A STATISTICAL STUDY OF THE NATURE OF PERIODONTAL DISEASE PROGRESSION

A thesis presented to the University of London

for the degree of Doctor of Philosophy

in the faculty of Science

by

Jonathan Abram Cutler Sterne

University College, London

ABSTRACT

The thesis is concerned with questions arising from recent disputes about the nature of periodontal disease progression. Relevant periodontal and statistical literature is reviewed.

Three models for periodontal disease progression, allowing for measurement error, are constructed; for (1) constant progression, (2) instantaneous bursts of activity and (3) varying, but non-instantaneous, rates of progression. When the covariance structures of the three models are examined, it becomes clear that they are hierarchical, with (1) being a limiting case of (2) and (2) a limiting case of (3).

The covariance structure estimation methods of Browne (*British Journal of Mathematical and Statistical Psychology* 37 62-83 (1984)) do not require restrictive assumptions about the underlying distribution. Software using these methods was written in APL. Simulation experiments were performed to examine the conditions under which it is possible to distinguish data from the three models.

The study of Löe et al. on Sri Lankan tea labourers (*Journal of Clinical Periodontology* 13 431-440 (1986)) is the largest longitudinal study of the natural history of periodontal disease. Preliminary analysis of these data reveals a large subject intraclass correlation, in contrast to recent claims made in the periodontal literature. The parameters and goodness of fit of models (1)-(3) for the data from this study are estimated, together with cross-validation coefficients and model fit indices to aid model selection.

A criterion for designing studies of periodontal disease progression might be the minimisation of the asymptotic variance of a parameter which distinguishes models (1) and (2). This is calculated, and the asymptotic properties of maximum-likelihood estimates under the minimising conditions are investigated.

Possibilities for further work are briefly discussed. It is to be hoped that the methods will be applied to other data collected in longitudinal studies of periodontal disease progression.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Professor Mervyn Stone, both for his help during the early stages of this thesis and also (and equally importantly) for his insistence that the problems addressed were those identified by me. I thank colleagues, particularly Dr Albert Kingman and Dr Harald Löe, at the National Institute for Dental Research, Bethesda, Maryland, for their hospitality during my visits there, and Professor Michael Browne for helpful comments and provision of software. I am grateful to Rob Hart for much helpful advice on APL programming and laser printing. I thank all my colleagues at the MRC Dental Research Unit for their friendship during my time there. Finally I thank Harriet Aston for her support and patience through the lengthy final stage of the work.

DEDICATION

To Max Aston Sterne, who came first.

CONTENTS

A - Appendix.

PAGE

PAGE

PAGE

TABLES

FIGURES

1

INTRODUCTION

Until the 1980s, the study of periodontal disease had attracted little interest from professional statisticians. The disease process was widely, though not universally, considered to be chronic and slowly progressing, and it was generally accepted that most tooth loss in people over the age of 35 years was caused by periodontal disease.

A series of publications by the group working at the Forsyth Dental Center in Boston, USA brought profound changes both in the view of the way in which the disease progressed, and types of data analysis considered to be appropriate to the study of the disease. They did this by the somewhat unfortunate means of proposing, using and publishing statistical methods so inappropriate that the periodontal journals ever since have been publishing papers by statisticians explaining the faults of these methods. There is now broad consensus on matters, such as allowing for subject effects when analyzing measurements made at different disease sites in the same mouth, which were for some time a matter of furious dispute.

Among the papers published by the Forsyth group between 1982 and 1985 was one entitled "New Concepts of Periodontal Disease Progression" (Socransky et al. 1984). This proposed that, rather than being a slow, chronic, continual process, the progression of the disease was in short, acute "bursts" of activity interspersed with periods of remission. The appealing nature of this idea was demonstrated by the speed with which the episodic nature of the disease became accepted by researchers. However the statistical basis for the observations on which the hypothesis was based was unsound.

The aim of the present study was to construct statistical models for periodontal disease progression, and to use these models as the basis for the development of methods which allow us to make inferences about the nature of disease progression. Three models for disease progression are constructed; respectively for constant progression, instantaneous bursts of activity and for varying, but non instantaneous progression. Their covariance structures are calculated and are shown, after simplification, to be hierarchical.

The estimation of covariance structures has been used for some years to make inferences about the effect of latent variables on multivariate data arising in psychology and econometrics, and a theoretical framework has recently been developed which allows inference to be made for a very general specification of covariance structure and in the absence of knowledge about the underlying distribution of the data. The estimation of covariance structures does not appear to have been used previously to distinguish between competing statistical models in the manner developed here.

The methods are applied to data from the study of Löe et al. (1978a) of the natural history of periodontal disease progression in Sri Lankan tea labourers. This is the largest long-term study of periodontal disease progression in the absence of therapy yet performed. A section on preliminary analysis of these data shows the inability of previously suggested methods to determine the nature of progression, and provides novel information on the magnitude of the subject intraclass correlation for this study.

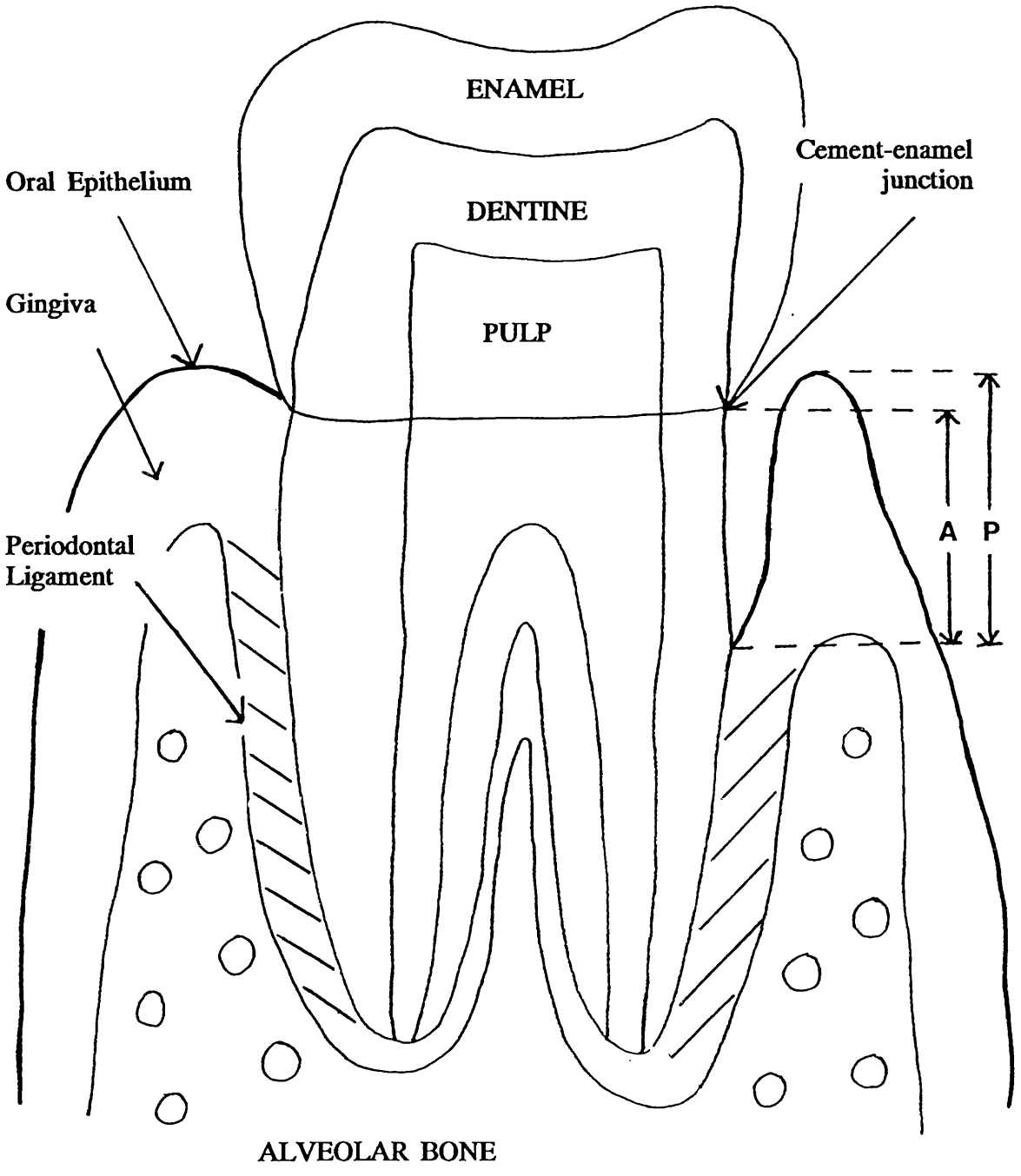
In the final sections we show how questions about the asymptotic properties of maximum likelihood estimation for non independently and identically distributed variables arise when criteria for the design of studies of periodontal disease progression are considered. Theorems for the consistency of maximum likelihood estimates in this case are presented and examples, based on the covariance structures of the models for disease progression, are given.

Throughout, column vectors are denoted by lower case Roman or Greek characters which are underlined. Matrices are denoted by upper case characters. The end of a proof is marked by the symbol ∎ at the right of the page.

2

REVIEW OF THE LITERATURE

FIGURE 2.1. THE PERIODONTIUM.



A - Attachment Level

P - Pocket Depth

2.1 THE NATURE OF THE PROGRESSION OF PERIODONTAL DISEASES

Periodontal diseases are characterised by the loss of the epithelial attachment of the tooth to the connective tissue of the gingivae, and the resorption of the alveolar bone which supports the tooth via the collagenous periodontal ligament. Figure 2.1 is a diagram of the periodontium, with the situation in health shown on the left side of the tooth and that in disease on the right.Clinically, the loss of attachment results in the apical extension of the junction epithelium, detachment of the collagen fibres of the ligament and thus an increasing distance between the cement-enamel junction on the tooth and nearest point of epithelial attachment. This attachment level is commonly measured by inserting a periodontal probe, graduated at 1mm intervals, into the space between the tooth and the gingivae. Change in attachment may be calculated by comparing attachment level measurements made on two or more occasions: if attachment level increases this change is described as loss of attachment. Periodontal disease may also be assessed by the measurement of alveolar bone levels using radiographs and sophisticated digital subtraction techniques are being developed for the assessment of the rate of bone loss over time. However we shall concentrate in this thesis on measurements made with periodontal probes, these being the source of most of the available data from longitudinal studies of periodontal disease progression.

The accepted view of the nature and prevalence of periodontal diseases has changed dramatically over the last decade. In a textbook published in 1984 (Carranza 1984), the following statements were made:

"All adults will at some point during their lifetime experience some deterioration of their periodontal structures"

"It has been shown that periodontal disease is responsible for approximately 50 per cent of the total tooth loss after age 15."

"As more people retain their teeth throughout their lifetime, and as the proportion of older people increases, more teeth will be at risk for periodontal disease. Hence the prevalence of destructive periodontal disease will likely increase in the future."

Six years later, it is generally accepted that periodontal diseases affect only a minority of the population, both in Western populations (Miller et al. 1987) and in populations not exposed to Western-style dental care (Baelum et al. 1988). There is now evidence that these diseases may be genetically linked (Michalowicz et al. 1989). Thus it may be that only individuals unable for some reason to mount an effective immune response to periodontal bacteria are at risk of developing forms of the disease severe enough to endanger their teeth.

The first longitudinal study of the natural history of periodontal disease was on two groups: of Norwegian academics and students and of Sri Lankan tea labourers. One of the first reports on this study (Löe et al 1978c), whose data will form the basis for the applied part of this thesis, stated that:

"It is apparent ... that the destruction of the periodontium progresses steadily over time. There may be periods of slow progress and periods in which the destructive

processes show acceleration. Indeed, different surfaces, teeth, dentitions and populations show different rates of progress during different age periods. These differences most likely reflect variations in both the pathogenic and the defense mechanisms. If these factors are not interfered with by treatment or home care, which was true for the Sri Lankan population, the disease progresses at a relatively even pace and there are no indications that this progress is not continuous."

It could not, however, be said that this view was universal. Hirschfield and Wasserman (1978), in a survey of tooth loss in periodontal patients who had been treated at least 15 years previously, stated that: "The disease process often followed a cyclically active pattern. Irregularly spaced cycles of destructive activity were evident in all response groups, even the well maintained group. Several advanced cases responded very well to treatment, with no teeth being lost for over 20 years, and then suffered rapid periodontal destruction, with the loss of many teeth. Many of the downhill and extreme downhill cases remained stable for years, with periods of destruction occurring sporadically."

Similarly, Newman (1979), in discussing the role of anaerobes in periodontal infections, proposed a model of quiescence/remission, then exacerbation, then quiescence/remission, and claimed that: "The cyclical nature of this disease has recently been substantiated by clinical observation and by cultural studies".

2.1.1 THE BURST HYPOTHESIS FOR PERIODONTAL DISEASE PROGRESSION

The paper in which the burst hypothesis was proposed (Socransky et al. 1984) can be seen as a culmination of a series of papers from the group at the Forsyth Dental Center, in which previously accepted paradigms, both for the nature of periodontal diseases and for appropriate statistical methods, were challenged. Whatever may be thought, after six years, of the quality of the views of the disease which were postulated, there can be no doubt that the statistical methods were seriously flawed.

At the root of the new ideas which were proposed was the concept that periodontal diseases are site-specific: that is that they occur independently at different disease sites in the same subject. Data from up to 192 sites in each subject (6 per tooth on 32 teeth) were therefore treated as independent: no allowance was made for subject effects. The papers were based mainly on a study in which 22 periodontal patients were studied at one-month intervals for between 9 and 23 months. At each examination repeat measurements were made at each site.

Goodson et al. (1982) performed linear regression on each of 1,155 sites (two per tooth) and, by calculating the ratio of the estimated slope to the standard error, determined whether the site had 'significantly' changed (using $p<0.01$ as the criterion). This was done by starting with the first three measurements, and increasing the number of points fitted until data from all appointments were included. They found that 82.8% of sites did not change, 5.7% became significantly deeper, and 11.5% became significantly shallower. They appeared unaware that by

performing multiple tests for each site they increased the probability of false positives. Interestingly, examination of the data which is presented by subject gives the clear impression that there is a marked subject effect, which could be tested by standard methods for the analysis of contingency tables. They concluded that "a dynamic condition of disease exacerbation and remission as well as periods of inactivity may be characteristic of periodontal disease.

Haffajee et al (1983a) noted that in most of the periodontal clinical trials which had been reported over the last 20 years, differences had been sought by looking for differences between mean values, for instance between groups of patients subjected to different therapies. They argued that this might obscure differences between different sites in the same patient. They presented data from a periodontal split-mouth clinical trial, and assessed which of several forms of statistical test was best by looking at which tests were sensitive to changes at a small number of sites. Different sites were treated as statistically independent.

This paper was typical of the series in that it raised genuine difficulties in the analysis of periodontal data but suggested answers which were statistically invalid. It is not the intention here to review the attempts which were made to correct these practices and to suggest alternatives. Sterne et al. (1990) review the subsequent attempts which have been made to provide appropriate methods for the analysis of these data.

Haffajee et al. (1983b) extended the idea of using linear regression to detect sites at which attachment level had changed, by comparing this with two further methods;

the tolerance method and the running median method. Using the tolerance method, the difference between pairs of measurements taken between successive visits is considered to be significantly different if three conditions are met. These are that first, a threshold difference of 2 population standard deviations is exceeded; second, a threshold difference of three standard deviations of the full mouth replicate measurements for the subject is exceeded and third, a threshold of three pooled standard deviations at the individual site is exceeded. For the method of running medians, successive medians of points taken three at a time were considered, and a difference was deemed to be shown by a change of greater than three population standard deviations between successive medians. Thus this paper attempted to find methods of data analysis which were able to detect sites at which changes had occurred over a short period of time. They concluded that the tolerance method was best suited to this purpose. Of course, they were only able to compare results between methods, having no external means of assessing whether a site had experienced breakdown.

Haffajee et al. (1983c) used the tolerance method to classify 3414 sites from the same group of 22 subjects as to whether they had showed disease activity. Various clinical signs were then assessed (again ignoring subject effects) for their association with disease activity. The introduction to this paper stated:

"In recent years, the concept of periodontal disease as a slow, progressive disease has been questioned. Longitudinal monitoring of individual sites has indicated that destruction occurs in relatively short periods of time which are followed by prolonged periods of inactivity."

The papers cited as proof of this claim were Goodson et al. (1982), Haffajee et al (1983b), and Socransky et al. (1984) (then in press). Thus the first two of these papers, which had no more than explored the possibility that this might be the case, were claimed as having shown something much stronger.

Socransky et al. (1984) provided the description of the burst model for periodontal diseases which was the starting point for this thesis. The basis of the paper was the comparison of the old concept of chronic destructive periodontitis which, it was suggested, brought to mind "terms such as slow, continual, progressive, inexorable and unremitting". This model was represented pictorially by constant progression whose rate varied from site to site. It was compared to a "random burst" model for periodontal diseases in which:

"certain sites within patients would be free of destructive periodontal disease throughout that individual's life. Other sites would demonstrate a brief active burst of destructive disease (which could take a few days to a few months) before going into a period of remission. The site may never demonstrate destructive activity again or could be subject to one or more bursts of activity at later time periods. The model suggests that prior history of disease would not necessarily make a site more likely to demonstrate further periods of activity nor would it exclude the occurrence of further destruction at that site."

A third, "asynchronous burst" model was also postulated. Here, it was supposed that multiple sites showed breakdown within a reasonably short period of an

individual's life with prolonged periods of remission, with the possibility of further waves of widespread destruction. It was emphasised that these models were as yet unproven hypotheses.

Four lines of evidence were put forward in support of the burst model hypothesis: 1) that some observed attachment loss rates were too fast or too slow to be consistent with the observed mean loss of attachment in individual subjects; 2) that large numbers of sites did not appear to change; 3) that animal studies indicate that disease does not progress in all lesions; and 4) that animal studies indicate that even severe experimentally induced perturbations at a site which induce rapidly destructive disease are soon brought under control.

Most of the evidence put forward in support of the first two lines of argument was, of course, based on Goodson et al. (1982) and Haffajee et al. (1983b), which we have already reviewed. An extra argument was adduced that measures of attachment level showed an exponential distribution, and that such distributions are used to model random processes such as radioactive decay. However papers by other authors were also cited by Socransky et al. (1984). Moskow (1978) reported on a patient who refused periodontal treatment and whose condition was substantially unchanged after ten years in which she had not received dental care. Selikowitz et al. (1981) examined routine bitewing radiographs taken at dental practices over a period of 10 years and concluded that in a majority of cases the rate of bone loss per year fluctuated, although it was not made clear how this conclusion had been reached. Hancock et al. (1981) noted that a cyclical nature of periodontitis had been proposed by Stanley (1955) and reviewed available methods

for the determination of periodontal disease activity. The relevance of animal models to the natural history of periodontitis in humans is uncertain (Page 1988), so it would clearly be unwise to base the 'burst' hypothesis solely on analogy with animal models.

It is not our purpose here to review in detail the battery of criticism of these papers which was advanced in subsequent years. Fidler (1984), Laster (1985), Blomqvist, (1985, 1987), Imrey (1986), Morrison and Kowalski (1986), Janssen et al. (1987), Osborne (1987), Fleiss et al (1987), Birkedal-Hansen (1988), Fleiss et al. (1988), Gunsolley and Best (1988) and Sterne (1988) are among those who have joined the attack.

We will, however, discuss two particular issues which will be of interest in sections 6 and 7: the issue of the standard deviation of measures of attachment level for subjects with periodontal disease, and the issue of the subject intraclass correlation coefficient for changes in attachment level at different sites. This is defined as $R_s = \sigma^2_s(\sigma^2_s+\sigma^2_{sws})^{-1}$, where σ^2_s is the component of variance between subjects and σ^2_{sws} is the component of variance for sites within subjects.

In a reproducibility study performed on 63 Danish men (Glavind and Löe 1967), the standard deviation of a single measurement of attachment level was estimated to be 0.41mm. Reproducibility data reported after the fourth examination in the Sri Lanka study (Löe et al 1978c) appears to give a standard deviation of around 0.53mm. Imrey (1986) took the standard deviation for subjects with periodontal disease based on a number of papers to be 0.8mm, although Goodson (1986)

claimed that this was an over estimate, with the true value being 0.55mm (Goodson et al 1986). Janssen et al. (1987) estimated the standard deviation of attachment level measurements in 13 patients with chronic periodontitis to be 0.74mm. They noted that the distribution was non-Gaussian, so that extreme values occurred more frequently than would be the case for the normal distribution, and that reproducibility was worse for measurements in deep than for shallow periodontal pockets.

The value of R_s was claimed by Haffajee et al. (1985) to have a median value over a number of groups of 0.07. This observation was used as justification for the practice in the papers discussed above of ignoring subject effects. It was pointed out by Fleiss et al. (1987) that Haffajee et al. had failed to take into account the component of variation due to measurement error. Fleiss et al. (1988) re-analysed the data of Goodson et al. (1982) by calculating the intraclass correlation coefficient where the data were sites, means of all sites for a tooth, and means of all sites for a quadrant. They showed that the intraclass correlations for change in attachment level increased when the observations used to calculate R_s were averages of increasing numbers of sites. This indicated that measurement error had obscured the values for individual sites. Fleiss et al. also found that the values of R_s increased as changes were calculated over longer time periods. They reported a value of 0.55 where R was calculated for the averages of all sites in a quadrant, and the time interval was for 9 months.

Criticism of the papers forming the basis of the burst hypothesis can be summarised in five points:

1. Sites within the same subject cannot be considered to be independent. This practice will lead to false positives in significance tests.

2. The substantial measurement error of attachment level measurements, and the fact that the frequency of extreme values will be higher than if the errors were normally distributed, must be taken into account when analyzing these data.

3. The rate of false positives (sites incorrectly classified as active) was substantially larger than assumed in these papers.

4. The methods of analysis used depended on assumptions about the nature of disease progression: they did not in themselves investigate the nature of disease progression.

5. There is a confusion between failure to achieve statistical significance (i.e. failure to be confident that the observation is not a false positive), and the assertion that the observation is not a false negative.

All this justifiable criticism of the statistical methods which were used in producing evidence means that one needs to be sceptical that the burst hypothesis has been demonstrated to be correct. However, the enthusiasm with which these ideas were taken up may mean that they fit in with the clinical impression of the disease which was held by many researchers (e.g. Hirsch and Clarke 1989). The plausibility of this theory and its wide acceptance is illustrated by the number of current publications

of longitudinal clinical studies which use methods of analysis which classify sites either as active or inactive, using the tolerance or similar methods. These are not only from the Forsyth group, (e.g. Haffajee et al. 1988, Dzink et al. 1988), but also from other researchers (most recently Bragd et al. 1987, Ramfjord et al. 1987, Jenkins et al. 1988, Listgarten et al. 1989, McCullough et al. 1989).

It is not sufficient, therefore, to dismiss the burst hypothesis on the grounds that it has not been proved to be true. Further, it seems clear that not all extreme changes observed in longitudinal studies are attributable to measurement error: Gunsolley and Best (1988) pointed out that the false positive rate for a population can be expressed as

$$P_{f+} = \frac{P(fp)\times(1-P(C+))}{P(T+)}$$

where P(fp) is the probability of false positive for an individual observation, P(T+) is the proportion of measured changes over time and P(C+) is the proportion of real changes. Thus an upper bound for the P_{f+} is given by P(fp)/P(T+) (this upper bound will be close to the true value if the proportion of true changes is close to zero. P(fp) for a 3mm loss of attachment level was estimated as 0.0074 for single measurements and 0.0014 for repeat measurements. This gave estimated false positive rates of 0.32 for single measurements and 0.15 for repeat measurements. P(T+) was estimated from the data of the Forsyth group. Thus, although the false positive rate was considerably higher than had been thought, it still appeared that at least 68% of measured changes in attachment level were due to real disease progression. Similarly, Lindhe et al. (1989) estimated the upper bound for the proportion of false positives in a study of untreated subjects as 53%. There is, therefore, evidence that real changes in attachment level took place during these

studies. To reject the constant progression model, however, one must also demonstrate that the underlying rate of progression at a site changed during the period of the study or, as suggested by Socransky et al. (1984) that the observed rate of progression at a number of sites, allowing for the effect of measurement error, is incompatible with the sites having retained the amount of attachment they still possess.

Even the observation that rates of observed attachment loss are inconsistent with a constant rate of progression need not lead to acceptance of an hypothesis that change occurs instantaneously or nearly so. Birkedal-Hansen (1988) noted the risk that methods of analysis such as the tolerance method might select for a particular disease entity by assuming that sites failing to reach the criterion for change had not undergone periodontal destruction. Similarly differing patterns of disease distribution (Haffajee and Socransky 1986) could be explained by a multitude of models for disease progression (Cohen and Ralls 1988). Manji and Nagelkerke (1989) noted that a Brownian motion model for changes in periodontal attachment might equally explain the observed changes in rates of progression. As discussed by Listgarten (1986) and Hausmann & Jeffcoat (1988), the change in attachment level between two time points represents the integrated sum of periodontal disease activity during the period. There is no available method for measuring the instantaneous rate of attachment loss.

Thus questions of how often the rate of progression changes, of how rapid is the rate of progression when a site experiences disease activity and of whether disease progression truly consists of short periods of activity and long periods of quiescence

remain open. The probability of sampling during a burst is in proportion to its duration, so that if they are indeed virtually instantaneous then they can never be observed. If the values of the explanatory variables change only during an episode of activity then the shorter the duration of the episodes, the smaller the chance of detecting associations of interest. Clarification of the nature of disease progression is clearly of enormous importance in the study of the aetiology of periodontal diseases.

2.2 ESTIMATION OF COVARIANCE STRUCTURES.

In this section we review procedures available for the estimation of the parameters of covariance structures. The notation used is based on that of Browne (1984) (see below). We consider a p×p population covariance matrix Σ obtained from N=n+1 independent observations on a p×1 vector variate $\underline{x}$. A structural model for a covariance matrix is defined as a p×p matrix-valued function

(2.1) $\Sigma=\Sigma(\underline{\gamma})$

where $\underline{\gamma}$ is a q×1 vector $\underline{\gamma}\in G$ and G is a parameter set contained in R^q. We may similarly specify a structural model for a correlation matrix $R(\underline{\gamma})$.

2.2.1 MODELS FOR COVARIANCE STRUCTURES.

Models for covariance structures have been studied predominantly in the social sciences, but interest in tests on the nature of Σ goes back to Wilks (1946) who considered hypotheses of equal means, equal variances and equal covariances of a multivariate normal sample, and Votaw (1948) who considered a more complicated set of twelve hypotheses of "compound symmetry" of the means, variances and covariances of a normal multivariate population. Campbell and Fiske (1959) considered how inference on the validity of tests might be derived from a multitrait multimethod correlation matrix, which presents the intercorrelations resulting when each of several different traits is measured by each of several methods. Bock (1960) showed how a covariance structure arises from a model for the response to two psychological tests, and gave a method for testing hypotheses about the model. He gave the name "structural analysis" to a method which attempts to make causal

statements about test performances by assigning to definite sources the covariation between psychological tests. Bock and Bargmann (1966) presented a general method for estimating variance components arising from latent variable models where the experimental design has one random way of classification and a possibly unbalanced fixed classification. Latent variable models were restricted by assuming that (I): the latent variables are uncorrelated, (II): (I) and the errors are uncorrelated but not homoscedastic, and (III): (II) and the errors are uncorrelated and homoscedastic. MLE's and likelihood ratio tests of the goodness of fit of the models were derived. Wiley et al. (1973) presented a sequence of eight models based on varying degrees of restrictive assumptions. These include the three restrictions suggested by Bock and Bargmann (1966).

Srivastava (1966) noted that if

(2.2) $\underline{x} = \alpha_1\underline{z}_1 + \alpha_2\underline{z}_2 + \ldots + \alpha_q\underline{z}_q$,

where $\underline{z}_i$ (with covariance matrix Σ_i) are mutually independent, and α_i are scalars, then Σ has the linear structure

(2.3) $\Sigma = b_1\Sigma_1 + \ldots + b_q\Sigma_q$,

where $b_i = \alpha_i^2$. He studied the problem of obtaining the likelihood ratio statistic for the testing the hypotheses that Σ_0 has linear structure. Anderson (1969) derived the MLE's and a likelihood ratio test for the hypothesis that either Σ or its inverse is a linear combination, while Anderson (1973) gave, in addition, a method for the iterative solution of these equations and conditions for the asymptotic efficiency of these estimates are given. McDonald (1974, 1975), gave MLE's of the free parameters, and an asymptotic likelihood-ratio test, for the hypothesis that one or more elements of a covariance or correlation matrix are zero, and/or that two or

more of its elements are zero. These hypotheses are contained in the class of linear covariance structures.

The major application of covariance structure estimation has been in the area of factor analysis and, more generally, structural equation modelling. The development of these models has been pioneered by Joreskog (1970, 1973, 1978, 1981). We make no attempt to review the vast literature on the application of these models in psychology and econometrics. However we will review the most general form of the linear structural relations (LISREL) model as described by Joreskog (1981). A computer programme, LISREL (Joreskog and Sorbom 1988) was devised to allow the use of these models and has been updated regularly to allow the use of ever more general methods.

The LISREL model considers random vectors $\underline{\eta}' = (\eta_1,\eta_2,...,\eta_m)$ and $\underline{\xi}' = (\xi_1,\xi_2,...,\xi_n)$ of latent dependent and independent variables respectively and the following system of linear structural relations:

(2.4) $B\underline{\eta} = \Gamma\underline{\xi} + \underline{\zeta}$

where B (m×m) and Γ(m×n) are coefficient matrices and $E[\underline{\zeta}] = \underline{0}$. It is further assumed that $\underline{\zeta}$ is uncorrelated with $\underline{\xi}$ and that B is nonsingular.

The vectors $\underline{\eta}$ and $\underline{\xi}$ are not observed but instead vectors $\underline{y}' = (y_1,y_2,...,y_p)$ and $\underline{x}' = (x_1,x_2,...,x_q)$ are observed such that

(2.5) $\underline{y} = \Lambda_y\underline{\eta} + \underline{\varepsilon}$,

(2.6) $\underline{x} = \Lambda_x\underline{\xi} + \underline{\delta}$

where $\underline{\varepsilon}$ and $\underline{\delta}$ are vectors of errors of measurement in $\underline{y}$ and $\underline{x}$ respectively. We take $\underline{y}$ and $\underline{x}$ to be measured as deviations from their means. The matrices Λ_y (p×m) and Λ_x (q×n) are regression matrices of $\underline{y}$ on $\underline{\eta}$ and of $\underline{x}$ on $\underline{\xi}$ respectively. The errors of measurement are assumed to be uncorrelated with $\underline{\eta}$, $\underline{\xi}$ and $\underline{\zeta}$ but may be correlated among themselves.

Let Φ (n×n) and Ψ (m×m) be the covariance matrices of $\underline{\xi}$ and $\underline{\zeta}$ respectively, and let Θ_ε and Θ_δ be the covariance matrices of $\underline{\varepsilon}$ and $\underline{\delta}$ respectively. Then it follows from the above assumptions that the covariance matrix Σ ((p+q)×(p+q)) of $\underline{z} = (\underline{y}',\underline{z}')'$ is

$$(2.7)\quad \Sigma = \begin{bmatrix} Q & \Lambda_y B^{-1}\Gamma\Phi\Lambda_z' \\ \Lambda_x\Phi\Gamma' B'^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_z' + \Theta_\delta \end{bmatrix}$$

where

$$(2.8)\quad Q = \Lambda_y(B^{-1}\Gamma\Phi\Gamma' B'^{-1} + B^{-1}\Psi B'^{-1})\Lambda_y' + \Theta_\varepsilon$$

The elements of Σ are functions of the elements of Λ_y, Λ_x, B, Γ, Φ, Ψ, Θ_ε and Θ_δ. In applications some of these elements are fixed and equal to assigned values. In particular, this is so for elements of Λ_y, Λ_x, B and Γ, but allowance is made for fixed values in the other matrices also. For the remaining nonfixed elements of the eight parameter matrices one or more subsets may have identical but unknown values. Thus the elements in Λ_y, Λ_x, B, Γ, Φ, Ψ, Θ_ε and Θ_δ are of three kinds:

(i) <u>fixed parameters</u> that have been assigned given values

(ii) <u>constrained parameters</u> that are unknown but equal to one or more other parameters, and

(iii) free parameters that are unknown and not constrained to be equal to any other parameter.

Applications of the LISREL model or restrictions of it include factor analysis, measurement models such as those described above and path analysis.

It has become clear that general structural models may be expressed in a variety of ways. Interrelations among these models were reviewed by Bentler and Weeks (1979). McDonald (1978) gave an alternative general form for covariance structural models in terms of matrix products, together with first and second derivatives with respect to the parameters which allow estimation and testing via a variety of methods. He showed that this form included as special cases many of the previously proposed models for covariance structures, including the LISREL model. However Bentler and Weeks (1980) noted that this specification in turn is a special case of LISREL, and that each may be considered as generalizations of another, previously presented, form for structural models. They stated that "there appears to be a circularity regarding the generality of models that allows different models to be considered as special cases of each other".

Probably the most general form for structural models was given by McArdle and McDonald (1984), who described a compact general form which they named the Reticular Action Model (RAM). They showed that this form includes and extends other general forms for structural models such as LISREL as well as those of McDonald (1978) and Bentler and Weeks (1980). There is however clearly

competition between the authors of different models: the claims of McArdle and McDonald are disputed by Bentler and Weeks (1985).

Having discussed the specification of models of covariance structures, we now examine the estimation and testing of these models.

2.2.2 MAXIMUM LIKELIHOOD ESTIMATION OF COVARIANCE STRUCTURES UNDER THE ASSUMPTION OF NORMALITY.

In this section we present estimates under the assumption that $\{\underline{x}_i\}$ (i=1,...,N) have a p-variate normal distribution with expectation $\underline{\mu}$, where $\underline{\mu}$ is a p×1 vector of unknown constants and variance-covariance matrix $\Sigma(\underline{\gamma})$. The log-likelihood is then given by:

(2.9) $L_N(\underline{\mu},\underline{\gamma}) = -\frac{1}{2}N\log|\Sigma_{(\underline{\gamma})}| - \frac{1}{2}N\mathrm{tr}(\Sigma(\underline{\gamma})^{-1}S) - \frac{1}{2}N(\bar{\underline{x}}-\underline{\mu})'\Sigma(\underline{\gamma})(\bar{\underline{x}}-\underline{\mu}) + \text{const}$

where

(2.10) $S = N^{-1}\sum_{i=1}^{N}(\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$, is the sample covariance matrix, and

(2.11) $\bar{\underline{x}} = N^{-1}\sum_{i=1}^{N}\underline{x}_i$

As is well known, (2.9) is maximised with respect to $\underline{\mu}$ at $\hat{\underline{\mu}} = \bar{\underline{x}}$, so that the reduced likelihood function is then given by:

(2.12) $L_N(\underline{\gamma}) = -\frac{1}{2}N\log|\Sigma_{(\underline{\gamma})}| - \frac{1}{2}N\mathrm{tr}(\Sigma(\underline{\gamma})^{-1}S) + \text{const}$

Maximisation of (2.12) is equivalent to maximisation of the Wishart likelihood function $f(\underline{\gamma})$ for the multivariate distribution of S, defined by:

(2.13) $f(\underline{\gamma})=\log|\Sigma(\underline{\gamma})|-\log|S|+\mathrm{tr}(S(\Sigma(\underline{\gamma})^{-1})$

Maximum-likelihood estimates (MLEs) for parameters of the multivariate normal distribution with covariance matrices which are linear combinations of known matrices were derived by Anderson (1969), while general solutions are given by (e.g.) Joreskog (1978). We write:

(2.14) $\partial\Sigma(\underline{\gamma})/\partial\gamma_i = \Sigma^{(i)}$.

and also denote the MLEs of $\underline{\gamma}$ by $\hat{\underline{\gamma}}$, $\Sigma(\underline{\gamma})$ evaluated at $\underline{\gamma} = \hat{\underline{\gamma}}$ by $\hat{\Sigma}$, $\Sigma^{(i)}$ evaluated at $\underline{\gamma} = \hat{\underline{\gamma}}$ by $\hat{\Sigma}^{(i)}$ etc. The maximum-likelihood estimates are the joint solutions of the equations:

(2.15) $0 = -\mathrm{tr}(\hat{\Sigma}^{-1}\hat{\Sigma}^{(i)}) + \mathrm{tr}(\hat{\Sigma}^{-1}\hat{\Sigma}^{(i)}\hat{\Sigma}^{-1}S)$ $(k=1,...q)$.

If $\Sigma(\underline{\gamma})$ is a linear combination of known matrices, $\Sigma(\underline{\gamma}) = \gamma_1 B_1 + ... + \gamma_q B_q$, then $\Sigma^{(i)} = B_i$. Anderson (1973) gives a numerical method for the calculation of the MLEs in this case.

Under the assumption that $\{\underline{x}_i\}$ are normally distributed, the estimates are consistent and asymptotically efficient as $N \to \infty$. That is:

(2.16) $\sqrt{N}\hat{\underline{\gamma}} \sim N(\underline{\gamma}, I^{-1}(\underline{\gamma}))$, where:

(2.17) $I(\underline{\gamma}) = E[\partial L_1(\underline{\gamma})/\partial\underline{\gamma}]^2 = -E[\partial^2 L_1(\underline{\gamma})/\partial(\underline{\gamma})^2]$ (Lehmann 1983).

The information matrix $I(\underline{\gamma})$ is given by:

(2.18) $(I(\underline{\gamma}))_{jk} = \frac{1}{2}\mathrm{tr}(\Sigma(\underline{\gamma})^{-1}\Sigma^{(j)}\Sigma(\underline{\gamma})^{-1}\Sigma^{(k)})$

Anderson (1973) gave this result for linear covariance structures: it holds for the more general definition (2.1).

These results will be used in section 8, when the asymptotic behaviour of the estimates of the parameters of models for the progression of periodontal disease will be investigated.

2.2.2.1 MAXIMUM LIKELIHOOD ESTIMATION WITH STRUCTURED MEANS.

More generally, we may assume that the mean vector $\underline{\mu}$ depends on a structural model $\underline{\mu} = \underline{\mu}(\underline{\varsigma})$ where $\underline{\varsigma}$ and $\underline{\gamma}$ are disjoint vectors of parameters, or that both $\underline{\mu}=\underline{\mu}(\underline{\gamma})$ and $\Sigma=\Sigma(\underline{\gamma})$ depend on the same parameter vector. This latter, most general case is treated by Browne (1988).

For the models of periodontal disease progression developed in subsequent sections we shall require only that $\underline{\mu} = \mu\underline{\tau}$, where $\underline{\tau}$ is a known vector and μ is an unknown parameter. The MLEs and information matrix in this case are easily derived. It is then found that the MLEs are the joint solutions of (2.19-21) where:

(2.19) $0 = -tr(\dot{\Sigma}^{-1}\dot{\Sigma}^{(i)}) + tr(\dot{\Sigma}^{-1}\dot{\Sigma}^{(i)}\dot{\Sigma}^{-1}\dot{C}^{*})$ $(k=1,...q)$.

(2.20) $\hat{\mu} = N^{-1}\sum_{i=1}^{N}\underline{\tau}'\dot{\Sigma}^{-1}\underline{x}_i(\underline{\tau}'\dot{\Sigma}^{-1}\underline{\tau})^{-1}$

(2.21) $C^{*} = \sum_{i=1}^{N}(\underline{x}_i - \hat{\mu}\underline{\tau})(\underline{x}_i - \hat{\mu}\underline{\tau})'$

Also, $\hat{\mu}$ and $\hat{\underline{\gamma}}$ are still asymptotically independent, so that the information matrix of $\underline{\gamma}$ is as in (2.18) and:

(2.22) $I(\mu) = \underline{\tau}'\Sigma(\underline{\gamma})^{-1}\underline{\tau}$

We have thus presented MLEs and the information matrix for the parameters of an arbitrary covariance structure, and have seen how these models can be extended to allow in addition for a structural model for the mean vector.

2.2.3 GENERALIZED LEAST SQUARES ESTIMATION OF COVARIANCE STRUCTURES

In a series of papers, Browne (1974, 1982, 1984) has proposed the estimation of covariance structures by generalized least squares. Since these are the methods used in subsequent sections to fit models for periodontal disease progression, we now summarise these results in some detail.

We have N=n+1 independently and identically distributed p-vectors $\underline{x}_k$ with mean $\underline{\mu}$ and variance-covariance matrix Σ_0. We denote by $\underline{r}$ = vecs(R) the $p^* \times 1$ vector formed from the $p^* = \frac{1}{2}p(p+1)$ non-duplicated elements of a symmetric matrix R. Thus $\underline{r}' = (r_{11}, r_{12}, r_{22}, r_{13}, r_{23}, r_{33} \ldots)$, where r_{ij} represents the (i,j)th element of R. We denote by S the usual unbiased estimator of the population covariance matrix,

$$S = n^{-1} \sum_{i=1}^{N} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$$

Write $\underline{s}$ = vecs(S), σ_0 = vecs(Σ_0). If $\underline{\delta}_* = \sqrt{n}(\underline{s} - \sigma_0)$, the finite sample distribution of $\underline{\delta}_*$ has a null mean vector and covariance matrix $Y = \text{cov}(\underline{\delta}_*, \underline{\delta}_*)$ with typical element:

$$(2.23) \quad (Y)_{ij,kl} = n \, \text{cov}((S-\Sigma_0)_{ij}, (S-\Sigma_0)_{kl})$$

$$= (\Sigma_0)_{ik}(\Sigma_0)_{jl} + (\Sigma_0)_{il}(\Sigma_0)_{jk} + (n/N)\kappa_{ijkl}$$

where κ_{ijkl} is a fourth-order cumulant given by

$$\kappa_{ijkl} = \sigma_{ijkl} - (\Sigma_0)_{ij}(\Sigma_0)_{kl} + (\Sigma_0)_{ik}(\Sigma_0)_{jl} + (\Sigma_0)_{il}(\Sigma_0)_{jk}$$

$$\sigma_{ijkl} = E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)] \quad (\mu_i = E[x_i])$$

In many situations the exact finite sample distribution of δ_* is not known but may be approximated by the asymptotic distribution as $n \to \infty$, which is multivariate normal with null mean vector and covariance matrix

(2.24) $\bar{Y} = \lim_{n\to\infty} \mathrm{cov}(\underline{\delta}_n, \underline{\delta}_n')$, with

(2.25) $(\bar{Y})_{ij,kl} = (\Sigma_0)_{ik}(\Sigma_0)_{jl} + (\Sigma_0)_{il}(\Sigma_0)_{jk} + \kappa_{ijkl}$

We write $\underline{\sigma} = \underline{\sigma}(\gamma)$ where $\underline{\sigma}(\gamma) = \mathrm{vecs}(\Sigma(\gamma))$. The Jacobian matrix of $\underline{\sigma}(\gamma)$ is the $p^* \times q$ matrix-valued function of γ:

(2.26) $\Delta = \Delta(\gamma) = \partial\underline{\sigma}/\partial\gamma'$.

The model (defined at the start of this section) is said to hold if there exists a $\gamma_0 \in G$ such that $\Sigma_0 = \Sigma(\gamma_0)$. Given a sample covariance matrix S, an estimate $\hat{\gamma}$ of γ may be obtained by minimising a discrepancy function. A discrepancy function (Browne 1982) is a scalar valued function $F(S,\Sigma)$ of two $p \times p$ symmetric matrices S and Σ such that:

(i) $F(S,\Sigma) \geq 0$

(ii) $F(S,\Sigma) = 0$ if and only if $S = \Sigma$

(iii) $F(S,\Sigma)$ is a twice continuously differentiable function of S and Σ.

We have

(2.27) $F(S,\Sigma(\hat{\gamma})) = \min_{\gamma \in G} F(S,\Sigma(\gamma))$

and we write $\hat{\Sigma} = \Sigma(\hat{\gamma})$. A more general definition of the population value, γ_0, of the parameter vector, γ, which is still valid if Σ_0 can only be approximated by the structural model, is to regard γ_0 as the value of γ which minimises $F(\Sigma_0,\Sigma(\gamma))$. We represent $\Sigma(\gamma_0)$ by $\dot{\Sigma}_0$ (which need not be equal to Σ_0), i.e.

(2.28) $\min_{\gamma \in G} F(\Sigma_0,\Sigma(\gamma)) = F(\Sigma_0,\Sigma(\gamma_0)) = F(\Sigma_0,\dot{\Sigma}_0)$

Browne deals predominantly with quadratic form discrepancy functions of the type:

(2.29) $F(S,\Sigma(\gamma)|U) = (\underline{s}-\sigma(\gamma))'U^{-1}(\underline{s}-\sigma(\gamma))$

where U is a $p^{*}\times p^{*}$ positive definite matrix. In many applications U is a stochastic matrix which converges in probability to a positive definite matrix $\bar{U}$ as $n\rightarrow\infty$. The matrix U is said to be correctly specified if $\bar{U} = \bar{Y}$. In this case generalized least squares (GLS) estimators have minimum asymptotic variances. The Wishart likelihood $f(\gamma)$ given in (2.13) is a discrepancy function: other examples were considered by Swain (1975) in the context estimating procedures for factor analysis.

2.2.3.1 GENERALIZED LEAST SQUARES ESTIMATION UNDER THE ASSUMPTION OF NO KURTOSIS.

Browne (1974) demonstrates various properties of generalized least squares estimators under the assumption that all fourth-order cumulants κ_{ijkl} are zero. This is the case if, but not only if, the $\underline{x}_i$ have a multivariate normal or elliptical distribution. In this case, we say that the distribution of $\underline{x}$ has no kurtosis, and we have:

(2.30) $Cov(s_{ij},s_{kl}) = n^{-1}(\sigma_{0ik}\sigma_{0jl} + \sigma_{0il}\sigma_{0jk})$

and

(2.31) $Y_{ij,kl} = (\Sigma_0)_{ik}(\Sigma_0)_{jl} + (\Sigma_0)_{il}(\Sigma_0)_{jk}$

It is then desirable for a typical element of U to have the corresponding form:

(2.32) $U_{ij,kl} = (V)_{ik}(V)_{jl} + (V)_{il}(V)_{jk}$

where V is a positive definite stochastic matrix converging in probability to a positive definite matrix $\bar{V}$ as $n \to \infty$. If this is the case then (Browne 1974, equation (24)) the quadratic form discrepancy function $F(S,\Sigma|U)$ is equal to:

(2.33) $F(S,\Sigma|V) = \frac{1}{2}tr[(S-\Sigma(\gamma))V^{-1}]^2$

which is more easily computed. A possible choice for V is V=S, so that $\bar{V}=\Sigma_0$. If the distribution of $\underline{x}$ has no kurtosis then also $\bar{U} = \bar{Y}$.

Browne proves the following propositions under the assumption of no kurtosis and the regularity conditions:

(a) All $\sigma_{ij}(\gamma)$ ($\sigma_{ij}(\gamma) = \Sigma(\gamma)_{ij}$) and partial derivatives of the first 3 orders with respect to γ exist and are continuous in a neighbourhood of $\gamma=\gamma_0$.

(b) $\Delta(\gamma_0)$ is of full column rank.

(c) γ_0 is identified, i.e. $\Sigma(\gamma_0) = \Sigma(\gamma_1)$ implies $\gamma_0 = \gamma_1$, and

(d) $\Sigma(\gamma_0)$ is positive definite.

Proposition 1.1: The GLS estimators are consistent

Proposition 1.2: The limiting distribution of a GLS estimator $\hat{\gamma}$ is multivariate normal with mean vector $E[\gamma] = \gamma_0$, and covariance matrix

(2.34) $Cov(\hat{\gamma},\hat{\gamma}')=2n^{-1}\{\Theta(\bar{V})\}^{-1}\{\Theta(\bar{V}\Sigma_0\bar{V})\}\{\Theta(\bar{V})\}^{-1}$

where $\Theta(\bar{V})$ is a q×q matrix function of $\bar{V}$ defined by $\Theta(\bar{V}) = \Delta'(\bar{V}\otimes\bar{V})\Delta$ with typical element $\Theta(\bar{V})_{ij} = tr(\Sigma_0^{(i)}\bar{V}\Sigma_0^{(j)}\bar{V})$.

Proposition 1.3: The asymptotic dispersion matrix of a GLS estimator is bounded below (in the sense that A≥B if A-B is positive semi-definite) by $2n^{-1}\{\Theta(\Sigma_0^{-1})\}^{-1}$. This bound is attained, and $\hat{\underline{\gamma}}$ is a best generalized least squares (BGLS) estimator, if $V = \kappa\Sigma_0^{-1}$ ($\kappa>0$).

Proposition 1.4: Let Ω denote the information matrix based on the limiting distribution of S. Then

$\lim_{n\to\infty} n[2n^{-1}\{\Theta(\Sigma_0^{-1})\}^{-1} - \Omega^{-1}]=0.$

Propositions 1.1-1.4 show that, as long as V is a consistent estimator of a multiple of Σ_0^{-1}, the asymptotic properties of the estimates are the same as those of MLEs based on the assumption of a normal distribution for $\{x_i\}$. The asymptotic dispersion matrix $2n^{-1}\{\Theta(\Sigma_0^{-1})\}^{-1}$ of a BGLS estimator is equal to that of the MLE (equation 2.18). However the asymptotic behaviour of BGLS estimators depends only on regularity conditions (a-d) and assumptions about the fourth-order moments of the distribution (equations 2.31 and 2.32) and the asymptotic normality of S. As pointed out above, the assumption of no kurtosis holds if, but not only if, the underlying distribution is normal.

Proposition 1.5: If $\bar{V} = \Sigma_0^{-1}$ and $\Sigma_0 = \Sigma(\underline{\gamma}_0)$, the limiting distribution of $nF(\underline{\gamma}|V) = 2^{-1}n\ tr[(S-\Sigma(\underline{\gamma}))V]^2$ is chi-square with p(p+1)/2 - q degrees of freedom.

It is also shown that GLS estimators converge in probability to Maximum Wishart Likelihood (MWL) estimators based on minimising $f(\underline{\gamma})$ as defined in (2.13), and that the functions $nF(\hat{\underline{\gamma}}|(\Sigma(\hat{\underline{\gamma}})^{-1}))$ and $nf(\underline{\gamma})$ have asymptotically equivalent properties. We have already noted that $f(\underline{\gamma})$ is itself a discrepancy function. These results are generalized by Shapiro (1985a) (see below). We will refer henceforth to these estimators as GLS (nk) estimators, in order to distinguish them from the asymptotically distribution-free GLS estimators which we now describe.

2.2.3.2 ASYMPTOTICALLY DISTRIBUTION-FREE ESTIMATION OF COVARIANCE STRUCTURES.

Browne (1982, 1984) presented results which generalized the above to the situation where the underlying distribution does not have the structure of (2.30). He uses the following regularity conditions:

(R1) $\bar{Y}$ is positive definite.

(R2) $F(\Sigma_0,\Sigma(\underline{\gamma}))$ has a unique minimum at $\underline{\gamma} = \underline{\gamma}_0$ ($\underline{\gamma}$ is 'conditionally identified' with respect to $F(.,.)$.

(R3) $\underline{\gamma}_0$ is an interior point of G.

(R4) $\Delta_0 = \Delta(\underline{\gamma}_0)$ is of full column rank.

(R5) $\|\Sigma_0-\Sigma(\underline{\gamma}_0)\|$ is $O(\sqrt{n^{-1}})$.

(R6) The parameter set G is closed and bounded. (implied by:

(R6*) Given any S, $F(S,\Sigma)\rightarrow\infty$ if $\|\Sigma\| =\rightarrow\infty$ ($\|\Sigma\|=\sqrt{(tr(\Sigma^2))}$) and $\|\Sigma(\underline{\gamma})\|\rightarrow\infty$ if $\|\underline{\gamma}\|\rightarrow\infty$ ($\|\underline{\gamma}\| = \sqrt{\underline{\gamma}'\underline{\gamma}}$)).

(R7) $\Delta(\underline{\gamma})$ and, consequently, $\Sigma(\underline{\gamma})$ are continuous functions of $\underline{\gamma}$.

Proposition 2.1: A minimum discrepancy function estimator defined in (2.27) is a consistent estimator for $\underline{\gamma}_0$ as defined by (2.28).

Shapiro (1984) investigated the consistency of estimators in the analysis of covariance structures $\underline{\Sigma}(\underline{\gamma})$, where given a sample of size n we have an estimate $\underline{S}_n$ of the population value $\underline{\Sigma}_0$. It was assumed that S_n converges to Σ_0 in probability. An example was given where in spite of the identifiability of the parameters. the associated estimator $\dot{\underline{\gamma}}$ which minimises the discrepancy function is inconsistent. Consistency was shown to hold under the additional condition of compactness of the parameter space G. In many practical situations the assumption of compactness of G does not hold (for instance if $G=R^q$). However this assumption can be replaced by the condition of inf-boundedness: There exist a number $\alpha>F(\underline{\Sigma}_0,\underline{\Sigma}(\underline{\gamma}_0))$ and a compact subset G^* of G such that $\{\underline{\gamma}\in G:F(S,\Sigma(\underline{\gamma}))<\alpha\} \subseteq G^*$. If the model holds, i.e. if $\underline{\Sigma}_0 = \underline{\Sigma}(\underline{\gamma}_0)$ then this condition assumes the simple form of boundedness: There does not exist an unbounded sequence $\{\underline{\gamma}_n\}\subset G$ such that $\underline{\Sigma}(\underline{\gamma}_n)\rightarrow\underline{\Sigma}(\underline{\gamma}_0)$.

