	0.1542

Chapter 5

Monte Carlo Simulations

5.1 Methodology

Monte Carlo simulations were performed using SAS PROC IML and the SAS macro procedure GEE written by R. Karim (1989) in order to compare the behavior of the OLS, MFA, EGLS and GEE methods. The SAS program is given in Appendix A. The three treatments three periods crossover design with all six possible sequences (ABC, ACB, BAC, BCA, CAB and CBA) was considered. This "uniform balance" design is known to have optimal properties when $\Sigma = \sigma^2 I$ (Jones and Kenward, 1989, p.209). Three different covariance structures were used in this simulation study and are presented in Table 5.1. The first covariance matrix has a sphericity structure, hence the OLS tests are exact. The second one has an auto-regressive-1 structure, and the third has no specific structure and is basically the estimated covariance matrix of the example considered in the previous chapter.

Table 5.1 *Covariances matrices used for the Monte Carlo simulations (variances are on the diagonal, the covariances are above and the correlations below the diagonal)*

Code	*Type*	*Covariance Matrix*		
		1.00	0.50	1.50
1	*Sphericity*	0.29	3.00	2.50
		0.67	0.65	5.00
		2.00	1.50	1.13
2	$AR-1$	0.75	2.00	1.50
		0.56	0.75	2.00
		1.12	0.99	0.91
3	*No structure*	0.92	1.04	1.19
		0.57	0.77	2.29

Four different sample sizes were used, namely 18, 36, 72 and 108 subjects per experiment, that is 3, 6, 12 and 18 subjects per sequence respectively. The last two sample sizes were used with the covariance matrix 3 only. Hence, a total of eight simulation patterns were run. For each simulation pattern, two thousand independent samples were generated following model (2.1) with a multivariate normal distribution for the response variable Y. For each sample, significance tests were carried out for carryover and treatment effects using OLS, MFA, EGLS and GEE methods. The five different working correlation structures described in section 3.4 were used for the latter method. The empirical percentage of Type I error for each test was defined as the propor-

tion of p-values smaller or equal to a specified nominal alpha. Three values, $\alpha = 0.01$, 0.05 and 0.10 were chosen. A summary of the simulation results are presented in the next section.

5.2 Results and comments

The simulation results for the 5% nominal level alpha are given in tables 5.2 and 5.3 for treatment and carryover effects respectively. The results for the 1% and 10% nominal level give similar conclusions and are presented in Appendix B. The standard error of the empirical level of Type I error for the nominal level α is given by

$$\text{s.e.} = [\tfrac{\alpha(1-\alpha)}{N}]^{\frac{1}{2}}$$

where N is the number of independent samples generated for the simulation. Hence, for the 5% nominal level and two thousand independent samples, we have

$$\text{s.e.} = [\tfrac{0.05(1-0.05)}{2000}]^{\frac{1}{2}} = 0.0049.$$

If we want a 95% confidence interval for the empirical level of Type I error ($\hat{\alpha}$) at the 5% nominal level, we first need to compute the accuracy which is equal to the standard error multiplied by the 97.5% quantile of the standard normal distribution,

$$\text{accuracy} = \text{s.e.} \cdot z_{(1-\frac{0.05}{2})} = 0.0049 \cdot 1.645 = 0.008$$

and the 95% confidence interval is then

$$\hat{\alpha} \pm 0.8\%.$$

The first three methods of analysis were already compared by Bellavance, Tardif and Stephens (1996) and they obtained very similar results. For the class of covariance structure for which the OLS is exact (sphericity), the 95% C.I. for the empirical level of Type I error includes 5% for the OLS method for both the treatment and carryover effects. The OLS approach performed also well when the covariance structure was of AR-1 type, especially for the carryover effect, but very badly with the no structure type.

For the MFA method, the case of three subjects per sequence for the test of treatment effects gives adequate control over Type I error but is a little liberal with six subjects per sequence and somewhat more liberal for the test of carryover effects for both sample sizes considered. For both the EGLS and the GEE methods, the results are very liberal for the two lowest sample sizes and are getting closer to the nominal level when the number of subjects per sequence becomes larger. The 95% C.I. for the empirical level of Type I error includes 5% for the EGLS method for both the treatment and carryover effects with twelve subjects per sequence. For the treatment effects with the GEE method, the 95% C.I. for the empirical level of Type I error includes 5% only with the identity working correlation matrix and the largest sample size. For the carryover effects with GEE method, this is the case for all working correlation matrices and the largest sample size.

Table 5.2 Empirical level of Type I error (%) for the test of treatment effects at the 5% nominal level

		Statistical methods							
Covariance Matrix Σ	*Number of subjects per sequence*	*OLS*	*MFA*	*EGLS*	*GEE with working correlation matrix*				
					Identity	*Exch.*	$2-$ *Dept.*	$AR-1$	*Unsp.*
Sphericity	3	5.10	5.00	9.90	11.10	13.40	16.00	13.80	19.10
	6	5.65	5.85	7.60	7.20	8.50	9.60	8.35	11.70
Type $AR-1$	3	6.35	4.60	9.70	11.35	13.40	16.40	12.30	20.90
	6	6.85	4.45	7.40	7.75	8.80	10.05	8.90	11.35
No structure	3	12.95	4.55	10.80	13.45	16.60	17.20	16.05	19.00
	6	13.35	5.50	6.65	9.40	10.20	10.60	10.15	10.25
	12	12.10	5.40	5.40	7.15	7.25	7.60	7.60	7.95
	18	--	--	--	4.70	6.30	6.15	6.05	6.55

Table 5.3 Empirical level of Type I error (%) for the test of carryover effects at the 5% nominal level

		Statistical methods							
Covariance Matrix Σ	*Number of subjects per sequence*	*OLS*	*MFA*	*EGLS*	*GEE with working correlation matrix*				
					Identity	*Exch.*	$2-$ *Dept.*	$AR-1$	*Unsp.*
Sphericity	3	5.65	6.80	9.70	9.25	15.55	16.10	16.20	20.35
	6	5.25	6.00	7.40	7.00	9.40	9.70	9.75	11.85
Type $AR-1$	3	4.60	5.00	8.70	10.35	14.60	16.55	14.00	19.70
	6	5.10	4.75	6.85	7.75	9.95	10.55	9.70	11.30
No structure	3	12.55	4.90	10.25	10.90	17.80	18.30	17.40	20.65
	6	13.05	4.55	7.05	7.00	9.35	10.30	10.25	11.35
	12	12.80	5.00	5.85	6.20	7.75	7.85	7.55	8.05
	18	--	--	--	5.25	5.30	5.40	5.25	5.35

In light of the results in tables 5.2 and 5.3, the OLS analysis is not robust to covariance structures that are not in the sphericity class. For the AR-1 type of structure, the empirical level of Type I error is accurate for the test of carryover effects and a little liberal for the test of treatment effects. Note however that the specific AR-1 matrix used here is very "close" to the

sphericity structure, so these findings are not surprising. For the unspecified structure the OLS method gives very liberal results even with as much as twelve subjects per sequence. All these observations about the behavior of the OLS method suggest that the OLS F-tests will be unreliable and could lead to serious errors in inference.

The EGLS method is almost always too liberal except in the case where there is a large number of subjects per sequence. The same conclusion can be drawn for the GEE method but the number of subjects per sequence has to be slightly larger than for the EGLS method. Also, a different choice of the working correlation matrix will give a different empirical level of Type I error. Even if the parameter estimates are expected to be equal asymptotically, the covariance matrix of the parameter estimates may change with the choice of the working correlation matrix, even asymptotically, therefore leading to different empirical level of Type I error (see section 3.3). The 2-dependent and unspecified working correlation matrices lead to the most liberal results. The identity working correlation matrix gave better results but they were still quite liberal.

