The analysis of repeated ordinal data using latent trends

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

• A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/13772

Publisher: © Justin Skinner

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
Pilkington Library

Author/Filing Title: Skinner

Vol. No. Class Mark

Please note that fines are charged on ALL overdue items.

26 JUL 2000

Loan copy
‘The Analysis of Repeated Ordinal Data using Latent Trends’

by

Justin Skinner B.Sc. (Honours)

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy of Loughborough University

February 19, 1999

© by Justin Skinner 1999
This thesis presents methodology to analyse repeated ordered categorical data (repeated ordinal data), under the assumption that measurements arise as discrete realisations of an underlying (latent) continuous distribution. Two sets of estimation equations, called quasi-estimation equations or QEEs, are presented to estimate the mean structure and the cutoff points which define boundaries between different categories. A series of simulation studies are employed to examine the quality of the estimation processes and of the estimation of the underlying latent correlation structure. Graphical studies and theoretical considerations are also utilised to explore the asymptotic properties of the correlation, mean and cutoff parameter estimates. One important aspect of repeated analysis is the structure of the correlation and simulation studies are used to look at the effect of correlation misspecification, both on the consistency of estimates and their asymptotical stability. To compare the QEEs with current methodology, simulations studies are used to analyse the simple case where the data are binary, so that generalised estimation equations (GEEs) can also be applied to model the latent trend. Again the effect of correlation misspecification will be considered. QEEs are applied to a data set consisting of the pain runners feel in their legs after a long race. Both ordinal and continuous responses are measured and comparisons between QEEs and continuous counterparts are made. Finally, this methodology is extended to the case when there are multivariate repeated ordinal measurements, giving rise to inter-time and intra-time correlations.
Dedicated to
Mum, Dad
and Nina
for their
love and
support.
Numerous people have been instrumental to the completion of this thesis and I would like to thank everyone of them wholeheartedly.

First and foremost, Ioannis Vlachonikolis. He has been a constant source of help and advice in every aspect of my research. He has also motivated me when required and given me an appreciation for the beauty of statistics. To Ioannis I owe this thesis.

Richard Buxton and David Green also deserve a big vote of thanks, since they have both been invaluable sources of help with the administrative side of things. I would also like to thank the Mathematical Sciences department of Loughborough University for giving me the resources to carry out this research and the EPSRC for providing the necessary funding.

On the non-academic side, I would like to thank my parents, for their support. Not just in the past few years but as long as I can remember. Also a big thank you to Nina Wright, friend and girlfriend, who has always been there when things were not going as they should. The final vote of thanks must go to my friends who have kept me in tune with the lighter side of life, thus reducing the stress of research.
# Contents

Abstract

Acknowledgements

Contents

List of Tables

List of Figures

1 Introduction

1.1 Models for continuous and categorical data .............................................. 2

1.1.1 Maximum Likelihood ................................................................. 3

1.1.2 Generalised Estimation Equations ................................................... 4

1.1.3 Random effects models ............................................................... 5

1.2 Models for categorical data ............................................................... 6
1.2.1 Markov models ............................................. 7
1.2.2 Multinomial models ........................................ 8
1.2.3 Generalised Logit Model ................................. 8
1.3 Methods for ordinal data ...................................... 9
  1.3.1 Proportional Odds ........................................ 9
  1.3.2 GEE marginal methods ................................... 10
1.4 Latent variables .............................................. 10
  1.4.1 Factor analysis .......................................... 11
  1.4.2 Latent trends ........................................... 11
1.5 Outline of thesis ............................................ 12

2 Likelihood Analysis ............................................ 14
  2.1 Notation .................................................... 14
  2.2 Maximum Likelihood estimation ............................ 15
  2.3 Restricted Maximum Likelihood Estimation ............... 17
  2.4 Models for the variance matrix ........................... 18
  2.5 Application - 'Milk data' .................................. 22

3 Generalised Estimation Equations (GEEs) ..................... 28
  3.1 Theory ................................................. 28
  3.2 Estimation using GEEs .................................... 30
  3.3 Different ‘working correlation’ matrices ............... 31
  3.4 An application of GEEs to latent trend analysis ....... 33
  3.5 Proof of the GEEs asymptotics ............................ 38

4 Cut-off point estimation ....................................... 42
  4.1 Quasi-estimation equations (QEEs) ....................... 43
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1</td>
<td>Modelling cut-off points</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Quasi-estimation equations for $\beta$</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Quasi-estimation equations for $\gamma$</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Non-linear links</td>
</tr>
<tr>
<td>4.1.5</td>
<td>Computation of QEEs estimates</td>
</tr>
<tr>
<td>4.1.6</td>
<td>Simulation studies</td>
</tr>
<tr>
<td>4.2</td>
<td>Asymptotics of the QEE estimates</td>
</tr>
<tr>
<td>4.2.1</td>
<td>$\beta$ estimates</td>
</tr>
<tr>
<td>4.2.2</td>
<td>$\gamma$ estimates</td>
</tr>
<tr>
<td>4.2.3</td>
<td>$\alpha$ estimate</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison with GEEs</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Correlation Miss-specification</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Stability</td>
</tr>
<tr>
<td>4.4</td>
<td>Runners example</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Graphical analysis</td>
</tr>
<tr>
<td>4.4.2</td>
<td>ANOVA analysis</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Correlation analysis</td>
</tr>
<tr>
<td>4.4.4</td>
<td>ML/REML analysis</td>
</tr>
<tr>
<td>4.4.5</td>
<td>GEEs analysis</td>
</tr>
<tr>
<td>4.4.6</td>
<td>QEEs analysis</td>
</tr>
<tr>
<td>4.4.7</td>
<td>Model comparisons</td>
</tr>
<tr>
<td>4.4.8</td>
<td>Model conclusions</td>
</tr>
<tr>
<td>5</td>
<td>Multivariate responses</td>
</tr>
<tr>
<td>5.1</td>
<td>Methodology</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Correlation structure</td>
</tr>
</tbody>
</table>
5.1.2 Model formulation .............................................. 123