Proposition 2.2: If $\underline{\gamma}$ is a GLS estimator obtained by minimising $F(S,\Sigma(\underline{\gamma}))$ in (2.29) then the asymptotic distribution of $\hat{\delta}_\gamma = \sqrt{n}(\dot{\underline{\gamma}}-\underline{\gamma}_0)$ is multivariate normal with a null mean vector and covariance matrix

(2.35) $\quad Lcov(\hat{\delta}_\gamma,\hat{\delta}_\gamma') = \{\Delta_0'\bar{U}^{-1}\Delta_0\}^{-1}\Delta_0'\bar{U}^{-1}\bar{Y}\bar{U}^{-1}\Delta_0\{\Delta_0'\bar{U}^{-1}\Delta_0\}^{-1}$

where $\Delta_0 = \Delta(\underline{\gamma}_0)$. Equivalently

(2.36) $\quad Lcov(\hat{\delta}_\gamma,\hat{\delta}_\gamma') = \{\Delta_0'\bar{U}^{-1}\Delta_0\}^{-1}H(\bar{U}^{-1}\bar{Y}\bar{U}^{-1},\Delta_0)\{\Delta_0'\bar{U}^{-1}\Delta_0\}^{-1}$

where H(.,.) is a q×q quadratic form defined by

(2.37) $H(U^{-1},\Delta) = \Delta'U^{-1}\Delta$

Corollary 2.1: If the matrix U of the discrepancy function $F(S,\Sigma(\gamma)|U)$ is correctly specified then the covariance matrix of the limiting multivariate normal distribution of $\hat{\delta}_\gamma$ is given by:

(2.38) $Lcov(\hat{\delta}_\gamma,\hat{\delta}_\gamma') = \{\Delta_0'\bar{Y}^{-1}\Delta_0\}^{-1} = \{H(\bar{Y}^{-1},\Delta_0)\}^{-1}$

Proposition 2.3: If $\bar{U} = \bar{Y}$ then $\hat{\gamma}$ is asymptotically efficient within the restricted class of estimators minimizing discrepancy functions of the form of $F(S,\Sigma(\gamma)|U)$ in (2.29) in the sense that the asymptotic variances of all its elements attain lower bounds for asymptotic variances of estimators in this class.

These estimators are again referred to as 'best' generalized least squares (BGLS) estimators. Note that, in contrast to the situation where the underlying distribution has no kurtosis, the BGLS estimators are not necessarily asymptotically efficient in that the asymptotic variance given in (2.38) is not necessarily equal to the Cramer-Rao lower bound where the underlying distribution is, for example, multivariate normal. Browne also observes that the term 'best' refers only to a specific asymptotic property which may not carry over to finite samples. It is possible that other estimators may have superior finite-sample properties.

Proposition 2.4: Suppose that $\hat{\gamma}$ is an estimator with the property that, as $n\rightarrow\infty$, $\hat{\delta}_\gamma = \sqrt{n}(\hat{\delta}-\gamma_0)$ has an asymptotic distribution with a null mean vector and a finite covariance matrix. Let $\Delta_c = \Delta_c(\gamma)$ be a $p^*\times(p^*-q)$ matrix-valued function of γ such that the rank of Δ_c is (p^*-q) and $\Delta_c'\Delta = 0$ when $\Delta = \Delta(\gamma)$ in (2.26). Also let

$\hat{\underline{e}} = \underline{s}-\hat{\underline{\sigma}} = \text{vecs}(S-\Sigma(\hat{\gamma}))$, $\underline{e}_0 = \underline{\sigma}_0-\hat{\underline{\sigma}}_0 = \text{vecs}(\Sigma_0-\Sigma(\gamma_0))$, $\hat{\Delta}_c = \Delta_c(\hat{\gamma})$ and $\Delta_{0c} = \Delta_c(\gamma_0)$.

If the positive definite matrix $\hat{Y}$ is a consistent estimator of $\bar{Y}$, then the asymptotic distribution of the quadratic form statistic

(2.39) $c(\hat{\gamma},S,\hat{Y}) = n\hat{\underline{e}}'\hat{\Delta}_c\{\hat{\Delta}_c\hat{Y}\hat{\Delta}_c\}^{-1}\hat{\Delta}_c'\hat{\underline{e}}$

as $n\rightarrow\infty$ is non-central chi-square with p^*-q degrees of freedom and non-centrality parameter

(2.40) $\lambda = n\underline{e}_0'\Delta_{0c}\{\Delta_{0c}'\bar{Y}\Delta_{0c}\}^{-1}\Delta_{0c}'\underline{e}_0$

Equivalent expressions for the quadratic form statistic and non-centrality parameter are

$c(\hat{\gamma},S,\hat{Y}) = n\hat{\underline{e}}'[\hat{Y}^{-1} - \hat{Y}^{-1}\hat{\Delta}\{\hat{\Delta}'\hat{Y}\hat{\Delta}\}^{-1}\hat{\Delta}'\hat{Y}^{-1}]\hat{\underline{e}}$

(2.41) $c(\hat{\gamma},S,\hat{Y}) = n\{\hat{\underline{e}}'\hat{Y}^{-1}\hat{\underline{e}}' - \hat{g}'\hat{H}\hat{g}\}$

where

(2.42) $\hat{g} = g(\hat{\gamma}|S,\hat{Y}) = \Delta'(\hat{\gamma})\hat{Y}^{-1}(\underline{s}-\underline{\sigma}(\hat{\gamma}))$,

the qx1 negative gradient of $\frac{1}{2}F(S,\Sigma(\gamma)|Y)$, $H = H(\hat{Y}^{-1},\Delta)$ in (2.37) and

$\lambda = n\underline{e}_0'[\bar{Y}^{-1} - \bar{Y}^{-1}\Delta_0\{\Delta_0'\bar{Y}^{-1}\Delta_0\}^{-1}\Delta_0'\bar{Y}^{-1}]\underline{e}_0$

(2.43) $\lambda = n\{\underline{e}_0'\bar{Y}^{-1}\underline{e}_0 - g_0'H^{-1}g_0\}$

where $g = g(\gamma_0|\Sigma_0,\bar{Y})$ and $H_0 = H(\bar{Y}^{-1}|\Delta_0)$

Corollary 2.2: If U is correctly specified the asymptotic distribution of $nF(S,\Sigma(\hat{\gamma})|U)$ is non-central chi-square with p^*-q degrees of freedom and non-centrality parameter $nF(\Sigma_0,S(\gamma_0)|\bar{Y})$.

These statistics will have asymptotic (central) chi-square distributions when $\Sigma_0=\Sigma(\gamma_0)$ so that $\underline{e}_0=0$ and $\lambda=0$. Consequently they may be employed to test the null hypothesis that the model holds. The non-central distribution is also of interest, as a measure of the goodness of the approximation of the model to reality.

2.2.3.3 ASYMPTOTICALLY DISTRIBUTION-FREE BGLS ESTIMATORS.

An estimator $\dot{\gamma}$ satisfying the conditions of proposition 2.3 may in general be obtained by defining a U that is a consistent estimator of $\bar{Y}$. This may be accomplished by substituting sample moments for the population moments in (2.25). Let

$$(2.44)\quad \bar{x}_i = N^{-1}\sum_{r=1}^{N} x_{ir}$$

$$(2.45)\quad w_{ij} = N^{-1}\sum_{r=1}^{N}(x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j) = nN^{-1}(S)_{ij}$$

$$(2.46)\quad w_{ijkl} = N^{-1}\sum_{r=1}^{N}(x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j)(x_{kr}-\bar{x}_k)(x_{lr}-\bar{x}_l)$$

In order for all sample fourth-order moments about the mean, w_{ijkl}, to be consistent estimators of the σ_{ijkl}, it is necessary that all eighth-order moments of the distribution of $\underline{x}$ be finite. This assumption also ensures that w_{ij} will be a consistent estimator of $(\Sigma_0)_{ij}$ so that

$$(2.47)\quad (\bar{U})_{ij} = w_{ijkl} - w_{ij}w_{kl}$$

will be a consistent estimator of $(\bar{Y})_{ij,kl}$ in (2.25). Since (2.47) represents the sample covariance between the product variables $(x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j)$ and $(x_{kr}-\bar{x}_k)(x_{lr}-\bar{x}_l)$ with means w_{ij} and w_{kl}, the matrix U defined by (2.47) will be positive definite with probability 1 provided that N is greater than p^* and $\bar{Y}$ is positive definite. Browne points out

that the number of non-duplicated w_{ijkl} is $p(p+1)(p+2)(p+3)/24$, which is considerably less than $p^*(p^*+1)/2$, the number of non-duplicated elements of U. The BGLS estimate, $\underline{\gamma}$, is obtained by minimising

(2.48) $F(S,\Sigma(\underline{\gamma})|U) = \frac{1}{2}(\underline{s}-\underline{\sigma}(\underline{\gamma}))'U^{-1}(\underline{s}-\underline{\sigma}(\underline{\gamma}))$

with respect to $\underline{\gamma}$, where U is defined by (2.47).

We shall refer henceforth to these asymptotically distribution-free GLS estimators as GLS (adf) estimators. These are distinguished from the GLS (nk) estimators described earlier.

Bentler (1983) pointed out that in general, the population counterpart to the optimal GLS weight matrix U will be a function of lower-order parameters. He argued that the use of fixed weight matrices as suggested by Browne (1982, 1984) and Shapiro (1983), may not in practice lead to the best possible (e.g. least biased or smallest variance or most accurate for small N) consistent estimates of the optimal U. If $U = U(\underline{\gamma})$, the minimisation of the GLS function can be accomplished by differentiating the function with respect to the parameters without assuming that $\partial U(\underline{\gamma})/\partial\underline{\gamma}=0$. However, the iterative nonlinear optimization involved in solving for $\hat{\underline{\gamma}}$ would be difficult. Thus a completely parametric GLS of this sort may be abandoned in favour of an iteratively reweighted GLS, in which the weight matrix U is updated at each step of an iterative procedure designed to minimise the GLS function.

Bentler notes that bias reduction may be achieved (Bentler and Dijkstra 1983, Shapiro 1983) by estimating and thus approximately eliminating the bias if an

estimator possesses an appropriate asymptotic expansion. The GLS function allows the appropriate expansion, and hence it may be of interest to obtain the bias adjusted estimator. After some lengthy algebra, using the methods of Shapiro (1983) the expression:

$b(\dot{\gamma}) = -N^{-1}\Delta\underline{\dot{\sigma}}'Vec(U\underline{\dot{\sigma}}X)$

can be derived. Here, in Bentler's notation, $\underline{\dot{\sigma}}'=\partial\underline{\sigma}/\partial\gamma$, $\Delta=(\underline{\dot{\sigma}}'U\underline{\dot{\sigma}})^{-1}$, $\underline{\dot{\sigma}}=\partial\underline{\dot{\sigma}}/\partial\gamma$, $X=\Delta\underline{\dot{\sigma}}'Y\underline{\dot{\sigma}}\Delta$, and $Y=\Sigma_{i,j}U_jV_{ij}U_i'$, all evaluated at $\gamma = \dot{\gamma}$. V_{ij} is the (i,j)th element of V, the limiting covariance matrix of $\underline{s}$. U_j is the jth column of U. It is noted that the calculation of the bias correction will take a substantial amount of computation.

2.2.3.4 ASYMPTOTICALLY DISTRIBUTION FREE MODELS FOR MULTIPLE GROUPS AND STRUCTURED MEANS

Browne's ADF approach has been extended to allow for estimation in multiple groups and for structured means. Bentler et al. (1987) considered multiple population analysis. They suppose that there are m populations, for each of which the GLS function is defined as above, with separate parameter vectors. The overall fit function is the sum of the population functions. Constraints are supposed to exist which may relate the parameter vectors for different populations. It is supposed that as N (the total sample size) tends to infinity the proportion from each population tends to a constant between zero and unity. The parameter estimates are then shown, under suitable regularity conditions, to be consistent with limiting normal distributions, and the asymptotic distribution of the overall fit function is shown to be chi-squared with degrees of freedom $\Sigma_g p_g(p_g+1)/2 -(q-r)$. Here the summation is over population groups g, p_g is the order of the sample covariance matrix S_g, q is

the sum of the dimensions of the population parameter vectors and r is the number of constraints. Estimation of covariance structures subject to functional constraints is discussed further below.

Muthen (1989) extended Browne's ADF approach to include structured means in multiple populations, following the work of Joreskog (1971) and Sorbom (1974), who considered simultaneous factor analysis in several groups. As applied to GLS, their approach results in minimising the fitting function which is the sum of the discrepancy functions for the different groups. Muthen extends the discrepancy function for a single group allow for structured means in the obvious way, so that the function vector is $\underline{\sigma}^{(g)\prime}=(\underline{\sigma}^{(g)}_1{}',\underline{\sigma}^{(g)}_2{}')$, where the first element is that of the parameters of the mean structure and the second that of the covariance structure. The discrepancy function for a single group is $(\underline{s}^{(g)}-\underline{\sigma}^{(g)})'W^{(g)-1}(\underline{s}^{(g)}-\underline{\sigma}^{(g)})$, where $\underline{s}'=(\underline{s}^{(g)}_1{}',\underline{s}^{(g)}_2{}')$ is the concatenation of the vectors of mean elements and covariance elements, and $W^{(g)}$ is chosen as a consistent estimator of the asymptotic covariance matrix of $s^{(g)}$.

Under multivariate normality, if we partition $W^{(g)}$ according to $\underline{s}^{(g)}$, then $W^{(g)}_{21}=0$. Since $W^{(g)}_{11}$ and $W^{(g)}_{22}$ are already known under non-normality, it remains to find $W^{(g)}_{21}$, a consistent estimator of the asymptotic covariances between $\underline{s}_1$ and $\underline{s}_2$, which are not zero under non-normality. Such a consistent estimator is formed as follows. Let μ_i denote the mean variable i and let μ_{ijk} denote a multivariate third-order moment about the mean. Note that $\mu_{ijk} = \kappa_{ijk}$ where κ_{ijk} is a third-order cumulant. Asymptotically N $\mathrm{cov}[(s_2)_{ij},(s_1)_k] = \kappa_{ijk}$. Let $y^{*}_i{}' = (y_{11}-\bar{y}_1,\ldots,y_{pi}-\bar{y}_p)$ and create the p^{*}

dimensional vector a_i of pairwise products of y^*_{ij}. Then, (deleting the group index) $W_{21} = N^{-2}\Sigma a_i y^{*}_{i}{}'$, where N is the sample size for the particular group.

Note that for multiple populations, the sample size must be larger than $p_g(p_g+1)/2$ (for the estimation of covariance structures), or $p_g+p_g(p_g+1)/2$ (for, additionally, structured means) to ensure a non-singular $W^{(g)}$ for each group.

2.2.4 MINIMUM DISCREPANCY FUNCTION ESTIMATION OF COVARIANCE STRUCTURES

In the previous section the only discrepancy functions considered were quadratic form functions as defined in (2.29). We now consider theory applicable to any discrepancy function.

Shapiro (1985a), following the definition given by Browne (1982, 1984) of a discrepancy function, defines a minimum discrepancy function (MDF) estimator of a q×1 parameter vector $\underline{\gamma}$ as that value of $\underline{\gamma}$ which minimises the discrepancy function $F(\underline{s},\underline{\sigma})$ ($\underline{s}$ = vecs(S) and $\underline{\sigma}$ = vecs($\Sigma(\underline{\gamma})$)). The MDF test statistic is given by $n\hat{F}$, where n is the sample size and $\hat{F}$ is the minimum of F. He proves that under the definition of a discrepancy function, there exists a continuous $p^*\times p^*$ symmetric valued function $V_{s,\sigma} = V(\underline{s},\underline{\sigma})$, such that:

(2.49) $F(\underline{s},\underline{\sigma})=(\underline{s}-\underline{\sigma})'V_{s,\sigma}(\underline{s}-\underline{\sigma})$.

This means that any discrepancy function can be represented in the form of an "extended" GLS function, where the word "extended" indicates that the corresponding weight matrix $V(\underline{s},\underline{\sigma}(\underline{\gamma}))$ may depend on $\underline{\gamma}$ as well as on $\underline{s}$. Another

important consequence of this theorem is that it follows that the second-order Taylor approximation of F at the (population) point $(\underline{\sigma}_0,\underline{\sigma}_0)$ is $(\underline{s}-\underline{\sigma})'V_0(\underline{s}-\underline{\sigma})$, where V_0 is the (matrix) value of $V(\underline{s},\underline{\sigma})$ at $(\underline{\sigma}_0,\underline{\sigma}_0)$. Alternatively the matrix V_0 is given by any one of the Hessian matrices $\frac{1}{2}\partial^2F/\partial\underline{\sigma}\partial\underline{\sigma}'$, $-\frac{1}{2}\partial^2F/\partial\underline{s}\partial\underline{\sigma}'$ or $\frac{1}{2}\partial^2F/\partial\underline{\sigma}\partial\underline{\sigma}'$. This implies that the second-order Taylor approximation of F at $(\underline{\xi}_0,\underline{\xi}_0)$ is symmetric in $\underline{s}$ and $\underline{\sigma}$ while the discrepancy function itself possibly is not. Finally the Hessian matrices of $F(\underline{s},\underline{\sigma}(\underline{\gamma}))$ are now readily available, since by the chain rule of differentiation we have that $\partial^2F/\partial\underline{\sigma}\partial\underline{\gamma}'$ and $\partial^2F/\partial\underline{\gamma}\partial\underline{\gamma}'$ are equal to $-2V_0\Delta$ and $2\Delta'V_0\Delta$ respectively, where Δ is defined as in (2.26).

The asymptotic distribution of the MDF estimator was investigated under the regularity conditions:

(S1) The parameter space G is compact.

(S2) The parameter vector $\underline{\gamma}$ is identified at $\underline{\gamma}_0$.

(S3) $\underline{\gamma}_0$ is an interior point of G.

(S4) The q×q matrix $\Delta'V_0\Delta$ is nonsingular.

As observed above, conditions (S1) and (S2) alone imply the consistency of the MDF estimator. If W is a symmetric matrix valued function of $\underline{s}$ converging to the matrix V_0 as $\underline{s}$ tends to $\underline{\sigma}_0$, and G is the GLS discrepancy function defined by $G(\underline{s},\underline{\sigma})=(\underline{s}-\underline{\sigma})'W(\underline{s}-\sigma)$, then F and G have the same second-order Taylor approximation at the point $(\underline{\sigma}_0,\underline{\sigma}_0)$. It follows from standard asymptotic theory that the estimates of $\underline{\gamma}$ which minimise F and G respectively have asymptotically equivalent properties.

Shapiro (1986) generalized the theory of MDF estimators still further by presenting a theory of overparametrized structural models. In such a model some "redundant" parameters are involved; the parameter vector is not identified, and the information matrix is not nonsingular. The MDF test statistic was shown to have an asymptotic chi-squared distribution almost everywhere for a wide class of discrepancy functions, and asymptotic distribution properties of the MDF estimator were investigated.

Shapiro (1987) studied robustness properties of the MDF analysis of moment structures. Generalizing the notation of Browne (1984), the discrepancy function is said to be correctly specified if V_0 (defined as in (2.49) is equal to a generalized inverse Y^- of Y (defined in (2.23)). The main purpose was to investigate conditions under which the standard or slightly modified MDF procedures result in a correct statistical inference for misspecified discrepancy functions.

Suppose that F is correctly specified with respect to a matrix Y_0. Necessary and sufficient conditions on Y with respect to Y_0 are given so that $n\hat{F}$ has an asymptotic non-central chi-squared distribution. Similarly, conditions on Y with respect to Y_0 are given so that the difference between the minimum discrepancy function under a restricted model and $\hat{F}$ has an asymptotic non-central chi-squared distribution, and under which this difference and $\hat{F}$ are asymptotically independent. Conditions on Y with respect to Y_0 are also given so that $\hat{\gamma}$ is asymptotically efficient.

Browne and Shapiro (1988) derived the structure of Y under the class of linear latent variable models is derived using properties of cumulants. An appendix to the paper provides a brief summary of relevant results from a general theory of

asymptotic robustness of minimum discrepancy methods in the analysis of covariance structures, derived from Browne (1984) and Shapiro (1984, 1985a, 1986, 1987).

These results show that any MDF estimator will have asymptotic properties equivalent to those of GLS estimator. The GLS (nk) estimates of section 2.2.3.1 have been shown to be asymptotically efficient where the underlying distribution has no kurtosis, while the best GLS (adf) estimates of section 2.2.3.3 have been shown to be asymptotically efficient in the class of GLS estimates. Since GLS estimates derived using these discrepancy functions may be taken as representing broad classes of estimators with asymptotically optimal properties. The studies of the robustness of MDF estimators which we have mentioned briefly give indications of the circumstances in which MDF methods will lead to correct inference despite relaxation of regularity conditions.

2.2.5 METHODS FOR THE ESTIMATION OF COVARIANCE STRUCTURES.

Lee and Jennrich (1979) considered the Fletcher-Powell, Gauss-Newton, Newton-Raphson, Fisher Scoring and Fletcher-Reeves algorithms for estimation of covariance structures using either maximum likelihood (ML) or generalized least squares assuming no kurtosis (GLS (nk)). From equations (2.12) and (2.33), the MLE of $\underset{\sim}{\gamma}$ is the vector which minimises the function

(2.50) $L(\underset{\sim}{\gamma}) = \log|\Sigma| + tr(S\Sigma^{-1})$

while the GLS (nk) estimate minimises

(2.51) $G(\underset{\sim}{\gamma}) = \frac{1}{2}tr((S-\Sigma)W)^2$

where W is a weight matrix usually chosen to be the inverse of S. We write $Q(\gamma)$ to represent L, G or any other discrepancy function.

A single step in the Newton-Raphson algorithm is defined by

(2.52) $\Delta\gamma = -H^{-1}g$,

where $g=g(\gamma)$ is the gradient vector for $Q(\gamma)$ and $H=H(\gamma)$ is the Hessian matrix for $Q(\gamma)$, both evaluated at γ. The main reason for considering other algorithms is that the Hessian matrix may be difficult to obtain, so that algorithms which require only first order derivatives may be of use.

If $Q(\gamma)$ is taken to be $L(\gamma)$ and the Hessian H in (2.52) is replaced by its expectation, one obtains the standard Fisher Scoring algorithm:

(2.53) $\Delta\gamma = -E[H]^{-1}g$

This algorithm is often more robust to bad starting values because E[H] is non-negative definite for all values of γ and is usually positive definite.

The Gauss-Newton algorithm can be applied when $Q(\gamma)$ is the weighted least squares function $G(\gamma)$. In this case, a basic step of the algorithm is

(2.54) $\Delta\gamma = -H^{*-1}g$, where

(2.55) $H^{*}_{ij} = trW\Sigma^{(i)}W\Sigma^{(j)}$

It is noted that the scoring algorithm is simply an iteratively reweighted Gauss-Newton algorithm, where the weight matrix W changes with γ from iteration to iteration, so that the Gauss-Newton algorithm may be used for both weighted least squares and maximum likelihood estimation. Lee and Jennrich recommended the

Gauss-Newton algorithm, because it is robust to poor starting values, converges quickly and conveniently produces consistent standard errors for both maximum-likelihood and weighted least squares problems.

Browne (1984) also recommends the Gauss-Newton algorithm for GLS estimation. This is modified so that

(2.56) $\Delta\underline{\gamma} = \alpha H^{*-1}\underline{g}$

as proposed by Jennrich and Sampson (1968). Here α ($0<\alpha<1$) is chosen so that the step always results in a reduction in the discrepancy function. Usually $\alpha=1$. For the general definition of GLS functions given in (2.29), the matrix H^* is as defined in (2.37).

2.2.6 ESTIMATION UNDER OF COVARIANCE STRUCTURES SUBJECT TO CONSTRAINTS.

In subsequent sections we will estimate covariance structures subject to two sorts of constraints. These may be broadly defined either as:

(2.57) minimise $G(\underline{\gamma})$ subject to $\underline{h}(\underline{\gamma})\geq 0$ (inequality constraints), or

(2.58) minimise $G(\underline{\gamma})$ subject to $\underline{h}(\underline{\gamma})=0$ (equality constraints)

where as above $G(\underline{\gamma})$ is any discrepancy function, and $\underline{h}(\underline{\gamma})= (h_1(\underline{\gamma}),...h_r(\underline{\gamma}))$, and $h_1,...,h_r$ are independent differentiable real-valued functions.

Lee (1980) investigated the use of the penalty function method to estimate parameters that are subject either to (2.57) or (2.58). This consists of sequential unconstrained minimisation of the function

$$G_k(\underline{\gamma}) = G(\theta) + c_k\Sigma_i\phi(h_i(\underline{\gamma}))$$

where (c_k) is a decreasing sequence of positive real numbers and ϕ is any real valued differentiable function which has huge value near the boundary of the feasible region (e.g $\phi(t) = -\ln(t)$). As the approximation is made more exact, by allowing c_k to tend to zero, the solution of the unconstrained problem converges to the solution of the original problem.

Lee and Poon (1985) presented a method for constrained estimation which handles inequality and equality constraints simultaneously by introducing 'slack' variables $\underline{z} = (z_1,...,z_r)$ so that (2.62) becomes

(2.59) minimise $G(\underline{\gamma})$ subject to $(\underline{h}(\underline{\gamma})-\underline{z})= 0$

The algorithm consists of sequential minimisation of an augmented Lagrangian function which includes $\underline{z}$. It was claimed that this method is better behaved than the penalty function method in the rate of convergence and numerical stability.

Lee and Bentler (1980) extended basic results by Browne (1974) on GLS estimation of covariance structures to covariance structures with parameters subject to arbitrary nonlinear constraints. They showed that the constrained estimators are consistent, asymptotically normally distributed and asymptotically equivalent to constrained maximum likelihood estimators. The relationships between the Lagrangian approach and the reparametrization approach were also discussed. They observed if one can impose equality constraints by finding a reparametrization it will result in a simpler and more efficient algorithm for obtaining estimates.

McDonald (1980) gave matrix identities and reparametrizations useful in imposing inequality constraints. A well known procedure for satisfying the requirement that the estimate of a variance be non-negative is to estimate the standard deviation (as a real number) and to compute the variance estimate as a parametric function, namely the square, of the estimated s.d. He showed how this procedure may be generalized to allow for, for instance, a sequence of inequalities $q_1 \leq q_2 \leq \ldots \leq q_k$.

It is clear that in order to impose constraints of the form (2.57) or (2.58), one may choose either to use a constrained estimation procedure or try to find a reparametrizations which allow unconstrained estimation of the reparametrized model. The constraints which will be imposed in subsequent sections are of the simple forms $\gamma_i=0$ or $\gamma_i \geq 0$. In either case, we choose to reparametrize the model rather than to attempt to implement a constrained estimation procedure.

Shapiro (1985b) considered the case where the population value of the parameter vector of a covariance structure is a boundary point of the feasible region. He showed that in this case the asymptotic distribution of the test statistic is a mixture of chi-square distributions; $n\hat{F} \sim \chi^2_\nu + \bar{\chi}^2$, where $\nu = p^*-q$, and $\bar{\chi}^2$ is a weighted sum of chi-square statistics, $pr(\bar{\chi}^2 \geq c^2) = \Sigma w_i pr(\chi^2_i \geq c^2)$ where the summation is over $i=0,\ldots,q$, χ^2_i is a chi-squared random variable with i degrees of freedom, $\chi^2_0 \equiv 0$, and w_i are non-negative weights such that $w_0+\ldots+w_q=1$. Formulae for the calculation of w_i were given for $q \leq 4$. These results explain why models with solutions at the boundary of the parameter space are rejected too often by the log likelihood ratio statistic.

2.2.7 SELECTION OF MODELS

In subsequent sections we will construct models which are intended to reflect possible properties of the true way in which periodontal disease progresses. We will wish to assess which of these models best describes the variation in the data. In this section we examine methods for the selection, assessment and comparison of models for covariance structures. We will be interested both in additional methods for assessing the best choice of model, and in methods which allow for the fact that none of our models is likely to be a complete description of the variation in the data. For example the assumption that disease progression is instantaneous (in our model for progression in 'bursts' of activity) is more realistically an approximation to the idea that periods of disease progression takes place over time periods which are short compared to the length of time between observations. Similarly in constructing our model for varying, non-instantaneous progression we hope to derive a covariance structure which will hold if disease progression varies with time. However we are unlikely to believe, as the model assumes, that disease progression is constant between time points which occur in a Poisson process, with instantaneous changes in rate which occur at those time points. Other assumptions made in the models which may well not hold in reality include the assumption the measurement error variance does not vary with attachment level.

We have already summarised the work of Browne (1974, 1984), which provides asymptotic chi-square tests for the goodness of fit of a maximum likelihood and covariance structural model. Steiger et al. (1985) investigated the multivariate

asymptotic distribution of sequential chi-square tests statistics. They were interested in procedures for choosing which of several nested structural model appear to be "best" for a population of interest. They discuss the sequential chi-square test (SCT) procedure, whereby one increases the complexity of the model until the hypothesis of perfect fit is not rejected. An alternative approach is the sequential chi-square difference test (SCDT), in which nested models are compared by treating the difference of their chi-square test statistics as a chi-square statistic with degrees of freedom equal to the difference in degrees of freedom for the individual chi-squares, and looking for a significant improvement in goodness of fit.

A sequence $M_1>M_2>...>M_r$ of nested models is considered, with parameter spaces $\{G_i\}$ such that G_2 is a subset of G_1 given by imposing k_1 equality constraints, G_3 is a subset of G_2 given by imposing k_2 equality constraints, and so on. In practice it is usually not reasonable to assume that a model is a precise representation of reality. A more reasonable view is that a model is an approximation. Following the formulation of Stroud (1972), a sequence $\{\underline{\sigma}_n\}$ of population values of $\underline{\sigma}$ converging to a point $\underline{\sigma}_0$ where all models hold is considered. That is, there exists $\gamma_0 \in G_r$ such that $\underline{\sigma}_0 = \underline{\sigma}(\gamma_0)$. The population badness of fit for models M_j, $j=1,...,r$ is defined as $\delta_n^{(i)} = \min F(\underline{\sigma}_n, \underline{\sigma}(\gamma))$, where the minimum is over all $\gamma \in G_j$. It is then proved under given regularity conditions that the test statistics $\{n\hat{F}_n^{(i)}\}$ $(i=1,..,r)$ have asymptotic non-central chi-square distributions with $\nu_i = m-q+k_1+...+k_i$ degrees of freedom, and non-centrality parameter δ_i. The differences in successive test statistics are mutually asymptotically independent with degrees of freedom and non-centrality parameters given by the difference for the two test statistics. A closed form is given for the

asymptotic correlation between the test statistics. Simulations confirm the results and show that the correlations between successive test statistics can be quite large.

Bentler and Bonnett (1980) observed that in large samples virtually any model tends to be rejected as inadequate, and in small samples various competing models, if evaluated, might be equally acceptable. They proposed a general null model based on modified independence among variables to provide an additional reference point for the statistical and scientific evaluation of covariance structure models. The model, M_0, is defined in the context of structural models as the severely restricted model which specifies that the measured variables are mutually independent. If M_t corresponds to a model of special interest, then a comparison of M_0 with M_t provides a test of whether the restrictions made in going from M_t to M_0 are reasonable. It was also suggested that statistical evaluation should be supplemented with incremental fit indices associated with the comparison of hierarchical models, and various such indices are described. A nonnormed fit index ρ_{kl} for a hierarchical models $M_0 \subset M_k \subset M_l$ is defined by:

(2.60) $\rho_{kl} = (Q_k - Q_l)/(Q_0 - 1)$

Here Q_* represents the ratio of a chi-squared variate to the degrees of freedom for model *. Thus ρ_{kl} represents the increment in fit in moving from M_k to the more general M_l. A more general normed fit index is given by:

(2.61) $\Delta_{kl} = (F_k - F_l)/F_0$

where F is any fit function such as a discrepancy function. Bentler and Bonnett noted that the scale of the fit indices is not necessarily easy to interpret.

Tanaka and Huba (1985) proposed a general fit index for GLS estimators of covariance structure models. This fit index is expressed as a function of the ratio of two trace functions. Bollen (1986) showed that Bentler and Bonnett's nonnormed fit index depends on the sample size, and proposed an alternative formulation:

(2.62) $\rho^*_{kl} = (Q_k - Q_l)/Q_0$.

For a given fit, ρ^*_{kl} will be the same regardless of the sample size N.

Cudeck and Browne (1983) examined methods for comparing the suitability of alternative models for covariance matrices. A model, M_k, for a population covariance matrix Σ between p variables is expressed as $M_k:\Sigma=\Phi_k(\gamma_k)$ (k=1,...g). Since statistical power theory virtually guarantees that, under the hypothesis testing approach to model selection, any model will be rejected if the sample size is sufficiently large, they prefer to suppose that $M_k:\Sigma\sim\Phi_k(\gamma_k)$, and to search not for the correct model, but for the best approximation.

A cross-validation procedure is suggested and its properties are examined. The sample is split randomly into two subsamples a and b, of equal size. The first sample is used to estimate the parameter values under model a, and the second is used as a validation sample to compute the "cross-validation index", $F(S_b;\hat{\Sigma}_{k|a})$. This process is repeated for each of the g models. While the smallest of the calibration-sample discrepancy function values $F(S_a;\hat{\Sigma}_{k|a})$ will usually correspond to the model with the largest number of parameters, this will frequently not be the case for the cross-validation indices $F(S_b;\hat{\Sigma}_{k|a})$. It is wise to carry out a "double cross-validation" by repeating the process with the roles of samples a and b reversed, yielding a second set of cross-validation indices $F(S_a;\hat{\Sigma}_{k|b})$.

In some situations if may not be possible to cross-validate, and an alternative is to use indices which incorporate penalty functions for the numbers of parameters (q_k for model k). The information criteria of Akaike (1974) and Schwarz (1978) can be rescaled to eliminate the effect of the sample size and expressed in terms of the MWL discrepancy function. This gives:

(2.63) $c_{Ak} = F(S;\dot{\Sigma}_k) + 2q_k/n$ and

(2.64) $C_{Sk} = F(S;\dot{\Sigma}_k) + q_k\ln(n)/n$

respectively.

Browne and Cudeck (1987) considered single-sample approximations for the cross-validation co-efficient in the analysis of covariance structures. The notation is similar to Cudeck and Browne (1983), with S_c and S_v being the sample covariance matrices for the calibration and validation samples respectively. The regularity and "population drift" conditions are the same as for Steiger et al. (1985). It is assumed that the calibration and validation sample sizes are equal. An asymptotic approximation, δ_k, is provided for the expected value of the difference between the cross-validation index and the calibration sample discrepancy function value:

(2.65) $\delta_k = E[F(S_v;\dot{\Sigma}_{k,c})-F(S_c;\dot{\Sigma}_{k,c})] = 2q_k/n + o(n^{-1})$.

We may therefore take $\tilde{\delta}_k = 2q_k/n$ so that the single sample cross-validation index for M_k is approximated by:

(2.66) $c_k = F(S_c;\dot{\Sigma}_{k,c}) + \tilde{\delta}_k = F(S_c;\dot{\Sigma}_{k,c}) + 2q_k/n$

Note that the same correction term $\tilde{\delta}_k$ is employed in conjunction with any correctly specified discrepancy function. Note also that where $F = F_{ML}$ the coefficient obtained

is the rescaled Akaike Information Criterion (AIC) (Cudeck and Browne 1983). This result is concordant with the finding of Stone (1977) that the "leaving one out at a time" method of cross-validation (Stone 1974) gives a log-likelihood measure of predictive validity which is asymptotically equivalent to the AIC. Results of a random sampling experiment are reported, and exact expressions for δ_k for F_{ML} and F_{GLS} are obtained for the saturated model.

Selection of one of a number of possible models for covariance structures presents the same difficulties as for model selection in any branch of statistics. As discussed by MacCallum (1986) in the context of structural equation models, the goodness of fit and meaning of models chosen as the best for a particular data set must be interpreted with caution. The aim of the procedures described above is to devise selection methods which are likely to identify models which will perform optimally in future samples rather than be the best for a particular sample. Although studies of the effect of misspecification and specification searches on covariance structural models have been performed (e.g. MacCallum 1986, Kaplan 1988), they are investigations of the effect of particular specifications and misspecifications on simulated data, so that it is difficult to draw general conclusions from them. This is illustrated by the fact that while MacCallum (1986) concluded that cross-validation following model searches was important, Kaplan specifically disagreed with this in concluding that chi-square tests should be used to test model specification.

2.2.8 SOFTWARE FOR THE ESTIMATION OF COVARIANCE STRUCTURES

In this section we review the currently available software for the estimation of covariance structures. We have already mentioned that the most widely used models are those which can be estimated using the package LISREL. The latest version of this package (Joreskog and Sorbom 1988), has many of the features of a standard statistical package. As well as maximum likelihood estimation under the assumption of multivariate normality, the programme now allows the use of the methods of Browne (1974, 1982, 1984) with general weight matrices. For models which can be specified in the LISREL form, this programme would clearly be the first choice.

Software has also been written to facilitate the estimation of the other major formulations of structural models. For instance the programme COSAN (Fraser and McDonald 1988) will fit models of the type specified by McDonald (1980), and the program EQS, which is now distributed as part of the statistical package BMDP, is designed for structural equations models as proposed by Bentler (1983).

More recently, Browne (1988) has developed AUFIT, a programme designed to provide both ML and GLS (nk) estimates. This programme uses the approach of Lord (1975), who provided a programme for automated hypothesis testing in which derivatives are evaluated numerically and thus need not be provided by the user. Generalized secant algorithms (Ralston and Jennrich 1978) are used to approximations to the Jacobian matrix which are then used in a modified Gauss-Newton algorithm. This programme will therefore fit nonstandard models which cannot be specified in McArdle and McDonald's (1984) RAM format, without the

need for the user to calculate the (possibly complex) Jacobian matrix. The same approach was used by Lee and Jennrich (1984), who used the BMDP programme PAR to estimate GLS functions using numerical derivatives.

2.2.9 SUMMARY AND CONCLUSIONS.

We have reviewed methods for the analysis of covariance structures. In all cases the stated properties of the estimates are asymptotic. Finite-sample properties of the estimators are virtually unknown, although Browne (1984), in a random sampling experiment, concluded that the asymptotically distribution-free methods of sections 2.2.3.1 and 2.2.3.2 showed unacceptable negative bias when used to estimate covariance structures.

For maximum likelihood and generalized least-squares methods, we have given the asymptotic distribution of the estimates (section 2.2.2, propositions 1.2-3, proposition 2.2 and corollary 2.1), and tests of goodness of fit (proposition 1.5 and discussion thereof for MLEs and GLS (nk) estimators), (proposition 2.4 and corollary 2.2 for GLS (adf) estimators). Because of the well-known asymptotic optimality of maximum-likelihood estimates, it would seem natural to use MLE's when the underlying distribution can be assumed to be normal. When the underlying distribution has no kurtosis, the GLS estimates have the same asymptotic properties as the MLEs. However, it is not clear how often in practice a non-normal distribution which has no kurtosis will occur, although the elliptical distribution is an example.

If the underlying distribution has non-zero fourth-order cumulants, then the asymptotically distribution-free methods of sections 2.2.3.2 and 2.2.3.3 provide a procedure for the estimation of covariance structures under weak assumptions. However, the estimates are not asymptotically efficient. Further the procedure will be impracticable for large p, since the matrix U, which has to be inverted, is of order p^*. Since large numbers of fourth order moments must be estimated, the sample size will need to be large for reliable estimates to be produced.

In subsequent sections, the following will be used in the assessment of models for covariance structures:

Discrepancy function: We will use the (correctly specified) functions defined in (2.33), with V=S, for the GLS (nk) procedure, and defined in (2.29), with U defined in (2.44) to (2.47), for the GLS (adf) procedure. As noted in section 2.2.4, estimates based on these functions have optimal properties over all discrepancy functions.

Parameter estimate: $\hat{\gamma}$ - the value of the parameter vector within the parameter space which minimises the discrepancy function (equation 2.27).

Standard errors: the square roots of the diagonal elements of the inverse of the asymptotic information matrix for the parameter estimates, defined in proposition 1.3 for the GLS procedure, and in corollary 2.1 for the GLS (adf) procedure.

Goodness of fit: nF, where F is the discrepancy function. By proposition 1.5 for the GLS (nk) procedure, and corollary 2.2 for the GLS (adf) procedure, when the

model holds this statistic has an asymptotic central chi-square distribution with p^*-q degrees of freedom. As discussed in section 2.2.7, the difference between the goodness of fit statistics for nested models has a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters.

We have described recent developments which include calculation of the asymptotic properties of these estimates for any discrepancy function and under inequality constraints, and have described computer software which may be used to fit covariance structural models. We have also described methods which may be used for selection of models. It is clear that a broad theoretical framework exists for our purpose, which will be to use a comparison of covariance structures to make inference about the suitability of different models for periodontal disease progression.

3

MODELS FOR PERIODONTAL DISEASE PROGRESSION.

3.1 DISTINCTION BETWEEN CONSTANT PROGRESSION AND 'BURSTS' OF ACTIVITY.

In section 2.1 we reviewed the current knowledge of the nature of periodontal disease progression and particularly the burst model which was postulated by Socransky et al. (1984). In this section we construct three models for periodontal disease progression and derive their covariance structures.

We start by constructing two models intended to represent the alternatives of constant progression and for instantaneous bursts of activity discussed by Socransky et al. Having constructed the models, we then compare the expectation and covariance structure of observed increments in attachment level. A feature of our approach is that we allow explicitly for measurement error. All sites are assumed to be independent; simply in order to make the mathematics tractable.

3.1.1 DEFINITION OF MODELS FOR DISEASE PROGRESSION.

Suppose we have N sites, at each of which p+1 measurements are made; at times $0,t_1,t_2,t_3, \ldots ,t_p$. Denote by $X^i(t_j)$ the attachment level at site i, time t_j. At time t_j we observe attachment level $y_{ij} = X^i(t_j) + e_{ij}$, where e_{ij} is the measurement error for observation j on site i, with mean 0 and variance σ^2, $e_{ij} \sim (0,\sigma^2)$ for short. Let:

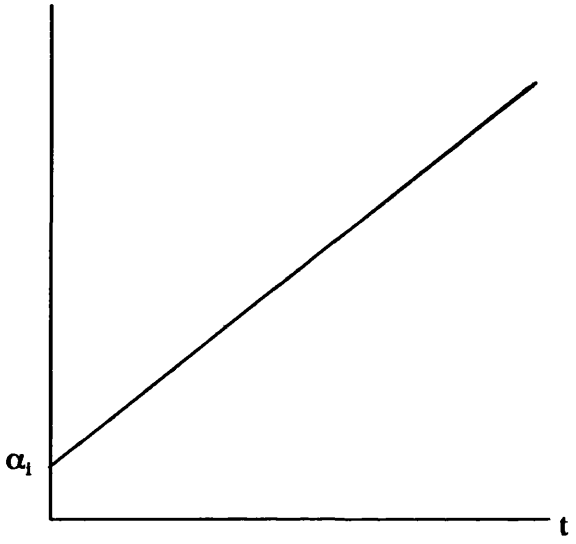
(3.1) $w_{ij} = y_{ij} - y_{i(j-1)} = X^i(t_j) - X^i(t_{(j-1)}) + e_{ij} - e_{i(j-1)}$

Thus y_{ij} is the observed attachment level at site i time t_j, and w_{ij} is the observed increment in attachment level at site i between times j-1 and j. Write $\tau_j = t_j - t_{j-1}$. The models for disease progression are defined as follows:

Model 1 - for constant disease progression

(3.2) $X^i(t) = \alpha_i + \beta_i t$

$X^i(t)$

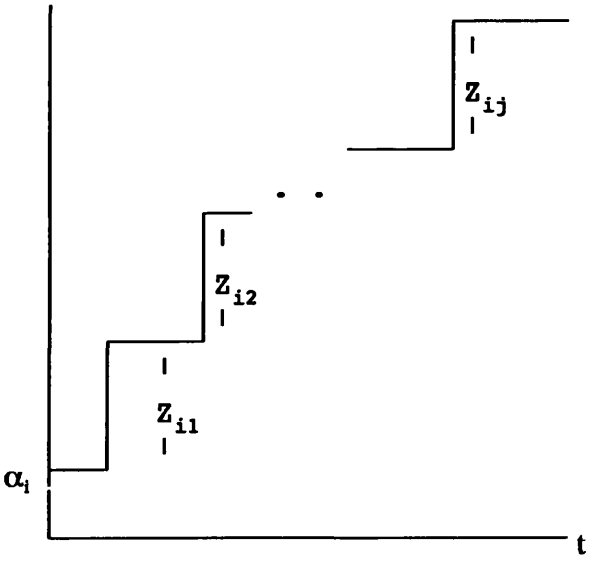


Disease progression is constant, with rate β_i for site i.

Model 2 - for bursts of activity

(3.3) $X^i(t) = \alpha_i + Z_{i1} + Z_{i2} + \ldots + Z_{ik}$, where $k \sim \text{Poisson}(\lambda_i t)$.

$X^i(t)$



The times at which the 'bursts' occur are a Poisson process with rate λ_i. The magnitudes Z_{ij} of the 'bursts' are independently and identically distributed with mean $\mu_{z(i)}$ and variance $\sigma^2_{z(i)}$.

($Z_{ij} \sim$ iid $(\mu_{z(i)}, \sigma^2_{z(i)})$).

VARIATION BETWEEN SITES

We suppose that each of the parameters for disease progression at a site is distributed independently between sites about a population mean. Thus:

Model 1:

(3.4) $\beta_i \sim (\beta,\sigma^2_\beta)$; $\alpha_i \sim (\alpha,\sigma^2_\alpha)$

where $-\infty < \beta, \alpha < \infty$ and $0 < \sigma^2_\beta, \sigma^2_\alpha < \infty$

Model 2:

(3.5) $\lambda_i \sim (\lambda,\sigma^2_\lambda)$; $\mu_{z(i)} \sim (\mu,\sigma^2_\mu)$; $\sigma^2_{z(i)} \sim (\sigma^2_z,\xi^2_z)$; $\alpha_i \sim (\alpha,\sigma^2_\alpha)$

where $-\infty < \mu, \alpha < -\infty$ and $0 < \lambda, \sigma^2_\lambda, \sigma^2_\mu, \sigma^2_Z, \xi^2_z, \sigma^2_\alpha, < \infty$

We also assume that λ_i, $\mu_{z(i)}$ and $\sigma^2_{z(i)}$ are mutually independent. This means that there will be non-zero covariances between different increments at the same site. The assumption of independence between sites means that increments at different sites will be uncorrelated.

3.1.2 CALCULATION OF EXPECTATION AND COVARIANCE STRUCTURES.

Note: For any random variables X (for which second moments exist), and A, the following hold:

(3.6) $E[X] = E_A[E[X|A=a]$

(3.7) $Var(X) = E_A[Var(X|A=a)] + Var_A(E[X|A=a])$

(3.8) $Cov(X,Y) = E_A[Cov(X,Y|A=a] + Cov_A(E[X|A=a],E[Y|A=a])$

For Model 1:

From (3.1) and (3.2) $w_{ij} = \tau_j\beta_i + e_{ij} - e_{i(j-1)}$. So:

(3.9) $E[W_{ij}|\beta_i] = \tau_j\beta_i$ and (applying (3.6) and using (3.4)):

(3.10) $\mathbf{E[w_{ij}] = \tau_j\beta}$

Applying (3.7) and using (3.5) and (3.9):

$Var(W_{ij}|\beta_i) = 2\sigma^2$

$Var(W_{ij}) = E_{\beta_i}[2\sigma^2] + Var_{\beta_i}(\tau_j\beta_i)$

(3.11) $\mathbf{Var(W_{ij}) = 2\sigma^2 + \tau_j^2\sigma_\beta^2}$

Applying (3.8) and using (3.5) and (3.9), for $j \neq k$:

$Cov(W_{ij},W_{ik}|\beta_i) = Cov(\beta_i\tau_j + e_{ij} - e_{i(j-1)}, \beta_i\tau_k + e_{ik} - e_{i(k-1)})$

$= -\sigma^2\delta'$, where:

(3.12) $\delta' = 1$ if $|j-k| = 1$, 0 otherwise.

$Cov(W_{ij},W_{ik}) = E_{\beta_i}[-\sigma^2\delta'] + Cov_{\beta_i}(\tau_j\beta_i,\tau_k\beta_i)$

(3.13) $\mathbf{Cov(W_{ij},W_{ik}) = -\sigma^2\delta' + \tau_j\tau_k\sigma_\beta^2}$

Clearly, $Cov(W_{ij},W_{i'k}) = 0$ if $i \neq i'$.

For Model 2:

Denote by N(ij) the number of 'bursts' between t_j and t_{j-1} for site i. (N(ij) has a Poisson distribution with rate $\lambda_i\tau_j$). Denote the magnitude of these 'bursts' by Z_{ij1}, Z_{ij2}, . . ., $Z_{ijN(ij)}$

Then, from (3.1):

(3.14) $w_{ij}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)},N(ij) = Z_{ij1}+Z_{ij2} + \ldots + Z_{ijNij} + e_{ij} - e_{i(j-1)}$

Thus:

(3.15) $E[W_{ij}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)},N(ij)] = N(ij)\mu_{z(i)}$

Applying (3.6) successively, and using (3.4), we have:

(3.16) $E[W_{ij}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)}] = E_{N(ij)}[N(ij)\mu_{z(i)}] = \lambda_i\tau_j\mu_{z(i)}$

(since $E[N(ij)] = \lambda_i\tau_j$)

(3.17) $E[W_{ij}|\mu_{z(i)},\sigma^2_{z(i)}] = E_{\lambda_i}[\lambda_i\tau_j\mu_{z(i)}] = \lambda\tau_j\mu_{z(i)}$

(3.18) $E[W_{ij}|\sigma^2_{z(i)}] = E_{\mu_{z(i)}}[\lambda\tau_j\mu_{z(i)}] = \lambda\tau_j\mu_z$

(3.19) $\mathbf{E[W_{ij}] = E_{\sigma^2_{z(i)}}[\lambda\tau_j\mu_z] = \lambda\tau_j\mu_z}$

Similarly, applying (3.7) successively and using (3.4) and (3.15-3.19):

$$Var(W_{ij}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)},N(ij)) = Var(Z_{ij1}+Z_{ij2}+...+Z_{ijN(ij)}+e_{ij}-e_{i(j-1)})$$
$$= 2\sigma^2 + N(ij)\sigma^2_{z(i)}$$

$$Var(W_{ij}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)}) = E_{N(ij)}[2\sigma^2 + N(ij)\sigma^2_{z(i)}] + Var_{N(ij)}(N(ij)\mu_{z(i)})$$
$$= 2\sigma^2 + \lambda_i\tau_j\sigma^2_{z(i)} + \lambda_i\tau_j\mu^2_{z(i)}$$
$$= 2\sigma^2 + \lambda_i\tau_j(\mu^2_{z(i)} + \sigma^2_{z(i)})$$

$$Var(W_{ij}|\mu_{z(i)},\sigma^2_{z(i)}) = E_{\lambda_i}[2\sigma^2 + \lambda_i\tau_j(\mu^2_{z(i)} + \sigma^2_{z(i)})] + Var_{\lambda_i}(\lambda_i\tau_j\mu_{z(i)})$$
$$= 2\sigma^2 + \sigma^2_\lambda\tau^2_j\mu^2_{z(i)} + \lambda\tau_j(\mu^2_{z(i)} + \sigma^2_{z(i)})$$

$$Var(W_{ij}|\sigma^2_{z(i)}) = E_{\mu_{z(i)}}[2\sigma^2+\sigma^2_\lambda\tau^2_j\mu^2_{z(i)}+\lambda\tau_j(\mu^2_{z(i)}+\sigma^2_{z(i)}] + Var_{\mu_{z(i)}}(\lambda\tau_j\mu_{z(i)})$$
$$= 2\sigma^2 + \sigma^2_\lambda\tau^2_j(\mu_z^2+\sigma^2_\mu + \lambda\tau_j(\mu_z^2+\sigma^2_\mu+\sigma^2_{z(i)}) + \lambda^2\tau^2_j\sigma^2_\mu$$
$$= 2\sigma^2 + \tau^2_j(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu) + \lambda\tau_j(\mu_z^2+\sigma^2_\mu+\sigma^2_{z(i)})$$

$$Var(W_{ij}) = E_{\sigma^2_{z(i)}}[2\sigma^2+\tau^2_j(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu)+\lambda\tau_j(\mu_z^2+\sigma^2_\mu+\sigma^2_{z(i)})] + Var_{\sigma^2_{z(i)}}(\lambda\tau_j\mu_z)$$

$= 2\sigma^2+\tau^2_j(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu)+\lambda\tau_j(\mu_z^2+\sigma^2_\mu+\sigma^2_z) + 0$

(3.20) $\mathbf{Var(W_{ij}) = 2\sigma^2 + \tau^2_j(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu) + \lambda\tau_j(\mu_z^2+\sigma^2_\mu+\sigma^2_z)}$

Finally, for $j\neq k$, applying (3.8) and using (3.4) and (3.15-19):

$Cov(W_{ij},W_{ik}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)},N(ij),N(ik))=Cov(Z_{ij1}+..+Z_{ijN(ij)}+e_{ij}-e_{i(j-1)},Z_{ik1}+..+Z_{ikN(ik)}+e_{ik}-e_{i(k-1)})$

$= -\sigma^2\delta'$ (with δ' defined as in 3.13)

$Cov(W_{ij},W_{ik}|\lambda_i,\mu_{z(i)},\sigma^2_{z(i)}) = E_{N(ij),N(ik)}[-\sigma^2\delta'] + Cov_{N(ij),N(ik)}(N(ij)\mu_{z(i)},N(ik)\mu_{z(i)})$

$= -\sigma^2\delta' + 0$ since N(ij) and N(ik) are independent

$= -\sigma^2\delta'$

$Cov(W_{ij},W_{ik}|\mu_{z(i)},\sigma^2_{z(i)}) = E_{\lambda_i}[-\sigma^2\delta'] + Cov_{\lambda_i}(\lambda_i\tau_j\mu_{z(i)},\lambda_i\tau_k\mu_{z(i)})$

$= -\sigma^2\delta'+ \tau_j\tau_k\sigma^2_\lambda\mu^2_{z(i)}$

$Cov(W_{ij},W_{ik}|\sigma^2_{z(i)}) = E_{\mu_{z(i)}}[-\sigma^2\delta'+\tau_j\tau_k\sigma^2_\lambda\mu^2_{z(i)}]+Cov_{\mu_{z(i)}}(\lambda\tau_j\mu_{z(i)},\lambda\tau_k\mu_{z(i)})$

$= -\sigma^2\delta'+ \tau_j\tau_k\sigma^2_\lambda(\mu_z^2+\sigma^2_\mu) + \tau_j\tau_k\lambda^2\sigma^2_\mu$

$= -\sigma^2\delta'+ \tau_j\tau_k(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu)$

$Cov(W_{ij},W_{ik}) = E_{\sigma^2_{z(i)}}[-\sigma^2\delta'+\tau_j\tau_k(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu)+Cov_{\sigma^2_{z(i)}}(\lambda\tau_j\mu_z,\lambda\tau_j\mu_z)$

(3.21) $\mathbf{Cov(W_{ij},W_{ik}) = -\sigma^2\delta'+ \tau_j\tau_k(\sigma^2_\lambda\mu_z^2+\sigma^2_\lambda\sigma^2_\mu+\lambda^2\sigma^2_\mu)}$

Clearly, $Cov(W_{ij},W_{i'k}) = 0$ if $i\neq i'$.