Regardless of the choice of the working correlation matrix, the same trend was observed : a larger number of subjects per sequence imply a more accurate empirical Type I error, and the identity working correlation matrix always performed better than the other working correlation matrices. Therefore, in the case of the three treatment three period crossover design, eighteen subjects per sequence (108 subjects total) are needed to get an empirical level of Type I error near 5% for the GEE method. This is a large sample size to use with crossover designs, especially in the medical area, since each subject

has multiple observations. Just adding few subjects per sequence can add considerable cost and can be very time consuming.

Finally, with the MFA method, all the 95% C.I. of the empirical Type I error include the nominal 5% level except for the sphericity case with six subjects per sequence where the lower limit is very close to 5% and somewhat less close with three subjects per sequence for the carryover effects. An interesting point is that the MFA method performes as well with small sample sizes as with larger sample sizes. Therefore, the MFA method is the one to be preferred over the other methods studied here. Moreover, this method is very easy to apply in practice. Its principal drawback is that predictions are not possible to compute since no β-coefficient estimates can be found with this method, only tests of the different effects can be conducted.

Figure 5.1 presents the scatter plot of the p-values obtained from the GEE and MFA methods with the identity working correlation matrix for the simulation case of three subjects per sequence and covariance matrix 1 (sphericity structure) for the test of treatment effects. In this graph it can be seen that the GEE method is more liberal than the MFA method. A majority of the points are below the 45° line $x = y$ and therefore implies that the p-values of the MFA method are larger than the one with the GEE method. Hence the GEE method will conclude to a treatment effects more often than the MFA method given H_0 = no treatment effects is exactly true. This figure represents very well the results found in tables 5.2 and 5.3.

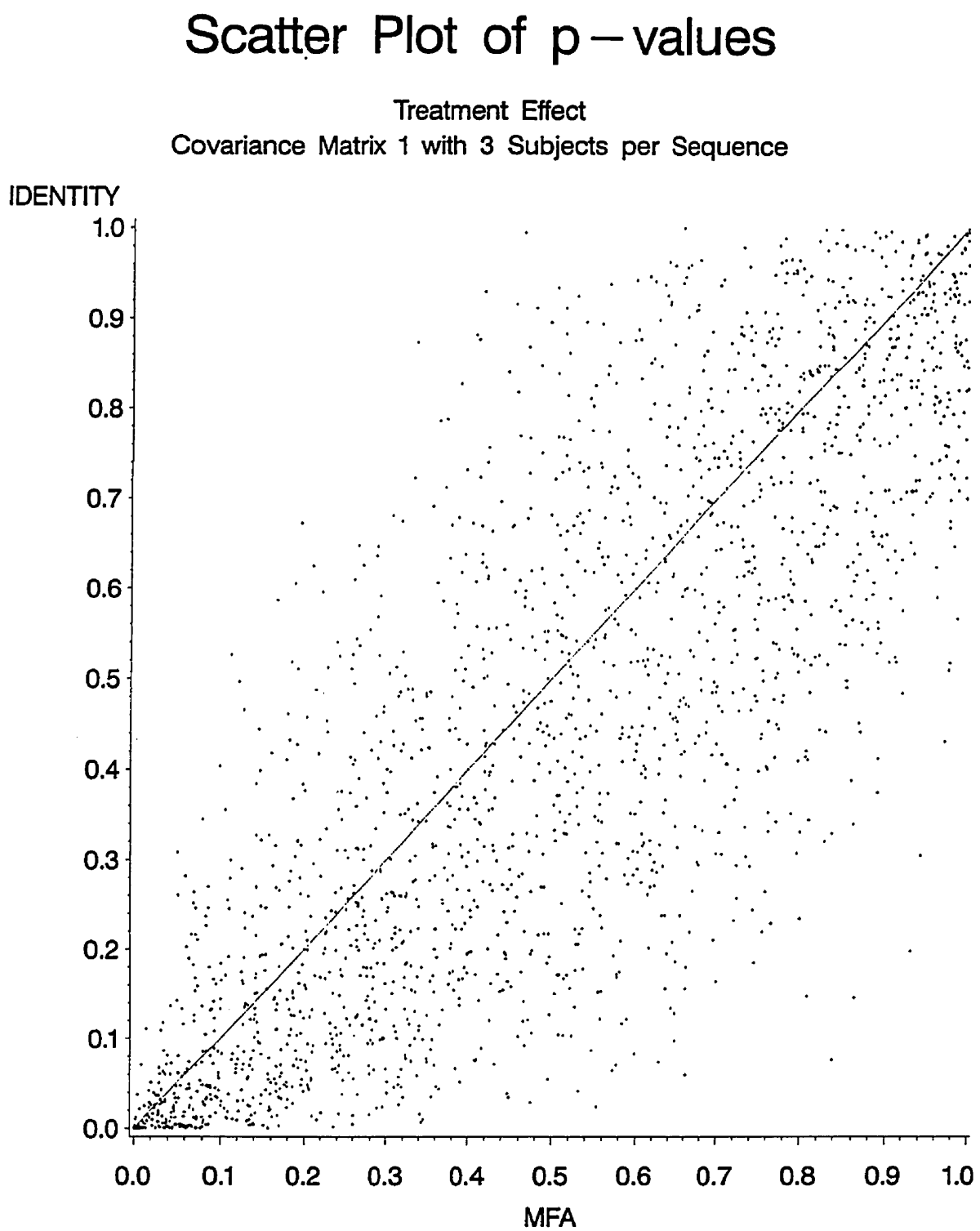


Figure 5.1 Scatter plot of the p-values of the GEE method with identity working correlation matrix vs the MFA .

Chapter 6

Concluding remarks

In this thesis, the analysis of crossover designs was studied. In this type of design, repeated measurements are observed for each subject. The difficulty arising from this type of data is that the responses within one subject may be correlated. Since the usual, likelihood based, method of analysis for crossover designs, the ordinary least squares (OLS) method, does not consider different covariance structures for the responses of one subject, other methods were also studied in this project. Two approximate likelihood based methods that take into consideration the covariance structure were considered, namely, the modified F-test approximation (MFA) and the empirical generalized least squares (EGLS) method. Bellavance, Tardif and Stephens (1996) examined tests of crossover design analysis with correlated errors for these three methods. A covariance matrix needs to be estimated for the use of the MFA and EGLS tests. A brief review of the three likelihood based methods and an estimate of the covariance matrix for the MFA and EGLS

methods were given in chapter 2.

Another type of method was also studied in this work : the generalized estimating equations (GEE) method. This method does not need a complete specification of the joint distribution of the responses since it is based on quasi-likelihood distributional assumptions. Also, GEE can be used when the distribution for the vector of responses has forms other than the multivariate normal distribution. This is not the case for the other three methods studied in this thesis. Moreover, different structures of the covariance matrix can be modeled with this method. The details of the GEE method and the covariance structures most commonly used were presented in chapter 3.

The performance of the four different methods was evaluated for the 3 period - 3 treatment - 6 sequence crossover design with multivariate normally distributed errors and small and medium sample sizes. The simulations performed were described in chapter 5. In these simulations the GEE method needed a large number of subjects per sequence to arrive to an adequate empirical level of Type I error, namely eighteen subjects per sequence. Since the sample sizes in many crossover trials are small and because large sample sizes can be very expensive and time consuming, the GEE method is not the best method for the type of crossover designs studied here. The EGLS method has similar sample size problems. Referring to the example presented in chapter 4, it was observed in table 4.1 that the EGLS method arrived to very different results than the other methods. This may be explained by the bad performance of the EGLS method with small sample sizes, which was the case in the example of chapter 4 with only twelve subjects for six sequences. These results suggest that the EGLS method and also the GEE

method are very sensitive to the accuracy of the estimate of the covariance matrix. The OLS analysis in cases other than when the covariance matrix has a sphericity structure, does not improve with larger sample sizes.

On the other hand, the MFA method has good Type I error accuracy in all cases considered, that is, with the three types of covariance structures and with both smaller and larger sample sizes. It is also easy to apply. The only down point of this method is that no estimates of the coefficients of the covariates can be calculated and therefore no predictions can be computed. In the case where this is of interest, a satisfactory method has yet to be developed. Where only the tests of different effects are of interest, the MFA method is to be preferred over all the other methods studied here, including the GEE method, unless a large number of observations is available to the analyst.

The simulations were made only considering the nominal level alpha. The MFA method was the only method performing well with regards to the nominal level in all cases. A power analysis would need to be performed to ensure that this method meets our expectations. If this is not the case, research should continue to find a method that will perform well with regards to both the nominal level and the power.

Appendix A

The following is the SAS program used to perform the Monte Carlo simulations presented in Chapter 5. The SAS macro GEE1 has been written by Mr. Rezaul Karim from Johns Hopkins University. These particular settings are for the simulations with the unspecified covariance structure and 3 subjects per sequence.

```
option pagesize=30500;
option linesize=80;
option nodate nonumber;
proc printto;run;

proc printto print='c:\litte\cross\seq18a.out'
             log='c:\litte\cross\log1.log';


/*************************************************************************
*************************************************************************
**                                                                     **
**  GEE1_PC  SAS:  PC  version  of  GEE1                               **
**                                                                     **
**  Due to the differences in ASCII and EBCDIC character sets, and     **
**  also due to the differences in translation tables used to convert  **
**  EBCDIC to ASCII (and ASCII to EBCDIC) in different computer        **
**  instalations, some 'special' characters may get changed            **
**  unexpectedly when you receive the SAS macro at your end.           **
**                                                                     **
**  An easy check against this problem is to make sure that the        **
**  following special characters are correctly represented on the      **
**  SAS file you have received. If not -- make global changes for      **
**  these characters with the help of any text editor.                 **
**                                                                     **
**  Following is a list of some special characters used in the macro.  **
**                                                                     **
**          '|'    ...    'vertical bar'                              **
**          '^'    ...    'NOT sign'                                  **
**          '['    ...    'left square bracket'                       **
**          ']'    ...    'right square bracket'                      **
**                                                                     **
*************************************************************************
*************************************************************************/
/*
                SAS Macro for Longitudinal Data Analysis:
                =========================================
```

GEE is a SAS macro for analyzing longitudinal data. This SAS IML macro uses the GEE approach of Liang and Zeger (1986) to model longitudinal data for a general class of outcome variables including gaussian, poisson, binary and gamma outcomes. The program uses an iterative procedure to estimate regression coefficients, treating the correlation among observation on the same individual as a nuisance. Final output from GEE includes: regression coefficients, naive and robust estimates of variance and z-score.