Summarising equations (3.10-13) and (3.19-21), we have thus shown:

Lemma 3.1: The expectations, variances and covariances of $\{w_{ij}\}$, (i=1,2,..,N); (j=1,2, . . ,p) are as follows:

$$E[W_{ij}] = \begin{cases} \tau_j\beta & \text{(model 1)} \\ \tau_j\lambda\mu_z & \text{(model 2)} \end{cases}$$

$$Var(W_{ij}) = \begin{cases} 2\sigma^2 + \tau_j^2\sigma_\beta^2 & \text{(model 1)} \\ 2\sigma^2 + \tau_j^2(\sigma_\lambda^2\mu_z^2+\sigma_\lambda^2\sigma_\mu^2+\lambda^2\sigma_\mu^2) + \lambda\tau_j(\mu_z^2+\sigma_\mu^2+\sigma_z^2) & \text{(model 2)} \end{cases}$$

$$Cov(W_{ij},W_{ik}) = \begin{cases} -\sigma^2\delta' + \tau_j\tau_k\sigma_\beta^2 & \text{(model 1)} \\ -\sigma^2\delta' + \tau_j\tau_k(\sigma_\lambda^2\mu_z^2+\sigma_\lambda^2\sigma_\mu^2+\lambda^2\sigma_\mu^2) & \text{(model 2)} \end{cases} \quad (j\neq k)$$

where $\delta'= 1$ if $|j-k| = 1$, 0 otherwise.

Cov $W(_{ij},W_{i'k}) = 0$ in each model, for $i\neq i'$. ∎

Observe that the expectations, variances and covariances have a simple structure which is similar for each model. Thus, we can write:

(3.22) $E(W_{ij}) = \mu\tau_j$ $(-\infty< \mu < \infty)$

(3.23) $Var(W_{ij}) = 2\theta_1 + \theta_2\tau_j^2 + \theta_3\tau_j$ $(0 < \theta_1,\theta_2,\theta_3 < \infty)$

(3.24) $Cov(W_{ij},W_{ik}) = -\theta_1\delta' + \theta_2\tau_j\tau_k$

where:

(3.25) in model 1; $\mu=\beta$, $\theta_2=\sigma_\beta^2$ and $\theta_3=0$

(3.26) in model 2; $\mu=\lambda\mu_z$, $\theta_2=\sigma_\lambda^2\mu_z^2+\sigma_\lambda^2\sigma_\mu^2+\lambda^2\sigma_\mu^2$ and $\theta_3=\lambda(\mu_z^2+\sigma_\mu^2+\sigma_z^2)$

In each model $\theta_1 = \sigma^2$.

In model 1, each parameter is determined by a single population parameter of the underlying model. In model 2, however, although θ_1 is given by the error variance (whose value does not affect the value of μ, θ_2 or θ_3), each of μ, θ_2 and θ_3 depends on a subset of the five population parameters λ, μ_z, σ^2_λ, σ^2_μ and σ^2_z. We wish to simplify the covariance structure of model 2 by treating $\{\mu,\theta_1,\theta_2,\theta_3\}$ as independently varying parameters. To justify this, we need the following lemma:

Lemma 3.2: In model 1, the range of values of $\{\mu,\theta_1,\theta_2,\theta_3\}$ as defined in (3.26) is $\mathbf{RxR^{+3}}$, given the ranges of the underlying parameters.

Proof: Fix $\mu = c > 0$, so that $\lambda = c/\mu_z$ where now

(3.27) $\mu_z > 0$, and therefore

(3.28) $\theta_2 = \sigma^2_\lambda\mu_z^2 + \sigma^2_\lambda\sigma^2_\mu + c^2\sigma^2_\mu/\mu_z^2$ and

(3.29) $\theta_3 = c/\mu_z(\mu_z^2+\sigma^2_\mu+\sigma^2_z)$

Fix $\theta_3 = d>0$, and write

(3.30) $A = \sigma^2_\mu = d\mu_z/c - \mu_z^2 - \sigma^2_z > 0$

Substituting (3.30) in (3.28), we have

$\theta_2 = \sigma^2_\lambda(d\mu_z/c - \sigma^2_z) + c^2/\mu_z^2(d\mu_z/c - \mu_z^2 - \sigma^2_z)$

where the conditions on the remaining free parameters σ^2_λ, μ_z and σ^2_z are given by (3.27) and (3.30). Write $B = \mu_z^2$. Then

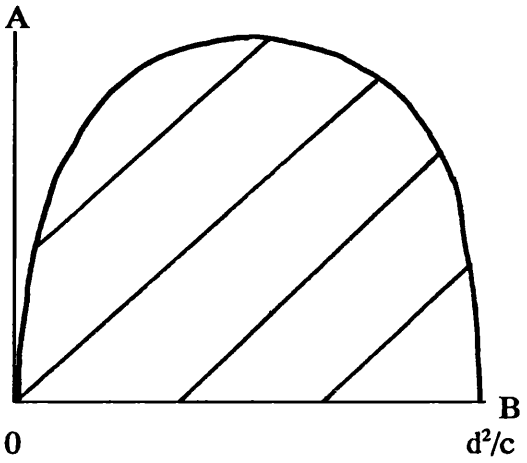
(3.31) $\theta_2 = \sigma^2_\lambda(A+B) + c^2A/B$

with $A = \sigma^2_\mu = d\mu_z/c - \mu_z^2 - \sigma^2_z > 0$ and $B = \mu_z^2$

We are free to vary σ^2_λ in (3.31) between 0 and infinity, while the ranges of A and B are constrained by A>0 and the fact that B is a component of A. The range of

possible values of (A,B) is shown below: clearly the maximum of B is when $\sigma^2_z = 0$.

Range of possible values of A and B



Clearly, θ_2 varies with the value of σ^2_λ between c^2A/B and infinity. Fixing B>0, and setting σ^2_λ to zero, we may clearly allow the value of c^2A/B to become arbitrarily small. Hence θ_2 is free to vary between 0 and infinity.

The argument proceeds exactly as above for the case $\mu=c<0$, in which case also $\mu_z<0$, so that the μ_z/c is positive as before.

We have thus shown that the range of values of θ_2 is not constrained by fixing μ and θ_3. Thus, the range of values of $\{\mu,\theta_1,\theta_2,\theta_3\}$ as $\mathbf{RxR^{+3}}$ ∎

From lemmas 3.1 and 3.2 we see that by combining all the parameters which make up the expectations, variances and covariances in the two models, we can express their covariance structures in the same form.

For each model μ is the average rate of change in attachment for the whole population, θ_1 is the variance due to measurement error, and θ_2 is due to variation in parameters of the model between sites. The parameter θ_3 is non-zero only in model 2, and thus if θ_3 is greater than zero then model 2 (the 'burst' model) will appear to be more appropriate than model 1.

Write $\underline{w}_i = (w_{i1}, w_{i2}, \ldots, w_{ip})'$, $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$, $\underline{\tau} = (\tau_1, \tau_2, \ldots, \tau_p)'$.

We have proved:

THEOREM 3.1: For each of models 1 and 2, $\{\underline{w}_i\}$ are independently and identically distributed with mean $\mu\underline{\tau}$, $(-\infty<\mu<\infty)$ and variance-covariance matrix

(3.32) $\Sigma_{(\theta)} = \theta_1 B_1 + \theta_2 B_2 + \theta_3 B_3$ $(0 \le \theta_3 < \infty,\ 0 < \theta_1, \theta_2, < \infty)$, where:

(3.33) B_1 is a pxp matrix whose diagonal elements equal 2 and off- diagonal elements equal -1.

(3.33) $B_2 = \underline{\tau}\underline{\tau}'$

(3.34) $B_3 = \text{diag}(\underline{\tau})$

and $\theta_3 = 0$ in model 1, $\theta_3 > 0$ in model 2 ∎

3.1.3 SIMULATIONS OF DATA FROM THE TWO MODELS.

In order to illustrate the difficulty in distinguishing data arising from the two models of disease progression, two simulation experiments were undertaken. In each experiment the parameters were chosen so that the mean and variance of an increment were the same for the two models. The simulations were performed using the statistical package MINITAB (Minitab Inc., 3081 Enterprise Drive, State

College, PA 16801, USA). The code for the simulations is shown in Appendix 3.1. Where distributions had not been specified in the formulation of the model, the normal distribution was used. The time intervals were set to 1 and p (the number of observations per site) to 10. Two experiments were performed, with differing frequencies of bursts per time interval in model 2.

Parameter values for experiment 1 - 2 bursts per time interval

Model 1: $\sigma^2 = 0.9506$, $\beta=1$, $\sigma^2_\beta=0.1619$

Model 2: $\sigma^2 = 0.64$, $\lambda=2$, $\sigma^2_\lambda=0.25$, $\mu_z=0.5$, $\sigma^2_\mu=0.0225$, $\sigma^2_z=0.04$, $\xi^2_z=0.0001$

		μ	θ_1	θ_2	θ_3	$E[W_{ij}]$	$Var(W_{ij})$
Parameter values:	Model 1	1	.9506	.1619	0	1	2.0631
	Model 2	1	.64	.1581	.625	1	2.0631

Parameter values for experiment 2 - 0.5 bursts per time interval

Model 1: $\sigma^2 = 1.69$, $\beta=1$, $\sigma^2_\beta=0.63$

Model 2: $\sigma^2 = 0.64$, $\lambda=0.5$, $\sigma^2_\lambda=0.01$, $\mu_z=2.0$, $\sigma^2_\mu=0.25$, $\sigma^2_z=1.0$, $\xi^2_z=0.04$

		μ	θ_1	θ_2	θ_3	$E[W_{ij}]$	$Var(W_{ij})$
Parameter values:	Model 1	1	1.69	.63	0	1	4.01
	Model 2	1	.64	.105	2.625	1	4.01

Plots of data arising from the simulations appear on following pages. Eight simulations appear on each plot; two plots were made for each model in each experiment.

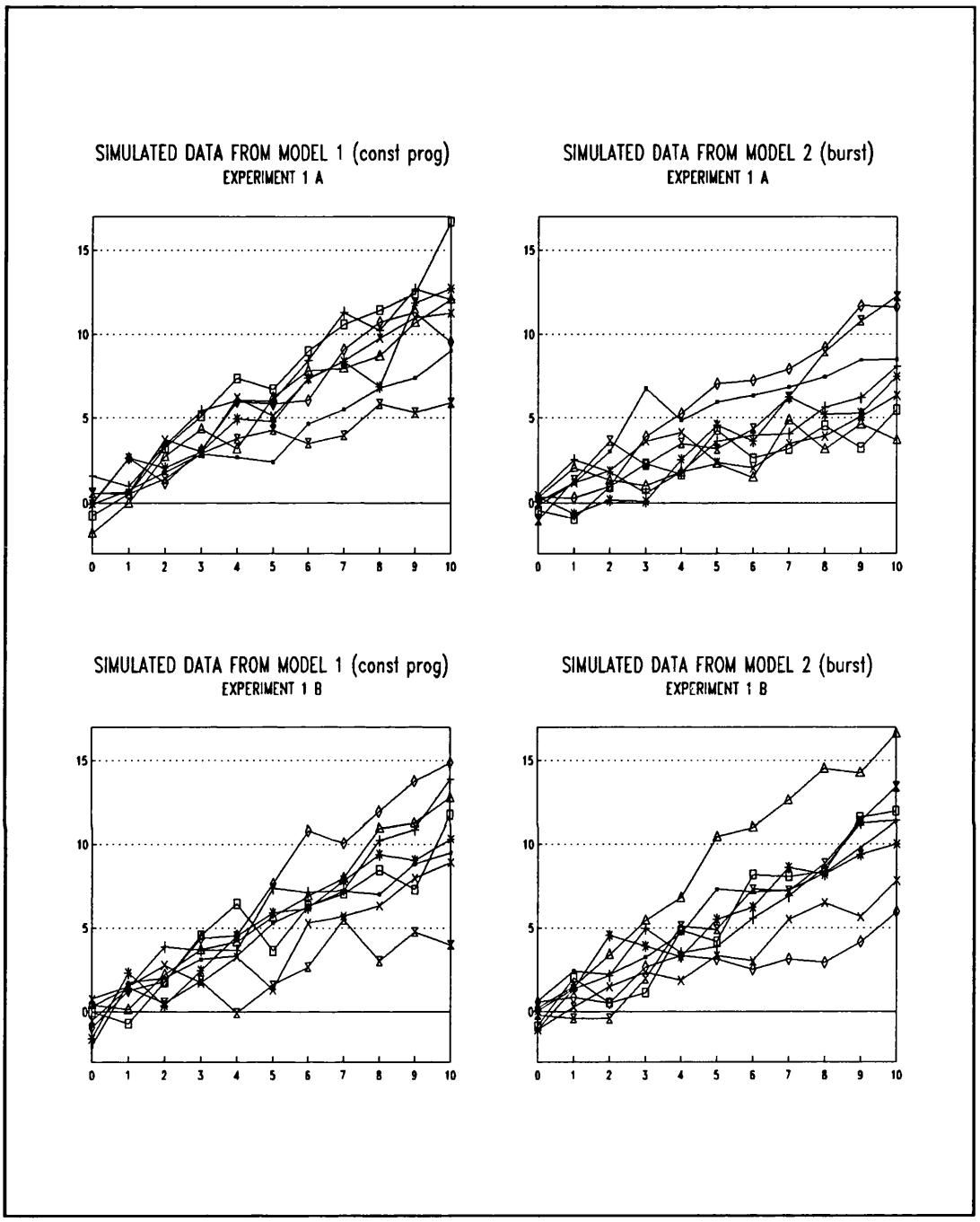


Figure 3.1. Simulations of data for experiment 1.

Twelve members of staff at the MRC Dental Research Unit took part in an experiment to test whether it was possible to distinguish the plots in figures 3.1 and 3.2. The nature of the two underlying processes was explained, and participants were asked to classify the four plots from each experiment into 2 from each process. The results were scored as -1 (all incorrect), 0 (2 incorrect, 2 correct) or

1 (all correct). Clearly, there are 4 possible ways in which the plots can be classified, so that the expected relative frequency under random choice of score 1 is 0.25.

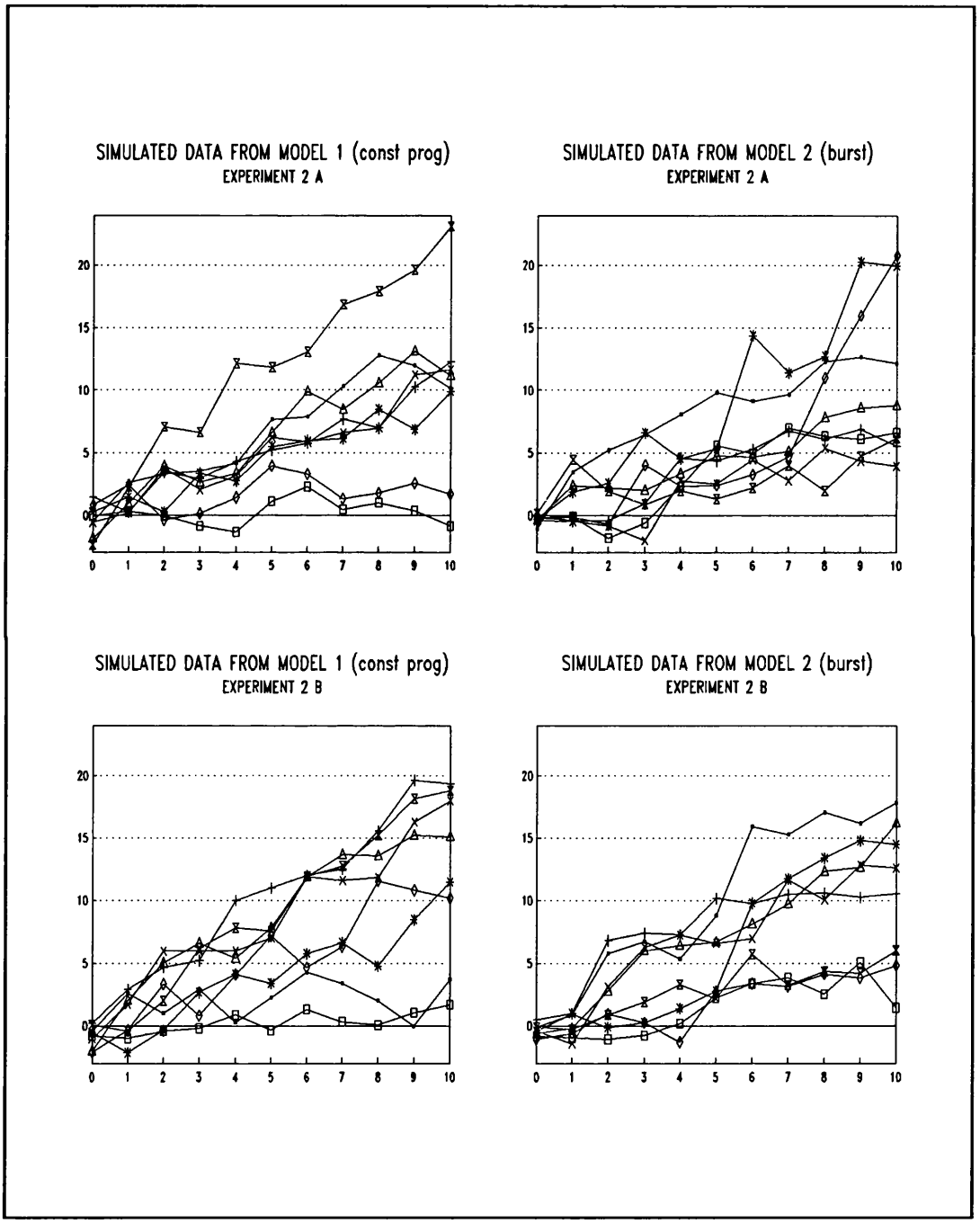


Figure 3.2. Simulations of data from experiment 2.

The actual frequencies of scores -1, 0 and 1 respectively were 2,10,0 for experiment 1 and 2,5,5 for experiment 2. For experiment 1, the frequency of correct results (based on the binomial distribution with n=12 and p=0.25) was worse than random (p=0.0317) while for experiment 2 the frequency was better than random (p=0.0544). The plot for experiment 1, model 1 A was consistently misclassified: this is not surprising if it is compared with those for experiment 1 model 2.

These results indicate that it can be impossible visually to distinguish between data arising from the two models. However an increase in the value of the parameter θ_3, achieved in this case by decreasing the expected frequency of bursts and increasing their expected magnitude, did enable visual distinction, albeit imperfect.

In subsequent sections we shall examine methods for the estimation of parameters of models 1 and 2. We will see whether statistical methods can distinguish data from the two models using the parameters of experiment 1, and we will examine the goodness of fit of the two models for data arising from periodontal research.

3.2 A MODEL FOR NON-INSTANTANEOUS CHANGE.

An interpretation of the covariance structures of models 1 and 2 is as follows. When disease progression is constant, the variation in true disease progression between observation periods at a disease site is necessarily zero. In the burst model, however, there is variation in the rate of disease progression between observation periods. This gives rise to variance component θ_3 in model 2.

Because disease progression is instantaneous, and because of the 'lack of memory' property of the Poisson process, changes in the rate of disease progression at a site are uncorrelated in model 2. Thus the off-diagonal elements of B_3 are zero. In biological terms, an interpretation is that the length of a period of disease activity is small compared to the interval between observation periods. However, the 'burst' model is clearly only one of many possible models in which the rate of disease progression varies. In order to show that the burst model for disease progression is that which best fits available data it will also be desirable to compare it with an alternative model in which the length of episodes of disease activity need not be small.

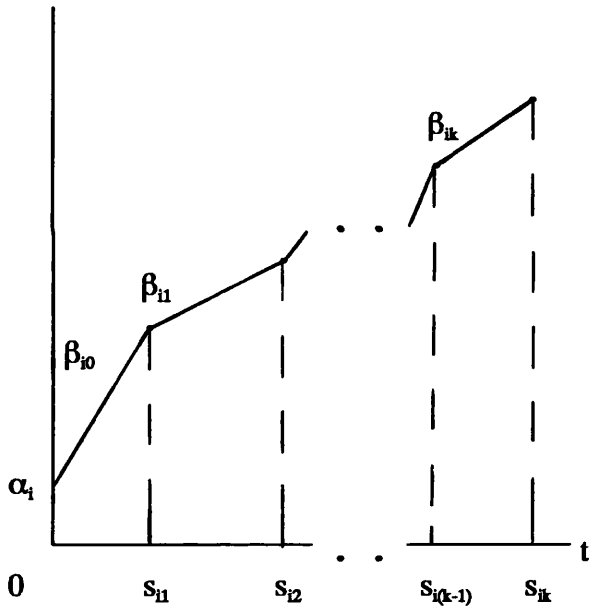
In this section, we specify a model for which there is variation in the rate of disease progression, but disease progression is not instantaneous. As will be seen, this gives rise to a covariance structure in which changes in the rate of disease progression at a site are correlated. We will thus be able to test an alternative to the 'burst' hypothesis in which the rate of disease progression varies, but disease progression cannot be said to be instantaneous.

3.2.1 DEFINITION OF MODEL FOR VARYING, NON-INSTANTANEOUS PROGRESSION.

We use the notation of section 3, so that N, p, $\underline{t}$, $\underline{\tau}$, $X^i(t)$, $\{\underline{y}_i\}$, $\{\underline{e}_i\}$ and $\{\underline{w}_i\}$ are defined as before.

Model 3

(3.35) $X^i(t)=\alpha_i+\beta_{i0}s_{i1}+\beta_{i1}(s_{i2}-s_{i1})+ \ldots +\beta_{ii}(s_{i2}-s_{i1})$, where $k \sim \text{Poisson}(\lambda_i t)$.



The times $\{s_{ij}\}$ at which the changes in rate occur are a Poisson process with rate λ_i for site i. The slope, or rate of disease progression, between s_{ij} and $s_{i(j+1)}$ is β_{ij}, where:

(3.36) $\{\beta_{ij}\} \sim \text{iid}(\beta_i, \sigma^2_{\beta(i)})$.

VARIATION BETWEEN SITES

Again, we suppose that each of the parameters for disease progression at a site is distributed independently between sites about a population mean. Thus:

(3.37) $\lambda_i \sim (\lambda,\sigma^2_\lambda)$; $\beta_i \sim (\beta,\sigma^2_\beta)$; $\sigma^2_{\beta(i)} \sim (\sigma^2_\delta,\xi^2_\delta)$; $\alpha_i \sim (\alpha,\sigma^2_\alpha)$, where

(3.38) $-\infty < \beta, \alpha < -\infty$ and $0 < \lambda, \sigma^2_\lambda, \sigma^2_\beta, \sigma^2_\delta, \xi^2_\delta, \sigma^2_\alpha, < \infty$

We also assume that λ_i, β_i and $\sigma^2_{\beta(i)}$ are mutually independent.

3.2.2 CALCULATION OF EXPECTATION AND COVARIANCE STRUCTURE.

Lemma 3.3: Let $t_1, t_2, \ldots, t_n$ be the order statistics of n i.i.d. variables uniformly distributed on [0,T]. Then:

(i) $E[t_1^2 + (t_1-t_2)^2 + (t_2-t_3)^2 + \ldots + (T-t_n)^2] = 2T^2(n+2)^{-1}$

(ii) $E[t_1] = E[T-t_n] = T(n+1)^{-1}$

Proof: The joint pdf of $t_1, \ldots, t_n$ is $\frac{n!}{T^n}$ $(0 < t_1 < t_2 < \ldots < t_n < T)$.

By symmetry, $E[t_1^2] = E[(t_1-t_2)^2] = E(t_2-t_3)^2] = \ldots = E[(T-t_n)^2]$, so that

$E[t_1^2 + (t_1-t_2)^2 + (t_2-t_3)^2 + \ldots + (T-t_n)^2] = (n+1)E[t_1^2]$

But $E[t_1^2] = \frac{n!}{T^n}\int_0^T \ldots \int t_1^2 \, dt_1 \, dt_2 \ldots dt_n$

$= \frac{n!}{T^n}\frac{2T^{n+2}}{(n+2)!} = \frac{2T^2}{(n+1)(n+2)}$ from which the result is immediate.

(ii) Similarly, $E[t_1^2] = \frac{n!}{T^n}\int_0^T \ldots \int t_2 t_1 \, dt_1 \, dt_2 \ldots dt_n = \frac{T}{n+1} = E[T-t_n]$

∎

Lemma 3.4: Suppose n~Poisson(λ). Then:

(i) $E[(n+2)^{-1}] = \lambda^{-2}(\lambda+e^{-\lambda}-1)$

(ii) $E[(n+1)^{-1}] = \lambda^{-1}(1-e^{-\lambda})$

Proof: We have that $\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1$ and $\sum_{k=0}^{\infty} \frac{k\lambda^k e^{-\lambda}}{k!} = \lambda$.

Thus (i) $E[(n+2)^{-1}] = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k+2)k!} = \sum_{k=0}^{\infty} \frac{(k+1)\lambda^k e^{-\lambda}}{(k+2)!} = \lambda^{-2}\sum_{k=0}^{\infty} \frac{[(k+2)-1]\lambda^{k+2}e^{-\lambda}}{(k+2)!}$

$= \lambda^{-2}[(\lambda-0-\lambda e^{-\lambda})-(1-e^{-\lambda}-\lambda e^{-\lambda})] = \lambda^{-2}(\lambda+e^{-\lambda}-1)$ ∎

Similarly (ii) $E[(n+1)^{-1}] = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k+1)k!} = \lambda^{-1}\sum_{k=0}^{\infty} \frac{\lambda^{k+1}e^{-\lambda}}{(k+1)!}$

$= \lambda^{-1}(1-e^{-\lambda})$ ∎

Recall:

(3.6) $E[X] = E_A[E[X|A=a]$

(3.7) $Var(X) = E_A[Var(X|A=a)] + Var_A(E[X|A=a])$

(3.8) $Cov(X,Y) = E_A[Cov(X,Y|A=a] + Cov_A(E[X|A=a],E[Y|A=a])$

For notational convenience, write the first time at which there is a slope-change after t_j to be r_{ij1}, the next r_{ij2}, and so on up to $r_{ijN(ij)}$, where N(ij) is the number of changes between t_j and t_{j+1} (thus $N(ij) \sim$ Poisson($\lambda_i\tau_j$). Let the corresponding slopes be γ_{ij0} (before r_{ij1}) up to $\gamma_{ijN(ij)}$ (between $r_{ijN(ij)}$ and t_{j+1}). Thus $\gamma_{ij0}=\gamma_{i(j-1)Ni(j-1)}$.

Then from (3.35):

(3.39) $w_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij},\underline{\gamma}_{ij} = \gamma_{ij0}(r_{ij1}-t_j)+\gamma_{ij1}(r_{ij2}-r_{ij1})+ \ldots$

$$+\gamma_{ij(N(ij)-1)}(r_{ijN(ij)}-r_{ij(N(ij)-1)})+\gamma_{ijN(ij)}(t_{j+1}-r_{ijN(ij)})+e_{ij}-e_{i(j-1)}$$

Applying (3.6) successively, and using (3.36) and (3.37), we have:

(3.40) $E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij},\underline{\gamma}_{ij}] = \gamma_{ij0}(r_{ij1}-t_j)+\gamma_{ij1}(r_{ij2}-r_{ij1})+ \ldots$

$$+\gamma_{ij(N(ij)-1)}(r_{ijN(ij)}-r_{ij(N(ij)-1)})+\gamma_{ijN(ij)}(t_{j+1}-r_{ijN(ij)})$$

$E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij}] = \beta_i(r_{ij1}-t_j)+\beta_i(r_{ij2}-r_{ij1})+ \ldots + \beta_i(t_{j+1}-r_{ijN(ij)})$

(3.41) $E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij}] = \beta_i(t_{j+1}-t_j) = \beta_i\tau_j$

(3.42) $E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij)] = E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)}] = E[W_{ij}|\lambda_i,\beta_i] = \beta_i\tau_j$

(3.43) $\mathbf{E[W_{ij}|\lambda_i] = E[W_{ij}] = \beta\tau_j}$

From (3.39), $Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij},\underline{\gamma}_{ij}) = 2\sigma^2$

Applying (3.7) successively, and using (3.36-7) and (3.40-43):

$Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij}) = E_{\underline{\gamma}_{ij}}[2\sigma^2]+Var_{\underline{\gamma}_{ij}}(E[W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij},\underline{\gamma}_{ij}])$

$= 2\sigma^2 + \sigma^2_{\beta(i)}[(r_{ij1}-t_j)^2+(r_{ij2}-r_{ij1})^2+ \ldots + (t_{j+1}-r_{ijN(ij)})^2]$

$Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij)) = E_{\underline{r}_{ij}}[2\sigma^2+\sigma^2_{\beta(i)}((r_{ij1}-t_j)^2+(r_{ij2}-r_{ij1})^2+..+(t_{j+1}-r_{ijN(ij)})^2)]+Var_{\underline{r}_{ij}}(\beta_i\tau_j)$

$= 2\sigma^2+\sigma^2_{\beta(i)}E_{\underline{r}_{ij}}[(r_{ij1}-t_j)^2+(r_{ij2}-r_{ij1})^2+..+(t_{j+1}-r_{ijN(ij)})^2]+Var_{\underline{r}_{ij}}(\beta_i\tau_j)$

Now, given N(ij), {rijk} (k=1,2,..,N(ij)) have the joint distribution of the order statistics of N(ij) iid variables uniformly distributed on $[t_j,t_{j+1}]$ (Grimmett and Stirzaker 1982). Using lemma 3.3 (i), we therefore have:

$Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij)) = 2\sigma^2 + 2\sigma^2_{\beta(i)}\tau_j^2(N(ij)+2)^{-1}$

$Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)}) = E_{N(ij)}[2\sigma^2+2\sigma^2_{\beta(i)}\tau_j^2(N(ij)+2)^{-1}] + Var_{N(ij)}(\beta_i\tau_j)$

Since $N(ij) \sim Poisson(\lambda_i\tau_j)$, we have (using lemma 3.4):

$Var(W_{ij}|\lambda_i,\beta_i,\sigma^2_{\beta(i)}) = 2\sigma^2 + 2\sigma^2_{\beta(i)}\tau_j^2(\lambda_i\tau_j)^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)$

$= 2\sigma^2 + 2\sigma^2_{\beta(i)}\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)$

$Var(W_{ij}|\lambda_i,\beta_i) = E_{\sigma^2_{\beta(i)}}[2\sigma^2+2\sigma^2_{\beta(i)}\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)] + Var_{\sigma^2_{\beta(i)}}(\beta_i\tau_j)$

$= 2\sigma^2 + 2\sigma^2_{\delta}\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)$

$Var(W_{ij}|\lambda_i) = E_{\beta_i}[2\sigma^2 + 2\sigma^2_{\delta}\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)] + Var_{\beta_i}(\beta_i\tau_j)$

$= 2\sigma^2 + 2\sigma^2_{\delta}\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1) + \sigma^2_{\beta}\tau_j^2$

$$\mathrm{Var}(W_{ij}) = E_{\lambda_i}\,[2\sigma^2+2\sigma^2_\delta\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)+\sigma^2_\beta\tau_j^2] + \mathrm{Var}_{\lambda_i}(\beta\tau_j)$$

$$\textbf{(3.44)}\quad \mathbf{Var(W_{ij}) = 2\sigma^2 + \sigma^2_\beta\tau_j^2 + 2\sigma^2_\delta E_{\lambda_i}[\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)]}$$

We now calculate the covariance between increments:

From (3.43), $\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij),\underline{r}_{ij,k},\underline{\gamma}_{ij,k}) = -\sigma^2\delta'$ $(j\neq k)$

where, as in section 3 equation (3.12), $\delta' = 1$ if $|j-k| = 1$, 0 otherwise.

Applying (3.8) successively, and using (3.36-7) and (3.40-43):

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij,k),\underline{r}_{ij,k}) = E_{\underline{\gamma}_{ij,k}}[-\sigma^2\delta']$$
$$+ \mathrm{Cov}_{\underline{\gamma}_{ij,k}}(\gamma_{ij0}(r_{ij1}-t_j)+..+\gamma_{ijN(ij)}(t_{j+1}-r_{ijN(ij)}),\gamma_{ik0}(r_{ik1}-t_k)+..+\gamma_{ikN(ik)}(t_{k+1}-r_{ikN(ik)}))$$

Now, $\gamma_{ijN(ij)} = \gamma_{ik0}$ if there are no slope changes in $[t_{j+1},t_k]$. Since the number of changes in this interval is distributed as Poisson$(\lambda_i(t_k-t_{j+1}))$;

Pr(no changes in $[t_{j+1},t_k]$) $= \exp(-\lambda_i(t_k-t_{j+1}))$.

Since $\mathrm{Cov}_{\underline{\gamma}_{ij,k}}(\gamma_{ijN(ij)},\gamma_{ik0}) = \sigma^2_{\beta(i)}\Pr(\gamma_{ijN(ij)}=\gamma_{ik0})$, we have:

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij,k),\underline{r}_{ij,k}) = -\sigma^2\delta'+\sigma^2_{\beta(i)}\exp(-\lambda_i(t_k-t_{j+1}))(t_{j+1}-r_{ijN(ij)})(r_{ik1}-t_k).$$

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij,k)) = E_{\underline{r}_{ij,k}}[-\sigma^2\delta'+\sigma^2_{\beta(i)}\exp(-\lambda_i(t_k-t_{j+1}))(t_{j+1}-r_{ijN(ij)})(r_{ik1}-t_k)]$$
$$+\mathrm{Cov}_{\underline{r}_{ij,k}}(\beta_i\tau_j,\beta_i\tau_k)$$

Using lemma 3.3 (ii):

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)},N(ij,k)) = -\sigma^2\delta'+\sigma^2_{\beta(i)}\exp(-\lambda_i(t_k-t_{j+1}))\tau_j\tau_k(N(ij)+1)^{-1}(N(ik)+1)^{-1}$$

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)}) = E_{N(ij,k)}[-\sigma^2\delta'+\sigma^2_{\beta(i)}\exp(-\lambda_i(t_k-t_{j+1}))\tau_j\tau_k(N(ij)+1)^{-1}(N(ik)+1)^{-1}]$$

Using lemma 3.4 (ii), and the fact that $\{N(ij)\}$ are independently distributed as Poisson$(\lambda_i\tau_j)$:

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i,\sigma^2_{\beta(i)}) = -\sigma^2_\delta + \sigma^2_{\beta(i)}\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))$$

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i,\beta_i) = E_{\sigma^2_{\beta(i)}}[-\sigma^2\delta'+\sigma^2_{\beta(i)}\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$$
$$+ \mathrm{Cov}_{\sigma^2_{\beta(i)}}(\beta_i\tau_j,\beta_i\tau_k)$$

$$= -\sigma^2\delta' + \sigma^2_\delta\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))$$

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i) = E_{\beta_i}\,[-\sigma^2\delta'+\sigma^2_\delta\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$$

$$+\mathrm{Cov}_{\beta_i}\,(\beta_i\tau_j,\beta_i\tau_k)$$

$$\mathrm{Cov}(W_{ij},W_{ik}|\lambda_i) = -\sigma^2\delta'+\sigma^2_\beta\tau_j\tau_k+\sigma^2_\delta\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))$$

$$\mathrm{Cov}(W_{ij},W_{ik}) = E_{\lambda_i}\,[-\sigma^2\delta'+\sigma^2_\beta\tau_j\tau_k+\sigma^2_\delta\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$$

$$+\,\mathrm{Cov}_{\lambda_i}\,(\beta\tau_j,\beta\tau_k)$$

(3.45) $$\mathrm{Cov}(W_{ij},W_{ik}) = -\sigma^2\delta' + \sigma^2_\beta\tau_j\tau_k$$

$$+\,\sigma^2_\delta E_{\lambda_i}\,[\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$$

Summarising equations (3.43-45), we have thus shown:

Lemma 3.5: The expectations, variances and covariances of $\{w_{ij}\}$, (i=1,2, . . ,N); (j=1,2, . . ,p) are as follows:

$E[W_{ij}] = \tau_j\beta$

$\mathrm{Var}(W_{ij}) = 2\sigma^2 + \sigma^2_\beta\tau_j^2 + 2\sigma^2_\delta E\lambda_i[\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)]$

$\mathrm{Cov}(W_{ij},W_{ik}) = -\sigma^2\delta'+\sigma^2_\beta\tau_j\tau_k+\sigma^2_\delta E\lambda_i[\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$

$\mathrm{Cov}\ W_{ij},W_{i'k}) = 0$ in each model, for $i\neq i'$. ∎

It is apparent that we cannot, as desired, calculate the covariance structure without specifying the distribution of λ_i in more detail. This is because, given λ_i, the covariances are not solely expectations of λ_i or λ_i^2. We therefore now choose a distribution for λ. We require a distribution which is continuous and strictly positive. In keeping with our general principle of specifying simple distributions

where it is necessary to specify them at all we therefore now assume that λ_i has a gamma distribution with parameters a and b, with a>2 and b>0 so that the required expectations exist. That is, λ are i.i.d. with:

(3.46) $f(\lambda) = (\Gamma(a)b^a)^{-1}\lambda^{a-1}e^{-\lambda/b}$.

Lemma 3.6: Suppose $\lambda \sim \Gamma(a,b)$. Then:

(i) $E[\lambda^{-1}] = [b(a-1)]^{-1}$

(ii) $E[\lambda^{-2}] = [b^2(a-1)(a-2)]^{-1}$

(iii) $E[\lambda^{-2}e^{-\lambda t}] = [b^2(a-1)(a-2)]^{-1}(1+bt)^{2-a}$

Proof: The moment generating function M(t) of $\lambda = E[e^{\lambda t}] = (1-bt)^{-a}$ $(t<b^{-1})$. Thus (i) = $\int M(0)dt$, (ii) = $\iint M(0)dt^2$, (iii) = $\iiint M(-t)dt^2$. ∎

We can now use lemma 3.6 to calculate the covariance structure for model 3.

From (3.44), $Var(W_{ij}) = 2\sigma^2 + \sigma^2_\beta\tau_j^2 + 2\sigma^2_\delta E_{\lambda_i}[\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)]$

(3.47) $Var(W_{ij}) = 2\sigma^2 + \sigma^2_\beta\tau_j^2 + 2\tau_j\sigma^2_\delta(b(a-1))^{-1} + 2((1+b\tau_j)^{2-a}-1)\sigma^2_\delta(b^2(a-1)(a-2))^{-1}$

From (3.45), and since $\tau_j = t_{j+1} - t_j$:

$Cov(W_{ij},W_{ik}) = E_{\lambda_i}[-\sigma^2_\delta+\sigma^2_\beta\tau_j\tau_k+\sigma^2_\delta\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$

(3.48) $Cov(W_{ij},W_{ik}) = -\sigma^2\delta' + \sigma^2_\beta\tau_j\tau_k$

$+ 2\tau_j\sigma^2_\delta((b^2(a-1)(a-2))^{-1}[(1+b(t_k-t_{j+1}))^{2-a}-(1+b(t_k-t_j))^{2-a}-(1+b(t_{k+1}-t_{j+1}))^{2-a}+(1+b(t_{k+1}-t_j))^{2-a}]$

Write:

(3.49) $\mu=\beta$, $\theta_1=\sigma^2$, $\theta_2=\sigma^2_\beta$, $\theta_3 =2\sigma^2_\delta(b(a-1))^{-1}$ $(\theta_3>0)$, $\theta_4=a-2$ $(\theta_4>0)$, $\theta_5=b$.

Then, substituting in (3.43), (3.47) and (3.48):

$E[W_{ij}] = \mu\tau_j$

(3.50) $\mathrm{Var}(W_{ij}) = 2\theta_1 + \theta_2\tau_j^2 + \theta_3\tau_j^2 + \theta_3\theta_4^{-1}\theta_5^{-1}((1+\theta_5\tau_j)^{-\theta_4}-1)$

(3.51) $\mathrm{Cov}(W_{ij},W_{ik}) = -\theta_1\delta' + \theta_2\tau_j\tau_k$

$+ \frac{1}{2}\theta_3\theta_4^{-1}\theta_5^{-1}[(1+\theta_5(t_k-t_{j+1}))^{-\theta_4}-(1+\theta_5(t_k-t_j))^{-\theta_4}-(1+\theta_5(t_{k+1}-t_{j+1}))^{-\theta_4}-(1+\theta_5(t_{k+1}-t_j))^{-\theta_4}]$

Write $\underline{w}_i = (w_{i1},w_{i2},\ldots,w_{ip})'$, $\underline{\theta}=(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5)$, $\underline{\tau}=(\tau_1,\tau_2,\ldots,\tau_p)'$. Note that given any value of $\{a,b\}$ (and thus $\{\theta_4,\theta_5\}$), the range of θ_3 varies with σ_δ^2 between 0 and ∞.

We have proved:

THEOREM 3.2: If $\{\theta_1$-$\theta_5\}$ are defined as in (3.49), then $\{w_{ij}\}$ are independently and identically distributed with mean $\mu\underline{\tau}$, $(-\infty<\mu<\infty)$ and variance-covariance matrix:

(3.52) $\Sigma_{(\underline{\theta})} = \theta_1B_1 + \theta_2B_2 + \theta_3B_3 + \theta_3\theta_4^{-1}\theta_5^{-1}B_4(\theta_4,\theta_5)$ $(0<\theta_1,\theta_2,\theta_3,\theta_4,\theta_5<\infty)$

where:

(3.53) B_1 is a pxp matrix whose diagonal elements equal 2 and off-diagonal elements equal -1.

(3.54) $B_2 = \underline{\tau}\underline{\tau}'$

(3.55) $B_3 = \mathrm{diag}(\underline{\tau})$

(3.56) $b_{4ij} = (1+\theta_5\tau_j)^{-\theta_4}-1$

(3.57) $b_{4jk} = \frac{1}{2}[(1+\theta_5(t_k-t_{j+1}))^{-\theta_4}-(1+\theta_5(t_k-t_j))^{-\theta_4}-(1+\theta_5(t_{k+1}-t_{j+1}))^{-\theta_4}-(1+\theta_5(t_{k+1}-t_j))^{-\theta_4}]$

where b_{4jk} is the (j,k)th element of $B_4(\theta_4,\theta_5)$ and $j\neq k$. ∎

3.2.3 PROPERTIES OF MODEL 3.

Clearly model 3 is nested within model 1. However its more complicated algebraic structure means that other of its properties are not obvious. We now consider the properties of that part of the covariance structure which is additional to that of model 1. We write addVar and addCov for the variance and covariances respectively so that:

(3.58) $\text{addVar}(w_{ij}) = \theta_3[\tau_j + \theta_4^{-1}\theta_5^{-1}((1+\theta_5\tau_j)^{-\theta_4}-1)]$

(3.59) $\text{addCov}(w_{ij},w_{ik}) = \theta_3\theta_4^{-1}\theta_5^{-1}b_{4jk}$, ($b_{4jk}$ defined as in (3.57))

Note that:

(3.60) $\text{addVar} = 2\sigma^2_\delta E_{\lambda_i}[\lambda_i^{-2}(\lambda_i\tau_j+\exp(\lambda_i\tau_j)-1)] = 2\sigma^2_\delta E_{\lambda_i} E_{N(ij)}[\tau_j^2(N(ij)+2)^{-1}]$

(3.61) $\text{addCov} = \sigma^2_\delta E_{\lambda_i}[\lambda_i^{-2}\exp(-\lambda_i(t_k-t_{j+1}))(1-\exp(-\lambda_i\tau_j))(1-\exp(-\lambda_i\tau_k))]$

(3.62) $= \sigma^2_\delta E_{\lambda_i} E_{N(ij,k)}[\exp(-\lambda_i(t_k-t_{j+1}))\tau_j\tau_k(N(ij)+1)^{-1}(N(ik)+1)^{-1}]$

Theorem 3.3:

(i) addVar and addCov are decreasing in θ_4. Hence model 3 is asymptotically equivalent to model 2 as θ_4 tends to ∞.

(ii) addVar > addCov if $\tau_j = \tau_k$.

(iii) addCov is decreasing in $t_j - t_{k+1}$.

Proof:

(i) Clearly, if x,c>0, we have that $(1+c)^{-x}$ is decreasing in x for all x. Therefore, from (3.58-59), addCov$\to$0 as $\theta_4\to\infty$ and addVar$\to\theta_3\tau_j$ as $\theta_4\to\infty$. ∎

(ii) If $t_j = t_k$, then (from 3.60 and 3.62):

addVar-addCov = $\sigma^2_\delta E_{\lambda_i} E_{N(ij,k)}[2\tau_j^2(N(ij)+2)^{-1}-\tau_j^2\exp(-\lambda_i(t_k-t_{j+1}))(N(ij)+1)^{-1}(N(ik)+1)^{-1}]$

$> \sigma^2_\delta E_{\lambda_i} E_{N(ij,k)}[2\tau_j^2(N(ij)+2)^{-1}-\tau_j^2(N(ij)+1)^{-1}] > 0.$

(since $\exp(-\lambda_i(t_k-t_{j+1})) < 1$ and $(N_{ik}+1)^{-1} < 1$). ∎

(iii) follows from (3.62). ∎

Thus, the matrix B_4 may be regarded as a generalisation, allowing for differing intervals between observation periods and scale parameters, of the usual autocorrelation matrix R ($r_{ij} = \rho^{|i-j|}$).

3.2.4 COMPARISON OF THE COVARIANCE STRUCTURES OF MODELS 1-3.

We have shown that the covariance structure of model 1 is a limiting case of that of model 2, whose covariance structure in turn is a limiting case of that of model 3. Our aim in formulating each of the models was to provide definitions which are both parsimonious and biologically reasonable. For instance, disease progression in each model is independent of the times at which observations are made: this would not have been true had we defined, for instance, polynomial models for disease progression.

For model 2, changes in the rate of progression are uncorrelated between time intervals. The derivation of the covariance structure of model 2 differs from those of models 1 and 3 in that for models 1 and 3 (see equations (3.25) and (3.49)) the parameters of the covariance structure can be transformed into the parameters of the original model, while for model 1 (see equation (3.26)) the four parameters of the covariance structure are combinations of the six parameters of the model. The

covariance structure of model 2 can in fact be derived from different formulations. For instance, if we modelled disease progression as a Brownian motion process (as the limit of a sequence of independent random changes occurring at short intervals) we would derive a structure where changes between observation periods were uncorrelated. Manji and Nagalkerke (1989) discussed this: their aim was to point out that the burst model was not the only one which could explain observed data. However such a model would fail to meet our second criterion: that of being biologically reasonable. Brownian motion has the property of being everywhere continuous but nowhere differentiable: this property seems unlikely to hold for a process consisting of the destruction of tissue. This also raises the point that our formulation of the burst model is only reasonable in that it represents the idea that disease progression takes place in a time which is short compared to the time between observation periods.

Our formulation of model 3 aimed to allow for non-instantaneous disease progression. Model 3 will be distinguishable from model 2 if covariances between successive increments in attachment level (the off-diagonal elements of B_4) are positive. Factors affecting the covariance between increments are firstly the variation in the rate of progression (the parameter θ_3 must be sufficiently large), and secondly the times between changes in the rate of progression, which could occur either too frequently or too infrequently compared to the time between observations to allow distinction of model 3 from models 1 and 2. In section 4 we examine discrimination between the three models, for simulated data.

3.3 COMPUTER SOFTWARE FOR THE ESTIMATION OF THE COVARIANCE STRUCTURES OF MODELS 1-3.

In section 2.2 we described generalised least-squares methods for the estimation of parameters of covariance structural models. We also reviewed software which is available for the estimation of these models. This software requires models to be expressed in the form of the LISREL model (equations 2.5-8), or an equivalent.

Although models 1 and 2 of have a simple linear covariance structures, that of model 3 is nonlinear and more complex. It does not appear possible to express it in a form which would allow the use of standard programmes. Computer programmes were therefore written in order to estimate the parameters of models 1 to 3; we now calculate the necessary derivatives and describe the programmes.

The discrepancy functions used were those of equation (2.33), with V=S, the sample covariance matrix, and equation (2.48), with U as defined in equations (2.44-47). Thus both the GLS (nk) and GLS (adf) methods were used. As discussed in section 2.2.5, the resulting estimators are representative of broad classes of MDF estimates with optimal properties.

3.3.1 METHOD OF ESTIMATION.

We noted in section 2.2 that the Gauss-Newton method has been recommended for the estimation of covariance structures. However, when this method was used for the estimation of the parameters of model 3, the matrix H (equation (2.37)

frequently became singular at the start of the process, causing computational errors. The programmes therefore employ the Newton-Raphson method (equation 2.57). For linear covariance structures the two methods are equivalent. The modification of Jennrich and Sampson (1968) is employed so that each step results in a reduction of the discrepancy function.

The gradient vector g is the q×1 negative gradient of ½F,

(3.63) $g_i = -\frac{1}{2}\partial F/\partial\gamma_i$

Recall:

(2.29) $F(S,\Sigma(\gamma)|U) = (\underline{s}-\underline{\sigma}(\gamma))'U^{-1}(\underline{s}-\underline{\sigma}(\gamma))$

As in section 2.2, we write $\Sigma^{(i)} = \partial\Sigma/\partial\gamma_i$, $\Sigma^{(ij)} = \partial^2\Sigma/\partial\gamma_i\partial\gamma_j$, and so on. Then for the GLS (adf) method (cf equations (2.26) and (2.42)):

(3.64) $g = \Delta'(\gamma)U^{-1}(\underline{s}-\underline{\sigma}(\gamma))$

while for the GLS (nk) method, g is given by the computationally more efficient form:

(3.65) $g_i = \frac{1}{2}tr(S^{-1}(S-\Sigma(\gamma)S^{-1}\Sigma^{(i)})$

The Hessian matrices of ½F are given by, for the GLS (adf) method:

(3.66) $H(\gamma|S,U) = \Delta'(\gamma)U^{-1}\Delta(\gamma) - \phi(\gamma)$ $(\phi(\gamma)_{ij} = (\sigma^{(ij)})U^{-1}(\underline{s}-\underline{\sigma}(\gamma)))$

and for GLS (nk) methods.

(3.67) $H_{ij} = \frac{1}{2}tr(S^{-1}\Sigma^{(i)}S^{-1}\Sigma^{(j)} - S^{-1}(S-\Sigma(\gamma))(\underline{\Sigma}^{(ij)})$

We discussed in section 2.2.6 the estimation of covariance structures subject to functional constraints. The parameters θ_i (i=1,..,5) of models 1-3 are all variance components which are constrained to be positive. Since the Newton-Raphson method

does not necessarily produce positive parameter estimates it was necessary to constrain the estimation procedure. This was done by defining $\theta_i = \gamma_i^2$ and estimating $\Sigma(\underline{\gamma})$ for each model. Clearly the minima of the structures $\Sigma(\underline{\theta})$ ($\underline{\theta}>0$) and $\Sigma(\underline{\gamma})$ are equal. As discussed in section 2.2.6, the choice between reparametrization and constrained estimation is one of computational convenience: the results are equivalent.

Such a reparametrization has disadvantages. Firstly, the procedure takes longer to converge and seems more sensitive to the initial values. Secondly, for linear covariance structures such as those of models 1 and 2 the Newton-Raphson algorithm converges in a single step. We can express any linear covariance structure in the form $\underline{\sigma}(\underline{\gamma}) = \Delta\underline{\gamma}$, so that setting $\underline{g} = 0$ in (3.64) gives:

(3.68) $\dot{\underline{\gamma}} = (\Delta' U^{-1}\Delta)^{-1}\Delta' U^{-1}\underline{s}$

Under the assumption of no kurtosis, the equivalent equation is:

(3.69) $\dot{\underline{\gamma}} = \{\theta(S^{-1})\}^{-1}\Delta\underline{s}$, where $\theta(V) = \Delta'(V\otimes V)\Delta$

In an attempt to take account of these disadvantages, the following procedure was devised. The process starts with the original parametrization. If, at any stage, the procedure produces a negative estimate for any element of $\underline{\theta}$, then the model is reparametrized by setting $\theta_i=\gamma_i^2$ (i=1,..,q, where q is the number of parameters of the model). From then on the value of $\underline{\gamma}$ is estimated until convergence is achieved.