Following command, with appropriate parameters, can be used to invoke the macro. All parameters have been assigned default values, so that they can be omitted if default values are acceptable. (Defaults are shown within {}). Parameters may be given in any order.

```
%GEE  ( DATA = SAS dataset,                              { _LAST_ }
        YVAR = y-variable,                               { Y }
        XVAR = x-variables,                              { X }
        ID   = id-variable,                              { ID }
        LINK = link function,                            { 1 }
        VARI = mean-variance relation,                   { 1 }
        N    = binomial denominator variable,            {_1_}
        CORR = correlation structure,                    { 1 }
        M    = dependence,                               { 1 }
        R    = given correlation matrix,                 { I }
        BETA = initial estimate of beta,                 { LSE }
        OFFS = offset variable,                          { _0_ }
        OUT  = output dataset,                           { _NULL_ }
        ITER = maximum iterations,                       { 20 }
        CRIT = convergence criterion                     { 0.001 }
      ) ;
```

```
=============================================================================
=
 Date: 7/4/89
 Author: M. Rezaul Karim
 Department of Biostatistics
 The Johns Hopkins University
=============================================================================
==*/

%MACRO GEE ( DATA =_last_,
             YVAR =y,
             XVAR =x,
             ID   =id,
             LINK =1,
             VARI =1,
             N    =_1_,
             CORR =1,
             M    =1,
             R    =I,
             BETA =0,
```

```
                OFFS =_0_,
                OUT  =_NULL_,
                ITER =20,
                CRIT =0.001) ;

OPTION nocenter;
PROC IML WORKSIZE=999;
RESET noname;

USE &DATA; SETIN &DATA NOBS nobs;

link={ &LINK };
vari={ &VARI };
corr={ &CORR };
m={ &M };
r={ &R };
n=INT(SQRT(NROW(r)#NCOL(r))); r=SHAPE(r,n,n);
beta={ &BETA };
labely={ &YVAR };
labelx={ &XVAR }`;
labelo={ &OFFS }; IF labelo={'_0_'} THEN offset=0;
labeln={ &N }; IF labeln={'_1_'} THEN n=1;

START init; /*************************************************************/

R1={'Data File:' &DATA};
*PRINT / 'Regression analysis using GEE:                ( Ver - 1.25
)',
*        '============================',,,
*         R1;

R1={'Outcome variable:'};
R2={'Covariates:'};
*PRINT    labely  [ROWNAME=R1],
*         {&XVAR} [ROWNAME=R2];
R1={'Offset:'};
*IF labelo^={'_0_'} THEN PRINT labelo [ROWNAME=R1];

_1={'(Identity)'};
_2={'(Logarithm)'};
_3={'(Logit)'};
_4={'(Reciprocal)'};
R1={'Link:'};
*IF (link<1 | link>4) THEN PRINT link [ROWNAME=R1 FORMAT=2.0]
*                              {'(Invalid Option !!!)'};
*ELSE PRINT link [ROWNAME=R1 FORMAT=2.0] _&LINK ;

_1={'(Gaussian)'};
_2={'(Poisson)'};
_3={'(Binomial)'};
_4={'(Gamma)'};
R1={'Variance:'};
*IF (vari<1 | vari>4) THEN PRINT vari [ROWNAME=R1 FORMAT=2.0]
*                              {'(Invalid Option !!!)'};
```

```
*ELSE PRINT vari [ROWNAME=R1 FORMAT=2.0] _&VARI ;
R1={'Denominator'};
*IF vari=3 THEN PRINT labeln [ROWNAME=R1];

FREE _1 _2 _3 _4;

R1={'Correlation:'};
*IF corr=1 THEN DO;
*    IF NCOL(r)=1 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0] {'(Independent)'};
*    ELSE PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0] {'(R  given):'},
*        r       [FORMAT=4.2];
*    END;
R2={'(Stationary'};
*IF corr=2 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0]
*        m       [ROWNAME=R2 FORMAT=2.0] {'- dependent)'};
R2={'(NonStationary'};
*IF corr=3 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0]
*        m       [ROWNAME=R2 FORMAT=2.0] {'- dependent)'};
*IF corr=4 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0] {'(Exchangeable)'};
R2={'(AR -'};
*IF corr=5 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0]
*        m       [ROWNAME=R2 FORMAT=2.0] {')'};
*IF corr=6 THEN PRINT
*        corr    [ROWNAME=R1 FORMAT=2.0] {'(Unspecified)'};

R1={'Total number of records read:'};
*PRINT nobs [ROWNAME=R1 FORMAT=8.0];

p=NROW(labelx);
dmean=0; imean=0;

READ VAR { &ID } INTO idk;
READ VAR labely  INTO yvar;
READ VAR labelx  INTO xvar;
IF labeln^={'_1_'} THEN DO; READ VAR labeln INTO n; END;
IF labelo^={'_0_'} THEN DO; READ VAR labelo INTO offset; END;

vsum=yvar||xvar;
IF NCOL(beta)=1 THEN DO;
xty=xvar`*(yvar/n-offset);
xtx=xvar`*xvar;
END;

war=J(1,p,1);
war=war#(war=xvar);

k=1; i=1; ni=0;
```

```
DO j=2 TO nobs;

READ VAR { &ID } INTO idj POINT j;
READ VAR labely  INTO yvar;
READ VAR labelx  INTO xvar;
IF labeln^={'_1_'} THEN DO; READ VAR labeln INTO n; END;
IF labelo^={'_0_'} THEN DO; READ VAR labelo INTO offset; END;

IF idk=idj THEN i=i+1;
ELSE DO;
imean=imean+vsum/i; dmean=dmean+vsum; vsum=0;
ni[k]=i; k=k+1; idk=idj; i=1; ni=ni//{0}; END;

war=war#(war=xvar);
vsum=vsum+(yvar||xvar);
IF NCOL(beta)=1 THEN DO;
xty=xty+xvar`*(yvar/n-offset);
xtx=xtx+xvar`*xvar;
END;
END;


ni[k]=i;
imean=(imean+vsum/i)/k;
dmean=((dmean+vsum)/nobs)//imean;
IF NCOL(beta)=1 THEN beta=SOLVE(xtx,xty);
ELSE beta=SHAPE(beta,p,1,0);
maxn=MAX(ni);
minn=MIN(ni);

R1={'Total number of clusters:'};
*PRINT  k      [ROWNAME=R1 FORMAT=5.0];
R1={'Maximum and minimum cluster size:'};
R2={'and'};
*PRINT  maxn   [ROWNAME=R1 FORMAT=5.0]
*       minn   [ROWNAME=R2 FORMAT=5.0];
R1={'Observations: ','Cluster Means:'};
C1={ &YVAR &XVAR };
*PRINT 'Averages of Outcome variable and Covariates (over all)',,
*       dmean [ROWNAME=R1 COLNAME=C1],;
*IF ALL(^war) THEN PRINT '*** WARNING: No intercept term in the model!';

*PRINT / 'Initial estimate of regression coefficients:', labelx beta;
FREE dmean i idj idk imean j vsum xtx xty xvar yvar;
*show names;
FINISH; /** INIT
*******************************************************/

START estb;
/*************************************************************/
us=J(p,1,0);
m0=J(p,p,0);

nx=0;
```

```
DO j=1 TO k;
nj=ni[j]; i=(nx+1):(nx+nj); nx=nx+nj;

READ VAR labely INTO yvar POINT i;
READ VAR labelx INTO xvar POINT i;
IF labeln^={'_1_'} THEN DO; READ VAR labeln INTO n POINT i; END;
IF labelo^={'_0_'} THEN DO; READ VAR labelo INTO offset POINT i; END;

*** Calculate ui and di;

lp=xvar*beta+offset;

     IF link=1 THEN DO; ui=lp; di=xvar; END;
ELSE IF link=2 THEN DO; ui=EXP(lp); di=ui#xvar; END;
ELSE IF link=3 THEN DO; ui=EXP(lp); ui=(n#ui)/(1+ui);
                        di=(ui#(1-ui/n))#xvar; END;
ELSE IF link=4 THEN DO; ui=1/(lp); di=-(ui#ui)#xvar; END;

zi=di*beta+(yvar-ui);

*** Calculate a*zi and a*di;

     IF vari=1 THEN ui=1;
ELSE IF vari=2 THEN ui=1/SQRT(ui);
ELSE IF vari=3 THEN ui=1/SQRT(ui#(1-ui/n));
ELSE IF vari=4 THEN ui=1/ABS(ui);

di=di#ui; zi=zi#ui;

vinv=INV(r[1:nj,1:nj]);
us=us+di`*vinv*zi;
m0=m0+di`*vinv*di;

END; *** End of beta estimation loop;

beta=solve(m0,us);
C1={'Estimate'};
*PRINT labelx beta [COLNAME=C1];
FREE di i j m0 nj nx us ui vinv xvar yvar zi;
*show names;
FINISH; /** ESTB
*******************************************************/

START estr;
/*************************************************************/
     IF corr=1 THEN DO; END;                        /* given correlation
*/
ELSE IF corr=2 THEN alp=J(1,m,0);              /* stationary m-dependent
*/
ELSE IF corr=3 THEN alp=J(maxn,maxn,0);        /* non stationary m-dept
*/
ELSE IF corr=4 THEN alp=0;                              /* exchangeable
*/
```

```
ELSE IF corr=5 THEN alp=J(1,m,0);                                    /* AR-m
*/
ELSE IF corr=6 THEN alp=J(maxn,maxn,0);                    /* unspecified
*/

sigma=0;
nx=0;
DO j=1 TO k;
nj=ni[j]; i=(nx+1):(nx+nj); nx=nx+nj;

READ VAR labely INTO yvar POINT i;
READ VAR labelx INTO xvar POINT i;
IF labeln^={'_1_'} THEN DO; READ VAR labeln INTO n POINT i; END;
IF labelo^={'_0_'} THEN DO; READ VAR labelo INTO offset POINT i; END;

lp=xvar*beta+offset;

     IF link=1 THEN DO; ui=lp; END;
ELSE IF link=2 THEN DO; ui=EXP(lp); END;
ELSE IF link=3 THEN DO; ui=EXP(lp); ui=(n#ui)/(1+ui); END;
ELSE IF link=4 THEN DO; ui=1/(lp); END;
ei=yvar-ui;

     IF vari=1 THEN ui=1;
ELSE IF vari=2 THEN ui=1/SQRT(ui);
ELSE IF vari=3 THEN ui=1/SQRT(ui#(1-ui/n));
ELSE IF vari=4 THEN ui=1/ABS(ui);
ei=ei#ui;
sigma=sigma+SSQ(ei)/nj;

     IF corr=1 THEN DO; END;
ELSE IF corr=2 THEN DO;
     i=nj/(nj+1-(1:m)); alp=alp+COVLAG(ei,-m)#i; END;
ELSE IF corr=3 THEN alp=alp+ei*ei`;
ELSE IF corr=4 THEN DO;
     IF (nj>1) THEN alp=alp+(SUM(ei*ei`)-SSQ(ei))/(nj#(nj-1)); END;
ELSE IF corr=5 THEN DO;
     i=nj/(nj+1-(1:m)); alp=alp+COVLAG(ei,-m)#i; END;
ELSE IF corr=6 THEN alp=alp+ei*ei`;


END; *** End of working covariance estimation loop;

     IF corr=1 THEN DO; END;
ELSE IF corr=2 THEN DO; alp=alp/sigma; alp[1]=1;
     r=SHAPE(alp,1,maxn,0); r=TOEPLITZ(r); END;
ELSE IF corr=3 THEN DO; r=alp/sigma;
     DO j=1 TO maxn; r[j,j]=1;
     DO i=j+m TO maxn; r[i,j]=0 ; r[j,i]=0 ; END; END;
     END;
ELSE IF corr=4 THEN DO; alp=alp/sigma;
     r=J(1,maxn,alp); r[1]=1; r=TOEPLITZ(r); END;
ELSE IF corr=5 THEN DO; alp=alp/sigma; alp[1]=1;
     r=SHAPE(alp,1,maxn,0);
```

```
     i=TOEPLITZ(alp[1:m-1]); alp=alp[1,2:m]*INV(i);
     DO i=m+1 TO maxn; DO j=1 TO m-1;
     r[i]=r[i]+alp[j]#r[i-j]; END; END;
     r=TOEPLITZ(r); END;
ELSE IF corr=6 THEN DO; r=alp/sigma;
     DO j=1 TO maxn; r[j,j]=1; END;
     END;


FREE alp ei i j nj nx ui xvar yvar;
*show names;
FINISH; /** ESTR
*********************************************************/


/*************************************************************************
/
/* Main Program:
*/
/*************************************************************************
/

RUN init;
IF NCOL(r)<=1 THEN r=I(maxn);

START; /** Check for consistency
*************************************/
crit=1;
IF corr=1 THEN DO;
   IF NCOL(r)<maxn THEN DO;
    * PRINT 'ERROR: Dimension of the given correlation matrix must be'
    *       'equal to the maximum cluster size';
     crit=0; END;
   END;
IF corr=2 | corr=3 THEN DO;
   IF m>=minn THEN DO;
    * PRINT 'ERROR: Gorup size too small for m-dependent correlation.';
    * crit=0; END;
   m=m+1; END;
IF corr=4 THEN DO; END;
IF corr=5 THEN DO;
   IF m>=minn THEN DO;
    * PRINT 'ERROR: Gorup size too small for AR-m correlation.';
    * crit=0; END;
   m=m+1; END;
IF corr=6 | corr=3 THEN DO;
   IF maxn=minn THEN DO; END;
   ELSE DO;
    *  PRINT 'ERROR: Unequal gorup size.';
      crit=0; END;
   END;
FINISH; RUN; /** End for consistency check
*****************************/
```

```
START; /** Main iteration
*****************************************/
IF crit=0 THEN STOP;

DO iter=1 TO &ITER WHILE(crit>&CRIT);

R1={'===> Iteration:'};
*PRINT iter [ROWNAME=R1 FORMAT=3.0];

save=beta;
IF corr>1 THEN RUN estr;
RUN estb;

crit=MAX(ABS(1-save/beta));

END; *** End of iterations;
*show names;

iter=iter-1;
*IF iter>=&ITER THEN PRINT ' ' /
*   {'No Convergence after'} {&ITER} [FORMAT=3.0] {'iterations.'};
*ELSE PRINT ' ' /
*   {'Convergence after'} iter [FORMAT=3.0] {'iteration(s).'};

IF maxn>10 THEN DO;
   save=r[1:10,1:10];
*   PRINT 'Working Correlation:', save;
END;
*ELSE PRINT 'Working Correlation:', r;
FREE save;

crit=1;
FINISH; RUN; /* End of iteration
*************************************/


START; /** Calculation of variance
***********************************/
IF crit=0 THEN STOP;
RUN estr;
sigma=SQRT(sigma/k);
m0=J(p,p,0); m1=J(p,p,0);
dev=0;

null={'_NULL_'};
C1={ FIT RES SRES      };
IF null = {&OUT} THEN DO; END;
ELSE DO;
  out={ 0 0 0    }; id={'12345678'};
  CREATE &OUT FROM out [ROWNAME=id COLNAME=C1];
  SETIN  &DATA;
END;

nx=0;
```

```
DO j=1 TO k;
nj=ni[j]; i=(nx+1):(nx+nj); nx=nx+nj;

READ VAR labely INTO yvar POINT i;
READ VAR labelx INTO xvar POINT i;
IF labeln^={'_1_'} THEN DO; READ VAR labeln INTO n POINT i; END;
IF labelo^={'_0_'} THEN DO; READ VAR labelo INTO offset POINT i; END;

*** Calculate ui and di;

lp=xvar*beta+offset;

     IF link=1 THEN DO; ui=lp; di=xvar; END;
ELSE IF link=2 THEN DO; ui=EXP(lp); di=ui#xvar; END;
ELSE IF link=3 THEN DO; ui=EXP(lp); ui=(n#ui)/(1+ui);
                        di=(ui#(1-ui/n))#xvar; END;
ELSE IF link=4 THEN DO; ui=1/(lp); di=-(ui#ui)#xvar; END;

ei=yvar-ui;
dev=dev+SSQ(ei)/nj;
IF null = {&OUT} THEN DO; END;
ELSE out=ui||ei;

*** Calculate a*ei and a*di;

     IF vari=1 THEN ui=1;
ELSE IF vari=2 THEN ui=1/SQRT(ui);
ELSE IF vari=3 THEN ui=1/SQRT(ui#(1-ui/n));
ELSE IF vari=4 THEN ui=1/ABS(ui);

di=(di#ui)/sigma; ei=(ei#ui)/sigma;


IF null = {&OUT} THEN DO; END;
ELSE DO;
  out=out||ei;
  id=J(nj,1,CHAR(j,8,0));
  SETOUT &OUT;
  APPEND FROM  out [ROWNAME=id];
  SETIN  &DATA;
END; /* End of output file process */

vinv=INV(r[1:nj,1:nj]);
m0=m0+di`*vinv*di;
i=ei`*vinv*di;
m1=m1+i`*i;

END; *** End of variance(beta) estimation loop;
nvar=INV(m0);
rvar=nvar*m1*nvar;

FREE di ei i j m0 m1 nj nx ui vinv xvar yvar;
*show names;
```

```
FINISH; RUN; /* End of variance
****************************************/

START; /* Outputs
*****************************************************/
IF crit=0 THEN STOP;
sigma=sigma#sigma;
dev=dev/k;
*ns=1/SQRT(VECDIAG(nvar));
rs=1/SQRT(VECDIAG(rvar));
rbeta=beta#rs;

*DO i=1 TO p;
*   DO j=i+1 TO p;
*      nvar[j,i]=nvar[i,j]#ns[i]#ns[j];
*      rvar[j,i]=rvar[i,j]#rs[i]#rs[j];
*      END;
*   END;

R1={'Scale parameter:'};
R2={'Mean Squared Error:'};
*PRINT  sigma  [ROWNAME=R1],
*       dev    [ROWNAME=R2] /;

*PRINT 'Variance estimate (naive):',,
*       nvar [ROWNAME=labelx COLNAME=labelx];

*PRINT 'Variance estimate (robust):',,
*       rvar [ROWNAME=labelx COLNAME=labelx],
*      'NOTE: Covariances are above diagonal and correlations are below'
*      'diagonal.',;

*ns=1/ns;
rs=1/rs;
pval=2*(1 - probnorm(abs(rbeta)));


C1={ 'Estimate' };
C2={ ' s.e.-Naive' };
C3={ ' z-Naive' };
C4={ 's.e.-Robust' };
C5={ 'z-Robust' };
C6={ 'p-value' };
*PRINT / 'Estimate, s.e. and z-score:',,
*         labelx beta  [COLNAME=C1]
*                ns    [COLNAME=C2 FORMAT=11.3]
*                rs    [COLNAME=C4 FORMAT=11.3]
*                rbeta [COLNAME=C5 FORMAT=8.2]
*                pval  [COLNAME=C6 FORMAT=8.4],;
*IF ALL(^war) THEN PRINT '*** WARNING: No intercept term in the model!';

pval1=t(pval);
```

```
cper1 = {0 1 0 0 0 0 0};
cper2 = {0 0 1 0 0 0 0};
ctrt1 = {0 0 0 1 0 0 0};
ctrt2 = {0 0 0 0 1 0 0};
ccar1 = {0 0 0 0 0 1 0};
ccar2 = {0 0 0 0 0 0 1};

cper = cper1 // cper2;
ctrt = ctrt1 // ctrt2;