3.3.2 CALCULATION OF DERIVATIVES.

To calculate the gradient vector and Hessian matrices we require the first and second derivatives of the covariance structure. We shall first calculate the derivatives with respect to $\underline{\theta}$. Then:

(3.70) $\Sigma^{(1)} = B_i$ (i=1,2, models 1, 2 and 3)

(3.71) $\Sigma^{(3)} = 0$ (model 1), B_3 (model 2), $B_3+\theta_4^{-1}\theta_5^{-1}B_4$

For model 3 we require also $\Sigma^{(4)}$ and $\Sigma^{(5)}$. Since B_4 is a function of θ_4 and θ_5, we need its partial derivatives. We shall define

$B_4^{(4)}$ and $B_4^{(5)}$ to be, respectively, the partial derivatives of B^4 with respect to θ_4 and θ_5. Thus:

(3.72) $\Sigma^{(4)} = 2\theta_3[-\theta_4^{-2}\theta_5^{-1}B_4 + \theta_4^{-1}\theta_5^{-1}B_4^{(4)}]$, and

(3.73) $\Sigma^{(5)} = 2\theta_3[-\theta_4^{-1}\theta_5^{-2}B_4 + \theta_4^{-1}\theta_5^{-1}B_4^{(5)}]$

To calculate $B_4^{(4)}$ and $B_4^{(5)}$ we note that if $f(x,y)=(1+ky)^{-x}$, then:

$\partial f/\partial x = -\log(1+ky)(1+ky)^{-x}$ and $\partial f/\partial y = -kx(1+ky)^{-(x+1)}$. Writing

(3.74) $k_1 = t_k-t_{j+1}$, $k_2 = t_k-t_j$, $k_3 = t_{k+1}-t_{j+1}$ and $k_4 = t_{k+1}-t_j$

we thus have:

(3.75) $b_{4jj}^{(4)} = -\log(1+\theta_5\tau_j)(1+\theta_5\tau_j)^{-\theta 4}$

(3.76) $b_{4jk}^{(4)} = -\frac{1}{2}[-\log(1+\theta_5k_1)(1+\theta_5k_1)^{-\theta 4} + \log(1+\theta_5k_2)(1+\theta_5k_2)^{-\theta 4}$
$+ \log(1+\theta_5k_3)(1+\theta_5k_3)^{-\theta 4} - \log(1+\theta_5k_4)(1+\theta_5k_4)^{-\theta 4}$

(3.77) $b_{4jj}^{(5)} = -\tau_j\theta_4(1+\theta_5\tau_j)^{-(\theta 4+1)}$

(3.78) $b_{4jk}^{(5)} = -\frac{1}{2}[-k_1\theta_4(1+\theta_5k_1)^{-(\theta 4+1)} + k_2\theta_4(1+\theta_5k_2)^{-(\theta 4+1)}$
$+ k_3\theta_4(1+\theta_5k_3)^{-(\theta 4+1)} - k_4\theta_4(1+\theta_5k_4)^{-(\theta 4+1)}$

Similarly:

(3.79) $\Sigma^{(11)} = \Sigma^{(22)} = \Sigma^{(33)} = \Sigma^{(12)} = \Sigma^{(13)} = \Sigma^{(23)} = 0$ (models 1, 2 and 3)

For model 3:

(3.80) $\Sigma^{(14)} = \Sigma^{(15)} = \Sigma^{(24)} = \Sigma^{(25)} = 0$

(3.81) $\Sigma^{(34)} = -\theta_4^{-2}\theta_5^{-1}B_4 + \theta_4^{-1}\theta_5^{-1}B_4^{(4)}$

(3.82) $\Sigma^{(35)} = -\theta_4^{-1}\theta_5^{-2}B_4 + \theta_4^{-1}\theta_5^{-1}B_4^{(5)}$

Then:

(3.83) $\Sigma^{(44)} = \theta_3\theta_4^{-1}\theta_5^{-1}[2\theta_4^{-2}B_4-2\theta_4^{-1}B_4^{(4)}+B_4^{(44)}]$

(3.84) $\Sigma^{(55)} = \theta_3\theta_4^{-1}\theta_5^{-1}[2\theta_5^{-2}B_4-2\theta_5^{-1}B_4^{(5)}+B_4^{(55)}]$

(3.84) $\Sigma^{(45)} = \theta_3\theta_4^{-1}\theta_5^{-1}[2\theta_4^{-1}\theta_5^{-1}B_4-\theta_4^{-1}B_4^{(5)}-\theta_5^{-1}B_4^{(4)}+B_4^{(45)}]$

To calculate $B_4^{(44)}$, $B_4^{(55)}$ and $B_4^{(45)}$ note that if $f(x,y)=(1+ky)^{-x}$, then:

$\partial^2 f/\partial x^2 = \{\log(1+ky)\}^2(1+ky)^{-x}$, $\partial^2 f/\partial y^2 = k^2x(x+1)(1+ky)^{-(x+2)}$ and $\partial^2 f/\partial x\partial y = k[x\log(1+ky)-1](1+ky)^{-(x+1)}$

Thus:

(3.85) $b_{4jj}^{(44)} = \{\log(1+\theta_5\tau_j)\}^2(1+\theta_5\tau_j)^{-\theta_4}$

(3.86) $b_{4jk}^{(44)} = \frac{1}{2}[\{\log(1+\theta_5k_1)\}^2(1+\theta_5k_1)^{-\theta_4} - \{\log(1+\theta_5k_2)\}^2(1+\theta_5k_2)^{-\theta_4}$
$- \{\log(1+\theta_5k_3)\}^2(1+\theta_5k_3)^{-\theta_4} - \{\log(1+\theta_5k_4)\}^2(1+\theta_5k_4)^{-\theta_4}]$

(3.87) $b_{4jj}^{(55)} = \tau_j^2\theta_4(\theta_4+1)(1+\theta_5\tau_j)^{-(\theta_4+2)}$

(3.89) $b_{4jk}^{(55)} = \frac{1}{2}[k_1^2\theta_4(\theta_4+1)(1+\theta_5k_1)^{-(\theta_4+2)} - k_2^2\theta_4(\theta_4+1)(1+\theta_5k_2)^{-(\theta_4+2)}$
$- k_1^2\theta_4(\theta_4+1)(1+\theta_5k_1)^{-(\theta_4+2)} + k_2^2\theta_4(\theta_4+1)(1+\theta_5k_2)^{-(\theta_4+2)}]$

(3.90) $b_{4jj}^{(45)} = \tau_j[\log(1+\theta_5\tau_j)-1](1+\theta_5\tau_j)^{-(\theta_4+1)}$

(3.91) $b_{4jk}^{(45)} = \tau_j[k_1\log(1+\theta_5\tau_j)-1](1+\theta_5k_1)^{-(\theta_4+1)} - k_2\log(1+\theta_5\tau_j)-1](1+\theta_5k_2)^{-(\theta_4+1)}$
$k_3\log(1+\theta_5\tau_j)-1](1+\theta_5k_3)^{-(\theta_4+1)} + k_4\log(1+\theta_5\tau_j)-1](1+\theta_5k_4)^{-(\theta_4+1)}]$

3.3.3 CALCULATION OF DERIVATIVES WITH RESPECT TO $\underline{\gamma}$ ($\underline{\gamma}^2=\underline{\theta}$).

Derivatives $\Sigma^{(i)}$ are now with respect to γ_i, but we continue to express the derivatives of B_4 with respect to $\underline{\theta}$, so that $\partial B_4/\partial\gamma_i = 2\gamma_i B_4^{(i)}$. Thus:

(3.92) $\Sigma^{(i)} = 2\gamma_i B_i$ (i=1,2; models 1, 2 and 3)

(3.93) $\Sigma^{(3)} = 0$ (model 1), $2\gamma_3 B_3$ (model 2), $2\gamma_3(B_3+\gamma_4^{-2}\gamma_5^{-2}B_4)$ (model 3)

For model 3:

(3.94) $\Sigma^{(4)} = 2\gamma_3^2[-\gamma_4^{-3}\gamma_5^{-2}B_4 + \gamma_4^{-1}\gamma_5^{-2}B_4^{(5)}]$ and

(3.95) $\Sigma^{(5)} = 2\gamma_3^2[-\gamma_4^{-2}\gamma_5^{-3}B_4 + \gamma_4^{-2}\gamma_5^{-1}B_4^{(5)}]$

Similarly:

(3.96) $\Sigma^{(ii)} = 2B_i$ (i=1,2; models 1, 2 and 3)

(3.97) $\Sigma^{(33)} = 0$ (model 1), $2B_3$ (model 2), $2(B_3+\gamma_4^{-2}\gamma_4^{-2}B_4)$ (model 3)

(3.98) $\Sigma^{(12)} = \Sigma^{(13)} = \Sigma^{(23)} = 0$

For model 3:

(3.99) $\Sigma^{(14)} = \Sigma^{(15)} = \Sigma^{(24)} = \Sigma^{(25)} = 0$

(3.100) $\Sigma^{(35)} = 4\gamma_3[-\gamma_4^{-3}\gamma_5^{-2}B_4 + \gamma_4^{-1}\gamma_5^{-2}B_4^{(4)}]$

(3.101) $\Sigma^{(35)} = 4\gamma_3[-\gamma_4^{-2}\gamma_5^{-3}B_4 + \gamma_4^{-2}\gamma_5^{-1}B_4^{(4)}]$

(3.102) $\Sigma^{(44)} = 4\gamma_3^2\gamma_5^{-2}[3\gamma_4^{-4}B_4 - 3\gamma_4^{-2}B_4^{(4)}+2B_4^{(44)}]$

(3.103) $\Sigma^{(55)} = 4\gamma_3^2\gamma_4^{-2}[3\gamma_5^{-4}B_4 - 3\gamma_5^{-2}B_4^{(5)}+2B_4^{(55)}]$

(3.104) $\Sigma^{(45)} = 4\gamma_3^2\gamma_4^{-1}\gamma_5^{-1}[\gamma_4^{-2}\gamma_5^{-2}B_4 - \gamma_4^{-2}B_4^{(5)} - \gamma_5^{-2}B_4^{(4)} + B_4^{(45)}]$

3.3.4 SOFTWARE FOR THE ESTIMATION OF COVARIANCE STRUCTURES.

Computer programmes to estimate parameters for each of models 1-3 of sections 1 and 2 were written in APL (STSC, Inc) running under the MS DOS operating

system. The APL equivalent of a 'programme' in other computer languages is a user-defined function. The software for estimation of covariance structures thus consists of a number of functions. The main function, RUN, calls other functions, which in turn may call still more functions.

Separate programmes were written for the two methods of generalised least-squares estimation described in section 3 - the method assuming that the underlying distribution has no kurtosis and the asymptotically distribution-free method. The two programmes have similar structures and similar code, but are sufficiently different to make writing them as one programme cause more problems than are saved by the reduction in duplication.

Flow charts for the two programmes are shown in Appendix 4.1. The only difference in structure is that the ADF procedure has the additional function SETU, called by the READ function, which calculates the $p^{*} \times p^{*}$ matrix U which is used to calculate the discrepancy function.

The structure shown is that for the estimation of parameters for model 3. Different functions (INIMOD1 or INIMOD2, etc.) are used for the estimation of parameters from models 1 and 2. The function RUN uses READ, INIMOD3 and ESTIM successively, as shown. Each of the functions INIMOD3 and ESTIM uses the functions SIGMOD3, DERIVM3, GRADM3 and HESMOD3 in that order. The function DERIVM3, which is used to calculate the derivatives of the matrix B_4, is not needed for the other two models.

The programmes read data in one of two formats. The time periods between observations can be specified in the data (this format was used in section 5) or may be specified in the READ function with only the data in the data file (this format was used for the simulation data analysed in the next section. In either case each data set is preceded in the data file by a line specifying N, the number of observations contained in the data set.

Estimates of the asymptotic covariance matrix of g are given (from propositions 1.2 and 2.2 of section 2.2) by $n^{-1}(\dot{\Delta}'U^{-1}\dot{\Delta})^{-1}$ for the asymptotically distribution-free method, and by $2n^{-1}\{\Theta(S)\}^{-1}$ for the method assuming no kurtosis, $((\Theta(S)_{ij} = tr(\dot{\Sigma}^{(i)}S\dot{\Sigma}^{(j)}S))$.

The function RUN outputs the site number, number of iterations, number of parameters, status ('OK' if no computational errors occur, 'ER' otherwise), parameter estimates and goodness of fit statistic to a file. Where no computational errors occur, the estimated covariance matrix is also output, otherwise the error message is output.

The complete programmes for the GLS (nk) procedure are shown in appendix 1, at the end of the thesis. The programmes for the GLS (adf) procedure, where these differ from the no kurtosis procedure, are shown in appendix 2.

APPENDIX 3.1 MINITAB INSTRUCTIONS TO SIMULATE DATA FROM MODELS 1 AND 2.

Code for model 1

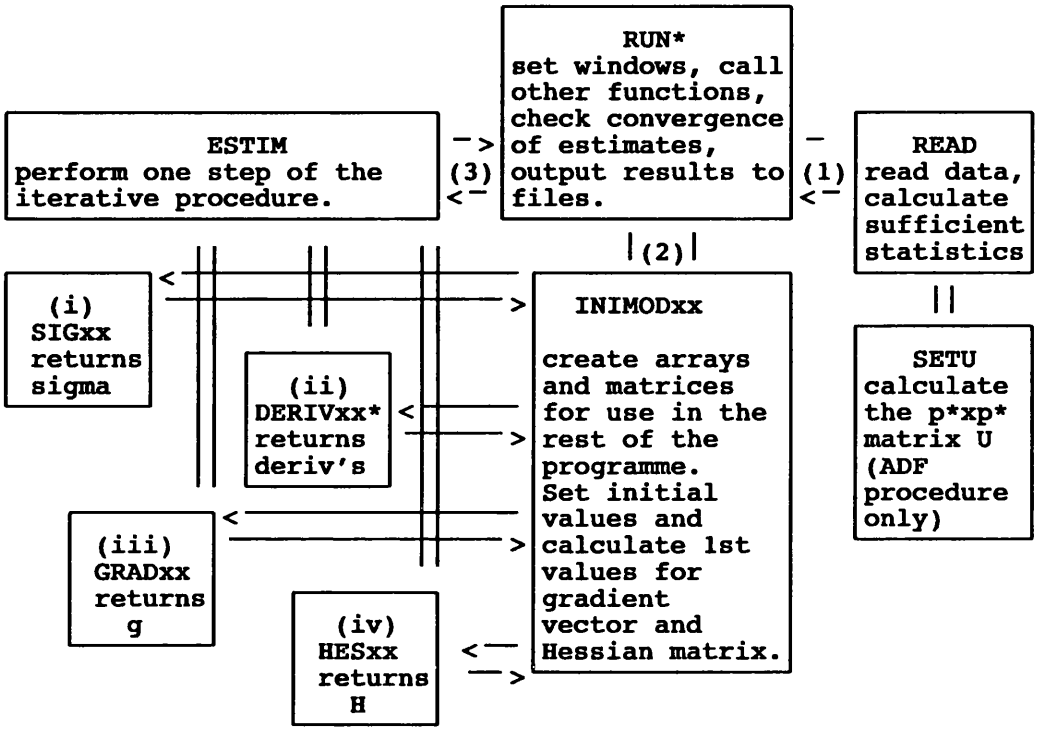
```
random 1 c25;
normal 2 0.5.
copy c25 k10
let k10=1/k10
NOTE K10 IS THE POISSON RATE OF OCCURRENCE OF BURSTS
random 1 c25;
normal 0.5 0.15.
copy c25 k11
NOTE K11 IS THE EXPECTED BURST MAGNITUDE
random 1 c25;
normal 0.2 0.01.
copy c25 k12
let k12=k12**0.5
NOTE K12 IS THE BURST STANDARD DEVIATION
random 50 c1;
expo k10.
NOTE GENERATE POISSON PROCESS
let k5=0
stack k5 c1 c1
parsum c1 c2
copy c2 c3;
omit c2 = 10:1000.
NOTE C2 CONTAINS ZERO, THEN TIMES AT WHICH BURSTS OCCUR
n c3 k1
random k1 c4;
normal k11 k12.
copy k5 c10
stack k5 c4 c4
copy c4 c4;
omit 2.
NOTE C4 CONTAINS MAGNITUDE OF BURSTS
random 11 c6;
normal 0 0.8.
NOTE C6 CONTAINS MAGNITUDE OF MEASUREMENT ERRORS
let k2=1
exec 'observe' 10
NOTE OBSERVE RETURNS TRUE ATTACHMENT LEVELS AT TIMES 0 TO
10, IN C10

copy c3 c4 c13 c14;
use c3=0:k2.
sum c14 k3
stack c10 k3 c10
let k2=k2+1
end
let c12=c10+c6
```

Code for model 2

```
random 1 c2;
normal 1 0.4023369.
copy c2 k2;
use 1.
random 11 c3;
normal 0 0.975.
let c4=(k2*c11)+c3
end
```

APPENDIX 3.2 STRUCTURE OF THE PROGRAMMES.



* - function is the same for the two procedures.

xx at the end of a function name means a different function is used for different models.

The programmes require the following variables to have values specified before starting:

BOX - is a character string which draws a box on the screen.
COL - specifies the number of columns to be skipped at the start of each line in the data file.
FORMAT - 1 if time intervals are read from the data file, 2 if not (for simulation data).
INIT - name of initialisation routine (INIMOD1, INIMOD2 or INIMOD3).
PMFILE - name of file to which output is to be written.
SCALE - Integer by which the time intervals are divided.

The APL command which starts the programmes is:

'XXX' RUN 'YYY'

where XXX is the name of the APL output file and YYY is the name of the file containing the data.

4

ESTIMATION OF COVARIANCE STRUCTURES FOR DATA FROM SIMULATIONS OF DIFFERENT MODELS FOR PERIODONTAL DISEASE PROGRESSION.

In section 2.2 we described methods for the analysis of covariance structures which, given suitable computer software, may be used to analyse data from studies of periodontal disease progression in order to assess which of models 1 to 3 (constructed in section 3) best describes the data. However, it was noted that the properties of the generalised least-squares estimates of covariance structures which were presented are asymptotic, and that little is known of their finite-sample behaviour.

It was also noted that the aim of constructing the three models is to derive covariance structures which reflect the characteristics of different postulated models for disease progression. We do not expect that any of the models will provide an exact description of the progression of disease. Further, the algebra used to derive the models is difficult to check and the software used to fit them is non-standard. It is therefore sensible to check that results are as expected when the true distribution of the data is known.

Before we analyse data arising from studies of periodontal disease progression, we therefore, in this section, analyse data from simulations of the three models for disease progression which were constructed in section 3, using the generalised least-squares methods of section 2.2.

In a recent issue of the Royal Statistical Society's News and Notes, S J Senn offered the following definition as part of a glossary of statistical terms.

Experimental Design: A subject whose findings every statistician believes any scientist should take into account when conducting an experiment unless the scientist is a statistician carrying out a simulation.

Taking note of this, we specify the aims of these experiments, which were:

1. To confirm that the algebra and the software do not contain errors.

2. To show that it is possible to distinguish the 3 models, given that we know the underlying distribution of the data.

3. To discover the extent of the finite-sample bias of parameter estimates.

4. To confirm the accuracy of the estimated standard errors of the parameter estimates.

Four simulation experiments were performed, with four data sets in each experiment. Parameters were estimated using each of models 1 to 3. The data were analysed using the programmes described in section 3.3, using both the GLS (nk) and the GLS (adf). The output from the APL programmes (see section 2.2.9 for summaries of the definitions of the reported statistics) was formatted using SAS: the results for each experiment are annotated as follows:

EXPT - experiment number (1 to 4)

METH - method (nk - assuming no kurtosis of underlying distribution, adf - asymptotically distribution free)

MOD - model (1 to 3)

DS - data set (1 to 4 for each experiment)

FROMMOD - Model used to simulate data (in experiments 1 and 2)

TYPE - F (full, unrounded data) or R (rounded data) (experiments 3 and 4).

ITER - number of iterations before termination of procedure (if 500 then convergence was not achieved)

STATUS - OK - no error occurred

ER - computational error occurred

Ti - $\hat{\theta}_i$

Si - standard error (= square root of estimated variance) of θ_i

CHIi - Chi-square goodness of fit statistic for model i on $p^* = p(p+1)/2-q$ degrees of freedom (53 df for model 1, 52 df for model 2 and 50 df for model 3.)

DIFFij - Difference between goodness of fit statistics for models i and j. Under regularity conditions given by Steiger et al. (1985) this statistic is distributed as chi-square with degrees of freedom equal to the difference in the number of parameters of the models.

Results are presented to six significant figures, although the output from the APL programmes is to a higher degree of accuracy.

It will be noticed as the results are presented that for model 3 they are not well-behaved. Firstly, the parameter estimates of θ_4 and θ_5 are highly unstable and the information matrix frequently near-singular. Secondly, the improvement in goodness of fit of model 3 over the other models 1 and 2 for data generated by these models is very small, rather than being distributed as at least χ^2_2, as might have been expected from the asymptotic theory. The increase in goodness of fit for data generated by model 3 is also, under certain conditions, smaller than expected.

It should be emphasised firstly that we do not believe that these results are due to computational or algebraic errors: extensive checks were made precisely because the results were puzzling and (see section 4.2.1) independently written software gave similar results for the GLS (nk) method. Secondly we are able to find conditions where the goodness of fit of model 3 over the other two models is markedly increased. The findings cannot therefore be blamed simply on the impossibility of distinguishing model 3 from model 2. We discuss the results for model 3 further at the end of this section.

4.1 SIMULATIONS OF DATA FROM MODELS 1 AND 2 (EXPERIMENTS 1 AND 2).

In section 3.1, we displayed plots of simulated data from models 1 and 2 which had been generated using MINITAB. Two sets of parameters were used; in each case they were specified so that the expectation and variance of an increment were the same in the two models. The parameter values used are shown in section 3.1.3, while the instructions used to generate the data are shown in Appendix 3.1. The

data were rounded to one decimal place, so that the results approximate the high degree of rounding seen in the measurement of periodontal disease.

The same sets of parameter values and the same MINITAB instructions were used to generate larger data sets for analysis using the software described in section 3.3. Two experiments (corresponding to the two sets of parameter values) were performed.

For experiments 1 and 2, four data sets were generated; data sets 1 and 2 using model 2 and data sets 3 and 4 using model 1, with sample size n=200 and the number of increments per "site", p=10. Parameters were estimated using the covariance structures for each of the three models.

Initial values

Since models 1 and 2 have linear covariance structures, the procedure will converge in one iteration providing the parameter estimates are all positive, whatever the initial values. The initial values did not affect the parameter estimates where the quadratic reparametrisation of model 2 was used (because the estimate of θ_3 given by the linear parametrisation was less than zero). However, the initial values did affect the parameter estimates for model 3, which is non-linear. This is discussed below.

EXPERIMENT 1

The results of experiment 1 are shown in appendix 4.1. Recall that where plots were made using simulated data with identical distributions to those used in experiment 1, it proved impossible to distinguish visually data from models 1 and 2.

For models 1 and 2, the procedure converged in two iterations, except for data set 4, adf method. For model 3, the estimates shown were (on all but one occasion) generated using initial values which were the true parameters for the model 2 parameters, and (30,.1) for (θ_4,θ_5). For data set 4, method nk, these values gave goodness of fit slightly worse than for model 2, while the estimates shown using initial values (6,.1) for (θ_4,θ_5), gave slightly better goodness of fit. Usually, where smaller initial values for θ_4 were used, the final goodness of fit was worse than for model 2.

Goodness of fit

For the data sets generated using model 2, there was a large improvement in the goodness of fit of model 2 over model 1. For the data sets generated using model 1, on the other hand, the improvement in goodness of fit when model 2 was assumed was small. Where the model used was correct for the data set the procedure always converged in two iterations. For model 3, small improvements in goodness of fit only were achieved over models 1 and 2.

Parameter Estimates

Where the model was that used to generate the data set, the parameter estimates were around the values expected, but appeared to show a negative bias compared to the true values. This applied to both estimation procedures. Where the parameters of model 2 were estimated from data generated using model 1, the estimate of θ_3 was close to zero. Where parameters for model 3 were estimated, the estimates varied widely, although θ_4 was often large.

Standard Errors

Using the approximate guide that 19 out of 20 parameter estimates should lie within two standard errors of the true value, it appeared from inspection of the duplicate data sets generated using each model that the estimated standard errors do give an approximate guide to the variability of the parameter estimates, given the negative bias of the parameter estimates.

Where parameter estimates were close to zero, or were large (for model 3) the estimated standard errors were large. For model 3, the information matrix was usually singular, so that standard errors could not be computed. This singularity suggests that one of this pair of parameters is redundant. However, fixing one always resulted in substantially worse goodness of fit than if both were allowed to vary.

EXPERIMENT 2

For experiment 2, the parameters used were the same as for the second half of the simulation experiment in section 3.1.3. Recall that θ_3 (the parameter which is non-zero only for model 2), is considerably larger than for experiment 1, and that when plots were made of data simulated using these parameter values, it proved possible to discriminate visually between data generated using the two models. For data generated using model 2, the mean frequency of bursts was 0.5 per time interval. The results of experiment 2 are shown in appendix 4.2.

Goodness of fit

For data sets 1 and 2, the increase in θ_3 is reflected in an increased goodness of fit of model 2 over model 1, compared to experiment 1. The increase in goodness of fit of model 3 over model 2 was even more minute than for experiment 1.For data sets 3 and 4 there was (to five decimal places) no increase in the goodness of fit of models 2 and 3 over model 1.

Parameter Estimates

Again, where the model was correct for the data set, the estimates were slightly negatively biased. The estimates of θ_4 and θ_5 for model 3 again varied widely.

Standard Errors

It appeared, as for experiment 1, that the estimated standard errors provided a reasonable indication of the variation in the parameter estimates, given the negative bias of the parameter estimates.

4.2 SIMULATIONS OF DATA FROM MODEL 3 (EXPERIMENTS 3 AND 4).

For experiments 3 and 4, data was simulated using model 3. The results are shown in appendices 4.3 and 4.4, and the MINITAB code used is shown in appendix 4.5. Because of the increased complexity and non-linearity of model 3, a sample size of n=500 was used, with p=10 increments, as in experiments 1 and 2. The initial values used for model 3 were the correct parameter values. The estimates for models 1 and 2 were all reached in two iterations, and thus did not depend on the initial values used.

It was noticed during the analysis of simulated data from model 3 that under some conditions the effect of rounding on the results could be substantial. Each data set was output to four decimal places, and then rounded to no decimal places. The covariance structures for the full and rounded data were then analysed separately.

EXPERIMENT 3.

This experiment confirmed that data from model 3 could be distinguished from models 1 and 2 using the GLS procedures.

Goodness of fit

There was an increase in goodness of fit of between 20.36 and 50.97 ($\sim\chi^2_2$) of model 3 over model 2. However, the increase in goodness of fit of model 2 over model 1 was much more marked (between 310.97 and 390.81 ($\sim\chi^2_1$). While the goodness of fit of model 3 was always reduced for the rounded data, that of model 2 was increased as often as it was reduced. Although the increase in goodness of fit of model 3 over model 2 was generally reduced for the rounded data, model 3 clearly fitted the data best on each occasion.

Parameter Estimates and standard errors.

For model 3, the estimates of θ_1 θ_2 and θ_3 were close to the true values. Although the estimates of θ_4 and θ_5 varied markedly, with large estimated standard errors, the estimate of λ (the rate of changes in progression, given by $(\theta_4+2)\theta_5$) was much more stable. The estimated standard errors of θ_3, θ_4 and θ_5 were substantially larger than those of θ_1 and θ_2.

It appears that for model 3, it is difficult to estimate θ_4 and θ_5 because the discrepancy function is "flat" for these parameters. This is possibly not surprising: these are parameters of the population distribution of the Poisson rate of occurrence of changes in the rate of attachment change, with the expected rate of change equal to ab (θ_4=a-2, θ_5=b), so that if the product ab is constant only higher moments of the distribution are affected by changes in the values of these parameters.

EXPERIMENT 4

For this experiment, the value of σ^2_δ, the population variance of the slope, was reduced from 1 in experiment 3 to 0.04. This meant that the value of θ_3 was reduced to 0.1143. The values of θ_4 and θ_5 were the same as for experiment 3.

Goodness of fit

The increase in goodness of fit of model 3 over model 2, for this experiment, was at most 2.44, while the 95th percentile of the χ^2_2 distribution is 5.99. For the rounded data, the increase in goodness of fit was in all but one case reduced by 75% compared to the unrounded data. The increase in goodness of fit of model 2 over model 1 was far larger: between 9.47 and 24.15 for the unrounded data and between 39.16 and 56.64 for the rounded data. Thus for these parameters, where the amount of variation in the rate of progression is relatively small, model 3 was effectively indistinguishable from model 2, even with the large sample sizes used in this experiment. Further, when the data were rounded the increase in goodness of fit of model 3 over model 2 was substantially decreased, while the increase in goodness of fit of model 2 over model 1 was substantially increased.

Parameter estimates

Again, the estimates of θ_4 and θ_5 showed substantial variation, although the estimate of λ was again much more stable. Where parameters were estimated using models

1 and 2, the procedure always converged in two iterations. Estimates of θ_1-θ_3 did not vary substantially between the two GLS methods.

4.2.1. RESULTS USING THE PROGRAMME AUFIT.

In section 2.2.8 we discussed available software for the estimation of covariance structures, and mentioned the programme AUFIT, which uses derivative-free algorithms to estimate the parameters of covariance structures. Professor Browne kindly adapted this programme to estimate the parameters of models 1, 2 and 3, although it was received after the APL programmes used above had been completed. The programme produces either the normal-theory maximum-likelihood, ordinary least-squares or GLS (nk) estimates.

For models 1 and 2, the GLS (nk) estimates calculated by AUFIT were identical to those calculated by the APL programmes. For model 3, the estimates differed. Appendix 4.6 shows the results produced by AUFIT for estimation of model 3 for the data from experiments 3 and 4, using the correct initial values in each case. It can be seen that the goodness of fit given by AUFIT was a marginal increase or decrease compared to the APL programmes. The exception was experiment 3 data set 2, where the estimate of θ_4 tended to zero, and the goodness of fit was improved compared to the APL programmes by 2.28 for the unrounded data and 1.81 for the rounded data. Except for this data set, the estimates of θ_1, θ_2 and θ_3 were similar to those produced by the APL programmes. The estimated standard errors varied somewhat, reflecting the near-singularity of the information matrix.

4.3 CONCLUSIONS.

We shall divide the conclusions of this chapter by the different aims stated at the start:

1. Errors in algebra or software

During the development of the software, both algebraic and software errors were detected because unexpected results were obtained. Because the results for model 3 were unexpected, both algebra and software were thoroughly checked. Further, the results accord with independently written programs. We are thus confident that they are not caused by errors of this nature.

2. Distinguishing the models.

For models 1 and 2, these experiments show that the models can be distinguished, even where (as for experiment 1) visual discrimination is not possible. Experiment 3 confirms that data from model 3 is distinguishable from the other two. However, experiment 4 shows that where the variation in slope for model 3 is relatively small, the increase in goodness of fit may never exceed that attributable to random error according the asymptotic chi-square test.

On the other hand, the increase in the goodness of fit of model 3 over models 1 and 2 for data from models 1 and 2 (experiments 1 and 2) was always much smaller, even than that achieved in experiment 4. The greatest increase in goodness

of fit of model 3 over model 2 was 0.0354. Although the asymptotic theory which we have presented (Steiger et al. 1985) suggests that the difference in goodness of fit should have a chi-squared distribution with two degrees of freedom, this theory depends on the existence of a limiting interior point of the parameter space at which all models hold. For data from models 1 or 2, there is no interior point of the parameter space of model 3 (see section 3.2.3) at which the model holds.

Our results are, therefore, consistent with the asymptotic theory of the estimation of covariance structures. We are not dealing here with standard asymptotic theory where, for instance, the degrees of freedom of the goodness of fit statistic depends on the number of observations: for the standard case increasing the number of parameters of a model by k will increase the goodness of fit of the model by an amount which has a chi-squared distribution with k degrees of freedom.

We noted in section 3.2.3 that the part of the covariance structure of model 3 which is in addition to that of model 2 has a specific form where, broadly, the magnitude of the (positive) elements decreases as we move from the diagonal to the one-off diagonal and so on. Where the true covariance structure of the data does not possess this form, it seems reasonable that only a very small increase in goodness of fit may be achieved.

What is clear from these simulations is that where the amount of variation in the rate of disease progression is comparatively small, as in experiment 4, it will not be possible to distinguish model 3 from model 2. Further, the substantial rounding of the data which occurs in the measurement of periodontal attachment artificially

decreases the goodness of fit of model 3 and increases the goodness of fit of model 2.

The nature of the "nesting" of the covariance structures of model 1 within model 2 within model 3 deserves careful consideration. Firstly, it should be emphasised that the word nesting applies only to the covariance structures rather than to the models themselves. Although it is clear that as the variation in the rate of progression in model 3 tends to zero model 3 tends to model 1, the parametrisations for which model 3 tends to model 2, or model 2 to model 1 are less obvious. It appears that as model 3 tends to model 1, and model 2 is "squeezed" between models 1 and 3, the covariance structures of models 2 and 3 become indistinguishable more rapidly than do those of model 1 and 2.

We must conclude that for data from studies of periodontal disease progression, it may not be possible, unless there is substantial variation in the rate of progression, to distinguish models 2 and 3. It will be much easier to establish whether there is variation in the rate of progression than to establish whether model 2 or model 3 is a more accurate description of this variation.

The final estimates sometimes depended on the initial values for the iteration; particularly for model 3, but also for model 2 where the quadratic estimation procedure has to be used because the estimate of a parameter would otherwise be negative. Thus, in the analysis of real rather than simulated data, it will be necessary to experiment with different initial values to ensure that the best goodness of fit has been achieved.

Generally similar results were obtained from the two methods (adf and nk), although the goodness of fit tended to be rather worse for the adf method.

3. Finite-sample bias.

It appeared that for both the adf and nk methods the parameter estimates tended to be somewhat smaller than their true values.

4. Standard errors.

For models 1 and 2, these reflected the observed variation in the parameter estimates. For parameters θ_4 and θ_5 of model 3, the large estimated standard errors reflected the instability of the estimates.

APPENDIX 4.1 RESULTS OF SIMULATION EXPERIMENT 1.

Model 1: $\sigma^2 = 0.9506$, $\beta=1$, $\sigma^2_\beta=0.1619$

Model 2: $\sigma^2 = 0.64$, $\lambda=2$, $\sigma^2_\lambda=0.25$, $\mu_z=0.5$, $\sigma^2_\mu=0.0225$, $\sigma^2_z=0.04$, $\xi^2_z=0.0001$

Parameter value	Model	Data set	μ	θ_1	θ_2	θ_3	$E[W_{ij}]$	$Var(W_{ij})$
for underlying	1	3,4	1	.9506	.1619	0	1	2.0631
distribution	2	1,2	1	.64	.1581	.625	1	2.0631

EXPT	METH	MOD	DS	FROMMOD	STATUS	ITER	STAT	T1	T2	T3	T4	T5
1	nk	1	1	2	2	OK		0.85165	0.21347	.	.	.
1	nk	1	2	2	2	OK		0.82762	0.14698	.	.	.
1	nk	1	3	1	2	OK		0.94643	0.16229	.	.	.
1	nk	1	4	1	2	OK		0.87387	0.16162	.	.	.
1	nk	2	1	2	2	OK		0.57809	0.15846	0.56412	.	.
1	nk	2	2	2	2	OK		0.55863	0.09738	0.56208	.	.
1	nk	2	3	1	2	OK		0.92380	0.15915	0.02764	.	.
1	nk	2	4	1	2	OK		0.86854	0.16105	0.00622	.	.
1	nk	3	1	2	500	ER		0.67212	0.15764	0.57391	35610.56	0.00008
1	nk	3	2	2	401	ER		0.56666	0.09738	0.56208	117.65	0.29745
1	nk	3	3	1	500	ER		0.93093	0.15892	0.03020	18411.06	0.00008
1	nk	3	4	1	7	OK		0.87064	0.15896	0.06250	1.87	0.04938

MOD	DS	FROMMOD	STATUS	S1	S2	S3	S4	S5
1	1	2	OK	0.021679	0.01612	.	.	.
1	2	2	OK	0.021448	0.01304	.	.	.
1	3	1	OK	0.022804	0.01265	.	.	.
1	4	1	OK	0.020976	0.01265	.	.	.
2	1	2	OK	0.032249	0.01703	0.0495	.	.
2	2	2	OK	0.030659	0.01378	0.0460	.	.
2	3	1	OK	0.031937	0.01304	0.0272	.	.
2	4	1	OK	0.028636	0.01265	0.0228	.	.
3	1	2	ER	.	.	.	.	.
3	2	2	ER	.	.	.	.	.
3	3	1	ER	.	.	.	.	.
3	4	1	OK	0.030496	1.10986	78.7977	9361.23	111.929

EXPT	METH	DS	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
1	nk	1	121.278	56.6656	56.6302	0.03543	64.6476	64.6122
1	nk	2	148.589	74.4267	74.4267	0.00000	74.1626	74.1626
1	nk	3	39.964	39.4524	39.4502	0.00225	0.5137	0.5115
1	nk	4	36.152	36.1152	36.1014	0.01376	0.0510	0.0372

EXPT	METH	MOD	DS	FROM MOD	ITER	STAT US	T 1	T 2	T 3	T 4	T 5
1	adf	1	1	2	2	OK	0.77650	0.17223	.	.	.
1	adf	1	2	2	2	OK	0.75271	0.09435	.	.	.
1	adf	1	3	1	2	OK	0.96893	0.15659	.	.	.
1	adf	1	4	1	2	OK	0.88440	0.16352	.	.	.
1	adf	2	1	2	2	OK	0.62554	0.14896	0.42052	.	.
1	adf	2	2	2	2	OK	0.46687	0.06378	0.63097	.	.
1	adf	2	3	1	2	OK	0.96112	0.15602	0.01142	.	.
1	adf	2	4	1	11	OK	0.88440	0.16352	0.00000	.	.
1	adf	3	1	2	442	ER	0.63218	0.14896	0.42052	104.685	0.30265
1	adf	3	2	2	390	ER	0.47571	0.06378	0.63097	116.108	0.30749
1	adf	3	3	1	4	OK	0.96337	0.15598	0.01167	25.062	0.08360
1	adf	3	4	1	21	ER	0.88440	0.16352	0.03845	0.007	0.00001

MOD	DS	FROM MOD	STAT US	S 1	S 2	S 3	S 4	S 5
1	1	2	OK	0.02588	0.016432	.	.	.
1	2	2	OK	0.02510	0.014832	.	.	.
1	3	1	OK	0.02739	0.014491	.	.	.
1	4	1	OK	0.02387	0.015811	.	.	.
2	1	2	OK	0.03376	0.016733	0.06	.	.
2	2	2	OK	0.03606	0.014832	0.06	.	.
2	3	1	OK	0.03302	0.014832	0.03	.	.
2	4	1	OK	0.01897	0.019494	10714495415.18	.	.
3	1	2	ER	.	.	.	.	.
3	2	2	ER	.	.	.	.	.
3	3	1	OK	0.36097	0.017607	0.12	163946.83	553.217
3	4	1	ER	.	.	.	.	.

EXPT	METH	DS	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
1	adf	1	118.918	70.044	70.044	.00000	48.875	48.875
1	adf	2	246.267	124.409	124.409	.00000	121.857	121.857
1	adf	3	58.211	58.034	58.032	.00236	0.179	0.176
1	adf	4	47.113	47.113	47.113	.00000	0.000	0.000

APPENDIX 4.2 RESULTS OF SIMULATION EXPERIMENT 2.

Model 1: $\sigma^2 = 1.69$, $\beta=1$, $\sigma^2_\beta=0.63$

Model 2: $\sigma^2 = 0.64$, $\lambda=0.5$, $\sigma^2_\lambda=0.01$, $\mu_z=2.0$, $\sigma^2_\mu=0.25$, $\sigma^2_z=1.0$, $\xi^2_z=0.04$

Parameter value	Model	Data set	μ	θ_1	θ_2	θ_3	$E[W_{ij}]$	$Var(W_{ij})$
for underlying	1	3,4	1	1.69	.63	0	1	4.01
distribution	2	1,2	1	.64	.105	2.625	1	4.01

EXPTH	METHOD	DS	FROMMOD	ITER	STATUS	T1	T2	T3	T4	T5
2	nk	1	2	2	OK	1.51588	0.33749	.	.	.
2	nk	2	2	2	OK	1.54174	0.27342	.	.	.
2	nk	3	1	2	OK	1.50300	0.51610	.	.	.
2	nk	4	1	2	OK	1.56992	0.55565	.	.	.
2	nk	1	2	2	OK	0.52682	0.09219	2.54575	.	.
2	nk	2	2	2	OK	0.66577	0.05852	2.21342	.	.
2	nk	3	1	8	OK	1.50300	0.51610	0.00000	.	.
2	nk	4	1	10	ER	1.56992	0.55565	0.00000	.	.
2	nk	1	2	500	ER	0.57359	0.09219	2.54575	100.07	0.27193
2	nk	2	2	500	ER	0.90265	0.05815	2.21797	1823.75	0.00252
2	nk	3	1	11	ER	1.50300	0.51610	0.00002	0.73	0.01441
2	nk	4	1	12	ER	1.56992	0.55565	0.00000	3.38	0.01863

MOD	DS	FROMMOD	STATUS	S1	S2	S3	S4	S5
1	1	2	OK	0.040620	0.026833	.	.	.
1	2	2	OK	0.041231	0.023875	.	.	.
1	3	1	OK	0.036469	0.040000	.	.	.
1	4	1	OK	0.038079	0.041352	.	.	.
2	1	2	OK	0.070781	0.030332	0.15	.	.
2	2	2	OK	0.067897	0.027203	0.14	.	.
2	3	1	OK	0.020736	0.027928	241646374.92	.	.
2	4	1	ER	.	.	.	.	.
3	1	2	ER	.	.	.	.	.
3	2	2	ER	.	.	.	.	.
3	3	1	ER	.	.	.	.	.
3	4	1	ER	.	.	.	.	.

EXPT	METH	DS	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
2	nk	1	200.657	55.8434	55.8434	.00000	144.814	144.814
2	nk	2	201.886	70.3990	70.3980	.00095	131.488	131.487
2	nk	3	47.340	47.3398	47.3398	.00000	0.000	0.000
2	nk	4	51.871	51.8706	51.8706	.00000	0.000	0.000

EXPT	METH	MOD	DS	FROM MOD	ITER	STAT US	T 1	T 2	T 3	T 4	T 5
2	adf	1	1	2	2	OK	1.13917	0.22233	.	.	.
2	adf	1	2	2	2	OK	1.05245	0.25624	.	.	.
2	adf	1	3	1	2	OK	1.47616	0.48904	.	.	.
2	adf	1	4	1	2	OK	1.57463	0.49917	.	.	.
2	adf	2	1	2	2	OK	0.60275	0.07362	2.02592	.	.
2	adf	2	2	2	2	OK	0.57793	0.06237	1.94885	.	.
2	adf	2	3	1	10	OK	1.47616	0.48904	0.00000	.	.
2	adf	2	4	1	9	OK	1.57463	0.49917	0.00000	.	.
2	adf	3	1	2	4	OK	0.76549	0.07366	2.02606	48.327	0.12872
2	adf	3	2	2	459	ER	0.60540	0.06237	1.94885	116.691	0.30392
2	adf	3	3	1	19	ER	1.47616	0.48904	0.03046	0.045	0.00000
2	adf	3	4	1	23	ER	1.57463	0.49917	0.00000	4.683	0.00823

MOD	DS	FROM MOD	STAT US	S 1	S 2	S 3	S 4	S 5
1	1	2	OK	0.054	0.026646	.	.	.
1	2	2	OK	0.057	0.024290	.	.	.
1	3	1	OK	0.041	0.048888	.	.	.
1	4	1	OK	0.048	0.044721	.	.	.
2	1	2	OK	0.069	0.029326	0.16	.	.
2	2	2	OK	0.070	0.029665	0.17	.	.
2	3	1	OK	0.024	0.035496	5371663398.18	.	.
2	4	1	OK	0.026	0.031937	657963633929.86	.	.
3	1	2	OK	274.657	0.071554	0.13	200044.84	10853.30
3	2	2	ER	.	.	.	.	.
3	3	1	ER	.	.	.	.	.
3	4	1	ER	.	.	.	.	.

EXPT	METH	DS	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
2	adf	1	222.837	68.305	68.299	.00579	154.538	154.532
2	adf	2	238.396	105.549	105.549	.00000	132.847	132.847
2	adf	3	85.351	85.351	85.351	.00000	0.000	0.000
2	adf	4	68.405	68.405	68.405	.00000	0.000	0.000

APPENDIX 4.3 RESULTS OF SIMULATION EXPERIMENT 3.

Model specification: Measurement error variance $\sigma^2 = 0.64$

Poisson rate $\lambda_i \sim \Gamma(a=8,b=0.1)$ Slope $\beta_i \sim N(\beta=1,\sigma^2_\beta=0.09)$

Slope variance $\sigma^2_{\beta(i)} \sim N(\sigma^2_\delta=1,\xi^2_\delta=0.01)$

From the definition of model 3:

$\mu=\beta=1$, $\theta_1=\sigma^2=0.64$, $\theta_2=\sigma^2_\beta=0.09$, $\theta_3=2\sigma^2_\delta(b(a-1))^{-1}=2.857143$ $\theta_4=a-2=6$, $\theta_5=b=0.1$.

EXPT	MOD	MOD	DS	TYP	ITER	STAT	T1	T2	T3	T4	T5
3	nk	1	1	F	2	OK	0.77963	0.30829	.	.	.
3	nk	1	1	R	2	OK	0.81879	0.32712	.	.	.
3	nk	1	2	F	2	OK	0.79332	0.30019	.	.	.
3	nk	1	2	R	2	OK	0.82233	0.31160	.	.	.
3	nk	1	3	F	2	OK	0.74844	0.34298	.	.	.
3	nk	1	3	R	2	OK	0.76340	0.34649	.	.	.
3	nk	2	1	F	2	OK	0.23444	0.18587	1.36291	.	.
3	nk	2	1	R	2	OK	0.24033	0.19585	1.44925	.	.
3	nk	2	2	F	2	OK	0.25046	0.17271	1.38295	.	.
3	nk	2	2	R	2	OK	0.24979	0.17572	1.46644	.	.
3	nk	2	3	F	2	OK	0.20844	0.21192	1.39325	.	.
3	nk	2	3	R	2	OK	0.18796	0.20271	1.50719	.	.
3	nk	3	1	F	500	ER	0.63160	0.07173	2.93212	43576.37	0.00001
3	nk	3	1	R	2	OK	0.64810	0.08036	3.07409	5.98	0.09688
3	nk	3	2	F	3	OK	0.63123	0.08497	2.61186	6.23	0.10263
3	nk	3	2	R	2	OK	0.65808	0.07165	2.93795	5.89	0.10086
3	nk	3	3	F	54	OK	0.57024	0.10783	3.06873	0.79	0.40645
3	nk	3	3	R	48	OK	0.56611	0.11179	2.88298	1.05	0.42417

MOD	DS	TYPE	STATUS	S1	S2	S3	S4	S5
1	1	F	OK	0.01304	0.01517	.	.	.
1	1	R	OK	0.01342	0.01612	.	.	.
1	2	F	OK	0.01304	0.01483	.	.	.
1	2	R	OK	0.01378	0.01549	.	.	.
1	3	F	OK	0.01225	0.01612	.	.	.
1	3	R	OK	0.01265	0.01643	.	.	.
2	1	F	OK	0.02530	0.01581	0.05459	.	.
2	1	R	OK	0.02683	0.01673	0.05788	.	.
2	2	F	OK	0.02449	0.01549	0.05282	.	.
2	2	R	OK	0.02569	0.01612	0.05568	.	.
2	3	F	OK	0.02387	0.01703	0.05254	.	.
2	3	R	OK	0.02470	0.01732	0.05532	.	.
3	1	F	ER	.	.	.	.	.
3	1	R	OK	0.02811	0.05586	1.10935	28.80	0.382
3	2	F	OK	0.01789	0.07874	0.26593	6.14	0.659
3	2	R	OK	0.02828	0.05254	1.03040	28.07	0.395
3	3	F	OK	0.03633	0.07057	2.18480	5.19	1.008
3	3	R	OK	0.04278	0.06099	1.61445	5.78	1.096

EXPT	METHOD	DS	TYPE	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
3	nk	1	F	422.496	111.525	60.5520	50.9727	361.944	310.972
3	nk	1	R	429.705	117.280	69.1680	48.1118	360.537	312.425
3	nk	2	F	441.893	99.723	73.8024	25.9204	368.091	342.170
3	nk	2	R	449.564	103.318	78.8483	24.4698	370.715	346.246
3	nk	3	F	435.655	84.161	59.2011	24.9599	376.454	351.494
3	nk	3	R	453.110	82.659	62.3027	20.3565	390.807	370.451

EXPT	METHOD	DS	TYPE	STATUS	T1	T2	T3	T4	T5
3	adf	1	F	OK	0.77771	0.25256	.	.	.
3	adf	1	R	OK	0.82174	0.27373	.	.	.
3	adf	2	F	OK	0.73930	0.31154	.	.	.
3	adf	2	R	OK	0.77482	0.31081	.	.	.
3	adf	3	F	OK	0.73437	0.32122	.	.	.
3	adf	3	R	OK	0.74880	0.30015	.	.	.
3	adf	1	F	OK	0.26766	0.17859	1.27374	.	.
3	adf	1	R	OK	0.27372	0.18870	1.36793	.	.
3	adf	2	F	OK	0.23555	0.15979	1.35295	.	.
3	adf	2	R	OK	0.24427	0.16283	1.41207	.	.
3	adf	3	F	OK	0.23791	0.21259	1.32485	.	.
3	adf	3	R	OK	0.21333	0.19534	1.44959	.	.
3	adf	1	F	OK	0.61707	0.08621	2.75575	5.96	0.10064
3	adf	1	R	ER	0.65192	0.10887	2.58585	22074.27	0.00003
3	adf	2	F	OK	0.63575	0.04723	3.06762	6.87	0.07758
3	adf	2	R	OK	0.65849	0.03931	3.33544	4.86	0.10214
3	adf	3	F	OK	0.59273	0.11099	2.70572	6.16	0.10071
3	adf	3	R	OK	0.58972	0.10700	2.61049	6.54	0.10913

MOD	DS	TYPE	STATUS	S1	S2	S3	S4	S5
1	1	F	OK	0.01703	0.02145	.	.	.
1	1	R	OK	0.01844	0.02324	.	.	.
1	2	F	OK	0.01703	0.02121	.	.	.
1	2	R	OK	0.01817	0.02168	.	.	.
1	3	F	OK	0.01703	0.02098	.	.	.
1	3	R	OK	0.01673	0.02121	.	.	.
2	1	F	OK	0.03521	0.02191	0.07694	.	.
2	1	R	OK	0.03715	0.02387	0.08056	.	.
2	2	F	OK	0.03391	0.02302	0.07874	.	.
2	2	R	OK	0.03564	0.02324	0.08155	.	.
2	3	F	OK	0.03225	0.02191	0.07348	.	.
2	3	R	OK	0.03317	0.02191	0.07772	.	.
3	1	F	OK	0.03578	0.06293	1.24825	36.84	0.511
3	1	R	ER	.	.	.	.	.
3	2	F	OK	0.02049	0.17649	0.44621	8.75	0.774
3	2	R	OK	0.02145	0.22159	0.52660	6.65	0.748
3	3	F	OK	0.03286	0.06237	1.18566	38.40	0.520
3	3	R	OK	0.03821	0.05683	0.99812	42.43	0.601

EXPT	METHOD	DS	TYPE	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
3	adf	1	F	398.421	125.108	84.4657	40.6428	313.956	273.313
3	adf	1	R	412.093	124.152	87.9645	36.1878	324.129	287.941
3	adf	2	F	412.311	117.523	91.0718	26.4508	321.239	294.789
3	adf	2	R	418.713	119.522	90.7237	28.7983	327.989	299.191
3	adf	3	F	428.704	104.359	66.8644	37.4949	361.839	324.345
3	adf	3	R	450.229	102.931	72.2993	30.6315	377.930	347.298

APPENDIX 4.4 RESULTS OF SIMULATION EXPERIMENT 4.

Model specification: Measurement error variance $\sigma^2 = 0.64$

Poisson rate $\lambda_i \sim \Gamma(a=8,b=0.1)$ Slope $\beta_i \sim N(\beta=1,\sigma^2_\beta=0.09)$

Slope variance $\sigma^2_{\beta(i)} \sim N(\sigma^2_\delta=0.04,\xi^2_\delta=0.0001)$

From the definition of model 3:

$\mu=\beta=1$, $\theta_1=\sigma^2=0.64$, $\theta_2=\sigma^2_\beta=0.09$, $\theta_3=2\sigma^2_\delta(b(a-1))^{-1}=0.114286$ $\theta_4=a-2=6$, $\theta_5=b=0.1$.