ccar = ccar1 // ccar2;

resp = j(nobs,1,0);
desi = j(nobs, p, 0);

do j=1 to nobs;
READ VAR { &ID } INTO idj POINT j;
READ VAR labely INTO yvar;
resp[j,] = yvar;
READ VAR labelx INTO xvar;
desi[j,] = xvar;
end;


*numper = t(cper*beta)*inv(cper*rvar*t(cper))*(cper*beta);
numtrt = t(ctrt*beta)*inv(ctrt*rvar*t(ctrt))*(ctrt*beta);
numcar = t(ccar*beta)*inv(ccar*rvar*t(ccar))*(ccar*beta);

numm = numtrt || numcar;

*PRINT 'Numtrt, Numcar';
*print numm;

den = t(resp - (desi*beta))*(resp - (desi*beta));

*pvalper = 1 - probchi(numper, 2);
pvaltrt = 1 - probchi(numtrt, 2);
pvalcar = 1 - probchi(numcar, 2);


pval2 = pvaltrt || pvalcar;

PRINT 'Pvaltrt, Pvalcar';
print pval2;


*PRINT ' ',,'(c)  M. Rezaul Karim, 1989',
*      'Department of Biostatistics, The Johns Hopkins University';
FINISH; RUN; /* End of outputs
*****************************************/
QUIT;

%MEND;
```

```
%MACRO simul(nombre);

  %do s=1 %to &nombre;


proc iml;

  p18=j(1,18,1);
  p1=j(1,6,1);
  p3=j(1,3,1);

  u1= j(324,1,1);

  res={0 0 0,
       1 0 0,
       0 1 0};

  abc={1 0 0,
       0 1 0,
       0 0 1};

  bca={0 1 0,
       0 0 1,
       1 0 0};

  cab={0 0 1,
       1 0 0,
       0 1 0};

  acb={1 0 0,
       0 0 1,
       0 1 0};

  bac={0 1 0,
       1 0 0,
       0 0 1};

  cba={0 0 1,
       0 1 0,
       1 0 0};

  d1=t(p3)@abc;
  d2=t(p3)@bca;
  d3=t(p3)@cab;
  d4=t(p3)@acb;
  d5=t(p3)@bac;
  d6=t(p3)@cba;

  r1=t(p3)@(res*abc);
  r2=t(p3)@(res*bca);
  r3=t(p3)@(res*cab);
  r4=t(p3)@(res*acb);
  r5=t(p3)@(res*bac);
  r6=t(p3)@(res*cba);
```

```
des=d1 // d2 // d3 // d4 // d5 // d6;

resi=r1 // r2 // r3 // r4 // r5 // r6;

periode = t(p18)@i(3);

sujet= i(18)@t(p3);

x= u1 || sujet || periode || des || resi;

x1= u1 || sujet || periode || des ;

x2= u1 || sujet || periode || resi;


pera={1, 0, -1};
perb={0, 1, -1};

treata1={1, 0, -1};
treata2={0, -1, 1};
treata3={-1, 1, 0};
treata4={1, -1, 0};
treata5={0, 1, -1};
treata6={-1, 0, 1};

treatb1={0, 1, -1};
treatb2={1, -1, 0};
treatb3={-1, 0, 1};
treatb4={0, -1, 1};
treatb5={1, 0, -1};
treatb6={-1, 1, 0};

carrya1={0, 1, 0};
carrya2={0, 0, -1};
carrya3={0, -1, 1};
carrya4={0, 1, -1};
carrya5={0, 0, 1};
carrya6={0, -1, 0};

carryb1={0, 0, 1};
carryb2={0, 1, -1};
carryb3={0, -1, 0};
carryb4={0, 0, -1};
carryb5={0, 1, 0};
carryb6={0, -1, 1};

periodea = pera;
periodeb = perb;

do i=1 to 17;
  periodea = periodea // pera;
  periodeb = periodeb // perb;
```

```
end;

trta1 = treata1;
trta2 = treata2;
trta3 = treata3;
trta4 = treata4;
trta5 = treata5;
trta6 = treata6;
trtb1 = treatb1;
trtb2 = treatb2;
trtb3 = treatb3;
trtb4 = treatb4;
trtb5 = treatb5;
trtb6 = treatb6;
cara1 = carrya1;
cara2 = carrya2;
cara3 = carrya3;
cara4 = carrya4;
cara5 = carrya5;
cara6 = carrya6;
carb1 = carryb1;
carb2 = carryb2;
carb3 = carryb3;
carb4 = carryb4;
carb5 = carryb5;
carb6 = carryb6;

do i=1 to 2;
  trta1 = trta1 // treata1;
  trta2 = trta2 // treata2;
  trta3 = trta3 // treata3;
  trta4 = trta4 // treata4;
  trta5 = trta5 // treata5;
  trta6 = trta6 // treata6;
  trtb1 = trtb1 // treatb1;
  trtb2 = trtb2 // treatb2;
  trtb3 = trtb3 // treatb3;
  trtb4 = trtb4 // treatb4;
  trtb5 = trtb5 // treatb5;
  trtb6 = trtb6 // treatb6;
  cara1 = cara1 // carrya1;
  cara2 = cara2 // carrya2;
  cara3 = cara3 // carrya3;
  cara4 = cara4 // carrya4;
  cara5 = cara5 // carrya5;
  cara6 = cara6 // carrya6;
  carb1 = carb1 // carryb1;
  carb2 = carb2 // carryb2;
  carb3 = carb3 // carryb3;
  carb4 = carb4 // carryb4;
  carb5 = carb5 // carryb5;
  carb6 = carb6 // carryb6;
end;
```

```
treata = trta1 // trta2 // trta3 // trta4 // trta5 // trta6;
treatb = trtb1 // trtb2 // trtb3 // trtb4 // trtb5 // trtb6;

carrya = cara1 // cara2 // cara3 // cara4 // cara5 // cara6;
carryb = carb1 // carb2 // carb3 // carb4 // carb5 // carb6;

id={1, 1, 1};
id1={1, 1, 1};
id2={1, 1, 1};

do i=2 to 18;
  id2=id1#i;
  id=id // id2;
end;


sigma3={1.12 0.99 0.91,
        0.99 1.04 1.19,
        0.91 1.19 2.29};



i18=i(18);
i=i(54);

sig=i18@sigma3;
rsig=root(sig);

xx=x*(ginv(t(x)*x))*t(x);
x1x1=x1*(ginv(t(x1)*x1))*t(x1);
x2x2=x2*(ginv(t(x2)*x2))*t(x2);

c=xx - x1x1;
t=xx - x2x2;
  e= i - xx;


rc= trace(c);
rt= trace(t);
re= trace(e);


yn=normal(repeat(0,54,1));
y= t(rsig)*yn ;

y1= y[1:3,1];
y2= y[4:6,1];
y3= y[7:9,1];
y4= y[10:12,1];
y5= y[13:15,1];
y6= y[16:18,1];
y7= y[19:21,1];
y8= y[22:24,1];
```

```
y9= y[25:27,1];
y10= y[28:30,1];
y11= y[31:33,1];
y12= y[34:36,1];
y13= y[37:39,1];
y14= y[40:42,1];
y15= y[43:45,1];
y16= y[46:48,1];
y17= y[49:51,1];
y18= y[52:54,1];

g1= (y1 + y2 + y3)/3;
g2= (y4 + y5 + y6)/3;
g3= (y7 + y8 + y9)/3;
g4= (y10 + y11 + y12)/3;
g5= (y13 + y14 + y15)/3;
g6= (y16 + y17 + y18)/3;

ss1= (y1 - g1)*t(y1 - g1);
ss2= (y2 - g1)*t(y2 - g1);
ss3= (y3 - g1)*t(y3 - g1);
ss4= (y4 - g2)*t(y4 - g2);
ss5= (y5 - g2)*t(y5 - g2);
ss6= (y6 - g2)*t(y6 - g2);
ss7= (y7 - g3)*t(y7 - g3);
ss8= (y8 - g3)*t(y8 - g3);
ss9= (y9 - g3)*t(y9 - g3);
ss10= (y10 - g4)*t(y10 - g4);
ss11= (y11 - g4)*t(y11 - g4);
ss12= (y12 - g4)*t(y12 - g4);
ss13= (y13 - g5)*t(y13 - g5);
ss14= (y14 - g5)*t(y14 - g5);
ss15= (y15 - g5)*t(y15 - g5);
ss16= (y16 - g6)*t(y16 - g6);
ss17= (y17 - g6)*t(y17 - g6);
ss18= (y18 - g6)*t(y18 - g6);

ss=(ss1+ss2+ss3+ss4+ss5+ss6+ss7+ss8+ss9+ss10+
      ss11+ss12+ss13+ss14+ss15+ss16+ss17+ss18)/12;

kk=t(root(ss));
ikk=inv(kk);

vv=i108@ikk;

z=vv*y;
b=vv*x;
e1=vv*x1;
e2=vv*x2;

ee=b*(ginv(t(b)*b))*t(b);
e1e1=e1*(ginv(t(e1)*e1))*t(e1);
e2e2=e2*(ginv(t(e2)*e2))*t(e2);
```

```
  bc=ee - ele1;
  bt=ee - e2e2;
  be= i - ee;

  brc= trace(bc);
  brt= trace(bt);
  bre= trace(be);

  v=i18@ss;

  trcv= trace(c*v);
  trtv= trace(t*v);
  trev= trace(e*v);

  hc= (trcv**2)/(trace (c*v*c*v));
  ht= (trtv**2)/(trace (t*v*t*v));
  he= (trev**2)/ (trace (e*v*e*v));

  ssbc= t(z)*bc*z;
  ssbt= t(z)*bt*z;
  ssbe= t(z)*be*z;

  ssc= t(y)*c*y;
  sst= t(y)*t*y;
  sse= t(y)*e*y;

  fec=(bre*ssbc)/(brc*ssbe);
  fet=(bre*ssbt)/(brt*ssbe);

  fc=(re*ssc)/(rc*sse);
  ft=(re*sst)/(rt*sse);

ratioe=ssbc/ssbe;
ratioc=ssc/sse;


  fac=((rc*trev)/(re*trcv))*fc;
  fat=((rt*trev)/(re*trtv))*ft;

  p_ec=1 - probf(fec,brc,bre);
  p_et=1 - probf(fet,brt,bre);

  p_c=1 - probf(fc,rc,re);
  p_t=1 - probf(ft,rt,re);

  p_ac=1 - probf(fac,hc,he);
  p_at=1 - probf(fat,ht,he);


  C1={ ID Y U1 PERIODEA PERIODEB
       TREATA TREATB CARRYA CARRYB       };
  out={0  0  0 0 0 0  0 0 0     };
  create donnee from out [colname=C1];
  out=id||y||u1||periodea||periodeb||treata||treatb||carrya||carryb;
```

```
  setout donnee;
  append from out;

title ' ';
reset noname;

print  p_ac p_ec p_c;
print  p_at p_et p_t;

quit;

%GEE (DATA=donnee, XVAR=u1 periodea periodeb treata treatb carrya
                        carryb);
run;

%GEE (DATA=donnee, XVAR=u1 periodea periodeb treata treatb carrya
                        carryb, CORR=2, M=2);
run;

%GEE (DATA=donnee, XVAR=u1 periodea periodeb treata treatb carrya
                        carryb, CORR=4);
run;

%GEE (DATA=donnee, XVAR=u1 periodea periodeb treata treatb carrya
                        carryb, CORR=5, M=1);
run;

%GEE (DATA=donnee, XVAR=u1 periodea periodeb treata treatb carrya
                        carryb, CORR=6);
run;

  %end;

%MEND simul;

%simul(2000);
```