EXPT	METHOD	MOD	DS	TYPE	STATUS	T1	T2	T3	T4	T5
4	nk	1	1	F	2 OK	0.63666	0.10761	.	.	.
4	nk	1	1	R	2 OK	0.68878	0.11567	.	.	.
4	nk	1	2	F	2 OK	0.64950	0.09191	.	.	.
4	nk	1	2	R	2 OK	0.70597	0.10149	.	.	.
4	nk	1	3	F	2 OK	0.63737	0.09571	.	.	.
4	nk	1	3	R	2 OK	0.68238	0.09901	.	.	.
4	nk	2	1	F	2 OK	0.57949	0.09967	0.07960	.	.
4	nk	2	1	R	2 OK	0.58761	0.09970	0.15870	.	.
4	nk	2	2	F	2 OK	0.60544	0.08581	0.05907	.	.
4	nk	2	2	R	2 OK	0.61231	0.08687	0.14477	.	.
4	nk	2	3	F	2 OK	0.59639	0.09020	0.05404	.	.
4	nk	2	3	R	2 OK	0.59233	0.08524	0.13699	.	.
4	nk	3	1	F	500 ER	0.61059	0.09033	0.22768	3305.64	0.00010
4	nk	3	1	R	74 OK	0.62820	0.09486	0.24841	0.00	1.23974
4	nk	3	2	F	500 ER	0.62174	0.08515	0.06698	460089.16	0.00000
4	nk	3	2	R	44 OK	0.63278	0.08671	0.14665	6.58	0.48898
4	nk	3	3	F	500 ER	0.61754	0.08143	0.20049	42335.52	0.00001
4	nk	3	3	R	3 OK	0.62949	0.08249	0.17066	8.36	0.13099

MOD	DS	TYPE	STATUS	S1	S2	S3	S4	S5
1	1	F	OK	0.00949	0.00548	.	.	.
1	1	R	OK	0.01049	0.00548	.	.	.
1	2	F	OK	0.01000	0.00447	.	.	.
1	2	R	OK	0.01095	0.00447	.	.	.
1	3	F	OK	0.00949	0.00447	.	.	.
1	3	R	OK	0.01049	0.00447	.	.	.
2	1	F	OK	0.01342	0.00548	0.01304	.	.
2	1	R	OK	0.01483	0.00548	0.01612	.	.
2	2	F	OK	0.01342	0.00447	0.01225	.	.
2	2	R	OK	0.01483	0.00548	0.01581	.	.
2	3	F	OK	0.01342	0.00447	0.01225	.	.
2	3	R	OK	0.01449	0.00548	0.01549	.	.
3	1	F	ER	.	.	.	.	.
3	1	R	OK	0.01342	0.01732	0.29766	8077.60	1.846
3	2	F	ER	.	.	.	.	.
3	2	R	OK	0.46822	0.00707	0.06033	1292.03	104.241
3	3	F	ER	.	.	.	.	.
3	3	R	OK	0.03507	0.00949	0.13450	299.33	4.382

EXPT	METHOD	DS	TYPE	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
4	nk	1	F	73.746	55.585	53.3790	2.2060	20.367	18.161
4	nk	1	R	97.542	49.574	49.0410	0.5331	48.501	47.968
4	nk	2	F	58.806	47.470	47.3202	0.1501	11.486	11.336
4	nk	2	R	91.944	49.644	49.6412	0.0032	42.303	42.300
4	nk	3	F	67.919	58.454	56.0159	2.4380	11.903	9.465
4	nk	3	R	96.335	57.177	56.1655	1.0118	40.169	39.157

EXPT	METHOD	MOD	DS	TYPE	STATUS	T1	T2	T3	T4	T5
4	adf	1	1	F	2 OK	0.63239	0.10446	.	.	.
4	adf	1	1	R	2 OK	0.67773	0.10902	.	.	.
4	adf	1	2	F	2 OK	0.64948	0.09169	.	.	.
4	adf	1	2	R	2 OK	0.70773	0.10088	.	.	.
4	adf	1	3	F	2 OK	0.63348	0.09864	.	.	.
4	adf	1	3	R	2 OK	0.68455	0.10190	.	.	.
4	adf	2	1	F	2 OK	0.57955	0.09901	0.07914	.	.
4	adf	2	1	R	2 OK	0.59284	0.09933	0.14789	.	.
4	adf	2	2	F	2 OK	0.60348	0.08660	0.05922	.	.
4	adf	2	2	R	2 OK	0.60703	0.08763	0.15121	.	.
4	adf	2	3	F	2 OK	0.58587	0.09183	0.06382	.	.
4	adf	2	3	R	2 OK	0.58981	0.08709	0.13799	.	.
4	adf	3	1	F	41 OK	0.61251	0.08656	0.32741	4.68	0.04373
4	adf	3	1	R	500 ER	0.63535	0.09717	0.17544	4209.52	0.00027
4	adf	3	2	F	3 OK	0.62350	0.08319	0.10429	5.60	0.09417
4	adf	3	2	R	500 ER	0.64584	0.08620	0.16632	22311.98	0.00007
4	adf	3	3	F	500 OK	0.61148	0.08498	0.16570	126.81	0.00282
4	adf	3	3	R	3 OK	0.62757	0.08432	0.17015	8.43	0.13200

MOD	DS	TYPE	STATUS	S1	S2	S3	S4	S5
1	1	F	OK	0.01265	0.00632	.	.	.
1	1	R	OK	0.01342	0.00632	.	.	.
1	2	F	OK	0.01304	0.00632	.	.	.
1	2	R	OK	0.01378	0.00632	.	.	.
1	3	F	OK	0.01265	0.00548	.	.	.
1	3	R	OK	0.01342	0.00548	.	.	.
2	1	F	OK	0.01643	0.00632	0.01612	.	.
2	1	R	OK	0.01761	0.00707	0.01975	.	.
2	2	F	OK	0.01844	0.00632	0.01673	.	.
2	2	R	OK	0.01975	0.00632	0.02121	.	.
2	3	F	OK	0.01761	0.00548	0.01673	.	.
2	3	R	OK	0.01924	0.00632	0.02049	.	.
3	1	F	OK	0.01140	0.29998	6.19149	144.42	9.775
3	1	R	ER	.	.	.	.	.
3	2	F	OK	0.01612	0.04827	0.96224	98.16	10.402
3	2	R	ER	.	.	.	.	.
3	3	F	OK	0.01265	0.12787	2.34878	7535.72	35.047
3	3	R	OK	0.04909	0.01304	0.17779	421.78	6.187

EXPT	METHOD	DS	TYPE	CHI1 (df 53)	CHI2 (df 52)	CHI3 (df 50)	DIFF32 (df 2)	DIFF31 (df 3)	DIFF21 (df 1)
4	adf	1	F	81.029	56.878	54.6737	2.2041	26.355	24.151
4	adf	1	R	113.652	57.015	56.4277	0.5877	57.224	56.637
4	adf	2	F	60.011	47.489	47.4103	0.0783	12.600	12.522
4	adf	2	R	107.707	57.107	56.8817	0.2253	50.826	50.600
4	adf	3	F	79.121	64.432	62.1196	2.3129	17.001	14.688
4	adf	3	R	108.668	62.973	62.7945	0.1788	45.874	45.695

APPENDIX 4.5. MINITAB INSTRUCTIONS TO SIMULATE DATA FROM MODEL 3.

```
random 1 c25;
gamma 8 0.1.
copy c25 k10
let k10=1/k10
NOTE K10 IS THE POISSON RATE OF OCCURANCE OF CHANGES IN
SLOPE
random 1 c25;
normal 1 0.3.
copy c25 k11
NOTE K11 IS THE MEAN SLOPE FOR THE 'SITE'
random 1 c25;
normal 1 0.1.
copy c25 k12
let k12=k12**0.5
NOTE K12 IS THE SLOPE STANDARD DEVIATION FOR THE 'SITE'
let k7=0
random 50 c1;
expo k10.
stack k7 c1 c1
parsum c1 c2
NOTE C2 CONTAINS TIME POINTS FOR CHANGES IN SLOPE
copy c1 c2 c3 c2;
NOTE C3 CONTAINS INTERVALS BETWEEN CHANGES IN SLOPE
omit c2 = 10:1000.
n c3 k1
let k1=k1+1
random k1 c4;
normal k11 k12.
NOTE C4 CONTAINS SLOPES
let c10=k7
random 11 c6;
normal 0 0.8.
NOTE C6 CONTAINS MEASUREMENT ERRORS
let k2=1
exec 'true' 10
```

```
copy c3 c13;
use c2=0:k2.
NOTE C13 CONTAINS INTERVALS UP TO TIME K2
sum c13 k3
NOTE K3 CONTAINS TIME OF LAST CHANGE BEFORE K2
let k4=k2-k3
NOTE K4 CONTAINS INTERVAL BETWEEN LAST CHANGE AND K2
stack c13 k4 c13
n c13 k5
copy c4 c14;
use 1:k5.
NOTE C14 CONTAINS SLOPES UP TO TIME K2
let c15=c13*c14
NOTE CHANGE = SLOPE * TIME INTERVAL
sum c15 k8
NOTE K8 = TRUE 'ATTACHMENT LEVEL' AT TIME K2
stack c10 k8 c10
let k2=k2+1
NOTE INCREASE K2 BY 1 AND REPEAT UP TO K2=10
end

let c22=c10+c6
NOTE C22 CONTAINS OBSERVED ATTACHMENT LEVELS
copy c22 c23;
omit 1.
copy c22 c24;
omit 11.
let c21=c23-c24
NOTE C21 CONTAINS INCREMENTS IN OBSERVED ATTACHMENT LEVELS
```

APPENDIX 4.6. RESULTS OF SIMULATION EXPERIMENTS 3 AND 4 USING THE PROGRAM AUFIT.

EXPT	METHOD	MOD	DS	TYPE	T1	T2	T3	T4	T5	CHI3
3	nk	3	1	F	0.63068	0.06961	2.97425	45.434	0.01355	60.6310
3	nk	3	1	R	0.65829	0.08250	2.98670	68.967	0.00951	68.4194
3	nk	3	2	F	0.62694	0.02241	4.46978	0.000	0.48817	71.5227
3	nk	3	2	R	0.64322	0.02510	4.50333	0.000	0.52127	77.0381
3	nk	3	3	F	0.57035	0.10809	3.05948	0.817	0.40185	59.2268
3	nk	3	3	R	0.56591	0.11173	2.88518	1.035	0.42703	62.3091
4	nk	3	1	F	0.61068	0.08976	0.24115	23.713	0.01320	53.2149
4	nk	3	1	R	0.62836	0.09499	0.24668	0.000	1.24785	48.9976
4	nk	3	2	F	0.62158	0.08508	0.06791	88.723	0.01408	47.2833
4	nk	3	2	R	0.63476	0.08660	0.14739	10.449	0.27827	49.4658
4	nk	3	3	F	0.61749	0.08135	0.20301	63.974	0.00427	56.0437
4	nk	3	3	R	0.63609	0.08154	0.18470	158.524	0.00528	55.7822

EXPT	MOD	DS	TYPE	S1	S2	S3	S4	S5
3	3	1	F	0.02424	0.03566	0.43350	7.466	0.00000
3	3	1	R	0.02604	0.03623	0.41581	11.187	0.00000
3	3	2	F	0.02842	0.04514	0.99092	0.000	0.13404
3	3	2	R	0.03034	0.04624	0.98475	0.000	0.14551
3	3	3	F	0.05128	0.09926	3.04670	7.431	1.41151
3	3	3	R	0.06063	0.08648	2.29555	8.121	1.55906
4	3	1	F	0.01650	0.02021	0.39469	36.504	0.00000
4	3	1	R	0.02941	0.01164	0.17017	0.000	2.26697
4	3	2	F	0.02448	0.00733	0.04676	204.361	0.00000
4	3	2	R	0.07905	0.00771	0.03889	0.000	1.21508
4	3	3	F	0.01636	0.02383	0.52060	138.674	0.00000
4	3	3	R	0.02105	0.00926	0.08464	137.949	0.00000

5

PRELIMINARY ANALYSIS OF THE SRI LANKA DATA.

In this section we analyse data from a study of the natural history of periodontal disease in a population of 480 male Sri Lankan tea labourers. Our aim is to examine how far existing methods such as those used recently in the periodontal disease literature allow us to make inferences about the nature of disease progression. These results are in preparation for the next section, in which we examine which of models 1-3 best fit these data. Our analyses will be solely of attachment level measurements. We make no attempt to associate changes in attachment level with other characteristics of the subject or site, except for the site type and the subject identity.

We may identify two questions about the nature of progression which are of interest:

a) What was the rate of progression? To what extent is variation in the observed rate of progression between sites accounted for by tooth type, subject effect or measurement error?

b) Did the rate of progression vary? If so, what inferences can we make about the nature of this variation?

We will see that although existing methods allow us to address both questions, it is difficult to use them to gain information on the nature of progression.

The Sri Lankan study is by far the most extensive study of the natural history of periodontal disease in the absence of any dental care. The subjects were examined

in 1970, and subsequently on five further occasions, in 1971, 1973, 1977, 1982 and 1985. The 1985 examination included 161 individuals who had participated in the first study. Results have already been presented in a series of publications.

Löe et al. (1978a) gave the study design and baseline data. This showed that the Sri Lankan population did not perform any conventional oral hygiene measures, and were not exposed to Western programmes of prevention or treatment of dental disease. They had abundant calculus, generalized gingivitis and showed loss of attachment by the age of thirty years. Löe et al. (1978b) presented the rate of tooth loss and showed that teeth with deep periodontal lesions started to exfoliate as subjects approached 40 years of age. Löe et al. (1978c) discussed the rate of periodontal destruction before 40 years of age. We have already quoted (page 4) their conclusion that disease progression was continuous. Löe et al. (1986) identified three subgroups of the subjects: a rapidly progressive group consisting of 8% of the population, a group showing moderate progression (81%) and a groups showing no progression (11%).

As with almost all large-scale studies of periodontal disease progression, disease levels were measured with a periodontal probe, to the nearest millimetre. Given that the total length of the attachment of a healthy tooth is between 10 and 15mm, the amount of rounding is substantial relative to the changes observed.

5.1 SUMMARY OF THE DATA.

As a starting point, data are presented from every site at which three or more observations were made. This is because a sensible minimum requirement for inference on the nature of progression at a site to be made is three observations, so that for the null model of constant progression we may estimate the initial attachment level, rate of progression and the measurement error variance. Only data from mesial sites are used - these sites showed the greatest progression of periodontal disease during the period of the study. Subjects therefore contributed up to 28 measurements per examination - one for each mesial site on each tooth, excluding third molars. By numbering these teeth as 1 to 28, we are able in the analyses below, to assess the effect of tooth type on disease progression.

Two tables of summary statistics for these data illustrate the enormous problem faced by a researcher attempting to use exploratory methods of analysis to gain an impression of how the disease progressed. Appendix 5.1 contains tables of the observed frequencies of attachment levels at each examination, and of the observed loss in attachment between successive examinations for each site. As can be seen, the data consist of 32907 individual observations, and 26075 observed increments in attachment level.

Table 5.4 shows that the range of measured attachment levels was 0 to 14mm, and that the measured attachment levels increased during the study. This is reflected both in the steady increase of the frequencies of extreme (>8mm) observations, and

also in the increase in the modal observations (0mm for examinations 1 and 2, 1mm for examination 3, 2mm for examinations 4, 5 and 6).

Table 5.5 contains the observed increments in attachment level between successive observations: an observation only appears in the column labelled 5-3 if the site was observed at examinations 5 and 3 but not at examination 4. Positive values for loss of attachment mean that disease worsened between observations. The range of observed changes in attachment level was -9mm to 12mm. Although the modal observation was 0mm, it is clear (and unsurprising) that there was a trend for more positive than negative observations. The occurrence of very large observed gains, as well as losses, in attachment indicates that very large measurement errors can occur, albeit infrequently.

We could, of course, divide these tables, for instance by subject or by tooth type, in an attempt to facilitate inference which might be made by inspection. Subdivision by tooth type would ensure that only one observation for each subject appeared in each table, so that the accumulated frequencies contained independent observations at each examination or interval. However even this subdivision will result in 28 tables for inspection. Subdivision by subject might be used to assess, for instance, which subjects had sites at which the observed loss of attachment had exceeded 2mm. Here the problem is even more acute - the number of individuals contributing data to tables 5.4 and 5.5 (that is, those examined on at least three occasions) was 259.

The usual numbers for tooth type (with 1 and 2 representing incisors and 6 and 7 molars) and their corresponding sequence numbers used here are shown below.

	Right							Left							
Tooth type	7	6	5	4	3	2	1	1	2	3	4	5	6	7	Upper
Sequence number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mouth
Tooth	7	6	5	4	3	2	1	1	2	3	4	5	6	7	Lower
Sequence number	15	16	17	18	19	20	21	22	23	24	25	26	27	28	mouth

A single, large change at a site may be due to measurement error or to real change. If it is due to real change, it is evidence that the constant progression model does not hold only if it is also possible to conclude that the site went through periods where disease progression was not so large. Even if it were possible to inspect the data for each of the 6832 sites, inference about an individual site would have to be tempered by the knowledge that, since a large number of measurements have been made, large errors will occur.

Inspection of the raw data, or of subsets of the raw data, is not, therefore, pursued further as a means of making inference about the nature of disease progression. We thus face a problem which is inevitable when dealing with such a substantial quantity of data - that data reduction which is necessary to make inference may obscure interesting features of the data. This was the problem which led Haffajee et al. (1983a) to criticise reduction of the data to subject means, and to propose instead to ignore subject effects. We now show that for these data such a practice would be wholly mistaken.

5.2 ANALYSIS OF THE OBSERVED RATES OF PROGRESSION AT EACH SITE.

In this section we present the results of methods which we have devised which use standard linear models to examine factors influencing the mean rate of disease progression. We discussed in section 2.1 the recent controversy over claims that disease sites progress effectively independently within a mouth. We have also seen (section 3) that the model for constant progression can be seen as a null model in

the sense that the covariance structures of models 2 and 3 can be reduced to the covariance structure of the constant progression model by fixing the values of certain of their parameters. Before examining whether the rate of progression varied, we now use the constant progression model to derive the mean rate of progression for each site. This was estimated using the usual least-squares methods. All the analyses in this section were performed using SAS (SAS Inc, Cary, NC). The times of the examinations were expressed in years by dividing the number of days by 365.25. This gave a total of 6832 estimated slopes (expressed in mm per year), from 259 subjects.

These estimated slopes were then used as the response variable in an analysis of the extent to which variation in the slopes were due to subject and to tooth type. The output from SAS is shown in Table 5.1. Type I sums of squares (SS) are the incremental improvement in the error SS as each effect is added to the model: type III sums of squares are the increase in the error SS when each effect, in turn, is omitted from the model.

TABLE 5.1. SAS OUTPUT FOLLOWING ANALYSIS OF VARIANCE ON ESTIMATED RATES OF PROGRESSION FOR EACH SITE.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YEARRATE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	285	114.24266687	0.40085146	13.48
ERROR	6546	194.69352311	0.02974237	
CORRECTED TOTAL	6831	308.93618998		

R-SQUARE	C.V.	ROOT MSE	YEARRATE MEAN
0.369794	75.3144	0.17245975	0.22898627

SOURCE	DF	TYPE I SS	F VALUE	PR > F
SUBJECT	258	110.83819764	14.44	0.0
TOOTH	27	3.40446923	4.24	0.0001

SOURCE	DF	TYPE III SS	F VALUE	PR > F
SUBJECT	258	110.88651394	14.45	0.0
TOOTH	27	3.40446923	4.24	0.0001

We thus see that there is a marked subject effect, and also that teeth of different types progress at different rates. The term ERROR (E) consists of the variation not explained by the model. We may suppose that this has two independent components; the true variation in the rate of progression of sites within subjects (SWS) and the variation in the estimated slopes caused by measurement error (ME). We may therefore write $\sigma^2_E = \sigma^2_{SWS} + \sigma^2_{ME}$.

We devised the following method to estimate σ^2_{ME}. Because the slopes were estimated from sites with at least three observations, it was possible (except where the observed progression was precisely linear) to estimate the measurement error and therefore the variance of the estimated slope. The estimate of the measurement error variance is given by usual unbiased mean square error estimate, s^2 for the site, and the covariance matrix of the estimates (of the intercept and the slope) is given by $s^2(X'X)^{-1}$, where X is the design matrix for the site. For each of the estimated

slopes, a corresponding estimated error variance was therefore derived. Because these estimates are unbiased, the expected value of their sum (assuming that the constant progression model holds) is σ^2_{ME}.

The value of this sum was 112.9571. Thus an unbiased estimate of σ^2_{SWS} is given by 194.6935-112.9571 = 81.7364. A modified analysis of variance, excluding measurement error, is therefore:

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	285	114.2427	0.40085
SUBJECT	258	110.8382	0.42961
TOOTH	27	3.4045	0.12609
SITES WITHIN SUBJECTS	6546	81.7364	0.012486

The use of the intraclass correlation coefficient is rather outdated. However we saw in section 2.1 that the value of R_S, the subject intraclass correlation coefficient for changes in attachment level, has been the subject of considerable interest in recent years. For comparative purposes, we therefore calculate R_S throughout this section. The estimate of R_S for these analyses is given by 0.42961/(0.42961+0.012486) = 0.972. This remarkably high value indicates that for this population most of the variation in the mean rate of disease progression at a site over the period of the study was explained by variation between subjects.

The mean rate of progression for the sites in each subject was therefore calculated and is shown in figure 5.1. Here we are of necessity ignoring the fact that subjects

had missing teeth of different types: the effect of tooth type is in any case comparatively small.

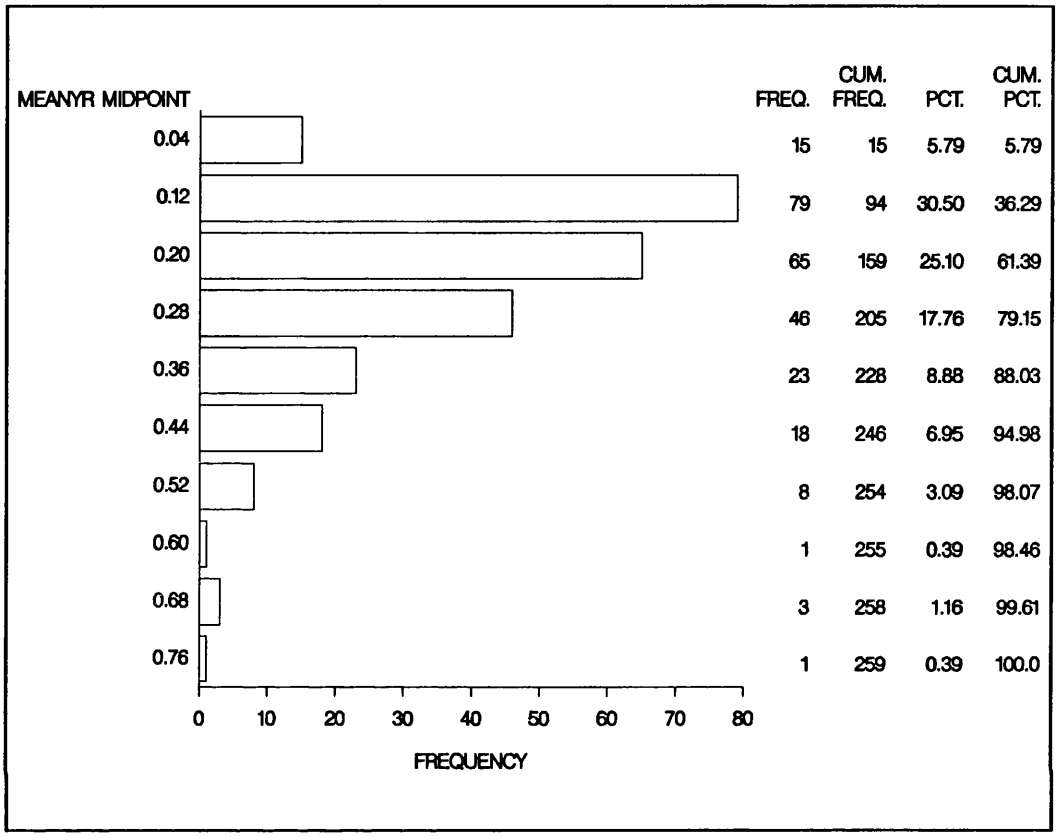


Figure 5.1. Subject mean rates of attachment loss (mm/year) for the complete Sri Lankan data set.

It is clear from figure 5.1 that the rate of disease progression showed a skewed distribution. While the modal mean rate of progression was 0.12mm/year, over 20% of the subjects experienced a mean rate of progression of greater than 0.3mm/year, with the maximum value being 0.77mm/year. These observations are entirely consistent with the conclusions of Löe et al (1986) although derived in a completely different manner. The large value of the intraclass correlation coefficient suggests that over the period of the study there was considerable homogeneity in the rate of disease progression within subjects, so that most of the progressing sites

belong to a minority of subjects. By far the most important factor which determines whether a site progresses is the subject to which the site belongs.

These results suggest that, even in a population as highly at risk as this one, periodontal disease progresses significantly (the word is used here in its biological sense) only in a minority of subjects. In asking questions about the rate of progression, it may therefore be that the only sites of interest are those in the minority of subjects for whom there was substantial disease progression. It may be that although for the majority of subjects there is no significant disease progression at any site, and therefore a very high subject intraclass correlation coefficient, the within subject variation is substantially higher for the minority of the subjects who experienced disease progression. To investigate whether this was the case, the above analyses were repeated for the 20% subjects for whom the mean of the estimated rate of progression for all their sites was highest (the quintile of subjects with greatest disease progression). The output from SAS is shown in table 5.2.

TABLE 5.2. SAS OUTPUT FOLLOWING ANALYSIS OF VARIANCE ON RATES OF PROGRESSION FOR EACH SITE IN THE QUINTILE OF SUBJECTS SHOWING THE MOST PROGRESSION.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YEARRATE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	77	14.33075778	0.18611374
ERROR	1227	103.87732462	0.08465960
CORRECTED TOTAL	1304	118.20808240	

MODEL F =	2.20		PR > F = 0.0001

R-SQUARE	C.V.	ROOT MSE	YEARRATE MEAN
0.121233	66.1533	0.29096322	0.43983199

SOURCE	DF	TYPE I SS	F VALUE	PR > F
SUBJECT	50	9.52975611	2.25	0.0001
TOOTH	27	4.80100167	2.10	0.0009

SOURCE	DF	TYPE III SS	F VALUE	PR > F
SUBJECT	50	9.58129989	2.26	0.0001
TOOTH	27	4.80100167	2.10	0.0009

As before, the component of σ^2_E due to measurement error was estimated from the sum of the estimated variances of the estimated rates of progression. This sum was 49.1359, so that an unbiased estimate of the sum of squares due to within subject variation is given by 103.8873-49.1359 = 54.7514. The amended analysis of variance, as before, becomes:

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	77	14.3308	0.18611
SUBJECT	50	9.5298	0.19060
TOOTH	27	4.8010	0.17781
SITES WITHIN SUBJECTS	1227	54.7514	0.04462

Hence the estimate of R_S is 0.19060/(0.19060+0.04462) = 0.81. This value is still substantial, and much higher than has been reported previously. Note also that the

between-tooth variance is now almost as high as the between subject variance, indicating that for subjects at risk, a large proportion of the difference in the rate of progression at different sites is explained by the tooth type.

5.3. SITES WITH SIX OBSERVATIONS.

In the next section we shall examine which of the covariance structures of models 1-3 best describes covariance structures of sample covariance matrices arising from the Sri Lankan data. We reviewed the theory of the estimation of covariance structures in section 2.2. One of the requirements is that observations be independently and identically distributed. Since the covariance structures of observations from each of models 1-3 depend both on the vector $\underline{\tau}$ of time intervals between observations and on the number of observations, we shall be able to use only those sites for which all six observations were made. It is therefore of interest to us to know how representative these sites are of the larger data set with which we have been dealing up till now.

Appendix 5.2 contains tables of attachment level and of increments in attachment level for this reduced data set, and also tables of increment in attachment level for each site. These 28 tables show the raw data used in the next section. There were 12354 observations of attachment levels, giving 10295 observed increments on 2059 sites. As for the whole data set, the measured attachment levels increased during the study. The range of measured attachment levels was 0 to 13mm, and the range of increments was -6 to 9mm. In each case the extreme values are smaller than for the complete data set.

Once again, an analysis of variance was performed on the estimated rates of progression for each site. The results are shown in Table 5.3.

TABLE 5.3. SAS OUTPUT FOLLOWING ANALYSIS OF VARIANCE ON RATES OF PROGRESSION FOR SITES WITH SIX OBSERVATIONS.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YEARRATE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	113	21.50630325	0.19032127	25.16
ERROR	1945	14.71011985	0.00756304	
CORRECTED TOTAL	2058	36.21642309		

R-SQUARE	C.V.	ROOT MSE	YEARRATE MEAN
0.593827	47.9826	0.08696576	0.18124432

SOURCE	DF	TYPE I SS	F VALUE	PR > F
SUBJECT	86	20.48109334	31.49	0.0
TOOTH	27	1.02520990	5.02	0.0001

SOURCE	DF	TYPE III SS	F VALUE	PR > F
SUBJECT	86	20.64420680	31.74	0.0
TOOTH	27	1.02520990	5.02	0.0001

We could proceed exactly as before to derive the subject intraclass correlation coefficient for these data. Because the data are identically distributed, however, an alternative method for estimating the component of variance due to measurement error may be used. The design matrix in the linear model for sites with 6 observations is:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 365/365.25 \\ 1 & 1095/365.25 \\ 1 & 2311/365.25 \\ 1 & 4257/365.25 \\ 1 & 5169/365.25 \end{bmatrix}, \quad \text{so that } X = \begin{bmatrix} 1 & 0 \\ 1 & 1.00 \\ 1 & 3.00 \\ 1 & 6.33 \\ 1 & 11.66 \\ 1 & 14.15 \end{bmatrix}$$

Hence:

$$X'X = \begin{bmatrix} 6 & 36.13 \\ 36.13 & 386.14 \end{bmatrix}, \quad \text{so that } (X'X)^{-1} = \begin{bmatrix} 0.3818 & -0.0357 \\ -0.0357 & 0.005933 \end{bmatrix}$$

The estimated component of the sum of squares due to measurement error is therefore given by $0.005933\sigma^2$, where σ^2 is the measurement error variance. Reasonable low and high estimates of the measurement error standard deviation, from the literature (as discussed in section 2), might be 0.55 and 0.8mm, giving the range of σ^2 as between 0.3 and 0.64mm^2. Estimates of the measurement error sum of squares are given by SSME = $2059 \times 0.005933 \times \sigma^2$ = between 3.665 and 7.818. Summation of the estimated variances of the slope estimates as before gives SSME= 8.524. We would expect the measurement error variance for a longitudinal study performed in difficult conditions to be towards the high of 0.64mm^2. This gives some confidence that the estimates of SSME used above are reasonable.

The amended analysis of variance, derived as before, is:

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
MODEL	113	21.5063	0.19032
SUBJECT	86	20.4811	0.23815
TOOTH	27	1.0252	0.03797
SITES WITHIN SUBJECTS	1945	6.1861	0.003181

The value of R_s for these data is thus 0.987.

Clearly, almost all the variation for these sites is between subject variation. The subject means were calculated, and are shown in Figure 5.2 below.

Inspection of Figure 5.2 shows that the mean rate of progression for the subjects contributing data at each examination was smaller than for the whole data set. This

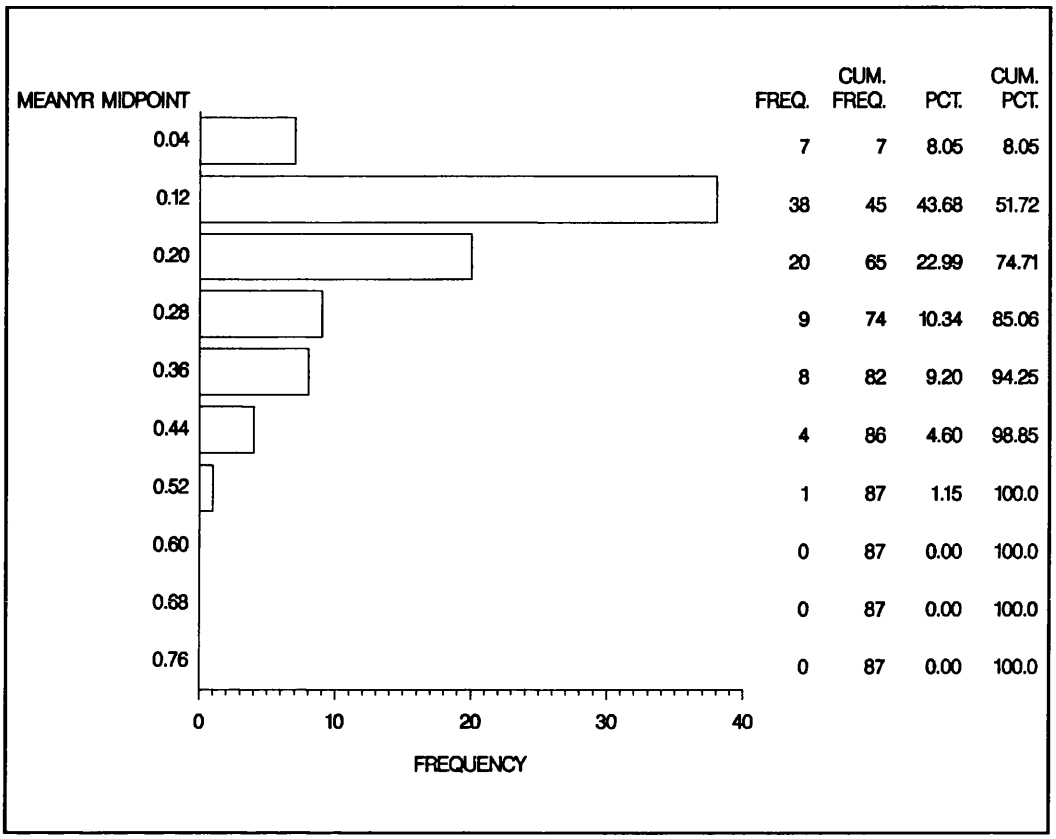


Figure 5.2. Subject mean rates of attachment loss for sites with six observations.

is unsurprising since some of the most rapidly progressing cases were reported to have lost many teeth through periodontal disease by the time of the sixth examination. However the removal from the data of some of the most rapidly progressing sites may make distinction of models 1-3 in the next section more difficult. It will be necessary to bear in mind the strong within subject correlations when interpreting the results of the next section.

Mean changes in attachment between examinations were calculated and are shown in Appendix 5.3. The mean changes for all sites were 0.2020, 0.6814, 0.6989, 0.6153 and 0.4313mm. The observed rates of progression were thus 0.5540, 0.9334, 0.5748, 0.3162 and 0.4729mm per 1000 days. These large fluctuations in the rate of progression are of concern since models 1-3 all assume a constant mean rate of

disease progression. They could be caused by changes in examiner bias between observation periods (if, in the notation of sections 3 and 4, $w_{ij} = x^i(t_j) + e_{ij} + b_j$, where b_j is the examiner bias for time j). If this is the case then b_j will be eliminated by allowing for differing rates of progression in the estimation procedure (i.e. setting the mean vector for the distribution equal to $\underline{\mu}$ ($\underline{\mu}$ unknown) rather than $\mu\underline{\tau}$ (μ unknown, $\underline{\tau}$ known). If, however, the rate of disease progression varies over time (for instance because the population is more susceptible at certain ages) then the covariance structures derived from the models may not describe the true underlying distribution.

5.4. VARIATION IN THE RATE OF PROGRESSION.

To test whether, as is assumed by our models for disease progression, the expected rate of progression was constant over the period of the study, the following procedure was devised. For each site with at least four observations, the usual least squares methods were used to fit a multiple regression of attachment level against time and time^2. The sign of the coefficient of the quadratic term was used as an indication of whether the rate of progression at the site had been greater at the beginning (negative) or the end (positive) of the study. The number of positive signs, out of the total number of sites, was calculated for each subject. Under the hypothesis that there is no tendency for disease progression to vary over the period of the study, the number of positive signs will show a binomial distribution with p=0.5. For each subject, the cumulative probability associated with the observed number of positives was calculated: a histogram of these probabilities is shown in

Figure 5.3. Probabilities near to zero or to one mean that the observation was respectively substantially smaller or larger than would be expected at random.

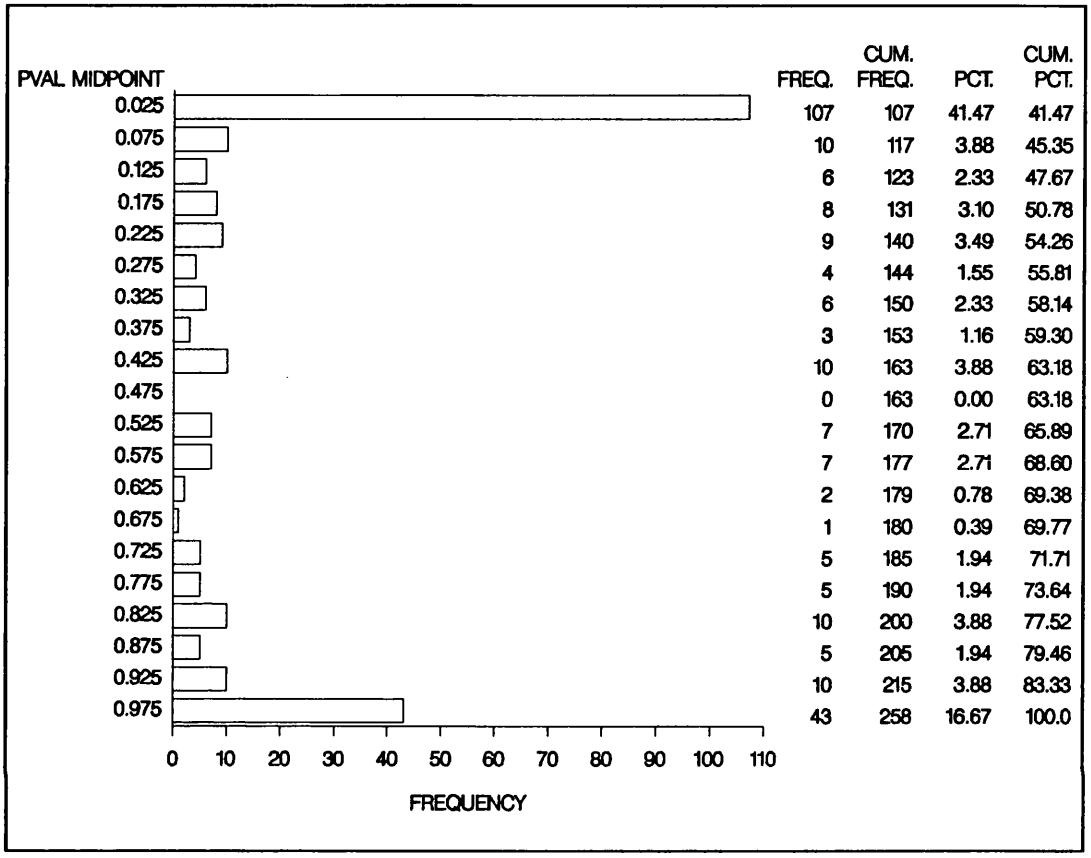


Figure 5.3. Probability of the observed number of positive quadratic terms for each subject, given that they arise from a binomial distribution with p=0.5.

It is clear from Figure 5.3 that a substantial number of subjects had proportions of positive signs which could not be ascribed to random error. Over twice as many had small numbers of positive signs than had large numbers of positive signs, indicating that disease progression tended to be higher at the beginning of the study. This accords with the observation (above) that the mean rate of progression was higher during the early part of the study. The large number of subjects showing non-random proportions of positive signs indicates that disease progression varied (though this method does not indicate the nature of the variation), and that variation

in the rate of progression at a site, as well as (as shown above) the estimated rate of progression at sites, was correlated within subjects. The observation that more subjects showed increased disease progression at the beginning of the study indicates (unless, as discussed above, the cause of this observation is changes in examiner bias) that the assumption of temporal homogeneity does not hold for these data. If the expected disease progression were constant for each subject then even if the rate of progression varied in a subject we would expect approximately equal numbers of subjects in whom there had been increased rates of progression at the beginning and the end of the study.

5.5. CONCLUSIONS

We have shown in this section that disease progression mainly occurred in a minority of subjects, and that the rates of progression in different sites in the same subject were strongly associated. This latter was also true in the minority of subjects showing rapid disease progression, in whom we have further shown that tooth type was also a major factor influencing the rate of progression at a site.

We have further provided evidence that the rate of progression varied, that variation in the rate of progression was correlated between different sites in the same subject and that the assumption of temporal homogeneity may not hold for these data. However the methods used in this section, which are based on existing statistical methods, are not designed to distinguish between different models for the nature of progression, and therefore gave limited information in this regard. In the next section, we use the estimation of covariance structures to compare how well models 1-3 of sections 3 and 4 describe these data.

APPENDIX 5.1 TABLE OF FREQUENCIES OF ATTACHMENT LEVEL AND OF LOSS OF ATTACHMENT.

Table 5.4. Frequencies of attachment level by examination number.

EXAMINATION	1	2	3	4	5	6
TIME (DAYS)	0	365	1095	2311	4257	5169

ATTACHMENT LEVEL	EXAMINATION NUMBER 1	2	3	4	5	6	TOTAL
0	3408	2736	1101	83	92	32	7452
1	2062	2116	2463	1509	786	494	9430
2	773	884	1484	1833	1068	1236	7278
3	252	274	704	793	577	684	3284
4	143	210	419	500	399	371	2042
5	85	168	281	392	422	385	1733
6	38	45	88	131	183	158	643
7	28	39	71	120	140	169	567
8	5	11	11	32	53	50	162
9	3	3	14	30	40	42	132
10	0	2	6	14	15	25	62
11	0	0	2	5	14	17	38
12	0	1	2	7	14	22	46
13	0	0	1	24	0	12	37
14	0	0	0	1	0	0	1
TOTAL	6797	6489	6647	5474	3803	3697	32907

Table 5.5. Frequencies of observed attachment loss by examination interval.

INTERVAL	TIME (DAYS)
2-1	365
3-1	1095
3-2	730
4-1	2311
4-2	1946
4-3	1216
5-1	4257
5-2	3892
5-3	3162
5-4	1946
6-2	4804
6-3	4074
6-4	2858
6-5	912

LOSS OF ATTACHMENT	TIME PERIOD														
	2-1	3-1	3-2	4-1	4-2	4-3	5-1	5-2	5-3	5-4	6-2	6-3	6-4	6-5	TOTAL
-9	0	0	0	1	0	1	0	0	0	0	0	0	0	0	2
-7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
-6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
-5	2	0	2	0	0	1	0	0	0	3	0	0	1	0	9
-4	7	1	8	0	0	5	0	0	0	8	0	0	2	3	34
-3	30	1	6	0	0	12	0	0	4	15	0	0	1	23	92
-2	149	2	50	0	0	94	1	0	10	100	0	1	6	141	554
-1	850	8	408	0	4	445	0	2	44	390	0	15	40	398	2604
0	3471	127	2300	2	13	1787	0	13	176	951	0	72	166	1100	10178
1	1428	100	2426	2	49	1889	0	23	289	782	0	108	144	815	8055
2	358	65	787	1	43	714	0	10	150	361	2	56	67	253	2867
3	100	16	233	0	4	257	0	1	78	178	0	27	24	92	1010
4	42	8	57	0	6	75	0	2	41	81	0	18	9	38	377
5	16	1	22	0	0	31	0	0	20	27	0	11	7	11	146
6	5	0	8	0	1	10	0	0	5	21	0	10	2	12	74
7	1	0	2	0	3	10	0	1	1	7	0	5	1	5	36
8	0	0	0	0	1	5	0	0	1	1	0	2	1	1	12
9	1	0	0	0	0	4	0	0	1	3	0	0	1	1	11
10	0	0	2	0	0	1	0	0	1	0	0	1	1	1	7
11	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2
12	0	0	0	0	0	1	0	0	0	1	0	1	0	0	3
	6461	329	6311	6	124	5344	1	52	821	2929	2	328	473	2894	26075

APPENDIX 5.2 TABLES OF ATTACHMENT LEVELS AND LOSS OF ATTACHMENT BETWEEN OBSERVATION PERIODS FOR SITES WITH SIX OBSERVATIONS.

TABLE OF EXAMINATION BY ATTACHMENT LEVEL

ALL DATA

EXAMINATION	ATTACHMENT LEVEL													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	TOTAL
1	1050	640	205	72	41	29	14	6	2	0	0	0	0	0	2059
2	852	723	283	82	48	43	12	12	2	1	1	0	0	0	2059
3	304	832	496	195	124	60	22	18	3	2	2	1	0	0	2059
4	24	630	708	280	184	140	42	30	9	7	3	0	0	2	2059
5	37	428	586	317	247	217	85	78	32	22	4	4	2	0	2059
6	10	211	682	409	222	224	81	104	36	33	18	13	10	6	2059
ALL	2277	3464	2960	1355	866	713	256	248	84	65	28	18	12	8	12354

TABLES OF TIME PERIOD BY INCREMENT IN ATTACHMENT LEVEL

TIME PERIOD 1 MEANS INTERVAL BETWEEN EXAMINATIONS 1 AND 2 ETC.

ALL DATA

ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	1	1	1	11	39	253	1125	501	101	18	5	1	1	1	0	0	2059
2	0	1	1	1	13	146	751	822	237	69	11	3	3	1	0	0	2059
3	0	1	0	6	36	171	705	747	265	97	21	9	0	0	1	0	2059
4	0	2	7	12	72	257	714	559	248	113	47	14	12	0	0	2	2059
5	0	0	2	16	102	296	743	581	196	70	28	10	9	4	1	1	2059
ALL	1	5	11	46	262	1123	4038	3210	1047	367	112	37	25	6	2	3	10295

TABLES BY SITE

SITE NUMBER 1
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	3	13	35	18	5	0	0	0	0	1	0	0	76
2	0	1	0	0	1	4	20	32	12	4	2	0	0	0	0	0	76
3	0	0	0	0	4	7	25	20	11	8	1	0	0	0	0	0	76
4	0	0	0	1	2	8	15	23	15	9	3	0	0	0	0	0	76
5	0	0	0	2	3	13	28	16	7	4	1	1	1	0	0	0	76
ALL	0	1	0	4	13	45	123	109	50	25	7	1	1	1	0	0	380

SITE NUMBER 2
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	11	25	20	5	1	2	0	0	0	0	0	65
2	0	0	0	0	0	6	18	30	11	0	0	0	0	0	0	0	65
3	0	0	0	0	0	5	25	23	8	3	1	0	0	0	0	0	65
4	0	0	0	0	2	4	15	19	18	5	1	1	0	0	0	0	65
5	0	0	0	0	2	13	24	16	4	5	0	1	0	0	0	0	65
ALL	0	0	0	0	5	39	107	108	46	14	4	2	0	0	0	0	325

SITE NUMBER 3
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	2	12	39	14	2	0	0	0	0	0	0	0	70
2	0	0	0	0	0	3	20	33	9	4	0	1	0	0	0	0	70
3	0	0	0	0	0	6	22	29	10	3	0	0	0	0	0	0	70
4	0	0	0	0	2	8	18	25	10	6	1	0	0	0	0	0	70
5	0	0	0	0	4	15	21	18	8	3	0	1	0	0	0	0	70
ALL	0	0	0	1	8	44	120	119	39	16	1	2	0	0	0	0	350

SITE NUMBER 4
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	1	0	0	0	12	34	25	3	0	0	0	0	0	0	0	75
2	0	0	0	0	0	5	36	22	9	2	1	0	0	0	0	0	75
3	0	0	0	0	1	4	19	40	8	2	1	0	0	0	0	0	75
4	0	0	1	0	2	10	24	23	7	2	5	1	0	0	0	0	75
5	0	0	0	1	7	10	26	21	7	2	1	0	0	0	0	0	75
ALL	0	1	1	1	10	41	139	131	34	8	8	1	0	0	0	0	375

SITE NUMBER 5
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	7	54	16	5	1	0	0	0	0	0	0	84
2	0	0	0	0	0	3	35	36	7	2	0	1	0	0	0	0	84
3	0	0	0	0	1	5	26	33	14	3	2	0	0	0	0	0	84
4	0	0	0	0	2	14	31	22	9	0	4	2	0	0	0	0	84
5	0	0	0	0	4	13	26	26	10	2	1	1	0	0	1	0	84
ALL	0	0	0	0	8	42	172	133	45	8	7	4	0	0	1	0	420

SITE NUMBER 6
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	15	46	15	3	0	0	0	0	0	0	0	80
2	0	0	0	0	1	4	26	39	7	2	0	0	1	0	0	0	80
3	0	0	0	0	3	6	22	33	8	7	0	1	0	0	0	0	80
4	0	0	0	0	4	7	34	21	5	6	1	2	0	0	0	0	80
5	0	0	0	1	5	6	24	33	4	6	0	1	0	0	0	0	80
ALL	0	0	0	1	14	38	152	141	27	21	1	4	1	0	0	0	400

SITE NUMBER 7
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	4	8	56	10	2	0	0	0	0	0	0	0	80
2	0	0	0	0	1	2	15	46	11	4	0	0	1	0	0	0	80
3	0	0	0	0	2	8	33	23	10	4	0	0	0	0	0	0	80
4	0	0	0	0	2	6	32	21	8	9	1	1	0	0	0	0	80
5	0	0	0	2	3	13	29	22	7	2	1	0	0	0	0	1	80
ALL	0	0	0	2	12	37	165	122	38	19	2	1	1	0	0	1	400

SITE NUMBER 8
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	2	7	53	15	0	0	0	0	0	0	0	0	77
2	0	0	0	0	0	2	27	35	9	2	2	0	0	0	0	0	77
3	0	0	0	1	1	5	32	29	7	1	1	0	0	0	0	0	77
4	0	0	1	1	3	5	30	21	9	5	1	0	1	0	0	0	77
5	0	0	0	0	3	11	28	26	4	2	0	0	3	0	0	0	77
ALL	0	0	1	2	9	30	170	126	29	10	4	0	4	0	0	0	385

SITE NUMBER 9
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	2	12	47	15	0	0	0	0	0	0	0	0	77
2	0	0	0	0	0	2	29	30	12	3	1	0	0	0	0	0	77
3	0	0	0	0	0	10	28	25	9	3	0	2	0	0	0	0	77
4	0	0	1	0	2	8	20	32	9	2	3	0	0	0	0	0	77
5	0	0	0	1	2	14	33	14	10	1	1	1	0	0	0	0	77
ALL	0	0	1	2	6	46	157	116	40	9	5	3	0	0	0	0	385

SITE NUMBER 10
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	9	50	19	2	1	0	0	0	0	0	0	82
2	0	0	0	0	0	2	37	28	10	4	1	0	0	0	0	0	82
3	0	0	0	0	2	5	27	31	10	5	1	0	0	0	1	0	82
4	0	0	1	0	2	13	33	16	8	6	2	0	1	0	0	0	82
5	0	0	0	2	6	11	31	24	5	2	0	1	0	0	0	0	82
ALL	0	0	1	2	11	40	178	118	35	18	4	1	1	0	1	0	410

SITE NUMBER 11
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	0	8	36	20	7	0	0	0	0	0	0	0	72
2	0	0	0	0	0	3	29	27	11	1	1	0	0	0	0	0	72
3	0	0	0	0	1	9	24	24	12	2	0	0	0	0	0	0	72
4	0	0	0	0	0	13	22	20	11	3	2	0	1	0	0	0	72
5	0	0	0	0	5	12	25	21	4	4	0	0	0	1	0	0	72
ALL	0	0	0	1	6	45	136	112	45	10	3	0	1	1	0	0	360

SITE NUMBER 12
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	3	9	36	19	4	1	0	0	0	0	0	0	72
2	0	0	0	0	1	5	26	32	7	1	0	0	0	0	0	0	72
3	0	0	0	0	1	3	27	27	8	5	0	1	0	0	0	0	72
4	0	0	0	1	0	15	13	28	9	5	1	0	0	0	0	0	72
5	0	0	0	0	4	12	29	20	5	0	2	0	0	0	0	0	72
ALL	0	0	0	1	9	44	131	126	33	12	3	1	0	0	0	0	360

SITE NUMBER 13
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	2	12	33	18	6	2	0	0	1	0	0	0	74
2	0	0	1	0	0	7	22	27	11	6	0	0	0	0	0	0	74
3	0	0	0	0	5	8	25	25	9	1	1	0	0	0	0	0	74
4	0	2	0	0	1	10	22	21	11	5	1	0	0	0	0	1	74
5	0	0	0	0	3	14	25	25	5	1	1	0	0	0	0	0	74
ALL	0	2	1	0	11	51	127	116	42	15	3	0	1	0	0	1	370

SITE NUMBER 14
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	2	11	44	11	1	2	1	0	0	0	0	0	73
2	0	0	0	0	0	6	20	27	10	9	1	0	0	0	0	0	73
3	0	0	0	1	3	11	16	26	9	5	1	1	0	0	0	0	73
4	0	0	0	2	4	7	21	24	9	5	0	1	0	0	0	0	73
5	0	0	0	0	0	11	29	24	6	2	1	0	0	0	0	0	73
ALL	0	0	0	4	9	46	130	112	35	23	4	2	0	0	0	0	365

SITE NUMBER 15
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	10	26	18	5	0	0	0	0	0	0	0	60
2	0	0	0	0	0	4	26	20	9	1	0	0	0	0	0	0	60
3	0	0	0	1	1	5	20	21	5	6	1	0	0	0	0	0	60
4	0	0	1	0	4	4	21	18	5	4	0	1	2	0	0	0	60
5	0	0	0	0	4	7	24	17	2	5	1	0	0	0	0	0	60
ALL	0	0	1	1	10	30	117	94	26	16	2	1	2	0	0	0	300

SITE NUMBER 16
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	1	0	0	0	0	5	19	22	5	1	0	0	0	0	0	0	53
2	0	0	0	0	1	10	22	13	4	2	0	0	1	0	0	0	53
3	0	0	0	0	1	4	20	18	8	2	0	0	0	0	0	0	53
4	0	0	0	0	1	9	18	13	7	3	2	0	0	0	0	0	53
5	0	0	0	0	3	5	20	15	8	1	1	0	0	0	0	0	53
ALL	1	0	0	0	6	33	99	81	32	9	3	0	1	0	0	0	265

SITE NUMBER 17
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	2	3	54	14	4	0	0	0	0	0	0	0	77
2	0	0	0	0	2	3	30	33	7	1	0	1	0	0	0	0	77
3	0	0	0	0	3	3	23	36	9	3	0	0	0	0	0	0	77
4	0	0	0	0	2	10	34	20	5	4	2	0	0	0	0	0	77
5	0	0	0	0	3	13	22	28	10	0	1	0	0	0	0	0	77
ALL	0	0	0	0	12	32	163	131	35	8	3	1	0	0	0	0	385

SITE NUMBER 18
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	1	4	39	28	7	1	0	0	0	0	0	0	81
2	0	0	0	0	1	11	33	29	5	1	1	0	0	0	0	0	81
3	0	0	0	0	0	8	27	33	9	2	0	2	0	0	0	0	81
4	0	0	0	2	2	6	29	31	9	2	0	0	0	0	0	0	81
5	0	0	0	0	5	8	34	23	9	1	1	0	0	0	0	0	81
ALL	0	0	0	3	9	37	162	144	39	7	2	2	0	0	0	0	405

SITE NUMBER 19
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	5	51	21	6	0	0	0	0	0	0	0	84
2	0	0	0	0	1	11	27	36	7	2	0	0	0	0	0	0	84
3	0	0	0	0	1	8	27	28	12	5	3	0	0	0	0	0	84
4	0	0	0	1	5	7	33	17	11	5	3	0	2	0	0	0	84
5	0	0	1	1	6	13	23	22	13	2	1	1	1	0	0	0	84
ALL	0	0	1	2	14	44	161	124	49	14	7	1	3	0	0	0	420

SITE NUMBER 20
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	1	0	2	12	31	19	4	1	0	0	0	0	0	0	70
2	0	0	0	0	0	6	24	27	10	2	1	0	0	0	0	0	70
3	0	0	0	0	1	7	32	18	9	3	0	0	0	0	0	0	70
4	0	0	0	0	1	4	32	14	13	4	1	1	0	0	0	0	70
5	0	0	0	0	1	15	22	22	7	1	1	1	0	0	0	0	70
ALL	0	0	1	0	5	44	141	100	43	11	3	2	0	0	0	0	350

SITE NUMBER 21
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	2	0	11	34	16	8	3	1	1	0	0	0	0	76
2	0	0	0	0	3	8	28	22	11	4	0	0	0	0	0	0	76
3	0	0	0	1	2	10	19	27	12	4	1	0	0	0	0	0	76
4	0	0	0	0	2	14	26	19	7	6	2	0	0	0	0	0	76
5	0	0	1	2	3	13	26	15	7	4	2	0	2	1	0	0	76
ALL	0	0	1	5	10	56	133	99	45	21	6	1	2	1	0	0	380

SITE NUMBER 22
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	11	34	21	6	2	0	0	0	0	0	0	75
2	0	0	0	0	0	9	20	37	6	3	0	0	0	0	0	0	75
3	0	0	0	1	0	10	24	23	12	5	0	0	0	0	0	0	75
4	0	0	0	0	3	15	25	19	6	4	1	2	0	0	0	0	75
5	0	0	0	0	6	7	27	17	10	5	2	0	0	1	0	0	75
ALL	0	0	0	1	10	52	130	117	40	19	3	2	0	1	0	0	375

SITE NUMBER 23
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	2	8	32	25	4	2	0	0	0	0	0	0	73
2	0	0	0	0	0	4	23	34	9	3	0	0	0	0	0	0	73
3	0	0	0	0	1	6	30	23	8	3	2	0	0	0	0	0	73
4	0	0	0	1	1	9	31	10	18	1	1	1	0	0	0	0	73
5	0	0	0	1	4	12	29	12	9	3	1	1	1	0	0	0	73
ALL	0	0	0	2	8	39	145	104	48	12	4	2	1	0	0	0	365

SITE NUMBER 24
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	1	5	53	22	2	0	0	0	0	0	0	0	83
2	0	0	0	0	0	2	32	40	6	3	0	0	0	0	0	0	83
3	0	0	0	1	0	4	33	24	14	4	3	0	0	0	0	0	83
4	0	0	0	1	8	10	34	15	7	3	3	0	1	0	0	1	83
5	0	0	0	2	3	6	31	23	9	1	6	0	1	1	0	0	83
ALL	0	0	0	4	12	27	183	124	38	11	12	0	2	1	0	1	415

SITE NUMBER 25
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	0	8	56	17	2	0	0	0	0	0	0	0	84
2	0	0	0	1	0	7	36	29	10	1	0	0	0	0	0	0	84
3	0	0	0	0	1	5	30	34	10	3	0	1	0	0	0	0	84
4	0	0	0	1	3	16	26	22	7	3	3	1	2	0	0	0	84
5	0	0	0	0	6	8	33	27	6	3	1	0	0	0	0	0	84
ALL	0	0	0	3	10	44	181	129	35	10	4	2	2	0	0	0	420

SITE NUMBER 26
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	1	8	47	14	1	0	0	0	0	0	0	0	72
2	0	0	0	0	1	5	36	23	7	0	0	0	0	0	0	0	72
3	0	0	0	0	0	4	23	33	11	1	0	0	0	0	0	0	72
4	0	0	0	1	3	10	30	18	7	3	0	0	0	0	0	0	72
5	0	0	0	0	2	7	36	17	5	4	1	0	0	0	0	0	72
ALL	0	0	0	2	7	34	172	105	31	8	1	0	0	0	0	0	360

SITE NUMBER 27
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	0	3	8	26	14	0	0	0	0	0	0	0	0	51
2	0	0	0	0	0	3	25	18	4	1	0	0	0	0	0	0	51
3	0	0	0	0	1	0	23	18	4	3	1	1	0	0	0	0	51
4	0	0	1	0	6	6	18	14	3	0	2	0	1	0	0	0	51
5	0	0	0	1	3	4	19	17	6	1	0	0	0	0	0	0	51
ALL	0	0	1	1	13	21	111	81	17	5	3	1	1	0	0	0	255

SITE NUMBER 28
ROWS: TIME PERIOD COLUMNS: INCREMENT

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	ALL
1	0	0	0	1	0	9	35	15	2	0	1	0	0	0	0	0	63
2	0	0	0	0	0	9	29	17	6	1	0	0	0	1	0	0	63
3	0	1	0	0	0	5	23	23	9	1	1	0	0	0	0	0	63
4	0	0	1	0	3	9	27	13	5	3	1	0	1	0	0	0	63
5	0	0	0	0	2	10	19	20	9	3	0	0	0	0	0	0	63
ALL	0	1	1	1	5	42	133	88	31	8	3	0	1	1	0	0	315

APPENDIX 5.3. MEAN CHANGES IN ATTACHMENT LEVEL BETWEEN OBSERVATION PERIODS.

ROWS: SITE COLUMNS: TIME

	1	2	3	4	5	ALL
1	0.1711	0.8553	0.7237	1.0132	0.4211	0.6368
2	0.4308	0.7077	0.7231	1.0923	0.4154	0.6738
3	-0.0143	0.9286	0.7429	0.7857	0.3571	0.5600
4	0.1867	0.6000	0.8000	0.6667	0.2400	0.4987
5	0.2381	0.6905	0.8452	0.5714	0.5714	0.5833
6	0.0500	0.7375	0.7875	0.6000	0.5625	0.5475
7	-0.0250	1.0250	0.5375	0.7875	0.3750	0.5400
8	0.0519	0.8442	0.5195	0.5974	0.5325	0.5091
9	-0.0519	0.8442	0.6753	0.6753	0.3247	0.4935
10	0.1829	0.7561	0.8415	0.5244	0.1951	0.5000
11	0.3194	0.7361	0.5972	0.7222	0.3611	0.5472
12	0.2083	0.5833	0.8056	0.6528	0.2500	0.5000
13	0.3514	0.7568	0.4324	0.6622	0.2973	0.5000
14	0.0685	0.9863	0.6575	0.5616	0.4795	0.5507
15	0.2667	0.6167	0.7167	0.6833	0.4167	0.5400
16	0.4528	0.3962	0.6415	0.6226	0.5094	0.5245
17	0.1948	0.6234	0.7013	0.4675	0.4286	0.4831
18	0.4444	0.4074	0.7284	0.4815	0.3704	0.4864
19	0.3095	0.5119	0.8214	0.6905	0.4405	0.5548
20	0.1429	0.7286	0.5143	0.7857	0.4429	0.5229
21	0.4342	0.5526	0.6579	0.5395	0.5132	0.5395
22	0.3467	0.6533	0.6533	0.4800	0.6400	0.5547
23	0.3699	0.7808	0.6575	0.6027	0.4247	0.5671
24	0.2289	0.7108	0.8313	0.4337	0.7590	0.5928
25	0.1190	0.5000	0.7262	0.5833	0.3810	0.4619
26	0.0417	0.4167	0.7500	0.3056	0.4444	0.3917
27	0.0000	0.5098	0.8235	0.2353	0.3725	0.3882
28	0.1746	0.4762	0.6032	0.3651	0.5238	0.4286
ALL	0.2020	0.6814	0.6989	0.6153	0.4313	0.5258

6

ESTIMATION OF COVARIANCE STRUCTURES FOR THE SRI LANKA DATA.

In this section we apply the methods described in section 2.2 in order to assess which of models 1-3 of section 3 best describes the progression of periodontal disease in the study. Because only limited assumptions were made about the various distributions which are defined in the formulation of the models, no distribution for the observed increments in attachment is available on which to base an estimation procedure. We can thus either make assumptions about the likelihood of the observed increments or use procedures which do not depend on specification of a particular distribution, but rather on assumptions about its moments. We choose the latter, and use the methods of Browne (1974, 1982, 1984) for the estimation of the parameters of the three models for disease progression.

The data analyzed are, as described in section 5.3, from those sites for which there were measurements at each of the six examinations. We have already ascertained that there is a strong within subject correlation of the observed rates of progression. Since estimation procedures require independent identically distributed observations, the data were partitioned by tooth type into 28 separate data sets, so that data from no more than one tooth per subject was contained in any data set. The raw data are shown in appendix 6.2. We thus have 28 data sets, whose distribution will be correlated. Although the estimates of parameters based on the different data sets will therefore also be correlated, consistency of parameter estimates between the data sets will provide evidence that the values reflect the nature of the underlying distribution rather than random variation. The number of teeth in each data set is shown below. The times between examinations (the vector $\underline{\tau}$ of section 3) were 365, 730, 1216, 1946 and 912 days.

6.1 SAMPLE COVARIANCE MATRICES

The sample covariance matrix $S = (N-1)^{-1}\sum_{i=1}^{N}(\underline{w}_i - \bar{\underline{w}})(\underline{w}_i - \bar{\underline{w}})'$ forms the basis for the GLS estimation procedures. A first step in the analysis is thus to inspect the sample covariance matrix. We have adopted the following form for the presentation of a real symmetric matrix S of dimension 5:

$$S = \begin{matrix} s_{11} & & & & \\ & s_{12} & & & \\ s_{22} & & s_{13} & & \\ & s_{23} & & s_{14} & \\ s_{33} & & s_{24} & & s_{15} \\ & s_{34} & & s_{25} & \\ s_{44} & & s_{35} & & \\ & s_{45} & & & \\ s_{55} & & & & \end{matrix}$$

This form enables the diagonals and one-off-diagonals of S, which are the important features of the covariance structures, to be easily discerned. We have:

$$B_1 = \begin{matrix} 2 & & & \\ & -1 & & \\ 2 & & 0 & \\ & -1 & & 0 \\ 2 & & 0 & \\ & -1 & & 0 \\ 2 & & 0 & \\ & -1 & & \\ 2 & & & \end{matrix}, \quad B_2 = \begin{matrix} \tau_1^2 & & & \\ & \tau_1\tau_2 & & \\ \tau_2^2 & & \tau_1\tau_3 & \\ & \tau_2\tau_3 & & \tau_1\tau_4 \\ \tau_3^2 & & \tau_2\tau_4 & & \tau_1\tau_5, \\ & \tau_3\tau_4 & & \tau_2\tau_5 & \\ \tau_4^2 & & \tau_3\tau_5 & & \\ & \tau_4\tau_5 & & & \\ \tau_5^2 & & & & \end{matrix}, \quad B_3 = \begin{matrix} \tau_1 & & & \\ & 0 & & \\ \tau_2 & & 0 & \\ & 0 & & 0 \\ \tau_3 & & 0 & & 0 \\ & 0 & & 0 & \\ \tau_4 & & 0 & & \\ & 0 & & & \\ \tau_5 & & & & \end{matrix}$$

The 28 sample covariance matrices are shown in Appendix 6.1. A "mean" sample covariance matrix, S, was derived by calculating the mean value of each element of these 28 matrices. The result, with the usual estimate of the standard error of each element in parentheses, was:

0.860(.071)
-0.450(.047)
1.007(.056) 0.002(.018)
-0.425(.038) -0.053(.027)
1.311(.060) 0.025(.030) 0.061(.033)
-0.373(.043) 0.006(.033)
2.086(.041) -0.019(.041)
-0.488(.052)
1.859(.116)

The component of the structure due to measurement error has the form $\theta_1 B_1$. By inspection, it is clear that the diagonal and off- diagonal elements of S contain, as expected, a large component due to measurement error.

We discussed the likely measurement error variance for these data in section 5.2. If we take the same high and low estimates (0.64 and 0.3mm^2) respectively, we may derive estimates of the underlying sample covariance matrix, excluding the effect of measurement error. Subtracting 0.64 B_1 from S gives:

$$S' = \begin{matrix} -0.420 & & & & \\ & 0.190 & & & \\ -0.273 & & 0.002 & & \\ & 0.215 & & -0.053 & \\ 0.031 & & 0.025 & & 0.061 \\ & 0.267 & & 0.006 & \\ 0.806 & & -0.019 & & \\ & 0.152 & & & \\ 0.579 & & & & \end{matrix}$$

Subtracting 0.3 B_1 from S gives:

$$S'' = \begin{matrix} 0.260 & & & & \\ & -0.150 & & & \\ 0.607 & & 0.002 & & \\ & -0.125 & & -0.053 & \\ 0.711 & & 0.025 & & 0.061 \\ & -0.073 & & 0.006 & \\ 1.486 & & -0.019 & & \\ & -0.188 & & & \\ 1.259 & & & & \end{matrix}$$

While some of the diagonal elements of S' are negative and all of the one-off-diagonal elements of S' are positive, all one-off-diagonal elements of S'' are still negative. This appears to confirm that the true measurement error variance for the study lies between the two values.

The six elements of S not affected by the value of θ_1 (S_{13} S_{14} S_{15} S_{24} S_{25} and S_{35}) are all small (absolute value 0.07 or less), with two being negative. These components, which are positive in each model, are affected by the value of θ_2 in all models, and of θ_4 and θ_5 in model 3. Thus the value of these elements suggests that the between-site variation is small, and that the correlations between different increments which may be observed under model 3 are small where the increments are separated by more than one examination.

6.2 SAMPLE KURTOSIS.

The choice of which of the generalised least-squares methods should be used for estimation of covariance structure parameters depends on whether the underlying distribution has excess kurtosis compared to the normal distribution. Mardia (1970) proposed $b_{2,p} = N^{-1}\sum_{i=1}^{N}\{(\underline{X}_i-\bar{\underline{X}})'S^{-1}(\underline{X}_i-\bar{\underline{X}})\}^2$ as a measure of kurtosis and showed that if the distribution of $\{\underline{X}_i\}$ is multivariate normal then $B = [b_{2,p} - \{p(p+2)(n-1)/(n+1)\}]/\sqrt{\{8p(p+2)/n\}}$ is asymptotically distributed as N(0,1). The values of $b_{2,p}$ and B were calculated for each of the 28 data sets. The results are shown in appendix 6.2. For each site the value of B clearly indicated that the underlying distribution possessed excess kurtosis. The

asymptotically distribution-free methods of section 2.2.3.2 may therefore be the most appropriate for these data.

6.3. ESTIMATION OF COVARIANCE STRUCTURES.

The generalized least squares methods of Browne described in section 2.2.3, and the computer software described in section 3.3, were used to estimate the parameters of models 1-3. Note that these methods assume $E[\underline{w}] = \underline{\mu}$ ($\underline{\mu}$ unknown), rather than $E[\underline{w}] = \mu\underline{\tau}$ (μ unknown, $\underline{\tau}$ known) as in each of our models. However, we clearly have $E[w_i] = \mu\tau_i$, so that S is a consistent estimator of $\Sigma(\underline{\gamma})$. Further, the possible variation in examiner bias between examinations (see section 5.3) suggests that it may be sensible to allow for differing observed rates of progression.

Both the GLS (adf) and GLS (nk) methods were used, since neither can be considered to be clearly more suited to the analysis of these data. While the GLS (nk) method assumes that the underlying distribution of the data has no excess kurtosis, the GLS (adf) methods are computationally expensive, and have been reported to produce parameter estimates with negative finite-sample bias (Browne 1984). The vector $\underline{\tau}$ was expressed in thousands of days, in order to ensure that the parameters were of similar magnitudes.

Parameters were estimated for each model, using each generalized least-squares method, on each data set (i.e. each site), using the software described in section 3.3. For model 3, a number of different sets of initial values were used. The sets of values were combinations of the means of the estimates of $\{\theta_1,\theta_2,\theta_3\}$ for model

1, the alternative values .5 for θ_2 and .8 for θ_3, and the pairs (5,5), (1,1) and (10,.1) for (θ_4,θ_5). This gave the following combinations.

SET NUMBER	METHOD	θ_1	θ_2	θ_3	θ_4	θ_5
1	nk	.314	.016	.314	5	5
2	nk	.314	.016	.314	1	1
3	nk	.314	.016	.8	5	5
4	nk	.314	.016	.8	1	1
5	nk	.314	.5	.314	5	5
6	nk	.314	.5	.314	1	1
7	nk	.314	.016	.314	10	.1
8	nk	.314	.5	.314	10	.1
9	nk	.314	.016	.8	10	.1
10	nk	.314	.5	.8	10	.1

SET NUMBER	METHOD	θ_1	θ_2	θ_3	θ_4	θ_5
1	adf	.315	.01	.235	5	5
2	adf	.315	.01	.235	1	1
3	adf	.315	.01	.8	5	5
4	adf	.315	.01	.8	1	1
5	adf	.315	.5	.235	5	5
6	adf	.315	.5	.235	1	1
7	adf	.315	.01	.235	10	.1
8	adf	.315	.5	.235	10	.1
9	adf	.315	.01	.8	10	.1
10	adf	.315	.5	.8	10	.1

The parameter estimates corresponding to the best goodness of fit achieved were then selected. The iterations for model 3 almost invariably ended when no value of the parameter gave a reduction in the discrepancy function, suggesting that the function was at a local minimum. Although there were sites for which all or nearly all of the sets of initial values gave the same goodness of fit, it was more common for between 1 and 4 of the sets to give the best goodness of fit. On 8 and 12 occasions for the nk and adf methods respectively, only 1 of the sets gave rise to the best value.

The results of Steiger et al. (discussed in section 2.2.7) suggest that, given that the covariance structures of models 1-3 are nested, it will be more important to examine the improvement in goodness of fit achieved by using model 3 over model 2 over model 1 than to test the goodness of fit of any of the models. We also calculate cross-validation indices and incremental fit indices as aids to model selection.

6.3.1 GOODNESS OF FIT.

The goodness of fit statistics and differences in goodness of fit between the three models, together with a table of summary statistics and p values, and a table of the relevant points of the χ^2 distribution are shown in Appendix 6.3. The values shown for model 3 are those for the best goodness of fit achieved.

The results for the two methods follow similar patterns, although the fit of the models is generally better for the ADF method than for the method assuming no kurtosis. Thus, taking p=0.05 as a cutoff; model 2 fails fully to account for the variation in 16 and 22 of the data sets for the adf and nk methods respectively.

Model 2 fits statistically significantly (again taking 0.05 as the cut off) better than model 1 in 16 and 11 of the data sets for the adf and nk methods respectively. However, it is noteworthy that for around half of the sites there is no evidence, even in a long term study such as this, that the rate of disease progression varied.

The improvement in goodness of fit of model 3 over model 2 was (to within five decimal places) zero for 16 of the data sets, for each GLS method. On no occasion

did the increase in goodness of fit reach the upper 10% point of the χ^2 (2 df) distribution. However, the results of section 5 must be borne in mind. The improvement in the goodness of fit of model 2 over model 1 was never as large as that achieved in simulation experiments 1 to 3, and the largest improvements were of the order of those achieved in experiment 4, where the data were in fact generated using model 3. It may therefore be that the power of the asymptotic chi-square tests is insufficient to distinguish between the two models, because of the relatively small variation in the rate of progression .

6.3.2 NULL MODELS.

In order to test the relative amounts of variation due to measurement error and between subject variation, two more models were fitted. These estimated the effect of θ_1 and θ_2 separately, by using a single matrix (B_1 and B_2 respectively) in the covariance structure. The results of fitting these models is shown in appendix 6.4. This shows that, although the omission of either component of model 1 caused a large decrease in goodness of fit, by far the largest source of variation is the measurement error. The mean goodness of fit statistic increased by around half when B_2 was omitted, but over fourfold when B_1 was omitted.

The presence of measurement error as a large proportion of the variation means that an appropriate "null" model, (cf Bentler and Bonnett (1980) and section 2.2.7) will be one where it is assumed that the only component of the covariance structure is the component $\theta_1 B_1$ due to measurement error.

6.3.3 INCREMENTAL FIT INDICES.

The null model M_0: $\Sigma(\underline{\theta}) = \theta_1 B_1$ was used to calculate incremental fit indices as discussed in section 2.2.7. The fit index chosen was Bollen's (1986) modification of Bentler and Bonnett's nonnormed fit index (equation 2.62). The chi square goodness of fit statistics and incremental fit indices, together with summary statistics for moving from model 0 to 1, 1 to 2 and 2 to 3 are shown in Appendix 6.5.

The greatest improvement in fit as measured by the nonnormed fit index was in moving from the null model to model 1. The mean of R_{12} was a little over half that for R_{01} for the adf method, a little under half for the nk method. The mean of R_{23} was close to zero for both methods.

6.3.4 CROSS VALIDATION COEFFICIENTS.

We also discussed in section 2.2.7 the use of cross-validation to aid the selection of models for covariance structures. Single-sample cross-validation indices c_k (equation 2.66) were calculated for each sample and each model. Appendix 6.6 contains the cross-validation coefficients (Ck for model k) and differences between the coefficients, together with summary statistics.

The single sample cross-validation index is simply the discrepancy function plus a penalty for the number of parameters, so that models with less parameters are favoured. This is reflected in that the values of c are similar for models 1 and 3. The smallest values of c are, on average, for model 2.

6.3.5 PARAMETER ESTIMATES FOR THE THREE MODELS.

The parameter estimates given by each of the GLS methods and each of the three models, together with summary statistics, are shown in appendix 6.7. For model 3, the parameters θ_4 and θ_5 are those of the distribution of λ, the rate of changes in slope (see section 2). This is in contrast to model 2, where the parameters of the distribution of λ are combined with other parameters. We can thus estimate the rate of changes in slope from model 3 as $(\theta_4+2)\theta_5$. This product is also shown, in the column labelled RATE, in appendix 6.7.

The parameter estimates are similar for the two methods, although the estimates for the ADF method are rather smaller. This may reflect the negative bias reported by Browne (1984). The mean estimate of θ_1 (the measurement error variance) for models 1, 2 and 3 respectively is (0.414, 0.314, 0.334)mm^2 for the nk method and (0.389, 0.315, 0.330)mm^2 for the adf method. These values appear entirely consistent with those reported in the literature (see section 2.1).

The estimated values of θ_2 (the between-subject variation) were surprisingly small for models 2 and 3 where the estimate was less than 10^{-5} for over half of the sites, given the heterogeneity of the subjects noted in section 5. in the study (Löe 1986) this was surprising. These models may be misspecified in a manner which tends to bias the estimate of θ_2 towards zero. When model 1 was used the estimate of θ_2 was zero only once.

For model 2, the estimates of θ_3 for the two methods were clearly strongly associated, although they were smaller for the adf method. Where the estimate was zero or near zero the improvement in the goodness of fit of model 2 over model 1 was negligible.

For model 3, the parameter estimates were as expected given the small increase in goodness of fit over model 2. The estimates of θ_1-θ_3 were similar for the two models. The estimates of θ_4 and θ_5 showed wide variation, as in section 5.

For data sets where the difference in goodness of fit of models 2 and 3 was greater than 0.01, the estimated rates of changes in slope, showed a surprising homogeneity, with every estimate for the nk method lying between 0.415 and 4.061 changes per 1000 days, and all but one of the estimates for the adf method lying between 0.397 and 3.105 per 1000 days. Note that the variance of this rate, which is given by $(\theta_4+2)\theta_5^2$, is almost always estimated to be near zero since the estimate of θ_5 is nearly always small. However, given the failure to demonstrate that model 3 holds, these results cannot be taken as more than a indication of the possible frequency of slope changes which requires more evidence.

6.4 DISCUSSION

These results do not lead to the unequivocal acceptance of any of models 1 to 3, although it seems clear that model 1 is accepted over the null model M_0.

Both the goodness of fit statistics and the cross-validation indices suggest that model 2 is a better description of the data than model 1. However, as noted above, there were a substantial number of sites where model 2 did not significantly improve on model 1. The main evidence against model 2 is that the estimates of the between-subject variance (θ_2) are nearly always zero. The results of section 5 where we showed both that there was a large subject effect and large intra-subject heterogeneity, indicate that θ_2 must be greater than zero. This suggests that model 2 is not an accurate description of the underlying distribution. It also suggests that the estimate of θ_3 (the variance component representing variation in the rate of progression) is artificially inflated by variation which is in reality between site variation.

We have provided no positive evidence that model 3 is the best description of the data. However the results of section 5 lead us to emphasise that failure to provide evidence for model 3 should not lead to unequivocal acceptance of model 2, leading to the conclusion that changes in the rate of disease progression occurred frequently in comparison to the time between observations. It certainly appears that changes in the rate of disease progression do not account for a large part of the observed variation, so that for these data, we are not likely to be able to distinguish models 2 and 3.

Each of the models assumes that the rate of progression of disease is constant over time. The breaching of this assumption for these data (which were collected over a substantial proportion of an adult life, may be an explanation for these results.

We noted above that calculation of the mean rates of progression between observations appeared to indicate that the rate of progression had changed between observation periods. This was also reported by Löe et al. (1978c), who reported that the rate of destruction showed a significant increase when subjects were in their late twenties and throughout the thirties. It may therefore be that variation in the rate of progression took place, but was correlated between subjects. This certainly might account for the differences between fitting models 1 and 2 discussed above.

6.5. CONCLUSIONS.

We have estimated the parameters of models 1, 2 and 3 using each of the GLS methods described in section 2.2. The estimates of the measurement error variance were consistent with the values reported in the literature. For models which assumed that the rate of progression varied, the estimates of the between-subject variation were surprisingly small given the heterogeneity of the sample population.

The improvement in goodness of fit of models 2 and 3 over model 1 showed that for around half the sites there was evidence that the rate of disease progression varied. However the small estimates of θ_2 indicate that neither of these models is a satisfactory description of the data. Because the variation in the rate of disease progression is a relatively small component of the total variation, models 2 and 3 are likely to be hard to distinguish.

We may conclude that we have provided evidence that the rate of disease progression varied, but that this variation is as likely to be due to changes in the rate of progression over time as to random bursts of activity.

APPENDIX 6.1 SAMPLE COVARIANCE MATRICES.

Site	Sample Covariance Matrix			
1	1.58			
	-1.00			
	1.67	-.11		
	-.44		-.11	
	1.80	.08		-.13
	-.93		.13	
	2.12	.16		
	-.65			
	2.49			
2	1.31			
	-.51			
	.74	.04		
	-.11		-.15	
	1.08	.17		.16
	-.35		-.08	
	1.74	.18		
	-.45			
	1.75			
3	.74			
	-.36			
	1.05	.03		
	-.32		-.09	
	.92	.04		.09
	-.37		.24	
	1.56	-.30		
	-.17			
	1.80			
4	.96			
	-.48			
	.95	.04		
	-.57		-.05	
	.89	.14		-.26
	-.32		.03	
	2.36	.01		
	-.66			
	1.70			
5	.62			
	-.37			
	.84	.06		
	-.36		-.11	
	1.17	.05		.13
	-.09		-.34	
	2.01	-.24		
	-.33			
	2.30			

Site Sample Covariance Matrix

6
.58
-.43
1.08
.07
-.58
-.21
1.54
.11
.16
-.30
-.15
1.96
-.16
-.18
1.74

7
.53
-.39
1.06
.03
-.36
-.12
1.16
.44
-.05
-.24
-.01
1.76
-.13
-.22
2.49

8
.39
-.27
.90
.01
-.40
-.10
1.04
.23
-.25
-.29
.24
2.22
.06
-.61
2.23

9
.58
-.36
.90
-.19
-.13
-.06
1.49
-.09
.04
-.38
.14
1.72
-.20
-.20
1.67

10
.57
-.23
.93
-.08
-.47
-.07
1.86
.02
.41
-.50
-.27
2.28
-.08
-.46
1.69

11
.81
-.24
.84
.07
-.45
-.19
1.09
.02
.01
-.18
-.09
1.89
-.01
-.17
2.04

Site Sample Covariance Matrix

12
.87
-.40
.75 -.09
-.42 -.14
1.26 .08 .13
-.55 -.27
1.67 -.12
-.18
1.35

13
1.46
-.79
1.45 -.03
-.39 -.40
1.37 -.12 .18
-.73 .03
3.27 .17
-.54
1.23

14
1.06
-.62
1.37 -.14
-.88 .28
2.06 -.05 -.02
-.67 .01
2.06 -.03
-.25
1.06

15
.81
-.29
.78 .01
-.26 .02
1.66 -.24 -.06
-.57 .30
3.03 .20
-.51
1.67

16
1.56
-1.16
1.71 -.03
-.62 -.13
1.04 .15 .00
-.37 .08
1.78 -.18
-.40
1.45

17
.50
-.29
1.00 .06
-.32 .01
1.00 .15 .03
-.40 -.13
1.44 .08
-.39
1.25

Site Sample Covariance Matrix

18
.83
-.51 .94
.16 -.29 1.28
.02 -.10 -.31 1.23
-.22 .06 .29 -.33 1.26

19
.55
-.29 .90
-.05 -.43 1.50
-.26 .26 -.15 2.67
.07 -.05 -.33 -.75 2.59

20
1.14
-.47 1.01
.04 -.39 1.07
.23 -.15 -.02 1.56
.14 .19 -.19 -.22 1.53

21
1.74
-.84 1.32
.04 -.65 1.61
.03 -.09 -.29 1.77
.13 .10 -.91 3.61

22
.93
-.42 .88
-.04 -.31 1.39
.09 -.30 .10 2.09
.03 .24 -.64 -.62 2.48

23
.90
-.21 .78
.09 -.38 1.28
-.06 .12 -.24 1.77
.58 -.31 .20 -.69 2.44

Site Sample Covariance Matrix

24

.42			
-.13			
.62	.03		
-.22	.06		
1.46	-.13	.07	
-.26	-.03		
3.30	.32		
-1.11			
2.89			

25

.49			
-.29			
.88	.06		
-.48	-.14		
1.09	-.02	-.05	
-.30	.14		
2.78	-.03		
-.42			
1.37			

26

.55			
-.33			
.67	.04		
-.27	-.06		
.70	.08	.05	
-.40	.00		
1.37	-.03		
-.45			
1.26			

27

.68			
-.22			
.65	-.24		
-.35	.04		
1.43	-.10	.20	
-.80	-.05		
2.74	.17		
-1.17			
1.36			

28

.89			
-.68			
1.51	.17		
-1.05	.18		
1.47	-.05	.12	
-.55	.00		
2.27	.16		
-.63			
1.35			

APPENDIX 6.2 MULTIVARIATE KURTOSIS.

$B = [b_{2,p} - \{p(p+2)(n-1)/(n+1)\}]/\sqrt{\{8p(p+2)/n\}}$

Site	$b_{2,p}$	B
1	46.75877	6.59981
2	39.84813	2.84690
3	41.52785	3.75688
4	48.20978	7.31340
5	50.20987	8.78185
6	50.18551	8.57893
7	51.93442	9.51376
8	59.68068	13.41328
9	48.84017	7.72847
10	55.86545	11.74801
11	45.97390	6.05104
12	40.21792	3.13222
13	56.48311	11.52401
14	37.40868	1.71288
15	44.14248	4.76335
16	46.04460	5.36915
17	47.89356	7.23206
18	48.04930	7.47775
19	41.15993	3.82500
20	44.32902	5.15747
21	41.10833	3.65599
22	39.79943	2.96063
23	41.51678	3.81048
24	43.95750	5.33065
25	48.32330	7.74854
26	41.83275	3.95109
27	40.61322	2.97014
28	62.48163	13.55449

APPENDIX 6.3. GOODNESS OF FIT OF MODELS 1-3.

SITE	METHOD	MODEL 1 (df 13)	MODEL 2 (df 12)	MODEL 3 (df 10)	DIFF 2-1 (df 1)	DIFF 3-1 (df 3)	DIFF 3-2 (df 2)
1	nk	18.1450	16.5177	16.5041	1.6274	1.6409	0.01359
2	nk	23.6708	21.5683	19.9641	2.1025	3.7068	1.60428
3	nk	21.4660	15.9048	15.9048	5.5612	5.5612	0.00000
4	nk	38.3423	38.3423	38.3423	0.0000	0.0000	0.00000
5	nk	43.8149	36.5947	36.2352	7.2202	7.5798	0.35959
6	nk	51.7203	38.2332	38.2332	13.4871	13.4871	0.00000
7	nk	45.2322	35.5888	35.5888	9.6434	9.6434	0.00000
8	nk	65.9509	42.8879	42.8879	23.0630	23.0630	0.00000
9	nk	46.7730	32.1678	32.1678	14.6052	14.6052	0.00000
10	nk	56.9228	30.8890	30.8890	26.0339	26.0339	0.00000
11	nk	25.3894	20.6480	20.5775	4.7414	4.8119	0.07047
12	nk	35.0191	33.6933	33.4607	1.3258	1.5584	0.23261
13	nk	21.5616	21.3767	21.3756	0.1849	0.1860	0.00107
14	nk	23.9305	23.5271	23.1656	0.4034	0.7649	0.36145
15	nk	25.9910	21.1699	20.9026	4.8211	5.0884	0.26733
16	nk	29.2324	29.2324	29.2324	0.0000	0.0000	0.00000
17	nk	24.1323	12.1352	12.1352	11.9971	11.9971	0.00000
18	nk	29.3471	21.5059	20.4712	7.8412	8.8759	1.03466
19	nk	57.3903	39.2618	39.2618	18.1285	18.1285	0.00000
20	nk	12.4326	11.5560	10.8946	0.8766	1.5379	0.66138
21	nk	22.3689	22.3689	22.3664	0.0000	0.0025	0.00253
22	nk	31.6753	25.3103	23.6271	6.3651	8.0482	1.68313
23	nk	35.7879	22.4290	22.4290	13.3590	13.3590	0.00000
24	nk	63.3936	35.5664	35.5664	27.8272	27.8272	0.00000
25	nk	45.3407	34.3083	34.3083	11.0324	11.0324	0.00000
26	nk	19.4860	17.2754	17.2754	2.2106	2.2106	0.00000
27	nk	25.3620	21.3255	21.3255	4.0365	4.0365	0.00000
28	nk	37.2938	37.2938	37.2938	0.0000	0.0000	0.00000

				p value				
	Minimum	Maximum	Mean	>0.1	<0.1	<0.05	<0.025	<0.01	<0.005
MODEL 1	12.4326	65.9509	34.8990	3	2	4	3	2	14
MODEL 2	11.5560	42.8879	27.0957	5	1	7	2	0	13
MODEL 3	10.8946	42.8879	26.8709	3	2	2	7	1	13
DIFF 2-1	0	27.8272	7.8034	11	0	3	2	2	10
DIFF 3-1	0	27.8272	8.0281	15	1	2	2	1	7
DIFF 3-2	0	1.6831	0.2247	28	0	0	0	0	0

Points of the χ^2 distribution

df	p 0.1	0.05	0.025	0.010	0.005
1	2.7055	3.8415	5.0239	6.6349	7.8794
2	4.6052	5.9915	7.3778	9.2103	10.5967
3	6.2514	7.8147	9.3484	11.3449	12.8382
10	15.9872	18.3070	20.4832	23.2093	25.1882
12	18.5493	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3620	24.7356	27.6882	29.8195

SITE	METHOD	MODEL 1 (df 13)	MODEL 2 (df 12)	MODEL 3 (df 10)	DIFF 2-1 (df 1)	DIFF 3-1 (df 3)	DIFF 3-2 (df 2)
1	adf	19.7298	19.2282	19.2280	0.5016	0.5018	0.00017
2	adf	32.4622	28.0504	26.7373	4.4119	5.7249	1.31307
3	adf	22.0897	13.0564	13.0506	9.0334	9.0391	0.00574
4	adf	41.1833	41.1833	41.1833	0.0000	0.0000	0.00000
5	adf	27.6042	19.3923	19.3923	8.2119	8.2119	0.00000
6	adf	43.3625	27.5473	27.5473	15.8151	15.8151	0.00000
7	adf	19.0775	13.3139	13.1209	5.7636	5.9566	0.19303
8	adf	31.6930	29.8102	29.8102	1.8827	1.8827	0.00000
9	adf	30.5279	21.2489	21.2489	9.2790	9.2790	0.00000
10	adf	31.7313	21.8643	21.8643	9.8670	9.8670	0.00000
11	adf	18.6846	15.3846	15.3846	3.3001	3.3001	0.00000
12	adf	27.1089	27.1089	23.7608	0.0000	3.3481	3.34812
13	adf	10.3278	10.2483	10.2483	0.0794	0.0794	0.00000
14	adf	31.9447	31.7577	31.1057	0.1870	0.8391	0.65205
15	adf	17.3332	17.3327	17.3314	0.0005	0.0019	0.00134
16	adf	23.3975	23.3975	23.3975	0.0000	0.0000	0.00000
17	adf	22.5639	10.9665	10.9665	11.5974	11.5974	0.00000
18	adf	27.1215	23.4602	22.3509	3.6613	4.7706	1.10922
19	adf	41.3090	28.2458	28.2458	13.0632	13.0632	0.00000
20	adf	21.7852	17.6210	16.6269	4.1642	5.1584	0.99419
21	adf	24.3588	24.3181	24.1148	0.0406	0.2440	0.20339
22	adf	22.9009	19.9129	18.9425	2.9880	3.9584	0.97035
23	adf	34.1740	24.1688	24.1688	10.0052	10.0052	0.00000
24	adf	48.6747	32.5448	32.5448	16.1299	16.1299	0.00000
25	adf	27.9697	26.3715	26.3715	1.5981	1.5981	0.00000
26	adf	14.6537	14.4772	14.4468	0.1766	0.2070	0.03040
27	adf	16.3992	13.4239	13.4239	2.9753	2.9753	0.00000
28	adf	21.6068	21.6068	21.6068	0.0000	0.0000	0.00000

	Minimum	Maximum	Mean	>0.1	p value <0.1	<0.05	<0.025	<0.01	<0.005
MODEL 1	10.3278	48.6747	26.8491	7	3	4	3	1	10
MODEL 2	10.2483	41.1833	22.0372	9	3	3	4	5	4
MODEL 3	10.2483	41.1833	21.7222	7	2	3	4	4	8
DIFF 2-1	0	16.1299	4.8119	12	4	2	1	0	9
DIFF 3-1	0	16.1299	5.1269	19	0	3	2	1	3
DIFF 3-2	0	3.3481	0.3150	28	0	0	0	0	0

Points of the χ^2 distribution

df	p 0.1	0.05	0.025	0.010	0.005
1	2.7055	3.8415	5.0239	6.6349	7.8794
2	4.6052	5.9915	7.3778	9.2103	10.5967
3	6.2514	7.8147	9.3484	11.3449	12.8382
10	15.9872	18.3070	20.4832	23.2093	25.1882
12	18.5493	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3620	24.7356	27.6882	29.8195

APPENDIX 6.4. RESULTS FROM FITTING MODELS WITH EITHER THE MATRIX B_1 OR THE MATRIX B_2 ONLY.

		MATRIX B_1 ONLY		MATRIX B_2 ONLY	
METHOD	SITE	THETA	CHISQ	THETA	CHISQ
nk	1	0.82212	27.3150	0.09958	150.00
nk	2	0.43823	46.6420	0.13834	128.00
nk	3	0.45317	38.7161	0.09468	138.00
nk	4	0.43887	40.3266	0.06165	148.00
nk	5	0.39170	59.9482	0.10376	166.00
nk	6	0.37001	67.1022	0.10533	158.00
nk	7	0.35638	65.3814	0.14413	158.00
nk	8	0.25596	80.8907	0.11057	152.00
nk	9	0.36538	60.7115	0.10592	152.00
nk	10	0.32343	85.1566	0.12509	162.00
nk	11	0.46393	44.8782	0.11936	142.00
nk	12	0.41481	38.5708	0.04837	142.00
nk	13	0.63458	36.2685	0.11426	146.00
nk	14	0.55582	34.2537	0.08743	144.00
nk	15	0.46240	43.2132	0.16399	118.00
nk	16	0.48436	29.9096	0.05370	104.00
nk	17	0.36354	49.8679	0.09619	152.00
nk	18	0.40091	47.6919	0.09659	160.00
nk	19	0.38333	74.7694	0.13062	166.00
nk	20	0.51036	33.7896	0.13394	138.00
nk	21	0.76050	34.5803	0.11584	150.00
nk	22	0.48163	52.1401	0.12613	148.00
nk	23	0.41595	57.3071	0.13864	144.00
nk	24	0.28172	91.7826	0.17776	164.00
nk	25	0.35425	62.3181	0.10321	166.00
nk	26	0.34382	33.2429	0.05115	142.00
nk	27	0.37439	34.1869	0.05730	100.00
nk	28	0.41677	40.5769	0.06052	124.00

Variable	Matrix	Method	Minimum	Maximum	Mean
THETA	B1	nk	0.2559628	0.8221204	0.4399401
THETA	B2	nk	0.0483733	0.1777558	0.1058592
CHISQ	B1	nk	27.3150370	91.7826312	50.4120688
CHISQ	B2	nk	100.0000000	166.0000000	145.0714286

		MATRIX B_1 ONLY		MATRIX B_2 ONLY	
METHOD	SITE	THETA	CHISQ	THETA	CHISQ
adf	1	0.73708	22.2910	0.09298	101.40
adf	2	0.39775	39.1305	0.14689	147.55
adf	3	0.44431	25.7363	0.08655	163.73
adf	4	0.38918	41.6117	0.04804	95.73
adf	5	0.35201	43.5704	0.05693	113.41
adf	6	0.29973	44.9195	0.00000	101.24
adf	7	0.35573	28.1782	0.00780	113.62
adf	8	0.18766	37.8547	0.03435	71.06
adf	9	0.36134	34.4503	0.07129	81.19
adf	10	0.25012	46.4624	0.06011	91.57
adf	11	0.43843	28.3649	0.03200	146.50
adf	12	0.41156	27.1089	0.03068	129.89
adf	13	0.50633	19.4085	0.04549	112.78
adf	14	0.47312	41.3106	0.08276	100.56
adf	15	0.40022	18.0869	0.02759	128.41
adf	16	0.44063	23.3975	0.03252	113.90
adf	17	0.36237	31.8975	0.02914	117.03
adf	18	0.33433	36.2728	0.04094	142.59
adf	19	0.35499	43.1556	0.12375	151.65
adf	20	0.40276	26.8335	0.00000	138.29
adf	21	0.66865	30.1703	0.03565	150.24
adf	22	0.42813	35.0944	0.08375	149.58
adf	23	0.34785	47.0004	0.06660	138.26
adf	24	0.26664	66.7639	0.08379	119.51
adf	25	0.39524	29.7257	0.03933	114.47
adf	26	0.35102	28.1715	0.05180	97.53
adf	27	0.33358	27.2725	0.04087	79.07
adf	28	0.34647	21.6274	0.02148	62.00

Variable	Matrix	Method	Minimum	Maximum	Mean
THETA	B1	adf	0.1876556	0.7370820	0.3941858
THETA	B2	adf	1.044049E-53	0.1468939	0.0526091
CHISQ	B1	adf	18.0868747	66.7639063	33.7809952
CHISQ	B2	adf	62.0015866	163.7294624	116.8838771

APPENDIX 6.5. INCREMENTAL FIT INDICES FOR MODELS 1-3.

SITE	METHOD	CHI0	CHI1	CHI2	CHI3	R01	R12	R23
1	adf	22.2910	19.7298	19.2282	19.2280	0.11490	0.02250	0.00001
2	adf	39.1305	32.4622	28.0504	26.7373	0.17041	0.11275	0.03356
3	adf	25.7363	22.0897	13.0564	13.0506	0.14169	0.35100	0.00022
4	adf	41.6117	41.1833	41.1833	41.1833	0.01029	0.00000	0.00000
5	adf	43.5704	27.6042	19.3923	19.3923	0.36645	0.18847	0.00000
6	adf	44.9195	43.3625	27.5473	27.5473	0.03466	0.35208	0.00000
7	adf	28.1782	19.0775	13.3139	13.1209	0.32297	0.20454	0.00685
8	adf	37.8547	31.6930	29.8102	29.8102	0.16277	0.04974	0.00000
9	adf	34.4503	30.5279	21.2489	21.2489	0.11386	0.26934	0.00000
10	adf	46.4624	31.7313	21.8643	21.8643	0.31705	0.21237	0.00000
11	adf	28.3649	18.6846	15.3846	15.3846	0.34128	0.11634	0.00000
12	adf	27.1089	27.1089	27.1089	23.7608	0.00000	0.00000	0.12351
13	adf	19.4085	10.3278	10.2483	10.2483	0.46787	0.00409	0.00000
14	adf	41.3106	31.9447	31.7577	31.1057	0.22672	0.00453	0.01578
15	adf	18.0869	17.3332	17.3327	17.3314	0.04167	0.00003	0.00007
16	adf	23.3975	23.3975	23.3975	23.3975	0.00000	0.00000	0.00000
17	adf	31.8975	22.5639	10.9665	10.9665	0.29261	0.36358	0.00000
18	adf	36.2728	27.1215	23.4602	22.3509	0.25229	0.10094	0.03058
19	adf	43.1556	41.3090	28.2458	28.2458	0.04279	0.30270	0.00000
20	adf	26.8335	21.7852	17.6210	16.6269	0.18813	0.15519	0.03705
21	adf	30.1703	24.3588	24.3181	24.1148	0.19263	0.00135	0.00674
22	adf	35.0944	22.9009	19.9129	18.9425	0.34745	0.08514	0.02765
23	adf	47.0004	34.1740	24.1688	24.1688	0.27290	0.21287	0.00000
24	adf	66.7639	48.6747	32.5448	32.5448	0.27094	0.24160	0.00000
25	adf	29.7257	27.9697	26.3715	26.3715	0.05907	0.05376	0.00000
26	adf	28.1715	14.6537	14.4772	14.4468	0.47984	0.00627	0.00108
27	adf	27.2725	16.3992	13.4239	13.4239	0.39869	0.10909	0.00000
28	adf	21.6274	21.6068	21.6068	21.6068	0.00095	0.00000	0.00000
1	nk	27.3150	18.1450	16.5177	16.5041	0.33571	0.05958	0.00050
2	nk	46.6420	23.6708	21.5683	19.9641	0.49250	0.04508	0.03440
3	nk	38.7161	21.4660	15.9048	15.9048	0.44555	0.14364	0.00000
4	nk	40.3266	38.3423	38.3423	38.3423	0.04921	0.00000	0.00000
5	nk	59.9482	43.8149	36.5947	36.2352	0.26912	0.12044	0.00600
6	nk	67.1022	51.7203	38.2332	38.2332	0.22923	0.20099	0.00000
7	nk	65.3814	45.2322	35.5888	35.5888	0.30818	0.14749	0.00000
8	nk	80.8907	65.9509	42.8879	42.8879	0.18469	0.28511	0.00000
9	nk	60.7115	46.7730	32.1678	32.1678	0.22959	0.24057	0.00000
10	nk	85.1566	56.9228	30.8890	30.8890	0.33155	0.30572	0.00000
11	nk	44.8782	25.3894	20.6480	20.5775	0.43426	0.10565	0.00157
12	nk	38.5708	35.0191	33.6933	33.4607	0.09208	0.03437	0.00603
13	nk	36.2685	21.5616	21.3767	21.3756	0.40550	0.00510	0.00003
14	nk	34.2537	23.9305	23.5271	23.1656	0.30137	0.01178	0.01055
15	nk	43.2132	25.9910	21.1699	20.9026	0.39854	0.11156	0.00619
16	nk	29.9096	29.2324	29.2324	29.2324	0.02264	0.00000	0.00000
17	nk	49.8679	24.1323	12.1352	12.1352	0.51607	0.24058	0.00000
18	nk	47.6919	29.3471	21.5059	20.4712	0.38465	0.16441	0.02169
19	nk	74.7694	57.3903	39.2618	39.2618	0.23244	0.24246	0.00000
20	nk	33.7896	12.4326	11.5560	10.8946	0.63206	0.02594	0.01957
21	nk	34.5803	22.3689	22.3689	22.3664	0.35313	0.00000	0.00007
22	nk	52.1401	31.6753	25.3103	23.6271	0.39250	0.12208	0.03228
23	nk	57.3071	35.7879	22.4290	22.4290	0.37551	0.23311	0.00000
24	nk	91.7826	63.3936	35.5664	35.5664	0.30931	0.30319	0.00000
25	nk	62.3181	45.3407	34.3083	34.3083	0.27243	0.17703	0.00000
26	nk	33.2429	19.4860	17.2754	17.2754	0.41383	0.06650	0.00000
27	nk	34.1869	25.3620	21.3255	21.3255	0.25814	0.11807	0.00000
28	nk	40.5769	37.2938	37.2938	37.2938	0.08091	0.00000	0.00000

	METHOD: GLS (adf)			METHOD: GLS (nk)		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
CHI1	10.3277805	48.6747197	26.8491311	12.4325634	65.9508942	34.8990284
CHI2	10.2483469	41.1833121	22.0372345	11.5559908	42.8879129	27.0956566
CHI3	10.2483469	41.1833121	21.7221955	10.8946150	42.8879129	26.8709392
CHI0	18.0868747	66.7639063	33.7809952	27.3150370	91.7826312	50.4120688
R01	0	0.4798375	0.2011033	0.0226419	0.6320598	0.3125251
R12	0	0.3635823	0.1257239	0	0.3057175	0.1253734
R23	0	0.1235063	0.0101108	0	0.0343956	0.0049601

APPENDIX 6.6. CROSS-VALIDATION COEFFICIENTS FOR MODELS 1-3.

SITE	METHOD	C1	C2	C3	C1-2	C1-3	C2-3
1	adf	0.31570	0.33532	0.38795	-0.01963	-0.07226	-0.052629
2	adf	0.56876	0.53059	0.57162	0.03817	-0.00286	-0.041022
3	adf	0.37728	0.27494	0.33200	0.10235	0.04529	-0.057060
4	adf	0.60986	0.63653	0.68986	-0.02667	-0.08000	-0.053333
5	adf	0.38020	0.30507	0.35269	0.07513	0.02751	-0.047619
6	adf	0.59889	0.42370	0.47370	0.17519	0.12519	-0.050000
7	adf	0.29149	0.24353	0.29109	0.04796	0.00040	-0.047557
8	adf	0.46896	0.47016	0.52211	-0.00120	-0.05315	-0.051948
9	adf	0.45363	0.35751	0.40946	0.09612	0.04417	-0.051948
10	adf	0.44053	0.34310	0.39188	0.09742	0.04864	-0.048780
11	adf	0.31872	0.30002	0.35557	0.01870	-0.03685	-0.055556
12	adf	0.43737	0.46515	0.47355	-0.02778	-0.03618	-0.008399
13	adf	0.19553	0.22147	0.27552	-0.02594	-0.07999	-0.054054
14	adf	0.49847	0.52327	0.56901	-0.02480	-0.07054	-0.045738
15	adf	0.36045	0.39377	0.46042	-0.03332	-0.09997	-0.066644
16	adf	0.52542	0.56316	0.63863	-0.03774	-0.11321	-0.075472
17	adf	0.34884	0.22222	0.27417	0.12662	0.07467	-0.051948
18	adf	0.38840	0.36733	0.40284	0.02108	-0.01444	-0.035517
19	adf	0.54532	0.41174	0.45936	0.13358	0.08596	-0.047619
20	adf	0.37287	0.34109	0.38383	0.03178	-0.01096	-0.042734
21	adf	0.37742	0.40319	0.45311	-0.02577	-0.07569	-0.049920
22	adf	0.36280	0.34909	0.38931	0.01371	-0.02651	-0.040221
23	adf	0.52943	0.41787	0.47266	0.11156	0.05677	-0.054795
24	adf	0.64179	0.46918	0.51737	0.17261	0.12442	-0.048193
25	adf	0.38460	0.38916	0.43678	-0.00456	-0.05217	-0.047619
26	adf	0.26195	0.28724	0.34236	-0.02529	-0.08042	-0.055127
27	adf	0.40642	0.38613	0.46456	0.02029	-0.05814	-0.078431
28	adf	0.41199	0.44373	0.50723	-0.03175	-0.09524	-0.063492
1	nk	0.29457	0.29918	0.35163	-0.00462	-0.05707	-0.052450
2	nk	0.43140	0.42931	0.46578	0.00208	-0.03439	-0.036472
3	nk	0.36824	0.31622	0.37336	0.05203	-0.00512	-0.057143
4	nk	0.57147	0.59814	0.65147	-0.02667	-0.08000	-0.053333
5	nk	0.57551	0.51233	0.55562	0.06318	0.01989	-0.043287
6	nk	0.70469	0.55897	0.60897	0.14572	0.09572	-0.050000
7	nk	0.62256	0.52549	0.57549	0.09707	0.04707	-0.050000
8	nk	0.91972	0.64224	0.69418	0.27749	0.22554	-0.051948
9	nk	0.66738	0.50118	0.55313	0.16620	0.11425	-0.051948
10	nk	0.75153	0.45452	0.50330	0.29702	0.24823	-0.048780
11	nk	0.41315	0.37415	0.42871	0.03900	-0.01556	-0.054563
12	nk	0.54878	0.55789	0.61017	-0.00911	-0.06138	-0.052279
13	nk	0.34942	0.37391	0.42795	-0.02449	-0.07853	-0.054039
14	nk	0.38716	0.40896	0.45873	-0.02179	-0.07157	-0.049774
15	nk	0.50719	0.45881	0.52095	0.04838	-0.01376	-0.062136
16	nk	0.63763	0.67537	0.75084	-0.03774	-0.11321	-0.075472
17	nk	0.36948	0.23760	0.28954	0.13188	0.07993	-0.051948
18	nk	0.41622	0.34290	0.37935	0.07332	0.03687	-0.036449
19	nk	0.73907	0.54446	0.59208	0.19461	0.14699	-0.047619
20	nk	0.23732	0.25319	0.30075	-0.01587	-0.06343	-0.047558
21	nk	0.35088	0.37720	0.42980	-0.02632	-0.07891	-0.052598
22	nk	0.48138	0.42203	0.45262	0.05935	0.02876	-0.030588
23	nk	0.55185	0.39371	0.44850	0.15814	0.10335	-0.054795
24	nk	0.82129	0.50603	0.55422	0.31526	0.26707	-0.048193
25	nk	0.59389	0.48478	0.53240	0.10911	0.06149	-0.047619
26	nk	0.33001	0.32665	0.38220	0.00336	-0.05220	-0.055556
27	nk	0.58567	0.54416	0.62259	0.04151	-0.03692	-0.078431
28	nk	0.66500	0.69675	0.76024	-0.03175	-0.09524	-0.063492

	METHOD: GLS (adf)			METHOD: GLS (nk)		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
C1	0.1955305	0.6417869	0.4240390	0.2373249	0.9197230	0.5318741
C2	0.2214694	0.6365312	0.3884023	0.2375962	0.6967506	0.4577182
C3	0.2741662	0.6898646	0.4392371	0.2895442	0.7602427	0.5098064
C1-2	-0.0377358	0.1751917	0.0356367	-0.0377358	0.3152595	0.0741559
C1-3	-0.1132075	0.1251917	-0.0151981	-0.1132075	0.2670667	0.0220677
C2-3	-0.0784314	-0.0083989	-0.0508348	-0.0784314	-0.0305883	-0.0520882

APPENDIX 6.7. PARAMETER ESTIMATES FOR MODELS 1, 2 AND 3.