Appendix B

Table B.1 Empirical level of Type I error (%) for the test of treatment effects at the 1% nominal level

Covariance Matrix Σ	Number of subjects per sequence	Statistical methods							
		OLS	MFA	EGLS	GEE with working correlation matrix				
					Identity	Excl.	2 – Dept.	AR – 1	Unsp.
Sphericity	3	1.05	1.15	3.35	4.05	5.55	7.05	5.20	9.75
	6	0.95	0.95	1.75	1.85	2.75	2.90	2.65	4.45
Type AR – 1	3	1.40	0.90	2.90	4.70	5.65	7.65	4.75	10.50
	6	1.60	1.25	1.65	2.25	2.55	2.95	2.30	4.05
No structure	3	3.95	1.00	3.45	5.65	7.25	7.55	7.00	10.40
	6	4.50	1.10	1.80	3.60	3.70	3.95	3.45	4.45
	12	3.80	0.65	1.40	1.50	1.60	1.40	1.50	1.65
	18	--	--	--	1.05	1.15	1.35	1.45	1.55

Table B.2 Empirical level of Type I error (%) for the test of treatment effects at the 10% nominal level

		Statistical methods							
Covariance Matrix Σ	*Number of subjects per sequence*	*OLS*	*MFA*	*EGLS*	*GEE with with correlation matrix*				
					Identity	*Exch.*	$2-$ *Dept.*	$AR-1$	*Unsp.*
Sphericity	3	10.10	10.70	17.60	17.95	20.95	23.80	21.70	27.75
	6	10.00	10.30	13.95	13.60	14.75	15.50	15.15	19.15
Type $AR-1$	3	12.95	9.65	15.85	18.30	20.15	24.30	19.00	28.65
	6	13.40	9.95	13.75	14.50	14.80	16.50	15.05	17.95
No structure	3	21.10	10.00	17.50	19.80	24.60	24.65	23.00	27.45
	6	21.10	11.05	12.60	15.25	17.85	17.40	16.60	18.05
	12	18.00	10.35	11.20	12.75	12.40	12.75	12.65	13.90
	18	--	--	--	10.45	11.15	11.35	10.95	11.70

Table B.3 Empirical level of Type I error (%) for the test of carryover effects at the 1% nominal level

		Statistical methods							
Covariance Matrix Σ	*Number of subjects per sequence*	*OLS*	*MFA*	*EGLS*	*GEE with working correlation matrix*				
					Identity	*Exch.*	$2-$ *Dept.*	$AR-1$	*Unsp.*
Sphericity	3	1.20	1.85	3.20	3.70	7.15	7.70	8.00	10.15
	6	0.95	1.55	1.95	2.00	3.05	3.40	2.85	4.45
Type $AR-1$	3	0.80	1.30	3.20	3.25	6.35	7.85	6.65	10.70
	6	1.00	1.10	1.65	2.60	2.70	3.85	3.00	4.75
No structure	3	4.55	1.05	3.60	3.60	8.00	9.00	8.10	11.10
	6	3.45	0.50	1.90	1.60	2.55	2.90	3.15	3.50
	12	3.85	0.65	1.55	1.60	1.80	1.75	1.55	1.75
	18	--	--	--	1.05	1.50	1.55	1.50	1.50

Table B.4 Empirical level of Type I error (%) for the test of carryover effects at the 10% nominal level

		Statistical methods							
Covariance Matrix Σ	*Number of subjects per sequence*	*OLS*	*MFA*	*EGLS*	*GEE with with correlation matrix*				
					Identity	*Exch.*	$2-$ *Dept.*	$AR-1$	*Unsp.*
Sphericity	3	11.10	11.95	16.60	15.65	21.75	23.25	23.45	27.60
	6	9.95	11.15	12.55	12.45	15.75	16.80	16.05	17.65
Type $AR-1$	3	8.45	9.75	14.60	17.20	21.75	25.40	21.85	27.60
	6	9.85	10.95	13.20	12.70	16.60	17.50	16.05	18.30
No structure	3	20.30	9.80	16.65	18.50	24.50	25.60	24.05	27.80
	6	21.10	9.75	12.25	12.60	16.65	17.20	17.45	17.95
	12	19.95	10.50	10.90	12.25	13.80	13.35	13.15	13.65
	18	--	--	--	10.35	10.20	10.30	10.05	10.55

Bibliography

[1] Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis,* Second Edition, Wiley.