SITE	INIT	METH	MOD	ITER	STATUS	T 1	T 2	T 3
1	0	nk	1	2	OK	0.77083	0.05089	.
2	0	nk	1	2	OK	0.42186	0.11761	.
3	0	nk	1	2	OK	0.42799	0.06775	.
4	0	nk	1	2	OK	0.41655	0.01556	.
5	0	nk	1	2	OK	0.36432	0.06612	.
6	0	nk	1	2	OK	0.34204	0.06730	.
7	0	nk	1	2	OK	0.33362	0.10430	.
8	0	nk	1	2	OK	0.23357	0.07124	.
9	0	nk	1	2	OK	0.33807	0.06579	.
10	0	nk	1	2	OK	0.30759	0.10497	.
11	0	nk	1	2	OK	0.43921	0.08929	.
12	0	nk	1	2	OK	0.38954	0.01637	.
13	0	nk	1	2	OK	0.59603	0.07384	.
14	0	nk	1	2	OK	0.51949	0.04821	.
15	0	nk	1	2	OK	0.43798	0.12636	.
16	0	nk	1	2	OK	0.46508	0.00963	.
17	0	nk	1	2	OK	0.34882	0.07951	.
18	0	nk	1	2	OK	0.37683	0.06638	.
19	0	nk	1	2	OK	0.35427	0.08636	.
20	0	nk	1	2	OK	0.48848	0.10601	.
21	0	nk	1	2	OK	0.71214	0.06776	.
22	0	nk	1	2	OK	0.45508	0.09473	.
23	0	nk	1	2	OK	0.39378	0.10805	.
24	0	nk	1	2	OK	0.26700	0.14872	.
25	0	nk	1	2	OK	0.32958	0.06737	.
26	0	nk	1	2	OK	0.32292	0.03245	.
27	0	nk	1	2	OK	0.34799	0.03491	.
28	0	nk	1	2	OK	0.38930	0.02114	.
1	0	nk	2	15	OK	0.65419	0.00000	0.28100
2	0	nk	2	2	OK	0.31536	0.06832	0.26528
3	0	nk	2	13	ER	0.30521	0.00000	0.35650
4	0	nk	2	10	OK	0.41655	0.01556	0.00000
5	0	nk	2	12	OK	0.24647	0.00000	0.36224
6	0	nk	2	10	OK	0.21261	0.00000	0.42467
7	0	nk	2	2	OK	0.20558	0.03709	0.38407
8	0	nk	2	10	OK	0.11422	0.00000	0.45801
9	0	nk	2	10	OK	0.20252	0.00000	0.44565
10	0	nk	2	10	OK	0.14150	0.00000	0.65979
11	0	nk	2	2	OK	0.31427	0.01542	0.37160
12	0	nk	2	13	OK	0.34985	0.00000	0.09874
13	0	nk	2	2	OK	0.55805	0.05711	0.09105
14	0	nk	2	2	OK	0.47482	0.02954	0.10387
15	0	nk	2	2	OK	0.29428	0.04093	0.43837
16	0	nk	2	11	OK	0.46508	0.00963	0.00000
17	0	nk	2	13	OK	0.21168	0.00000	0.43305
18	0	nk	2	13	OK	0.23591	0.00000	0.38342
19	0	nk	2	10	OK	0.20529	0.00000	0.53423
20	0	nk	2	2	OK	0.42436	0.07716	0.15243
21	0	nk	2	11	OK	0.71214	0.06776	0.00000
22	0	nk	2	2	OK	0.29043	0.00617	0.48091
23	0	nk	2	2	OK	0.21875	0.00110	0.56463
24	0	nk	2	11	OK	0.08671	0.00000	0.77422
25	0	nk	2	12	ER	0.22229	0.00000	0.35925
26	0	nk	2	2	OK	0.26242	0.00251	0.15831
27	0	nk	2	12	OK	0.27273	0.00000	0.21457
28	0	nk	2	10	OK	0.38930	0.02114	0.00000

SITE	INIT	METH	MOD	ITER	STATUS	T1	T2	T3	T4	T5	RATE
1	7	nk	3	500	ER	0.69070	0.00000	0.28476	437212.51	0.00001	3.377
2	10	nk	3	500	ER	0.37221	0.00000	0.84075	101268.32	0.00001	0.617
3	8	nk	3	320	ER	0.30712	0.00000	0.35650	127.29	0.73085	94.490
4	7	nk	3	48	ER	0.41655	0.01556	0.00019	20182.41	0.00000	0.000
5	8	nk	3	500	ER	0.31010	0.00000	0.41172	326144.86	0.00001	1.726
6	8	nk	3	244	ER	0.21488	0.00000	0.42467	127.93	0.72950	94.787
7	7	nk	3	246	ER	0.20765	0.03709	0.38407	127.28	0.72788	94.102
8	8	nk	3	247	ER	0.11669	0.00000	0.45801	125.38	0.73918	94.154
9	8	nk	3	259	ER	0.20492	0.00000	0.44565	133.45	0.69549	94.206
10	7	nk	3	334	ER	0.14493	0.00000	0.65979	128.93	0.74493	97.530
11	7	nk	3	27	ER	0.36650	0.00000	0.53036	0.31	1.75547	4.061
12	2	nk	3	13	ER	0.36943	0.00000	0.13107	0.00	1.76631	3.533
13	3	nk	3	3	OK	0.56125	0.05706	0.09138	3.76	3.58745	20.680
14	10	nk	3	20	OK	0.49650	0.00002	0.40751	32.69	0.01384	0.480
15	8	nk	3	500	ER	0.34855	0.03054	0.50718	838755.19	0.00000	2.672
16	10	nk	3	30	ER	0.46508	0.00963	0.00000	14.64	0.00000	0.000
17	8	nk	3	328	ER	0.21391	0.00000	0.43305	138.51	0.70300	98.778
18	8	nk	3	500	ER	0.31757	0.00000	0.43923	656556.30	0.00000	1.436
19	8	nk	3	311	ER	0.20806	0.00000	0.53423	130.34	0.74069	98.025
20	9	nk	3	500	ER	0.45497	0.00948	0.81462	6597.29	0.00006	0.415
21	7	nk	3	10	OK	0.70832	0.05793	0.10931	8.84	0.02942	0.319
22	8	nk	3	500	ER	0.38441	0.00000	0.59351	821390.37	0.00000	1.480
23	5	nk	3	355	ER	0.22033	0.00110	0.56463	21.26	8.36868	194.636
24	9	nk	3	379	ER	0.09120	0.00000	0.77422	104.68	0.82360	87.858
25	5	nk	3	147	ER	0.22327	0.00000	0.35925	21.05	8.68683	200.270
26	5	nk	3	201	ER	0.26322	0.00251	0.15831	68.26	1.44819	101.748
27	5	nk	3	180	ER	0.27331	0.00000	0.21457	19.39	9.60649	205.491
28	10	nk	3	36	ER	0.38930	0.02114	0.00000	1.23	0.00000	0.000

Variable	Model	Minimum	Maximum	Mean
T1	1	0.2335743	0.7708282	0.4139272
T2	1	0.0096288	0.1487155	0.0717366
T1	2	0.0867112	0.7121430	0.3143772
T2	2	1.365589E-33	0.0771631	0.0160514
T3	2	1.86583E-28	0.7742242	0.3141374
T1	3	0.0912014	0.7083244	0.3336050
T2	3	0	0.0579317	0.0086447
T3	3	9.844385E-11	0.8407529	0.3903048
T4	3	3.7895282E-9	838755.19	114622.95
T5	3	5.944275E-10	9.6064864	1.4963534
RATE	3	1.8038888E-7	205.4914764	57.0311093

SITE	INIT	METH	MOD	ITER	STATUS	T1	T2	T3
1	0	adf	1	2	OK	0.68477	0.02706	.
2	0	adf	1	2	OK	0.37628	0.07402	.
3	0	adf	1	2	OK	0.43556	0.04861	.
4	0	adf	1	2	OK	0.37955	0.01135	.
5	0	adf	1	2	OK	0.34361	0.05033	.
6	0	adf	1	2	OK	0.31383	0.03404	.
7	0	adf	1	2	OK	0.38959	0.08046	.
8	0	adf	1	2	OK	0.18620	0.03272	.
9	0	adf	1	2	OK	0.34069	0.03999	.
10	0	adf	1	2	OK	0.25234	0.06233	.
11	0	adf	1	2	OK	0.45129	0.05132	.
12	0	adf	1	11	ER	0.41156	0.00000	.
13	0	adf	1	2	OK	0.49911	0.03911	.
14	0	adf	1	2	OK	0.42578	0.04561	.
15	0	adf	1	2	OK	0.40075	0.03428	.
16	0	adf	1	10	OK	0.44063	0.00000	.
17	0	adf	1	2	OK	0.37832	0.05083	.
18	0	adf	1	2	OK	0.33627	0.04415	.
19	0	adf	1	2	OK	0.33777	0.03058	.
20	0	adf	1	2	OK	0.43404	0.06063	.
21	0	adf	1	2	OK	0.66677	0.03359	.
22	0	adf	1	2	OK	0.41163	0.05913	.
23	0	adf	1	2	OK	0.36494	0.10744	.
24	0	adf	1	2	OK	0.27284	0.09196	.
25	0	adf	1	2	OK	0.38964	0.02303	.
26	0	adf	1	2	OK	0.31735	0.03139	.
27	0	adf	1	2	OK	0.31599	0.03094	.
28	0	adf	1	2	OK	0.34346	0.00118	.
1	0	adf	2	15	OK	0.61185	0.00000	0.161
2	0	adf	2	2	OK	0.27005	0.01447	0.311
3	0	adf	2	9	OK	0.29064	0.00000	0.360
4	0	adf	2	10	ER	0.37955	0.01135	0.000
5	0	adf	2	11	OK	0.27185	0.00000	0.284
6	0	adf	2	8	OK	0.24140	0.00000	0.436
7	0	adf	2	2	OK	0.24159	0.01121	0.375
8	0	adf	2	2	OK	0.14367	0.00534	0.181
9	0	adf	2	10	OK	0.21891	0.00000	0.353
10	0	adf	2	11	OK	0.16374	0.00000	0.410
11	0	adf	2	2	OK	0.33121	0.00483	0.306
12	0	adf	2	11	OK	0.41156	0.00000	0.000
13	0	adf	2	2	OK	0.47709	0.02951	0.048
14	0	adf	2	2	OK	0.39131	0.03163	0.075
15	0	adf	2	2	OK	0.39965	0.03348	0.004
16	0	adf	2	11	OK	0.44063	0.00000	0.000
17	0	adf	2	10	OK	0.23088	0.00000	0.445
18	0	adf	2	2	OK	0.23935	0.00352	0.261
19	0	adf	2	10	OK	0.21165	0.00000	0.474
20	0	adf	2	11	OK	0.30538	0.00000	0.322
21	0	adf	2	2	OK	0.64844	0.02745	0.035
22	0	adf	2	12	OK	0.31492	0.00000	0.304
23	0	adf	2	2	OK	0.21839	0.02993	0.479
24	0	adf	2	12	OK	0.13498	0.00000	0.584
25	0	adf	2	13	OK	0.33884	0.00000	0.137
26	0	adf	2	2	OK	0.29482	0.02237	0.051
27	0	adf	2	13	OK	0.24340	0.00000	0.181
28	0	adf	2	12	OK	0.34346	0.00118	0.000

SITE	INIT	METH	MOD	ITER	STATUS	T1	T2	T3	T4	T5	RATE
1	3	adf	3	8	OK	0.61549	0.00000	0.161	4.89	4.3969	30.279
2	8	adf	3	500	ER	0.33355	0.00000	0.485	133.32	0.0084	1.135
3	7	adf	3	500	ER	0.33718	0.00000	0.368	145.98	0.0237	3.504
4	10	adf	3	24	ER	0.37955	0.01135	0.000	5.50	0.0474	0.355
5	5	adf	3	176	ER	0.27224	0.00000	0.284	11.10	33.3282	436.619
6	5	adf	3	223	ER	0.24232	0.00000	0.436	16.06	14.6815	265.176
7	4	adf	3	6	OK	0.30492	0.00000	0.525	0.65	1.1712	3.105
8	5	adf	3	194	ER	0.14416	0.00534	0.180	22.64	8.1590	201.056
9	5	adf	3	136	ER	0.21920	0.00000	0.353	8.82	68.5931	742.034
10	7	adf	3	267	ER	0.16598	0.00000	0.410	137.19	0.6688	93.088
11	5	adf	3	213	ER	0.33200	0.00483	0.306	20.67	9.3737	212.502
12	9	adf	3	31	ER	0.42808	0.00000	521.960	0.00	0.0000	0.000
13	5	adf	3	110	ER	0.47727	0.02951	0.048	29.98	4.4863	143.482
14	10	adf	3	500	ER	0.40718	0.00000	0.413	126.07	0.0031	0.397
15	10	adf	3	20	ER	0.40080	0.03432	17.560	0.00	0.0001	0.000
16	10	adf	3	21	ER	0.44063	0.00000	0.000	16.19	0.0000	0.000
17	5	adf	3	203	ER	0.23129	0.00000	0.445	10.16	53.3888	649.398
18	7	adf	3	500	ER	0.30646	0.00000	0.340	134.92	0.0083	1.138
19	5	adf	3	211	ER	0.21212	0.00000	0.474	10.03	50.3875	606.012
20	8	adf	3	500	ER	0.38557	0.00000	0.394	136.13	0.0099	1.373
21	9	adf	3	4	OK	0.63387	0.00296	0.209	9.29	0.0970	1.095
22	9	adf	3	500	ER	0.37639	0.00000	0.369	130.19	0.0105	1.389
23	5	adf	3	283	ER	0.21971	0.02993	0.479	20.91	8.7157	199.707
24	8	adf	3	283	ER	0.13815	0.00000	0.584	119.66	0.7696	93.635
25	5	adf	3	154	ER	0.33938	0.00000	0.137	36.96	3.4299	133.629
26	7	adf	3	500	ER	0.30169	0.01981	0.069	179920.34	0.0000	1.799
27	5	adf	3	155	ER	0.24385	0.00000	0.181	16.73	12.2071	228.629
28	2	adf	3	20	ER	0.34346	0.00118	0.132	0.01	0.0000	0.000

Variable	Model	Minimum	Maximum	Mean
T1	1	0.1861963	0.6847705	0.3893061
T2	1	2.317071E-34	0.1074371	0.0427174
T1	2	0.1349816	0.6484418	0.3146147
T2	2	1.284429E-30	0.0334827	0.0080813
T3	2	3.36887E-30	0.5842294	0.2348848
T1	3	0.1381500	0.6338700	0.3297318
T2	3	0	0.0343200	0.0049725
T3	3	0	521.9601700	19.5464711
T4	3	0	179920.34	6472.30
T5	3	0	68.5930800	9.7844932
RATE	3	0	742.0337661	144.6620523

7

THE ASYMPTOTIC VARIANCE OF $\hat{\theta}_3$ AS A CRITERION FOR STUDY DESIGN.

In previous sections, we have developed models for periodontal disease progression and have investigated the goodness of fit of these models for real and simulated data. We now examine how an asymptotic property of these models might guide us in the design of studies of periodontal disease progression. Because this involves a substantial amount of algebraic manipulation, it was necessary to make more restrictive assumptions than have been made in previous sections. We consider only the conditions under which model 2 may be distinguished from model 1, and examine the properties of maximum-likelihood, rather than generalised least-squares, estimates. We also assume that time intervals are constant ($\underline{\tau} = t\underline{1}$, with t known). Thus we have (using the notation of section 3), that:

(7.1) $B_2 = t^2\underline{1}\underline{1}'$, where $\underline{1}$ is the unit p-vector and

(7.2) $B_3=tI$, where I is the identity matrix of dimension p.

As has been seen, model 2 is distinguished from model 1 according to whether θ_3 is greater than zero. In order to assess whether this is the case, a measure of the precision of $\hat{\theta}_3$ (the estimate of θ_3) is needed. A criterion for the design of a study to determine which of models 1 and 2 better describes the nature of periodontal disease progression might therefore be to minimise the variance of $\hat{\theta}_3$ using the factors over which we have control; namely n (the number of sites), p (the number of observations per site), and t (the time between observations).

The null hypothesis that model 1 holds will be rejected if $\hat{\theta}_3$ differs from zero by an amount which can be considered to be unlikely to arise from random variation

in the data, given that model 1 holds. We thus minimise the asymptotic variance of $\dot{\theta}_3$ given that $\theta_3 = 0$ ($\text{var}(\dot{\theta}_3|\theta_3=0)$).

The information matrix $I(\underline{\theta})$ for a covariance structure $\Sigma(\underline{\theta})$ where the underlying distribution is normal was shown in section 2 to be given by

$$(7.3)\quad I(\underline{\theta})_{ij} = \tfrac{1}{2}\text{tr}\left(\Sigma_{(\underline{\theta})}^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma_{(\underline{\theta})}^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right).$$

The asymptotic variance-covariance matrix of $\dot{\theta}$ is given by the inverse of $NI(\underline{\theta})$. Under model 2, we have that:

$$(7.4)\quad \Sigma_{(\underline{\theta})} = \theta_1B_1 + \theta_2B_2 + \theta_3B_3. \text{ Thus}$$

$$(7.5)\quad \frac{\partial\Sigma_{(\underline{\theta})}}{\partial\theta_i} = B_i \ (i=1,2).$$

The elements of $I(\underline{\theta})$ under model 1 (for which $\theta_3 = 0$) are therefore given by:

$$(7.6)\quad I(\underline{\theta})_{ij} = \tfrac{1}{2}\text{tr}[(\theta_1B_1+\theta_2B_2)^{-1}B_i(\theta_1B_1+\theta_2B_2)^{-1}B_j] \ (i,j = 1,2,3)$$

7.1 CALCULATION OF THE INVERSE OF $\Sigma_{(\underline{\theta})}$ UNDER MODEL 1.

Lemma 7.1: The inverse of B_1 as defined in section 3 is given by:

$$(7.7)\quad (B_1^{-1})_{ij} = \begin{cases} (p+1)^{-1}i(p-j+1) & (i\le j) \\ (p+1)^{-1}j(p-i+1) & (j\le i) \end{cases}$$

Proof: Recall that $b_{1ij} = \begin{cases} 2 & (i=j) \\ -1 & (|i-j|=1) \\ 0 & \text{otherwise} \end{cases}$

The product (P, say) of the two matrices is given by:

$$p_{ij} = \sum_{k=1}^{j} b_{1ik}(p+1)^{-1}k(p-j+1) + \sum_{k=j+1}^{p} b_{1ik}(p+1)^{-1}j(p-k+1)$$

$$= (p+1)^{-1}\left[\sum_{k=1}^{j} b_{1jk}k(p-j+1) + \sum_{k=j+1}^{p} b_{1jk}j(p-k+1)\right]$$

It is now straightforward to check, for the three cases i<j, i=j, and i>j, that the three terms involving non zero elements of B_1 sum to 0, 1 and 0 respectively. ∎

Lemma 7.2: $\Sigma_{(\theta)}^{-1} = (\theta_1^{-1}B_1^{-1} - t^2\theta_2\theta_1^{-2}(1+t^2\theta_2\theta_1^{-1}\sum_{i,j=1}^{p}(B_1^{-1})_{ij})^{-1}C$ where $C = B_1^{-1}\underline{1}\underline{1}'B_1^{-1}$

Proof: $\Sigma_{(\theta)}^{-1} = (\theta_1B_1 + \theta_2B_2)^{-1}$

$= (\theta_1B_1 + t^2\theta_2\underline{1}\underline{1}')^{-1}$

$= \theta_1^{-1}B_1^{-1} - t^2\theta_2\theta_1^{-2}(1+t^2\theta_2\theta_1^{-1}\underline{1}'B_1^{-1}\underline{1})^{-1}B_1^{-1}\underline{1}\underline{1}'B_1^{-1}$ (by a standard formula)

$= \theta_1^{-1}B_1^{-1} - t^2\theta_2\theta_1^{-2}(1+t^2\theta_2\theta_1^{-1}\sum_{i,j=1}^{p}(B_1^{-1})_{ij})^{-1}C$ as required. ∎

Note that $C_{ij} = (\sum_{k=1}^{p}(B_1^{-1})_{ik})(\sum_{k=1}^{p}(B_1^{-1})_{kj})$.

Now $\sum_{k=1}^{p}(B_1^{-1})_{ik} = \sum_{k=1}^{i}(p+1)^{-1}k(p-i+1) + \sum_{k=i+1}^{p}(p+1)^{-1}i(p-k+1)$

$= (p+1)^{-1}[(p-i+1)\sum_{k=1}^{i}k + i\sum_{l=1}^{p-i}l]$

$= (p+1)^{-1}[\frac{1}{2}(p-i+1)i(i+1)+\frac{1}{2}(p-i+1)i(p-i)]$

whence

(7.8) $\sum_{k=1}^{p}(B_1^{-1})_{ik} = \frac{1}{2}i(p-i+1)$

Thus, $\sum_{i,j=1}^{p}(B_1^{-1})_{ij} = \sum_{i=1}^{p}\frac{1}{2}i(p-i+1)$

$= \frac{1}{2}[(p+1)\frac{1}{2}p(p+1) - 6^{-1}p(p+1)(2p+1)]$

so that

(7.9) $\sum_{i,j=1}^{p}(B_1^{-1})_{ij} = 12^{-1}p(p+1)(p+2)$

So, using lemmas 7.1 and 7.2 and equations (7.8-9):

$\Sigma_{(\theta)}^{-1} = \theta_1^{-1}B_1^{-1} - t^2\theta_2\theta_1^{-2}(1+t^2\theta_2\theta_1^{-1}12^{-1}p(p+1)(p+2))^{-1}C$

(7.10) $\Sigma_{(\theta)}^{-1} = \theta_1^{-1}B_1^{-1} - 12\theta_2t^2[\theta_1(12\theta_1+\theta_2t^2p(p+1)(p+2)]^{-1}C$, where

(7.11) $C_{ij} = \frac{1}{4}i(p-i+1)j(p-j+1)$

Write

(7.12) $\phi = 12\theta_2 t^2[\theta_1(12\theta_1+\theta_2 t^2 p(p+1)(p+2))]^{-1}$; then we have proved

Theorem 7.1.

(7.13) $\Sigma_{(\underline{\theta})}^{-1} = \theta_1^{-1}B_1^{-1} - \phi C$

with B_1^{-1}, C and ϕ defined in (7.7), (7.11) and (7.12) respectively. ∎

7.2 CALCULATION OF THE ELEMENTS OF $I(\underline{\theta})$

From (7.6) and (7.13), the elements of $I(\underline{\theta})$ are given by

(7.14) $I(\underline{\theta})_{ij} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_i(\theta_1^{-1}B_1^{-1}-\phi C)B_j]$ (i,j = 1,2,3)

We now calculate the six non-duplicated elements of $I(\underline{\theta})$.

$I(\underline{\theta})_{1,1} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)B_1]$

$= \frac{1}{2}tr[(\theta_1^{-1}I - \phi CB_1)(\theta_1^{-1}I - \phi CB_1)]$

$= \frac{1}{2}tr[\theta_1^{-2}I - 2\theta_1^{-1}\phi CB_1 + \phi^2 CB_1CB_1]$

$= \frac{1}{2}tr[\theta_1^{-2}I - 2\theta_1^{-1}\phi B_1^{-1}\underline{1}\underline{1}' + \phi^2 B_1^{-1}\underline{1}\underline{1}'B_1^{-1}\underline{1}\underline{1}']$ (since $C=B_1^{-1}\underline{1}\underline{1}'B_1^{-1}$)

$= \frac{1}{2}tr(\theta_1^{-2}I) - \theta_1^{-1}\phi tr(\underline{1}'B_1^{-1}\underline{1}) + \frac{1}{2}\phi^2 tr(\underline{1}'B_1^{-1}\underline{1}\underline{1}'B_1^{-1}\underline{1})$

(7.15) $I(\underline{\theta})_{1,1} = \frac{1}{2}\theta_1^{-2}p - \theta_1^{-1}\phi\sum_{i,j=1}^{p}(B_1^{-1})_{ij} + \frac{1}{2}\phi^2[\sum_{i,j=1}^{p}(B_1^{-1})_{ij}]^2$

Using equation (7.9), we therefore have that

$I(\underline{\theta})_{1,1} = \frac{1}{2}\theta_1^{-2}p - \theta_1^{-1}[12^{-1}p(p+1)(p+2)] + \frac{1}{2}\phi^2[144^{-1}p^2(p+1)^2(p+2)^2]$

$= \frac{1}{2}\theta_1^{-2}p - \theta_1^{-2}\theta_2 t^2 p(p+1)(p+2)[12\theta_1+\theta_2 t^2 p(p+1)(p+2)]^{-1}$

$+ \frac{1}{2}\theta_2^2 t^4 p^2(p+1)^2(p+2)^2[12\theta_1+\theta_2 t^2 p(p+1)(p+2)]^{-2}$

$$= \tfrac{1}{2}\theta_1^{-2}p[12\theta_1+\theta_2t^2p(p+1)(p+2)]^{-2}[144\theta_1^2+24\theta_1\theta_2t^2p(p+1)(p+2)+\theta_2t^4p^2(p+1)^2(p+2)^2-$$
$$2\theta_2t^2(p+1)(p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))+\theta_2^2t^4p(p+1)^2(p+2)^2]$$

(7.16) $I(\underline{\theta})_{1,1} = \tfrac{1}{2}\theta_1^{-2}p[12\theta_1+\theta_2t^2p(p+1)(p+2)]^{-2}[144\theta_1^2+24(p-1)(p+1)(p+2)\theta_1\theta_2t^2$
$$+(p-1)p(p+1)^2(p+2)^2\theta_2t^4]$$

$I(\underline{\theta})_{1,2} = \tfrac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)B_2]$

$= \tfrac{1}{2}t^2tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)\underline{1}\underline{1}']$

$= \tfrac{1}{2}t^2\underline{1}'(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)\underline{1}$

$= \tfrac{1}{2}t^2\underline{1}'(\theta_1^{-2}B_1^{-1} - 2\phi\theta_1^{-1}C + \phi^2B_1^{-1}\underline{1}\underline{1}'B_1^{-1}\underline{1}\underline{1}'B_1^{-1})\underline{1}$

$= \tfrac{1}{2}t^2\underline{1}'(\theta_1^{-2}B_1^{-1} - 2\phi\theta_1^{-1}C + \phi^2(\underline{1}'B_1^{-1}\underline{1})(B_1^{-1}\underline{1}\underline{1}'B_1^{-1}))\underline{1}$

(7.17) $I(\underline{\theta})_{1,2} = \tfrac{1}{2}t^2\sum_{i,j=1}^{p}[\theta_1^{-1}(B_1^{-1})_{ij} + (\phi^2\sum_{k,m=1}^{p}(B_1^{-1})_{km}-2\phi\theta_1^{-1})C_{ij}]$

Lemma 7.3: $\sum_{i,j=1}^{p}C_{ij} = (144)^{-1}p^2(p+1)^2(p+2)^2$

Proof: $\sum_{i,j=1}^{p}C_{ij} = \sum_{i,j=1}^{p}\tfrac{1}{4}i(p-i+1)j(p-j+1)$

$= \tfrac{1}{4}[\sum_{i,j=1}^{p}(p+1)i - i^2]^2$

$= \tfrac{1}{4}(\tfrac{1}{2}p(p+1)^2 - 6^{-1}p(p+1)(2p+1))^2$

$= (144)^{-1}p^2(p+1)^2(p+2)^2$ as required. ∎

Using equation (7.9) and lemma 7.3, we have

$I(\underline{\theta})_{1,2} = \tfrac{1}{2}t^2[\theta_1^{-2}12^{-1}p(p+1)(p+2)+(\phi^2 12^{-1}p(p+1)(p+2)-2\phi\theta_1^{-1})(144)^{-1}p^2(p+1)^2(p+2)^2]$

$= 24^{-1}t^2p(p+1)(p+2)[\theta_1^{-2}-2\phi\theta_1^{-1}12^{-1}p(p+1)(p+2)+\phi^2 144^{-1}p^2(p+1)^2(p+2)^2]$

$= 24^{-1}\theta_1^{-2}t^2p(p+1)(p+2)[(12\theta_1+\theta_2t^2p(p+1)(p+2))]^{-2}[144\theta_1^2+24\theta_1\theta_2t^2p(p+1)(p+2)$

$+\theta_2^2t^4p^2(p+1)^2(p+2)^2-24\theta_1\theta_2t^2p(p+1)(p+2)-2\theta_2^2t^4p^2(p+1)^2(p+2)^2+2\theta_2^2t^4p^2(p+1)^2(p+2)^2]$

(7.18) $I(\underline{\theta})_{1,2} = 6\theta_1^{-2}t^2p(p+1)(p+2)[(12\theta_1+\theta_2t^2p(p+1)(p+2))]^{-2}$

$I(\underline{\theta})_{1,3} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)B_3]$

$= \frac{1}{2}t\ tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_1(\theta_1^{-1}B_1^{-1}-\phi C)I]$

$= \frac{1}{2}t\ tr[\theta_1^{-2}B_1^{-1} - 2\phi\theta_1^{-1}C + \phi^2B_1^{-1}C]$

$= \frac{1}{2}t\ tr[\theta_1^{-2}B_1^{-1} - 2\phi\theta_1^{-1}C + \phi^2B_1^{-1}\underline{1}\underline{1}'B_1^{-1}\underline{1}\underline{1}'B_1^{-1}]$

$= \frac{1}{2}t\ [tr(\theta_1^{-2}B_1^{-1}) - 2\phi\theta_1^{-1}tr(C) + \phi^2(\underline{1}B_1^{-1}\underline{1})tr(C)]$

(7.19) $I(\underline{\theta})_{1,3} = \frac{1}{2}t\ [tr(\theta_1^{-2}B_1^{-1}) - [2\phi\theta_1^{-1} + \phi^2(\sum_{i,j=1}^{p}(B_1^{-1})_{ij})]tr(C)]$

Lemma 7.4: $tr(B_1) = 6^{-1}p(p+2)$

Proof: $tr(B_1) = \sum_{j=1}^{p}(p+1)^{-1}j(p-j+1)$

$= 6^{-1}(p+1)^{-1}[3p(p+1)^2 - p(p+1)(2p+1)$

$= 6^{-1}p(p+2)$ as required. ∎

Lemma 7.5: $tr(C) = 120^{-1}p(p+1)(p+2)(p^2+2p+2)$

Proof: $tr(C) = \sum_{j=1}^{p}\frac{1}{4}j^2(p-j+1)^2$

$= \frac{1}{4}\sum_{j=1}^{p}(j^4 - 2(p+1)j^3 + (p+1)^2j^2$

$= \frac{1}{4}[30^{-1}p(p+1)(2p+1)(3p^2+3p-1) - 2(p+1)4^{-1}p^2(p+1)^2 + (p+1)^26^{-1}p(p+1)(2p+1)]$

$= \frac{1}{4}30^{-1}p(p+1)[p^3 + 4p^2 + 6p + 4]$

$= 120^{-1}p(p+1)(p+2)(p^2+2p+2)$ as required ∎

Using (7.19), (7.9), lemma 7.4 and lemma 7.5, we thus have

$I(\underline{\theta})_{1,3} = \frac{1}{2}t[\theta_1^{-2}6^{-1}p(p+2) - [2\phi\theta_1^{-1} + \phi^2(12^{-1}p(p+1)(p+2))]120^{-1}p(p+1)(p+2)(p^2+2p+2)]$

$= 60^{-1}tp(p+2)[5\theta_1^{-2} - \frac{1}{2}\phi\theta_1^{-1}(p+1)(p^2+2p+2) + \phi^2 48^{-1}p(p+1)^2(p+2)(p^2+2p+2)]$

$= 60^{-1}tp(p+2)\theta_1^{-2}(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[5(12\theta_1+\theta_2t^2p(p+1)(p+2))^2$

$- 6\theta_2t^2(p+1)(p^2+2p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2)) + 3\theta_2^2t^4p(p+1)^2(p+2)(p^2+2p+2)]$

$= 60^{-1}tp(p+2)\theta_1^{-2}(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[720\theta_1^2+\theta_1\theta_2t^2(p+1)(48p^2+96p-144)$

$+\theta_2^2t^4p(p+1)^2(p+2)(2p^2+4p-6)]$

(7.20) $I(\underline{\theta})_{1,3} = 30^{-1}tp(p+2)\theta_1^{-2}(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[360\theta_1^2$

$+24\theta_1\theta_2t^2(p-1)(p+1)(p+3)+\theta_2^2t^4(p-1)p(p+1)^2(p+2)(p+3)]$

$I(\underline{\theta})_{2,2} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_2(\theta_1^{-1}B_1^{-1}-\phi C)B_2]$

$= \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)t^2\underline{1}\underline{1}'(\theta_1^{-1}B_1^{-1}-\phi C)t^2\underline{1}\underline{1}']$

$= \frac{1}{2}t^4[\underline{1}'(\theta_1^{-1}B_1^{-1}-\phi C)\underline{1}]^2$

(7.21) $I(\underline{\theta})_{2,2} = \frac{1}{2}t^4[\sum_{i,j=1}^{p}(\theta_1^{-1}(B_1^{-1})_{ij}-\phi C_{ij})]^2$

Using (7.9) and lemma 7.3, we have

$I(\underline{\theta})_{2,2} = \frac{1}{2}t^4[\theta_1^{-1}12^{-1}p(p+1)(p+2)-\phi(144)^{-1}p^2(p+1)^2(p+2)^2]^2$

$= 288^{-1}\theta_1^{-2}t^4p^2(p+1)^2(p+2)^2(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[12\theta_1$

$+\theta_2t^2p(p+1)(p+2)-\theta_2t^2p(p+1)(p+2)]^2$

(7.22) $I(\underline{\theta})_{2,2} = \frac{1}{2}t^4p^2(p+1)^2(p+2)^2(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}$

$I(\underline{\theta})_{2,3} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_2(\theta_1^{-1}B_1^{-1}-\phi C)B_3]$

$= \frac{1}{2}t^3tr[(\theta_1^{-1}B_1^{-1}-\phi C)\underline{1}\underline{1}'(\theta_1^{-1}B_1^{-1}-\phi C)I]$

$= \frac{1}{2}t^3\underline{1}'(\theta_1^{-1}B_1^{-1}-\phi C)^2\underline{1}$

$= \frac{1}{2}t^3[\theta_1^{-2}tr(B_1^{-1}\underline{1}'\underline{1}B_1^{-1}) -2\theta_1^{-1}\phi\underline{1}'B_1^{-1}\underline{1}\underline{1}'B_1^{-1}B_1^{-1}\underline{1} + \phi^2\underline{1}'B_1^{-1}\underline{1}\underline{1}'B_1^{-1}B_1^{-1}\underline{1}\underline{1}'B_1^{-1}\underline{1}]$

(7.23) $I(\underline{\theta})_{2,3} = \frac{1}{2}t^3[\theta_1^{-2}tr(C) - 2\theta_1^{-1}\phi(\underline{1}'B_1^{-1}\underline{1})tr(C) + \phi^2(\underline{1}'B_1^{-1}\underline{1})^2tr(C)]$

Using (7.9) and lemma 7.5, we have that

$I(\underline{\theta})_{2,3} = \frac{1}{2}t^3[\theta_1^{-2}120^{-1}p(p+1)(p+2)(p^2+2p+2)-720^{-1}\phi\theta_1^{-1}p^2(p+1)^2(p+2)^2(p^2+2p+2)$

$+\phi^2 17280^{-1}p^3(p+1)^3(p+2)^3(p^2+2p+2)]$

$= 240^{-1}t^3\theta_1^{-2}p(p+1)(p+2)(p^2+2p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[(12\theta_1+\theta_2t^2p(p+1)(p+2))^2$

$- 2\theta_2t^2p(p+1)(p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2)) + \theta_2^2t^4p^2(p+1)^2(p+2)^2]$

(7.24) $I(\underline{\theta})_{2,3} = 3(5^{-1})t^3p(p+1)(p+2)(p^2+2p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}$

$I(\underline{\theta})_{3,3} = \frac{1}{2}tr[(\theta_1^{-1}B_1^{-1}-\phi C)B_3(\theta_1^{-1}B_1^{-1}-\phi C)B_3]$

$= \frac{1}{2}t^2tr[(\theta_1^{-1}B_1^{-1}-\phi C)^2]$

(7.25) $I(\underline{\theta})_{3,3} = \frac{1}{2}t^2\sum_{i,j=1}^{p}((B_1^{-1})_{ij}-\phi C_{ij})^2$

We therefore require

(i) $\sum_{i,j=1}^{p}(B_1^{-1})_{ij}^2$

(ii) $\sum_{i,j=1}^{p}(C_{ij})^2$, and

(iii) $\sum_{i,j=1}^{p}(B_1^{-1})_{ij}C_{ij}$

Now, (i) $= [2\sum_{j=1}^{p}\sum_{i=1}^{j-1}(p+1)^{-2}(p-j+1)^2i^2] + [\sum_{j=1}^{p}(p+1)^{-1}j^2(p-j+1)^2]$

where the sum in the first bracket is of the off-diagonal elements, and that in the second bracket is of the diagonal elements of B_1.

$= (p+1)^{-2}\sum_{j=1}^{p}(p+1-j)^2[\sum_{j=1}^{p}2i^2 + j^2]$

$$= (p+1)^{-2}\sum_{j=1}^{p}((p+1)^2-2(p+1)j+j^2)[3^{-1}(j-1)j(2j-1) + j^2]$$

$$= 3^{-1}(p+1)^{-2}\sum_{j=1}^{p}(2j^5-4(p+1)j^4+(2(p+1)^2+1)j^3-2(p+1)j^2+(p+1)^2j)$$

$$= 3^{-1}(p+1)^{-2}[6^{-1}p^2(p+1)^2(2p^2+2p-1) - 2.15^{-1}p(p+1)^2(2p+1)(3p^2+3p+1)$$

$$+ \tfrac{1}{4}(2p^2+4+3)p^2(p+1)^2 - 3^{-1}p(p+1)^2(2p+1) + \tfrac{1}{2}p(p+1)^3$$

$$= 180^{-1}p[2p^3+8p^2+17p+18]$$

(7.26) $$\sum_{i,j=1}^{p}(B_1^{-1})_{ij}^2 = 180^{-1}p(p+2)(2p^2+4p+9)$$

Similarly, (ii) $$= (\sum_{i=1}^{p}(\tfrac{1}{2}i(p-i+1))^2)^2$$

$$= 16^{-1}(\sum_{i=1}^{p}[(p+1)^2i^2-2(p+1)i^3+i^4])^2$$

$$= 16^{-1}[6^{-1}p(p+1)^3(2p+1) - \tfrac{1}{2}p^2(p+1)^3 + 30^{-1}p(p+1)(2p+1)(3p^2+3p-1)]^2$$

$$= 14400^{-1}p^2(p+1)^2[p^3+4p^2+6p+4]^2$$

(7.27) $$\sum_{i,j=1}^{p}(C_{ij})^2 = 14400^{-1}p^2(p+1)^2(p+2)^2(p^2+2p+2)^2$$

Finally, (iii) $$= \tfrac{1}{4}\sum_{j=1}^{p}j(p-j+1)[\sum_{i=1}^{j}i(p-j+1)i(p-i+1) + \sum_{i=j+1}^{p}j(p-i+1)i(p-i+1)]$$

$$= \tfrac{1}{4}(p+1)^{-1}\sum_{j=1}^{p}j(p-j+1)[(p-j+1)\sum_{i=1}^{j}((p+1)i^2-i^3) + j\sum_{k=1}^{p-j}((p+1)k^2-k^3)]$$

$$= \tfrac{1}{4}(p+1)^{-1}\sum_{j=1}^{p}j(p-j+1)[(p-j+1)((p+1)6^{-1}j(j+1)(2j+1)-\tfrac{1}{4}j^2(j+1)^2)$$

$$+ j(6^{-1}(p+1)(p-j)(p-j+1)(2p-2j+1) - \tfrac{1}{4}(p-j)^2(p-j+1)^2)]$$

$$= 48^{-1}\sum_{j=1}^{p}[-j^6+3(p+1)j^5-(2p^2+4p+1)j^4-(p+1)(p^2+2p+3)j^3+(p+1)^2(p^2+2p+2)j^2]$$

$$= 48^{-1}[-42^{-1}p(p+1)(2p+1)(3p^4+6p^3-3p+1)+\tfrac{1}{4}p^2(p+1)^3(2p^2+2p-1)$$

$$-30^{-1}p(p+1)(2p+1)(3p^2+3p-1)(2p^2+4p+1)-\tfrac{1}{4}p^2(p+1)^3(p^2+2p+3)$$

$$+6^{-1}p(p+1)^3(2p+1)(p^2+2p+2)]$$

(7.28) $$\sum_{i,j=1}^{p}(B_1^{-1})_{ij}C_{ij} = 20160^{-1}p(p+1)(p+2)[17p^4+68p^3+133p^2+130p+72]$$

Substituting (7.26-28) in (7.25), we therefore have

$$I(\underline{\theta})_{3,3} = \tfrac{1}{2}t^2[\theta_1^{-2}180^{-1}p(p+2)(2p^2+4p+9)$$

$$- 2\theta_1^{-1}\phi 20160^{-1}p(p+1)(p+2)(17p^4+68p^3+133p^2+130p+72)$$

$$+ \phi^2 14400^{-1}p^2(p+1)^2(p+2)^2(p^2+2p+2)^2]$$

$$= 201600^{-1}t^2p(p+2)[560\theta_1^{-1}(2p^2+4p+9) - 10\theta_1^{-1}\phi(p+1)(17p^4+68p^3+133p^2+130p+72)$$

$$+ 7\phi^2p(p+1)^2(p+2)(p^2+2p+2)^2]$$

$$=201600^{-1}\theta_1^{-2}t^2p(p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[560(2p^2+4p+9)(12\theta_1+\theta_2t^2p(p+1)(p+2))^2$$

$$- 120\theta_2t^2(p+1)(17p^4+68p^3+133p^2+130p+72)(12\theta_1+\theta_2t^2p(p+1)(p+2))$$

$$+ 1008\theta_2^2t^4p(p+1)^2(p+2)(p^2+2p+2)^2]$$

$$= 25200^{-1}\theta_1^{-2}t^2p(p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[10080(2p^2+4p+9)\theta_1^2$$

$$+60(p+1)(5p^4+20p^3+77p^2+114p-216)\theta_1\theta_2t^2+p(p+1)^2(p+2)(11p^2+44p^3+203p^3+318p-576)\theta_2^2t^4]$$

(7.29) $$I(\underline{\theta})_{3,3} = 25200^{-1}\theta_1^{-2}t^2p(p+2)(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}[10080(2p^2+4p+9)\theta_1^2$$

$$+60(p-1)(p+1)(p+3)(5p^2+10p+72)\theta_1\theta_2t^2+(p-1)p(p+1)^2(p+2)(p+3)(11p^2+22p+192)\theta_2^2t^4]$$

We have thus completed calculation of the elements of $I(\underline{\theta})$, which are given in equations (7.16), (7.18), (7.20), (7.22), (7.24) and (7.29). We write

(7.30) $$I(\underline{\theta}) = \tfrac{1}{2}\theta_1^{-2}p(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

where

(7.31) $$a = 144\theta_1^2+24(p-1)(p+1)(p+2)\theta_1\theta_2t^2+(p-1)p(p+1)^2(p+2)^2\theta_2^2t^4$$

(7.32) $$b = 12(p+1)(p+2)\theta_1^2t^2$$

(7.33) $c = 15^{-1}t(p+2)[360\theta_1^2+24(p-1)(p+1)(p+3)\theta_1\theta_2t^2+(p-1)p(p+1)^2(p+2)(p+3)\theta_2^2t^4]$

(7.34) $d = p(p+1)^2(p+2)^2\theta_1^2t^4$

(7.35) $e = 6(5^{-1})(p+1)(p+2)(p^2+2p+2)\theta_1^2t^3$

(7.36) $f = 12600^{-1}(p+2)t^2[10080(2p^2+4p+9)\theta_1^2$

$+60(p-1)(p+1)(p+3)(5p^2+10p+72)\theta_1\theta_2t^2+(p-1)p(p+1)(p+2)^2(p+3)(11p^2+22p+192)\theta_2^2t^4]$

7.3 CALCULATION OF THE ASYMPTOTIC VARIANCE OF $\hat{\theta}_3$.

As stated earlier, the asymptotic variance-covariance matrix of $\hat{\theta}$ is given by the inverse of $NI(\underline{\theta})$. We thus now calculate $I^{-1}(\underline{\theta})$. Write

(7.37) $\psi = \frac{1}{2}\theta_1^{-2}p(12\theta_1+\theta_2t^2p(p+1)(p+2))^{-2}$

With $I(\underline{\theta})$ defined in (7.30), we then have

(7.38) $|I(\underline{\theta})| = \psi^3(adf+2cbe-ae^2-dc^2-fb^2)$

so that

(7.39) $I^{-1}(\underline{\theta}) = \psi^{-1}(adf+2cbe-ae^2-dc^2-fb^2)^{-1}\begin{pmatrix} df-e^2 & ec-bf & be-cd \\ ec-bf & af-c^2 & bc-ae \\ be-cd & bc-ae & ad-b^2 \end{pmatrix}$

Now, $var(\hat{\theta}_3|\theta_3=0)$ is given by the (3,3) element of $N^{-1}I^{-1}(\underline{\theta})$, so that

(7.40) $var(\hat{\theta}_3|\theta_3=0) = n^{-1}\psi^{-1}(adf+2cbe-ae^2-dc^2-fb^2)^{-1}(ad-b^2)$

Since we are interested in asymptotic properties of $var(\hat{\theta}_3|\theta_3=0)$, and since the amount of algebraic manipulation required to calculate its exact value is substantial, we simplify by writing

(7.41) π_i means "polynomial of order i"

Thus, from (7.31-36)

(7.42) $a= 144\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4$

(7.43) $b = 12\pi_2\theta_1^2t^2$

(7.44) $c = 15^{-1}t[360\pi_1\theta_1^2+24\pi_4\theta_1\theta_2t^2+\pi_7\theta_2^2t^4]$

(7.45) $d = \pi_5\theta_1^2t^4$

(7.46) $e = 6(5^{-1})\pi_4\theta_1^2t^3$

(7.47) $f = 12600^{-1}t^2[20160\pi_3\theta_1^2+300\pi_6\theta_1\theta_2t^2+11\pi_9\theta_2^2t^4]$

where $\pi_1,...,\pi_9$ are implicitly defined by the previously given formulae.

For large p we can use the approximations that $\pi_i\pi_j = \pi_{i+j}$, and $\pi_i + \pi_j = \pi_i$ if i>j. We therefore have, from (7.42-47),:

(7.48) $adf = 12600^{-1}\pi_8\theta_2^2t^6(29030400\theta_1^4+527040\pi_3\theta_1^3\theta_2t^2+28944\pi_6\theta_1^2\theta_2^2t^4$
$+564\pi_9\theta_1\theta_2^3t^6+11\pi_{12}\theta_2^4t^8)$

(7.49) $2cbe = 48(25)^{-1}\pi_7\theta_1^4t^6(360\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)$

(7.50) $ae^2 = 36(25)^{-1}\pi_8\theta_1^4t^6(144_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)$

(7.51) $dc^2 = 225^{-1}\pi_7\theta_1^2t^6(129600\theta_1^4+17280\pi_3\theta_1^3\theta_2t^2+1296\pi_6\theta_1^2\theta_2^2t^4+48\pi_9\theta_1\theta_2^3t^6$
$+\pi_{12}\theta_2^4t^8)$

(7.52) $fb^2 = 175^{-1}\pi_7\theta_1^4t^6(20160\theta_1^2+300\pi_3\theta_1\theta_2t^2+11\pi_6\theta_2^2t^4)$

(7.53) $ad = \pi_5\theta_1^2t^4(144\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)$

(7.54) $b^2 = 144\pi_4\theta_1^4t^4$

Thus

(7.55) $(adf+2cbe-ae^2-dc^2-fb^2) = 12800^{-1}\pi_8\theta_1^2t^6(29030400\theta_1^4+527040\pi_3\theta_1^3\theta_2t^2$
$+28944\pi_6\theta_1^2\theta_2^2t^4+564\pi_9\theta_1\theta_2^3t^6+11\pi_{12}\theta_2^4t^8)$

and

(7.56) $ad-b^2 = \pi_5\theta_1^2t^4(144\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)$

Also, from (7.37)

(7.57) $\psi = \frac{1}{2}\pi_1\theta_1^{-2}(12\theta_1+\pi_3\theta_2t^2)^{-2}$

Thus, from (7.40) and (7.55-7)

$$\text{var}(\hat{\underline{\theta}}_3 \mid \theta_3=0) = \frac{n^{-1}[2\pi_1^{-1}\theta_1^2(12\theta_1+\pi_3\theta_2t^2)^2\pi_5\theta_1^2t^4(144\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)}{(adf+2cbe-ae^2-dc^2-fb^2)}$$

$$(7.58) = \frac{6400n^{-1}\pi_4^{-1}t^{-2}\theta_1^2(12\theta_1+\pi_3\theta_2t^2)^2(144\theta_1^2+24\pi_3\theta_1\theta_2t^2+\pi_6\theta_2^2t^4)}{(29030400\theta_1^4+527040\pi_3\theta_1^3\theta_2t^2+28944\pi_6\theta_1^2\theta_2^2t^4+564\pi_9\theta_1\theta_2^3t^6+11\pi_{12}\theta_2^4t^8)}$$

We now examine (7.58) to ascertain asymptotic properties of $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$. By subtracting the maximum power of the term of interest in the denominator from that in the numerator we find that:

(i) keeping all other parameters fixed and varying p, $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$ is proportional to $p^{-4} + O(p^{-5})$.

(ii) keeping tp (the total time of the study) constant, and varying p, $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$ is proportional to $p^{-2} + O(p^{-3})$

(iii) keeping all other parameters fixed and varying t, $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$ is proportional to $t^{-2} + O(t_{-3})$

(iv) varying θ_2 has no overall effect on $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$.

(v) keeping all other parameters fixed and varying n, $\text{var}(\hat{\theta}_3 \mid \theta_3=0)$ is proportional to n^{-1}.

(vi) keeping all other parameters fixed and varying θ_1, $var(\hat{\theta}_3 | \theta_3=0)$ is proportional to $\theta_1^2 + O(\theta_1)$.

Thus the variance of $\hat{\theta}_3$ will be minimised faster by increasing p, the number of observations per site, rather than n, the number of sites, and will decrease if t, the time between observations, increases. We therefore see that in designing a study whose aim is to estimate θ_3 we will try to make as many observations as possible per site, rather than to observe a large number of sites, and will try to maximise the time between these observations.

It may at first seem odd that we detect whether bursts of activity occur by increasing, rather than decreasing the time between observations. The reason for this is that the longer the time between observations, the greater the average change in attachment, and the larger, therefore the proportion of observed change which is due to true change rather than to measurement error.

It should also be remembered that we do not attempt either to identify at which sites bursts of activity occur, or to identify when a burst of activity has taken place. We simply assess whether the whole data set appears to be described better by the 'burst' model or the 'steady progression' model.

8

MAXIMUM LIKELIHOOD ESTIMATION FOR DEPENDENT OBSERVATIONS.

We saw in the last section that the properties of maximum-likelihood (ML) estimates of the parameters of covariance structures as p, rather than n, tends to infinity may be of interest. Under such conditions, the assumption that the observations are independently and identically distributed (iid) no longer holds, so that the classical asymptotic properties of maximum likelihood estimates (MLEs) need no longer be valid. In this section we review the literature on ML estimation where these assumptions are relaxed, and present sufficient conditions for the estimates to be consistent as $p \to \infty$.

Most of the literature on asymptotic properties of MLEs is based on the work either of Cramér (1946) or of Wald (1949). While Cramér (1946) assumed that the log-likelihood is three-times differentiable, Wald (1949) made no differentiability assumptions to prove consistency. Each applied the strong law of large numbers, using the fact that the log-likelihood and its derivatives are the sum of independently and identically distributed random variables.

We will use the following notation. Let $X_1,...,X_n$ be random variables taking values $x_1,...,x_n$ in a sample space, with joint distribution $p(\underline{x}_n,\underline{\theta})$ ($(\underline{x}_n = (x_1,...,x_n)')$ which depends on a $q \times 1$ parameter vector $\underline{\theta}$ belonging to a parameter space Ω. We denote by $L(\underline{\theta},\underline{x}_n)$ the log-likelihood of $\underline{\theta}$ and by $\underline{\theta}_0$ the true value of $\underline{\theta}$.

A number of authors (Prasad and Prakasa Rao 1976, Bhat 1979, Bad 1979, Sarma 1986) make assumptions about the conditional density, defined by

(8.1) $p^n(\underline{x}_n,\underline{\theta}) = p(\underline{x}_n,\underline{\theta})/p(\underline{x}_{n-1},\underline{\theta})$.

Since the log-likelihood can be expressed as

(8.2) $L(\underline{\theta},\underline{x}_n) = \sum_{i=1}^{n} \log(p^i(\underline{x}_i,\underline{\theta}))$

suitable regularity conditions can be imposed to allow the application of martingale limit theorems in a manner analogous to the application of the strong law of large numbers in the iid case. Our approach, on the other hand, will be to impose conditions on $L(\underline{\theta},\underline{x}_n)$. This approach was taken by, for instance, Crowder (1976) and Heijmans and Magnus (1986a). The theorems are adaptations of those of Lehmann (1983) which apply to the iid case: we replace the application of the strong law of large numbers by assumptions about the expectation and variance of the log likelihood and its derivatives. We make the following assumptions:

(A1) For each (fixed) n, the distributions $p(\underline{x}_n,\theta)$ and $p(\underline{x}_n,\theta_0)$ are distinct for $\underline{\theta}$ not equal to $\underline{\theta}_0$.

(A2) The support of $p(\underline{x}_n,\underline{\theta})$ does not depend on $\underline{\theta}$.

8.1 ASYMPTOTIC CONSISTENCY FOR THE SINGLE PARAMETER CASE.

Here we assume that q=1. We now give conditions for the consistency of the MLE for the single parameter case. We need the further assumptions:

(A3) $P[L(\theta_0,\underline{X}_n) - L(\theta,\underline{X}_n) > 0] \rightarrow 1$

(A4) The parameter space Ω contains an open interval ω of which the true parameter value θ_0 is an interior point.

Theorem 8.1. For each (fixed) n, let $\underline{X}_n$ satisfy (A1)-(A4) and suppose that, for all $\underline{X}_n$, $L(\theta,\underline{x}_n)$ is differentiable with respect to θ in ω. Then with probability tending to 1 as $n\to\infty$, the likelihood equation

(8.3) $(\partial/\partial\theta)L(\theta,\underline{x}_n) = 0$

has a root $\hat{\theta}_n = \hat{\theta}(\underline{x}_n)$ such that $\hat{\theta}(\underline{x}_n)$ tends to the true value θ_0 in probability.