[2] Bellavance, F. (1994). *Étude des présupposés sous normalité dans les plans croisés suivie d'une analyse non paramétrique de certains d'entre eux,* Thèse de doctorat, Département de mathématiques et de statistique, Université de Montréal.

[3] Bellavance, F., Tardif, S. & Stephens, M. A. (1996). *Tests for the Analysis of Variance of Crossover Designs with Correlated Errors,* Biometrics, 52. 607-612.

[4] Box, G. E. P. (1954a). *Some theorems on quadratic forms applied in the study of analysis of variance problems. Effect of inequality of variance in the one-way classification,* Annals of Mathematical Statistics, 25, 290-302.

[5] Box, G. E. P. (1954b). *Some theorems on quadratic forms applied in the study of analysis of variance problems. Effects of inequality of variance and of correlations between errors in the two-way classification,* Annals of Mathematical Statistics, 25, 484-498.

[6] Crowder, M.J., (1995). *On the use of a working correlation matrix in using generalized linear models for repeated measures,* Biometrika, 82. 407-410.

[7] Crowder, M. J. & Hand, D. J. (1990). *Analysis of Repeated Measures,* Chapman and Hall.

[8] Diggle, P.J., Liang, K.-Y. & Zeger, S.L. (1994). *Analysis of Longitudinal Data,* Oxford Science Publications.

[9] Hyunh, H. & Feldt, L. S. (1970). *Conditions under which mean square ratios in repeated measurements designs have exact F-distributions,* Journal of the American Statistical Association, 65, 1582-1589.

[10] Jones, B. & Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials,* Chapman and Hall.

[11] Karim, R. (1989). *S.A.S. Macro GEE1,* Department of Biostatistics, Johns Hopkins University.

[12] Kleinbaum, D. G. & Kupper, L. L. (1993). *Analysis of Longitudinal Data with Epidemiologic Applications,* Course Manual.

[13] Laird, N.M. & Ware, J.H. (1982). *Random-effects models for longitudinal data,* Biometrics, 33. 133-158.

[14] Liang, K.-Y. & Zeger, S. L. (1986). *Longitudinal data analysis using generalized linear models,* Biometrika, 73, 13-22.

[15] McCullagh, P. (1983). *Quasi-likelihood functions,* Annals of Statistics, 11, 59-67.

[16] McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models,* Chapman and Hall.

[17] Rubin, D. B. (1976). *Inference and missing data,* Biometrika, 63, 81-92.

[18] Searle, S.R. (1987). *Linear models for unbalanced data,* Wiley.

[19] Serfling, R.J. (1980). *Approximation theorems of mathematical statistics,* Wiley series in probability and mathematical statistics.

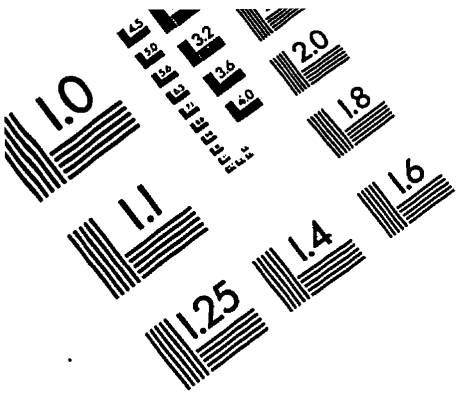
[20] Stiratelli, R., Laird, N. & Ware, J. (1984). *Random effects models for serial observations with binary responses,* Biometrics, 40, 961-971.

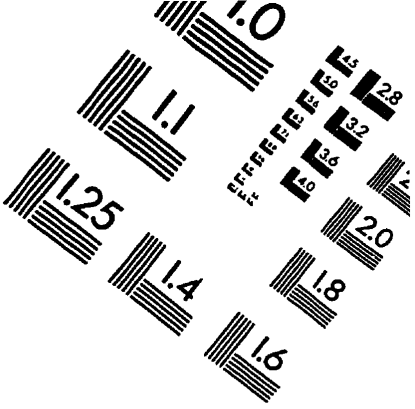
[21] Ware, J.H. (1985). *Linear models for the analysis of longitudinal studies,* American Statistician, 39. 95-101.

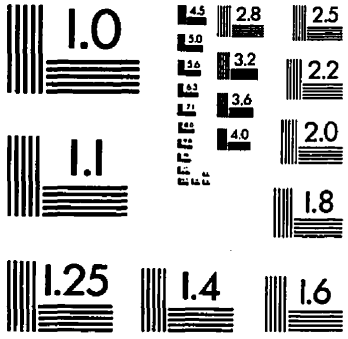
[22] Wedderburn, R. W. M. (1974). *Quasi-likelihood functions, generalized linear models, and Gauss-Newton method,* Biometrika, 61, 439-447.

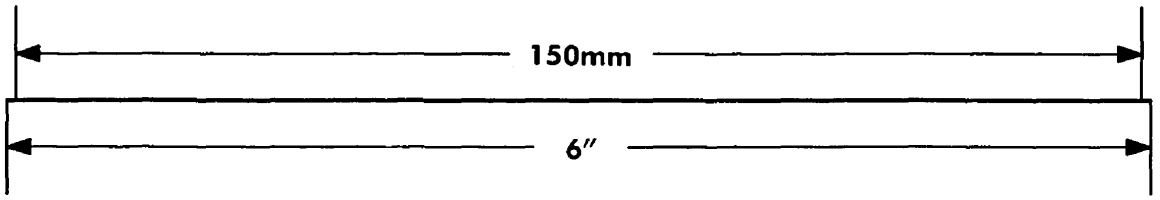
[23] Zeger, S. L. & Liang, K.-Y. (1986). *Longitudinal Data Analysis for Discrete and Continuous Outcomes,* Biometrics, 42, 121-130.

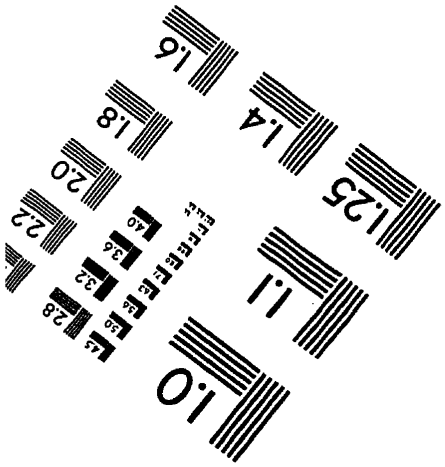
[24] Zeger, S. L., Liang, K.-Y. & Albert, P. S. (1988). *Models for Longitudinal Data: A Generalized Estimating Equation Approach,* Biometrics, 44, 1049-1060.

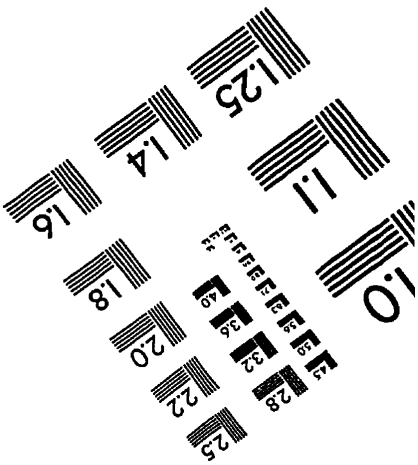












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989