Proof: Let δ be small enough so that $(\theta_0-\delta,\theta_0+\delta)\subset\omega$, and let

$S_n=\{\underline{x}_n: L(\theta_0,\underline{x}_n)>L(\theta_0-\delta,\underline{x}_n)$ and $L(\theta_0,\underline{x}_n)>L(\theta_0+\delta,\underline{x}_n)\}$

By (A3), $P(\underline{x}_n\in S_n)\to 1$. For any $\underline{x}_n$ in S_n, there exists a value $\hat{\theta}_n(\delta)$ with $\theta_0-\delta<\hat{\theta}_n(\delta)<\theta_0+\delta$ at which $L(\theta_0,\underline{x}_n)$ has a local maximum and at which, therefore, $(\partial/\partial\theta)L(\theta,\underline{x}_n)$ evaluated at $\theta=\hat{\theta}_n(\delta)$ is equal to zero. Hence for any small $\delta>0$,

(8.4) P[(8.3) has a root in $(\theta_0-\delta,\theta_0+\delta)$] $\to$ 1 as $n\to\infty$.

Let Θ_n be the set of roots of $L(\theta,\underline{x}_n)$ (if there is no root let $\Theta_n=\{1\}$). Let $\hat{\theta}_n^*$ be the element of Θ_n closest to θ_0 [this exists because the limit of a sequence of roots is again a root by the continuity of $L(\theta)$]. If there is a root in $(\theta_0-\delta,\theta_0+\delta)$ then $\hat{\theta}_n^*\in(\theta_0-\delta,\theta_0+\delta)$, and therefore from (8.4) $P(\hat{\theta}_n^*\in(\theta_0-\delta,\theta_0+\delta))\to 1$ as $n\to\infty$. Choosing $\hat{\theta}_n = \hat{\theta}_n^*$ establishes the theorem. ∎

The argument in the last paragraph is, we believe, an improvement on that presented by Lehmann. For the iid case condition (A3) can be established from the strong law of large numbers: we present sufficient conditions for (A3) to hold in the non iid case below.

Corollary 8.1. Under the assumptions of theorem 8.1, if the likelihood equation has a unique root δ_n for each n and for all $\underline{x}_n$, then $\{\delta_n\}$ is a consistent estimator sequence for θ_0. If, in addition, the parameter space is an open interval (a,b) then with probability tending to 1, δ_n maximises the likelihood, i.e. δ_n is the MLE, which is therefore consistent.

Proof: The first statement is clearly true. To prove the second, suppose the probability of δ_n being the MLE does not tend to 1. Then, for arbitrarily large n, the likelihood has a supremum at a or b with probability ϕ with $\liminf_{n\to\infty} \phi>0$. We suppose (without loss of generality) that this supremum is at b.

From theorem 8.1, as $n\to\infty$, there is a root $\hat{\theta}_n$ such that for any $\varepsilon>0$ which is sufficiently small $\hat{\theta}_n\in(\theta_0-\varepsilon,\theta_0+\varepsilon)$ and $\hat{\theta}_n$ is a local maximum of $L(\theta,\underline{x}_n)$. Since δ_n is unique, $\delta_n=\hat{\theta}_n$ and therefore δ_n is a local maximum. Since $L(b,\underline{x}_n)>L(\delta_n,\underline{x}_n)$ for arbitrarily large n, L has a local minimum for arbitrarily large n, contradicting the uniqueness of δ_n. ∎

We now give a condition can be used to prove consistency in the non iid case. Assume:

(B1) There exists a sequence $a_n = a_n(\theta)$ with $a_n(\theta)>0$ and $\liminf_{n\to\infty} a_n>0$ such that for each (fixed) θ not equal to θ_0:

(i) $\limsup_{n\to\infty} E_\theta[a_n^{-1}(L(\theta,\underline{X}_n) - L(\theta_0,\underline{X}_n))] = \mu(\theta) < 0.$

(ii) $Var_\theta(a_n^{-1}(L(\theta,\underline{X}_n)-L(\theta_0,\underline{X}_n)) \to 0$ as $n\to\infty$.

Theorem 8.2. Given assumptions A1, A2 and B1, $P[L(\theta_0,\underline{X}_n) - L(\theta,\underline{X}_n) \geq 0] \to 1$ as $n\to\infty$ for any fixed θ not equal to θ_0.

Proof: Suppose not. Then there exists some ε such that for arbitrarily large n, $P[L(\theta_0,\underline{X}_n)-L(\theta,\underline{X}_n) \leq 0] > \varepsilon$. Thus

$P[a_n^{-1}(L(\theta_0,\underline{X}_n) - L(\theta,\underline{X}_n)) \leq 0] > \varepsilon$

But from (i) there exists some N such that for n>N

$E_{\theta}[a_n^{-1}(L(\theta,\underline{X}_n) - L(\theta_0,\underline{X}_n))] < 0.5\times\mu(\theta)$ implying that for n>N

$Var_{\theta}[a_n^{-1}(L(\theta_0,\underline{X}_n) - L(\theta,\underline{X}_n))] > 0.25\times\varepsilon\mu(\theta)^2$, contradicting B1 (ii) above.

(since if $E[Y] = \mu < 0$ and $p(Y>0) > \varepsilon$ then

$Var(Y) = \int_{-\infty}^{\infty}(y-\mu)^2dF(y) \geq \int_{0}^{\infty}(y-\mu)^2dF(y) \geq \mu^2\int_{0}^{\infty}dF(y) > \mu^2\varepsilon$) ∎

Since assumption B1 implies assumption A3, it can be used to prove consistency of the MLE for particular examples.

8.2 ASYMPTOTIC CONSISTENCY FOR MULTIPLE PARAMETERS.

It was shown by Crowder (1976) that where the likelihood depends on multiple parameters (i.e. q>1) and the observations are not iid, it is possible either that the MLEs are consistent or that the MLEs for a subset of the parameters are consistent while the MLEs for "transient" parameters (as Crowder calls them) are not consistent. Clearly, it is possible that the MLEs of all parameters are not consistent; for instance in the non iid case where the density of X_1 is $p(x,\underline{\theta})$, while $X_i=0$ with probability 1 for i=2,3,... It is therefore clear that conditions which ensure the consistency of the MLEs for non iid variables, will not be as simple, general or easy to specify as for the iid case.

Intuitively, it seems likely that the MLEs will be consistent if an analogue of the Fisher information (for instance the matrix $-E[\partial^2 L(\underline{\theta},\underline{x}_n)/\partial\underline{\theta}\partial\underline{\theta}']$) given by $\underline{x}_n$ tends to infinity in some specified sense as $n\to\infty$. However, such a condition does not appear to be sufficient in itself to prove consistency of the MLE: we need also to ensure convergence in probability of the first three derivatives of the likelihood function. In the theorem below we have again adapted the proof of Lehmann by specifying conditions sufficient for consistency in the non iid case, which replace the use of the strong law of large numbers in the iid case.

We make the following assumptions, in addition to (A1) and (A2).

(C1) There exists an open subset ω of the parameter space Ω containing the true parameter value $\underline{\theta}_0$ such that for all $\underline{x}_n$ the likelihood admits all third derivatives.

(C2) There exists a sequence $\{a_n\}$ $(a_n>0)$ $(n=1,2,...)$ which may depend on $\underline{\theta}$ such that the following hold:

(i) $E_{\underline{\theta}}[\partial L(\underline{\theta},\underline{x}_n)/\partial\theta_j] = 0$ $(j=1,..,q)$ for all $\underline{\theta}\in\omega$.

(ii) Define the $q\times q$ matrix $J_n(\underline{\theta})$ with (i,j)th element

(8.5) $J_{ijn} = E_{\underline{\theta}}[a_n^{-1}(\partial L(\underline{\theta},\underline{x}_n)/\partial\theta_i)(\partial L(\underline{\theta},\underline{x}_n)/\partial\theta_j)] = E_{\underline{\theta}}[-a_n^{-1}\partial^2 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j]$

Since $J_n(\underline{\theta})$ is a covariance matrix, it is positive semi-definite. We assume that, as $n\to\infty$, for any $q\times 1$ vector $\underline{y}$, with at least one non-zero element, we have for all $\underline{\theta}\in\omega$:

$\underline{y}'J_n(\underline{\theta})\underline{y} > 0$, and $\lim_{n\to\infty}\inf \underline{y}'J_n(\underline{\theta})\underline{y}/\underline{y}'\underline{y} = \kappa > 0$.

(iii) For all $\underline{\theta}\in\omega$, $|E_{\underline{\theta}}[a_n^{-1}\partial^3 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j\partial\theta_k]|<\infty$ (i,j,k=1,...,q) and is bounded (less than b_{ijk}, say) as $n\to\infty$.

(iv) $\mathrm{Var}(a_n^{-1}\partial L(\underline{\theta},\underline{x}_n)/\partial\theta_i)\to 0$ as $n\to\infty$ (i=1,...,q) for all $\underline{\theta}\in\omega$.

(v) $\mathrm{Var}(a_n^{-1}\partial^2 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j)\to 0$ as $n\to\infty$ (i,j=1,...,q) for all $\underline{\theta}\in\omega$.

(vi) $\mathrm{Var}(a_n^{-1}\partial^3 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j\partial\theta_k)\to 0$ as $n\to\infty$ (i,j,k=1,...,q) for all $\underline{\theta}\in\omega$.

Theorem 8.3. If $L(\underline{\theta},\underline{x}_n)$ is defined for all n and satisfies (A1), (A2) and (C) then, with probability tending to 1 as $n\to\infty$ there exists a solution $\hat{\theta}_n=\hat{\theta}_n(\underline{x}_n)$ of the likelihood equations such that $\{\hat{\theta}_n\}$ is consistent for $\underline{\theta}_0$.

Proof: We consider the behaviour of the log-likelihood $L(\underline{\theta},\underline{x}_n)$ on the sphere Q_δ with radius δ and centre at the true point $\underline{\theta}_0$. We will show that for any sufficiently small δ the probability tends to 1 that:

(8.6) $L(\underline{\theta},\underline{x}_n) < L(\underline{\theta}_0,\underline{x}_n)$

at all points $\underline{\theta}$ on the surface of Q_δ, and hence that $L(\underline{\theta},\underline{x}_n)$ has a local maximum in the interior of Q_δ. Since at a local maximum the likelihood equations must be satisfied it will follow that for any $\delta>0$ with probability tending to 1 as $n\to\infty$, the likelihood equations have a solution $\hat{\theta}_n(\delta)$ within Q_δ. We can then complete the proof as in the 1-dimensional case.

We write $L^{(i)}(\underline{\theta}',\underline{x}_n) = \partial L(\underline{\theta},\underline{x}_n)/\partial\theta_i$, $L^{(i,j)}(\underline{\theta}',\underline{x}_n) = \partial^2 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j$, etc., and use $L^{(i)}(\underline{x}_n,\underline{\theta}^*)$ to mean $\partial L(\underline{\theta},\underline{x}_n)/\partial\theta_i$ evaluated at $\underline{\theta}=\underline{\theta}^*$. Expand $L(\underline{\theta},\underline{x}_n)$ about $\underline{\theta}_0$ and divide by a_n:

$$a_n^{-1}L(\underline{\theta},\underline{x}_n)-a_n^{-1}L(\underline{\theta}_0,\underline{x}_n) = a_n^{-1}\sum_{i=1}^{q}L^{(i)}(\underline{\theta}_0,\underline{x}_n)(\theta_i-\theta_{0i})+(2a_n)^{-1}\sum_{i=1}^{q}\sum_{j=1}^{q}L^{(i,j)}(\underline{\theta}_0,\underline{x}_n)(\theta_i-\theta_{0i})(\theta_j-\theta_{0j})$$
$$+(6a_n)^{-1}\sum_{i=1}^{q}\sum_{j=1}^{q}\sum_{k=1}^{q}L^{(i,j,k)}(\underline{\theta}^*,\underline{x}_n)(\theta_i-\theta_{0i})(\theta_j-\theta_{0j})(\theta_k-\theta_{0k})$$

where $\underline{\theta}^*$ is in the interior of Q_δ. Write $a_n^{-1}L(\underline{\theta},\underline{x}_n)-a_n^{-1}L(\underline{\theta}_0,\underline{x}_n) = S_1 + S_2 + S_3$. We now show that, with probability tending to 1 as $n\to\infty$, if δ is sufficiently small then the maximum S_2 is negative while S_1 and S_3 are small compared to S_2.

We begin with S_1. By (C2) (i) and (iv), and using Tchebychev's inequality, $a_n^{-1}L^{(i)}(\underline{\theta}_0,\underline{x}_n)$ tends to zero in probability, so that for any δ^2, with probability tending to 1, $|a_n^{-1}L^{(i)}(\underline{\theta}_0,\underline{x}_n)|<\delta^2$. On Q_δ we have $|S_1| \le a_n^{-1}\delta\Sigma|L^{(i)}(\underline{\theta},\underline{x}_n)|$. It follows that:

(8.7) $|S_1| \le q\delta^3$ with probability 1 as $n\to\infty$.

For S_2, we have from (C2) (ii) and (v) and Tchebychev's inequality that $a_n^{-1}L^{(i,j)}(\underline{\theta}_0,\underline{x}_n)$ tends in probability to $-J_n(\underline{\theta}_0)$ as $n\to\infty$. We have:

$$2S_2 = \sum_{i=1}^{q}\sum_{j=1}^{q}[-J_{ijn}(\underline{\theta}_0)(\theta_i-\theta_{0i})(\theta_j-\theta_{0j})] + \sum_{i=1}^{q}\sum_{j=1}^{q}[(a_n^{-1}L^{(i,j)}(\underline{\theta}_0,\underline{x}_n)-(-J_{jkn}(\underline{\theta}_0)))(\theta_i-\theta_{0i})(\theta_j-\theta_{0j})]$$

For the second term, for an argument analogous to that for S_1, it follows that its absolute value is less than $s^2\delta^3$ with probability tending to 1 as $n\to\infty$. The first term is a negative (nonrandom) quadratic form in the variables $(\theta_j-\theta_{0j})$. Write $\underline{y}=(\underline{\theta}-\underline{\theta}_0)$. From condition (C2) (ii) we have:

$-\underline{y}'J_n(\underline{\theta})\underline{y}/\underline{y}'\underline{y} > 0$, and $\lim_{n\to\infty}\inf -\underline{y}'J_n(\underline{\theta})\underline{y}/\underline{y}'\underline{y} = -\kappa < 0$.

Hence there exist $k>0$ and N such that $-\underline{y}'J_n(\underline{\theta})\underline{y}/\underline{y}'\underline{y} < -k$ for all $n>N$. On Q_δ $\underline{y}'\underline{y}=\delta^2$, and therefore the first term has absolute value greater than $k\delta^2$ for $n>N$.

Combining the first and second terms, we see that there exist $c>0$, $\delta_0>0$ such that for $d<\delta_0$

(8.8) $S_2 < -c\delta^2$ as $n\rightarrow\infty$.

By (C2) (iii) and (vi) and Tchebychev's inequality:

$a_n^{-1}L^{(i,j,k)}L(\underline{\theta}^*,\underline{x}_n)$ tends in probability to $E[a_n^{-1}L^{(i,j,k)}(\underline{\theta}^*,\underline{x}_n)] < b_{ijk}$ as $n\rightarrow\infty$. Hence:

(8.9) $S_3<b\delta^3$

on Q_δ, where $b=(q^3/6)\Sigma\Sigma\Sigma b_{ijk}$.

Combining (8.7), (8.8) and (8.9):

$\max (S_1 + S_2 + S_3) < -c\delta^2 + (b+q)\delta^3$

which is less than zero if $\delta < c/(b+q)$. ∎

We present two conditions for (C2) (ii) to hold. These are:

(D1) As $n\rightarrow\infty$, $J_n(\underline{\theta})\rightarrow I(\underline{\theta})$, where $I(\underline{\theta})$ is a positive definite matrix.

(D2) $\liminf_{n\rightarrow\infty} \lambda_n^*(\underline{\theta}) > 0$, where $\lambda_n^*(\underline{\theta})$ is the minimum eigenvalue of $J_n(\underline{\theta})$.

Each follows immediately from the fact that for any positive definite matrix A and vector $\underline{x}$, $\lambda^* \leq (\underline{x}'A\underline{x})(\underline{x}'\underline{x})^{-1}$, where λ^* is the minimum eigenvalue of A.

8.3 EXAMPLES.

In this section we give examples, based on simplifications of the covariance structures of models 1 and 2, firstly where the conditions of theorem are satisfied

and secondly where they are not satisfied. We will assume that $\underline{X}\sim N(\underline{0},\Sigma_n(\underline{\theta}))$, so that:

(8.10) $L_n(\underline{\theta},\underline{x}_n) = \text{const} - \frac{1}{2}\log|\Sigma_n(\underline{\theta})| - \frac{1}{2}\text{tr}(\Sigma_n(\underline{\theta})\underline{x}_n\underline{x}_n')$

and that $\Sigma_n(\underline{\theta}) = \theta_1 B_{n1} + \ldots + \theta_q B_{nq}$, where B_{ni} (i=1,,,q) are known (n×n) matrices. Thus (cf sections 2.2.1 and 7) we find that

(8.11) $L^{(i)}(\underline{\theta},\underline{x}_n) = -\frac{1}{2}\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}) + \frac{1}{2}Q_i$

(8.12) $L^{(i,j)}(\underline{\theta},\underline{x}_n) = \frac{1}{2}\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}) - Q_{ij}$

(8.13) $L^{(i,j,k)}(\underline{\theta},\underline{x}_n) = -\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nj}) + 3Q_{ijk}$

where

(8.14) $Q_i = \underline{x}'\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}\underline{x}$

(8.15) $Q_{ij} = \underline{x}'\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}\underline{x}$

(8.16) $Q_{ijk} = \underline{x}'\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nk}\Sigma_n(\underline{\theta})^{-1}\underline{x}$

Since if $\underline{X} \sim N(\underline{0},\Sigma)$ then $E[\underline{x}'A\underline{x}] = \text{tr}(A\Sigma)$ and $\text{Var}(\underline{x}'A\underline{x}) = 2\text{tr}(A\Sigma)^2$, we have:

(8.17) $E[Q_i] = \text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni})$

(8.18) $\text{Var}(Q_i) = 2\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni})^2$

(8.19) $E[Q_{ij}] = \text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj})$

(8.20) $\text{Var}(Q_{ij}) = 2\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj})^2$

(8.21) $E[Q_{ijk}] = \text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nk})$

(8.22) $\text{Var}(Q_{ijk}) = 2\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nk})^2$

Hence:

(8.23) $E[L^{(i)}(\underline{\theta},\underline{x}_n)] = 0$

(8.24) $\text{Var}(L^{(i)}(\underline{\theta},\underline{x}_n)) = \frac{1}{2}\text{tr}(\Sigma_n(\underline{\theta})^{-1}B_{ni})^2$

(8.25) $E[L^{(i,j)}(\underline{\theta},\underline{x}_n)] = -\frac{1}{2}tr(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj})^2$

(8.26) $Var(L^{(i,j)}(\underline{\theta},\underline{x}_n)) = 2tr(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj})^2$

(8.27) $E[L^{(i,j,k)}(\underline{\theta},\underline{x}_n)] = 2tr(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nj})$

(8.28) $Var(L^{(i,j,k)}(\underline{\theta},\underline{x}_n)) = 18tr(\Sigma_n(\underline{\theta})^{-1}B_{ni}\Sigma_n(\underline{\theta})^{-1}B_{nj}\Sigma_n(\underline{\theta})^{-1}B_{nj})^2$

8.3.1 ASYMPTOTIC CONSISTENCY FOR DEPENDENT VARIABLES.

Suppose that we sample only one site, at unit time intervals, and that the mean rate of progression is zero. Then (see section 3.1.2) the covariance structure is $\Sigma(\underline{\theta})=\theta_1B_1$ for model 1, and $\Sigma(\underline{\theta})=\theta_1B_1+\theta_3B_3$ for model 2 ($B_3=I$), while $E[\underline{X}]=\underline{0}$. Here we are estimating the value of θ_3 for the site, rather than for the population.

If A is a tri-diagonal matrix, that is with diagonal elements equal to a, off-diagonal elements equal to b and other elements equal to zero, then the eigenvalues of A are given (Press 1972) by

(8.29) $\lambda_{in} = a + 2b\cos i\pi/(n+1)$ $(i=1,...,n)$

with corresponding eigenvectors

(8.30) $\underline{x}_{in}' = (\sin i\pi/(n+1), \sin 2i\pi/(n+1), ..., \sin 2i\pi/(n+1))$

Since $\{\underline{x}_{in}\}$ do not depend on a and b, the $(n\times n)$ orthogonal matrix $P_n=(\underline{x}_{1n},...,\underline{x}_{nn})$ diagonalises any matrix of the form of A. In particular, it diagonalises $\Sigma(\underline{\theta})$, which has eigenvalues

(8.31) $\phi_{in} = \theta_3+\theta_1\lambda_{in} = \theta_3 + \theta_1(2-2\cos(i\theta/(n+1))) = \theta_3\lambda_{3in} + \theta_1\lambda_{1in}$

where

(8.32) $\lambda_{1in} = 2-2\cos(i\theta/(n+1))$ $(i=1,...,n)$

(8.33) $\lambda_{3in} = 1$ $(i=1,...,n)$

so that $\lambda_{1in}>0$, $\lambda_{1in}<4$, $\phi_{in}>\theta_3$, $\phi_{in}<\theta_3+4\theta_1$ for i=1,...,n, n=1,...,∞.

Define:

(8.34) $\Phi_n = P_n'\Sigma(\underline{\theta})P_n = \text{diag}(\phi_1,...,\phi_n)$

(8.35) $\Lambda_{jn} = P_n'B_{jn}P_n = \text{diag}(\lambda_{j1},...,\lambda_{jn})$ (j=1,3).

Using (8.23) to (8.28), we therefore have

(8.36) $\text{Var } L^{(i)}(\underline{\theta},\underline{x}_n)) = \frac{1}{2}\text{tr}(P_n'\Phi_n^{-1}P_nP_n'\Lambda P_n) = \frac{1}{2}\sum_{j=1}^{n}\lambda_{ijn}^2\phi_{in}^{-2}$

(8.37) $E[L^{(i,j)}(\underline{\theta},\underline{x}_n)] = -\frac{1}{2}\sum_{i=1}^{n}\lambda_{ijn}\lambda_{ijn}\phi_{in}^{-2}$

(8.38) $\text{Var}(L^{(i,j)}(\underline{\theta},\underline{x}_n)) = 2\sum_{i=1}^{n}\lambda_{ijn}^2\lambda_{ijn}^2\phi_{in}^{-4}$

The conditions are satisfied, as follows:

(C2) (i) holds for any choice of a_n, from (8.23).

(C2) (ii) $J_n = (2a_n)^{-1}\begin{bmatrix} \sum_{i=1}^{n}\lambda_{i1n}^2\phi_{in}^{-2} & \sum_{i=1}^{n}\lambda_{i1n}\lambda_{i3n}\phi_{in}^{-2} \\ \sum_{i=1}^{n}\lambda_{i1n}\lambda_{i3n}\phi_{in}^{-2} & \sum_{i=1}^{n}\lambda_{i3n}^2\phi_{in}^{-2} \end{bmatrix}$

We verify (D1). From the definition of the Riemann integral, the limit as $n\to\infty$ of J_{ijn} is given by:

(8.39) $I_{ij} = (2a_n)^{-1}\int_0^1 nf_{ij}(y)dy$

where

(8.40) $f_{11}(y) = (2-2\cos(y))^2(\theta_3+\theta_1(2-2\cos(y)))^{-2}$

(8.41) $f_{12}(y) = (2-2\cos(y))(\theta_3+\theta_1(2-2\cos(y)))^{-2}$

(8.42) $f_{22}(y) = (\theta_3+\theta_1(2-2\cos(y)))^{-2}$

Taking $a_n=n$, we have thus shown that the limit as $n\to\infty$ of J_n is the positive definite matrix I with elements defined in (8.39-42), and therefore that (D1), which implies (C2) (ii), is satisfied.

(C2) (iii) We require that for all $\underline{\theta}\in\omega$, $|E_{\underline{\theta}_0}[a_n^{-1}\partial^3 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j\partial\theta_k]|<\infty$ (i,j,k=1,...,q) and is bounded (less than b_{ijk}, say) as $n\to\infty$.

Now $E_{\underline{\theta}_0}[a_n^{-1}\partial^3 L(\underline{\theta},\underline{x}_n)/\partial\theta_i\partial\theta_j\partial\theta_k] = 2a_n^{-1}\sum_{l=1}^{n}\lambda_{il_n}\lambda_{jl_n}\lambda_{kl_n}\phi_{l_n}^{-3}$

$< 2a_n^{-1}p(\max_{i=1,3;\ j=1,..,n}\lambda_{ijn})^3\ (\min_{i=1,3;\ j=1,..,n}\phi_{ijn})^{-3}$

$= 2na_n^{-1}\times 64\theta_3^{-3}$

Thus (C2) (iii) is satisfied for $a_n = n$.

(C2) (iv) $\mathrm{Var}(a_n^{-1}L^{(i)}(\underline{\theta},\underline{x}_n))=\tfrac{1}{2}a_n^{-2}\sum_{i=1}^{n}\lambda_{ijn}^2\phi_{in}^2$

$< a_n^{-2}p(\lambda_{ijn\ max})^2(\phi_{jn\ max})^2 \to 0$ as $n\to\infty$ for $a_n=n$.

(C2) (v) and (vi) follow similarly.

Hence the conditions for theorem 8.3 are satisfied and the MLEs of θ_1 and θ_3 are consistent. ∎

8.3.2 FAILURE TO MEET CONDITIONS FOR ASYMPTOTIC CONSISTENCY.

Suppose now that, in the notation of section 3, we have p observations on m sites, so that the total number of observations is given by n = mp. Suppose also that we

make observations at unit time intervals, that the population mean rate of progression is zero and that there is no measurement error. It follows from standard theory of maximum likelihood estimation for iid variables that the estimates are consistent as $m\to\infty$; we will investigate whether this holds for $p\to\infty$. The covariance matrix for model 2 has block diagonal structure and is given by:

(8.43) $\Sigma(\underline{\theta}) = \text{diag}(A_1,...,A_m)$

where $\Sigma(\underline{\theta})$ has dimension n and A_i has dimension p with

(8.44) $A_i = \theta_2\underline{11}' + \theta_3 I \ (i=1,...,m)$

where $\underline{1}'=(1,...,1)$ is a $(p\times 1)$ vector and I is the identity matrix of dimension p. If we write $\underline{x}_n' = (\underline{x}_{1n},...,\underline{x}_{mn}) = (x_{11},...,x_{1p},x_{21},...,x_{2p},...,x_{m1},...,x_{mp})$

then clearly $L(\underline{\theta},\underline{x}_n) = \sum_{i=1}^{m} L(\underline{\theta},\underline{x}_i)$

Since $(A_i)^{-1} = \theta_3^{-1}I - \theta_3^{-1}\theta_2\underline{1}(\theta_1+p\theta_2)^{-1}\underline{1}'$, we have

(8.45) $\lim_{p\to\infty}(A_i)^{-1} = \theta_3^{-1}(I - p^{-1}\underline{11}')$

and using (8.25) we find:

$E[-L^{(2,2)}(\underline{\theta},\underline{x}_n)] = \tfrac{1}{2}m\ \text{tr}[A_i^{-1}\underline{11}'A_i^{-1}\underline{11}'] = m(\underline{1}'A_i^{-1}\underline{1})^2$

(8.46) $\lim_{p\to\infty} E[-L^{(2,2)}(\underline{\theta},\underline{x}_n)] = \tfrac{1}{2}\theta_3^{-2}m[p - 1]^2$

$E[-L^{(2,2)}(\underline{\theta},\underline{x}_n)] = \tfrac{1}{2}m\ \text{tr}[A_i^{-1}\underline{11}'A_i^{-1}I] = m\ \text{tr}[\underline{1}'A_i^{-2}\underline{1}]$

(8.47) $\lim_{p\to\infty} E[-L^{(2,3)}(\underline{\theta},\underline{x}_n) = \tfrac{1}{2}\theta_3^{-2}m[p - 2p + p] = 0$

$E[-L^{(3,3)}(\underline{\theta},\underline{x}_n) = \tfrac{1}{2}m\ \text{tr}[A_i^{-2}]$

(8.48) $\lim_{p\to\infty} E[-L^{(3,3)}(\underline{\theta},\underline{x}_n)] = \tfrac{1}{2}\theta_3^{-2}m[p - 2 + 1] = \tfrac{1}{2}\theta_3^{-2}m[p - 1]$

while similarly, from (8.26):

(8.49) $\lim_{p\to\infty} \mathrm{Var}L^{(2,2)}(\underline{\theta},\underline{x}_n)) = 2m(\underline{1}'A_i^{-1}\underline{1})^4 = 2\theta_3^{-4}m[p-1]^4$

(8.50) $\lim_{p\to\infty} \mathrm{Var}L^{(2,3)}(\underline{\theta},\underline{x}_n)) = 2m(\underline{1}'A_i^{-2}\underline{1})^2 = 0$

(8.51) $\lim_{p\to\infty} \mathrm{Var}L^{(3,3)}(\underline{\theta},\underline{x}_n)) = 2m\ \mathrm{tr}(A_i^{-4}) = 2m(p-4+6-4+1) = 2m(p-1)$

We thus see that as $p\to\infty$ $E[-L^{(i,j)}(\underline{\theta},\underline{x}_n)])^2 = O(\mathrm{Var}(L^{(i,j)}(\underline{\theta},\underline{x}_n)))$. Since the off-diagonal elements of J_n tend to zero as $p\to\infty$, the eigenvalues of J_n are equal to the diagonal elements of J_n. Thus, for any fixed m, and n=mp,

$a_n^{-1}E[-L^{(3,3)}(\underline{\theta},\underline{x}_n)])^2 = O(\mathrm{Var}(a_n^{-1}L^{(i,j)}(\underline{\theta},\underline{x}_n)))$

Hence conditions (C2) (ii) and (C2) (v) cannot be simultaneously satisfied for any choice of a_n, as $p\to\infty$. ∎

8.4 DISCUSSION.

We have given conditions under which MLEs for non iid variables are consistent, and have presented examples where these conditions are, and are not, satisfied.

Our conditions present the same problems as others given in the literature: that they are difficult to verify in practice. The paper by Heijmans and Magnus (1986a) contains conditions which may be more easily verified since they do not depend on derivatives of the likelihood function with respect to the parameter vector. This paper appeared after the completion of the theorems presented in this section: the examples given are still relatively simple and depend on knowledge of the eigenvalues of the covariance matrix. The treatment for the normal distribution is

more general than for the present study in that the case is treated where the mean vector and covariance matrix depend on the same parameter vector.

We have given simplifications of the covariance structures of models 1 and 2 where the MLEs are and are not consistent as $p \to \infty$. For the second example, it is unsurprising that the estimates are not consistent as $p \to \infty$: since increasing the number of observations in this way provides limited extra information on θ_2, the between-site variance.

9

DISCUSSION AND CONCLUSIONS.

The major part of this thesis is the development of methods which may be used to provide information in the nature of the progression of periodontal diseases. Although the nature of periodontal disease progression has been the subject of controversy and interest during recent years, statistical methods for determining their nature have not, until now, been proposed. We have shown that data consisting of successive measurements of attachment level and subject to measurement error may be used to determine which of three models for disease progression is the most appropriate. Our observation that the covariance structure for the burst model is a limiting case of that of a general model for varying rates of progression serves to remind periodontal researchers that the rejection of the constant progression model does not lead to the acceptance of any specific model in which the rate of progression varies. However we have also seen that it may not be possible to distinguish our formulation of this general model from the burst model when the amount of variation in the rate of progression is small.

We have examined the nature of disease progression in the most extensive long-term longitudinal study of untreated periodontal disease. Using standard linear regression, we have shown that there is a marked subject effect, in contrast to claims made recently in the literature. The observation that over a long time period the rate at which sites progressed was determined substantially by the subject in whom the site was found might in itself be thought to be evidence against the random burst model as proposed by Socransky et al. As we observed in section 2.1, the burst hypothesis was developed as part of a view of the disease that the nature of the microflora at a particular disease site was much more important than the subject effect. Although the large subject effect which we have observed could be

due either to the increased susceptibility of a minority of subjects or to the infection of the whole mouth over a long period with microorganisms causing periodontal disease, neither of these explanations sits easily with the view of the disease underlying the burst hypothesis.

Application of our methods for determining disease progression to these data has met with mixed success. None of the models could be said unambiguously to be the most appropriate for the data. Although there is evidence that the rate of progression changed, the estimates of the parameters representing between-subject variation in the burst and varying rates models were zero or close to zero, in contrast to the marked between-subject variation which was shown to exist in the preliminary analysis of the data. A possible explanation is that the variation in the rate of progression was due to differing rates of progression at different ages, and is therefore correlated between subjects.

It is to be hoped, therefore, that after the methods have been published it will be possible to apply them to other, possibly more appropriate data sets. For instance the assumption of constant mean rate of progression may be more realistic for a study in which observations were made at shorter intervals and over a shorter total time period. Similarly, given that a majority of subjects in the Sri Lanka study did not experience disease progression, studies of subject with pre-existing disease might provide more sites which show progression and therefore be more powerful for making inference about the nature of progression, where it occurs.

However an ideal data set probably does not exist. For instance the data of Goodson et al. were collected on a total of 22 subjects, with the aim of examining each subject every 28 days. We might require, for inclusion in data sets suitable for our methods, that a subject provide five successive increments in attachment level, with between 14 and 56 days between each observations. Sixteen of the subjects provided data meeting this criterion. There seems little chance of making reliable inference about the covariance structure of data sets based on a maximum of sixteen observations.

Another problem which will make accumulation of a suitable data set difficult is, for populations receiving Western-style dental care, the necessity of treating any patient for whom it is believed that disease is progressing. Clearly, the assumptions of the models could not hold for subjects for whom there has been major intervention in the disease process.

It is to be hoped that our critique of the evidence so far presented for differing theories on the nature of periodontal disease progression, together with the provision of methods which can distinguish between these different possibilities, and application of these methods to data from a major study will provoke further investigation into the nature of the disease.

Various possibilities for further work arise. An obvious possible extension is to use estimation procedures which do not require identical time intervals for each site. We reviewed in section 2.2.3.4 the estimation of covariance structures for multiple groups. This theory is easily applicable to the situation where different

subjects have been measured at different time intervals: a separate sample covariance matrix is formed for each group of sites analysed at the same time intervals. This does however lead to the major restriction that each sample covariance matrix must be nonsingular, so that the number of sites in each group must be at the very least greater than the dimension of the sample covariance matrix. The construction of the GLS estimator is based around the sample covariance matrix so that for the general case where each subject is examined at different time intervals no obvious analogue to the GLS estimator exists. It would, however be possible in this situation to derive MLEs, although the procedures would be computationally rather heavy because of the lack of a sufficient statistic for the parameter vector.

We have shown that, for the GLS (nk) method, parameters estimates produced by the program AUFIT were identical to those of the APL programs for models 1 and 2, which have linear covariance structures. For model 3 the goodness of fit was similar or slightly better than for the APL programs, while the parameter estimates differed (as they did for different starting values for the APL programs. These results confirmed the potential of AUFIT to save the substantial amount of time involved in calculating and writing routines for the derivatives of nonlinear covariance structures.

In the final sections we showed that the power of a study to discriminate between models 1 and 2 was increased faster as p, the number of observations per site, rather than n, the number of sites, increased. This led us to examine the conditions under which MLEs for dependent variables are consistent. We presented such

conditions, gave an example where they are satisfied, and showed that they were not satisfied for an example where a component of the covariance structure is the between-subject variation, but the number of subjects is fixed.

The main problem for our, and others' theorems on asymptotic consistency is the verification of the conditions, which is time-consuming even for simple examples. It would be of interest to provide more easily verifiable conditions. A possibility for this would be to show that the likelihood satisfied conditions for the convergence of dependent variables, by showing that it meets the conditions of one of a number of papers which have recently appeared, for instance that of Andrews (1988), who provided laws of large numbers for L^1 mixingales.

Asymptotic consistency is a weak condition for estimating sequences. It would be of interest to generalise the theorems of section 8 by presenting conditions where the MLEs for dependent variables are asymptotically normally distributed, such as were given by Heijmans and Magnus (1986b) and Sarma (1986).

LITERATURE CITED

Akaike, H., (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19** 716-723.

Anderson, T.W., (1969). Statistical Inference for Covariance Matrices with Linear Structure. *In: P.R. Krishnaiah, (ed.) Multivariate Analysis II.* New York: Academic Press. 55-66.

Anderson, T.W., (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics* **1** 135-141.

Bad, M., (1979). The asymptotic behaviour of the maximum-likelihood estimate for dependent observations and of the Bayes generalized estimate. *Rev. Roum. Math Pures. et Appl.* **24** 847-862.

Badersten, A., Nilveus, R. and Egelberg, J., (1987). 4-year observations of basic periodontal therapy. *Journal of Clinical Periodontology* **14** 448-444.

Baelum, V., Fejerskov., O. and Manji, F., (1988). Periodontal diseases in adult Kenyans. *Journal of Clinical Periodontology* **15** 445-452.

Bentler, P.M., (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures. *Psychometrika* **48** 493-517.

Bentler, P.M. and Bonnett, D.G., (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* **88** 588-606.

Bentler, P.M. and Dijkstra, T., (1985). Efficient estimation via linearization in structural models. *In: Krishnaiah, P.R. (ed.) Multivariate Analysis VI.* Amsterdam: North-Holland. 9-42.

Bentler, P.M. and Weeks, D.G., (1979). Interrelations among nodels for the analysis of moment structures. *Multivariate Behavioral Research* **14** 169-186.

Bentler, P.M. and Weeks, D.G., (1980). Linear structural equations with latent variables. *Psychometrika* **45** 289-308.

Bentler, P.M. and Weeks, D.G., (1985). Some comments on structural equation models. *British Journal of Mathematical and Statistical Psychology* **38** 120-121.

Bentler, P.M., Lee, S.Y., and Weng, L.J., (1987). Multiple population covariance structure analysis under arbitrary distribution theory. *Communications in Statistics A* **16** 1951-1964.

Bhat, B.R., (1979). Strong consistency of maximum likelihood estimator for dependent observations. *Journal of the Indian Statistical Association* **17** 27-39.

Birkedal-Hansen, H., Lantz, M.S. and Cogen, R.B., (1988). A critical assessment of the burst hypothesis. *In: Guggenheim, B. (ed.) Periodontology Today.* Basel: Karger. 227-235.

Blomqvist, N., (1985). On the choice of computational unit in statistical analysis. *Journal of Clinical Periodontology* **12** 873-876.

Blomqvist, N., (1987). On the bias caused by regression toward the mean in studying the relation between change and initial value. *Journal of Clinical Periodontology* **14** 34-37.

Bock, R.D., (1960). Components of variance analysis as a structural and discriminal analysis for psychological tests. *British Journal of Statistical Psychology* **13** 151-163.

Bock, R.D. and Bargmann, R.E., (1966). Analysis of covariance structures. *Psychometrika* **31** 507-534.

Bollen, K.A., (1986). Sample size and Bentler and Bonnett's nonnormed fit index. *Psychometrika* **51** 375-377.

Bragd, L., Dahlén G., Wikström M. and Slots J., (1987). The capacity of Actinobacillus actinomycetemcomitans, Bacteroides gingivalis and Bacteroides intermedius to indicate progressive periodontitis; a retrospective study. *Journal of Clinical Periodontology* **14** 95-99.

Browne, M.W., (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* **8** 1-24.

Browne, M.W., (1982). Covariance Structures. *In: Topics in Applied Multivariate Analysis.* Cambridge: Cambridge University Press. 72-141.

Browne, M.W., (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **37** 62-83.

Browne, M.W., (1988). Automated fitting of nonstandard models. *Technical report - University of South Africa.*

Browne, M.W. and Cudeck, R., (1987). Single sample cross-validation indices for covariance structures. *Research Report 87/5 Department of Statistics, University of South Africa.*

Browne, M.W. and Shapiro, A., (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology* **41** 193-208.

Campbell, D.T. and Fiske, D.W., (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin* **56** 81-105.

Carranza, F.A., (1984). *Glickman's Clinical Periodontology, §23.* London: W.B. Saunders. 309-341.

Cohen, M.E. and Ralls, S.A., (1988). Distributions of periodontal attachment levels, mathematical models and implications. *Journal of Periodontology* **59** 254-258.

Cramér, H., (1946). Asymptotic properties of maximum likelihood estimates. *In: Mathematical methods of statistics, Princeton: Princeton University Press.* 500-506.

Crowder, M.J., (1976). Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society Series B* **38** 45-53.

Cudeck, R., and Browne, M.W., (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research* **18** 147-167.

Dzink, J.L., Socransky, S.S. and Haffajee, A.D., (1988). The predominant cultivable microbiota of active and inactive lesions of destructive periodontal disease. *Journal of Clinical Periodontology* **15** 316-323.

Fidler, V., (1984). Letters to the editor. *Journal of Clinical Periodontology* **11** 138-140,141.

Fleiss, J.L., Park, M.H. and Chilton, N.W., (1987). Within-mouth correlations and reliabilities for probing depth and attachment level. *Journal of Periodontology* **58** 460-463.

Fleiss, J.L., Wallenstein, S. Chilton, N.W. and Goodson, J.M., (1988). A re-examination of within-mouth correlations of attachment level and of change in attachment level. *Journal of Clinical Periodontology* **15** 411-414.

Fraser, C. and McDonald, R.P., (1988). COSAN: Covariance Structure Analysis. *Multivariate Behavioural Research* **23** 263-265.

Glavind, S. and Löe, H., (1967). Errors in the clinical assessment of periodontal destruction. *Journal of Periodontal Research* **2** 180-184.

Goodson, J.M., (1986). Clinical measurements of periodontitis. *Journal of Clinical Periodontology* **13** 446-455.

Goodson, J.M., Tanner, A.C.R., Haffajee, A.D., Sornberger, G.C. and Socransky S.S., (1982). Patterns of progression and regression of advanced destructive periodontal disease. *Journal of Clinical Periodontology* **9** 472-481.

Grimmett, G. and Stirzaker, D., (1982). *Probability and Random Processes.* Oxford: Clarendon Press 164-166.

Gunsolley, J.C. and Best, A.M., (1988). Change in attachment level. *Journal of Periodontology* **59** 450-456.

Haffajee, A.D. and Socransky, S.S., (1986). Frequency distributions of periodontal attachment loss: clinical and microbiological features. *Journal of Clinical Periodontology* **13** 625-637.

Haffajee, A.D., Socransky, S.S. and Lindhe J., (1983a). Comparison of statistical methods of analysis of data from clinical periodontal trials. *Journal of Clinical Periodontology* **10** 247-256.

Haffajee, A.D., Socransky, S.S. and Goodson, J.M., (1983b). Comparison of different data analyses for detecting changes in attachment level. *Journal of Clinical Periodontology* **10** 298-310.

Haffajee, A.D., Socransky, S.S. and Goodson, J.M., (1983c). Clinical parameters as predictors of destructive periodontal disease activity. *Journal of Clinical Periodontology* **10** 257-265.

Haffajee, A.D., Socransky, S.S., Goodson, J.M. and Lindhe, J., (1985). Intraclass correlations of periodontal measurements. *Journal of Clinical Periodontology* **12** 216-224.

Haffajee, A.D., Socransky, S.S., Dzink, J.L., Taubman, M.A., Ebersole, J.L. and Smith, D.J., (1988). Clinical, microbiological and immunological features of subjects with destructive periodontal disease. *Journal of Clinical Periodontology* **15** 240-246.

Hancock, E.B., (1981). Determination of periodontal disease activity. *Journal of Periodontology* **52** 492-499.

Hausmann, E. and Jeffcoat. M., (1988). A perspective on periodontal disease activity measurements. *Journal of Clinical Periodontology* **15** 134-136.

Heijmans, R.D.H. and Magnus, J.R., (1986a). Consistent maximum likelihood estimations with dependent observations. The general (non normal) case and the normal case. *Journal of Econometrics* **32** 253-285.

Heijmans, R.D.H. and Magnus, J.R., (1986b). Asymptotic normality of estimators obtained from normally distributed but dependent observations. *Econometric Theory* **2** 374-412.

Hirsch, R.S. and Clarke, N.G., (1989). Infection and periodontal diseases. *Reviews of infectious diseases* **11** 707-715.

Hirschfield, L. and Wasserman, B., (1978). A long-term survey of tooth loss in 600 treated periodontal patients. *Journal of Periodontology* **49** 225-237.

Imrey, P.B., (1986). Considerations in the statistical analysis of clinical trials in periodontitis. *Journal of Clinical Periodontology* **13** 517-528.

Janssen, P.T.M., Faber, J.A.J. and von Palenstein Helderman, W.H., (1987). Non-Gaussian distribution of differences betweem duplicate probing depth measurements. *Journal of Clinical Periodontology* **14** 345-349.

Jenkins, W.M.M., Macfarlane, T.W. and Gilmour, W.H., (1988). Longitudinal study of untreated periodontitis. (I). Clinical findings. *Journal of Clinical Periodontology* **15** 324-330.

Jennrich, R.I. and Sampson, P.F., (1968). Application of Stepwise Regression to Non-Linear Estimation. *Technometrics* **10** 63-72.

Joreskog, K.G., (1970). A general method for the analysis of covariance structures. *Biometrika* **57** 239-251.

Joreskog, K.G., (1971). Simultaneous factor analysis in several populations. *Psychometrika* **36** 409-426.

Joreskog, K.G., (1973). Analysis of covariance structures. *In: P.R. Krishnaiah (ed.) Multivariate Analysis III.* New York: Wiley. 263-285.

Joreskog, K.G., (1978). Structural analysis of covariance and correlation matrices. *Psychometrika* **43** 443-477.

Joreskog, K.G., (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics* **8** 65-92.

Joreskog, K.G. and Sorbom, D., (1988). *LISREL VII: A guide to the Program and its Applications.* SPSS International B.V. P.O.Box 115, 4200 AC Gorinchem, The Netherlands.

Kaplan, D., (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioural Research* **23** 69-86.

Laster, L. L., (1985). The effect of subsampling sites within patients. *Journal of Periodontal Research* **20** 91-96.

Lee, S.Y. and Bentler, P.M., (1980). Some asymptotic properties of constrained least squares estimation in covariance structure models. *South African Statistical Journal* **14** 121-136.

Lee, S.Y. and Jennrich, R.I., (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika* **44** 99-113.

Lee, S.Y. and Jennrich, R.I., (1984). The analysis of structural equation models by means of derivative free nonlinear least squares. *Psychometrika* **49** 521-528.

Lee, S.Y. and Poon, W.Y., (1985). Further developments on constrained estimation in the analysis of covariance structures. *The Statistician* **34** 305-316.

Lehmann, E.L., (1983). Theory of point estimation. *New York: Wiley.*

Lindhe, J., Okamoto, H., Yoneyama, T., Haffajee, A. and Socransky, S.S., (1989). Periodontal loser sites in untreated adult subjects. *Journal of Clinical Periodontology* **16** 671-678.

Listgarten, M.A., (1986). A perspective on periodontal diagnosis. *Journal of Clinical Periodontology* **13** 175-198.

Löe, H., Anerud, A., Boysen, H., and Smith, M.R., (1978a). The natural history of periodontal disease in man. Study design and baseline data. *Journal of Periodontal Research* **13** 550-562.

Löe, H., Anerud, A., Boysen, H., and Smith, M.R., (1978b). The natural history of periodontal disease in man. Tooth mortality rates before 40 years of age. *Journal of Periodontal Research* **13** 563-572.

Löe, H., Anerud, A., Boysen, H., and Smith, M.R., (1978c). The natural history of periodontal disease in man. The rate of destruction before 40 years of age. *Journal of Periodontology* **49** 607-620.

Löe, H., Anerud, A., Boysen, H., and Morrison, E., (1986). Natural history of periodontal disease in man. Rapid, moderate and no loss of attachment in Sri Lankan Laborers 14 to 46 years of age. *Journal of Clinical Periodontology* **13** 431-440.

Lord, F.M., (1975). Automated hypothesis tests and standard errors for nonstandard problems. *The Amerian Statistician.* **29** 56-59.

MacCallum, R., (1986). Specification searches in covariance structure modeling. *Psychological Bulletin* **100** 107-120.

Manji, F. and Nagelkerke, N., (1989). A stochastic model for periodontal breakdown. *Journal of Periodontal Research* **24** 279-281.

Mardia, K.V., (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57** 519-530.

McArdle, J.J. and McDonald, R.P., (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology* **37** 234-251.

McCullough, C.A.G., Birek, P., Aitken, S., and Lee, W., (1989). Efficacy of Doxycycline in prevention of recurrent periodontitis. *Journal of Dental Research* **68** 916.

McDonald, R.P., (1974). Testing pattern hypothesis for covariance matrices. *Psychometrika* **39** 189-201.

McDonald, R.P., (1975). Testing pattern hypothesis for correlation matrices. *Psychometrika* **40** 253-255.

McDonald, R.P., (1978). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **31** 59-72.

McDonald, R.P., (1980). A simple comprehensive model for the analysis of covariance structures. Some remarks on applications. *British Journal of Mathematical and Statistical Psychology* **33** 161-183.

Michalowicz, B., Virag, J., Klump, D., Hinrichs, J., Aeppli, D. and Philstrom, B., (1989). Periodontal findings in adult twins. *Journal of Dental Research* **68** 954.

Miller, A.J., Brunelle, J.A., Carlos, J.P., Brown, L.J. and Löe, H., (1987). *Oral Health of United States Adults.* Washington: National Institutes of Health Publication Number 87-2868.

Morrison, E.C. and Kowalski, C.J., (1986). Discussion: Clinical measurements of periodontitis. *Journal of Clinical Periodontology* **13** 456-460.

Moscow, B.S., (1978). Spontaneous arrest of advanced periodontal disease without treatment: an interesting case. *Journal of Periodontology* **49** 465-468.

Muthen, B., (1989). Multiple-group structural modelling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology* **42** 55-62.

Newman, M.G., (1979). The role of Bacteroides melaninongenicus and other anaerobes in periodontal infections. *Reviews of infectious diseases* **1** 313-323.

Osborne, J., (1987). The choice of computational unit in the statistical analysis of unbalanced clinical trials. *Journal of Clinical Periodontology* **14** 519-523.

Page, R.C., (1988). Are there convincing animal models for periodontal diseases? *In: Guggenheim, B. (ed.) Periodontology Today.* Basel: Karger. 112-122.

Prasad, M.S. and Prakasa Rao, B.L.S., (1976). Maximum likelihood estimation for dependent random variables. *Journal of the Indian Statistical Association* **14** 75-97.

Press, S.J. (1972). Applied multivariate analysis. *New York: Holt, Rinehart and Winston, inc.* 29.

Ralston, M.L. and Jennrich, R.I., (1978). Dud, a derivative-free algorithm for nonlinear least squares. *Technometrics* **20** 7-14.

Ramfjord, S.P., Caffesse R.G., Morrison, E.C., Hill, R.W., Kerry, G.J., Appleberry, E.A., Nissle, R.R. and Stults, D.L., (1987). 4 modalities of periodontal treatment compared over 5 years. *Journal of Clinical Periodontology* **14** 445-452.

Sarma, Y.R., (1986). Asymptotic properties of maximum likelihood estimates from dependent observations. *Statistics and Probability Letters* **4** 309-311.

Schwarz, G., (1978). Estimating the parameters of a model. *Annals of Statistics* **6** 461-464.

Selikowitz, H.S., Sheiham, A., Albert, D., and Williams, G.M., (1981). Retrospective longitudinal study of the rate of alveolar bone loss in humans using bite-wing radiographs. *Journal of Clinical Periodontology* **8** 431-438.

Shapiro, A., (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal* **17** 33-81.

Shapiro, A., (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology* **37** 84-88.

Shapiro, A., (1985a). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal* **19** 73-81.

Shapiro, A., (1985b). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72** 133-144.

Shapiro, A., (1986). Asymptotic theory of overparametrized structural models. *Journal of the American Statistical Association* **81** 142-149.

Shapiro, A., (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical Journal* **21** 39-62.

Socransky, S.S., Haffajee, A.D., Goodson, J.M. and Lindhe, J., (1984). New concepts of destructive periodontal disease. *Journal of Clinical Periodontology* **11** 21-32.

Sorbom, D., (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology.* **27** 229-239.

Srivastava, J.N., (1966). On testing hypotheses regarding a class of covariance structures. *Psychometrika* **31** 147-164.

Stanley, H.R. Jr., (1955). The cyclic phenomenon of periodontitis. *Oral Surgery* **8** 598-610.

Steiger, J.H., Shapiro, A. and Browne, M.W., (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika* **50** 253-264.

Sterne, J.A.C., (1988). Application of statistical modelling to periodontal diseases as viewed by the statistician. *In: Guggenheim, B. (ed.) Periodontology Today.* Basel: Karger. 86-94.

Sterne, J.A.C., Curtis, M.A., Gillett, I.R., Griffiths, G.S., Maiden, M.F.J., Wilton, J.M.A. and Johnson, N.W., (1990). Statistical models for data from periodontal research. *Journal of Clinical Periodontology* **17** in press.

Stone, M., (1974). Cross-validatory choice and assessment of statistical predictions (with Discussion). *Journal of the Royal Statistical Society, Series B* **36** 111-147.

Stone, M., (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series B* **39** 41-47.

Stroud, T.W., (1972). Fixed alternatives and Wald's formulation of the noncentral asymptotic behaviour of the likelihood ratio statistic. *Annals of Mathematical Statistics* **43** 447-454.

Swain, A.J., (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika* **40** 315-335.

Tanaka, J.S. and Huba, G.J., (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology* **38** 197-201.

Votaw, D.F, (1948). Testing compound symmetry in a normal multivariate population. *Annals of Mathematical Statistics* **19** 447-473.

Wald, A., (1949). Note on the consistency of the maximum likelihood estimator. *Annals of Mathematical Statistics* **20** 595-601.

Wiley, D.E., Schmidt, W.H. and Bramble, W.J., (1973). Studies of a class of covariance structure models. *Journal of the American Statistical Association* **68** 317-323.

Wilks, S.S, (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics* **17** 257-281.