5.2 Simulation studies ............................................... 124
  5.2.1 Correlation Miss-specification ............................ 133

5.3 Asymptotics ....................................................... 140
  5.3.1 \( \beta \) estimates ........................................... 141
  5.3.2 \( \gamma \) estimates ......................................... 145
  5.3.3 \( \alpha \) estimates ......................................... 147

5.4 Runners example ................................................. 151

6 Conclusions and Further Work ................................. 153
  6.1 Chapter 3 ....................................................... 153
  6.2 Chapter 4 ....................................................... 154
  6.3 Chapter 5 ....................................................... 155
  6.4 Further work ................................................... 156

A Results from the 'milk' data .................................. 158
  A.1 No missing value imputation ................................. 159
  A.2 Time averaged imputation ................................... 162
  A.3 Moving average imputation .................................. 165

B Multi-dimensional calculus results ................................ 168
List of Tables

1.1 Contingency table showing probabilities of different joint outcomes 8

2.1 Grouping of the cows to different diets 22

2.2 Mean models for cows from different groups 25

3.1 GEE estimates of the underlying exponential distribution 35

3.2 GEE estimates of the independent underlying exponential distribution under correlation miss-specification 36

3.3 GEE estimates of the 1-dependent underlying exponential distribution under correlation miss-specification 36

3.4 GEE estimates of the exchangeable underlying exponential distribution under correlation miss-specification 37

3.5 GEE estimates of the exponential underlying exponential distribution under correlation miss-specification 37

4.1 Specification of the simulation study for QEEs 53
4.2 QEEs estimates of the underlying multivariate normal distribution, using ML for $\alpha$ and QEEs for $\beta$ and $\gamma$ ................................................................. 54
4.3 QEEs estimates of the underlying independent multivariate normal distribution under correlation miss-specification ......................................................... 66
4.4 QEEs estimates of the underlying 1-dependent multivariate normal distribution under correlation miss-specification ......................................................... 67
4.5 QEEs estimates of the underlying exchangeable multivariate normal distribution under correlation miss-specification ......................................................... 68
4.6 QEEs estimates of the underlying exponential multivariate normal distribution under correlation miss-specification ......................................................... 69
4.7 QEEs and GEEs estimates using binary realisations from an underlying multivariate normal distribution ................................................................. 88
4.8 QEEs and GEEs estimates using binary realisations from an underlying independent multivariate normal distribution ......................................................... 90
4.9 QEEs and GEEs estimates using binary realisations from a one-dependent underlying multivariate normal distribution ......................................................... 91
4.10 QEEs and GEEs estimates using binary realisations from an exchangeable underlying multivariate normal distribution ......................................................... 92
4.11 QEEs and GEEs estimates using binary realisations from an exponential underlying multivariate normal distribution ......................................................... 93
4.12 Likert Scale of Muscle Soreness ............................................................. 101
4.13 ANOVA table for the 1st VAS score with the full model .......................... 110
4.14 ANOVA table for the 1st VAS score with the reduced model .................. 110
4.15 ANOVA table for the VAS score (assuming independence) with the full model 111
4.16 ANOVA table for the VAS scores (assuming independence) with the reduced model ................................................................. 111
4.17 Parameter estimates using GEEs .......................................................... 113
4.18 Parameter estimates using ML .............................................................. 114
4.19 Parameter estimates using REML .......................................................... 115

xii
4.20 Parameter estimates using GEEs ........................................... 116
4.21 Parameter estimates using QEEs ........................................... 117
4.22 VAS score cut-off points when compared to the Likert scores .... 118
4.23 Parameter estimates using QEEs when transformed to fit the first and last VAS cut-off value ........................................... 119

5.1 Specification of the simulation study for repetitive measurements ......... 124
5.2 Simulation parameter estimates of the underlying distribution, using an independent (uncorrelated) repetitive structure .............. 125
5.3 Simulation parameter estimates of the underlying distribution, using a correlated repetitive structure ........................................ 126
5.4 Simulation parameter estimates of the underlying distribution, using an independent (uncorrelated) repetitive structure and exchangeable repeated correlation ........................................ 133
5.5 Simulation parameter estimates of the underlying distribution, using a correlated repetitive structure and exchangeable repeated correlation ........................................ 134
5.6 Parameter estimates treating the ordinal and continuous data as having the same factor effects ........................................ 152
List of Figures

2.1 Shadowed time series plots for the three groups of cows .......................... 23

2.2 Typical mean trend for a typical cow ...................................................... 25

3.1 Representation of discrete data arising as realisations of a continuous distribution 33

4.1 Stability plots for the QEEs estimates of the correlation, mean and cut-off point parameters of the underlying exchangeable multivariate normal distribution ............................................................... 58

4.2 Consistency plots for the QEEs estimates of the correlation, mean and cut-off point parameters of the underlying exchangeable multivariate normal distribution ............................................................... 64

4.3 Consistency plots for the QEEs estimates of the mean and cut-off point parameters of the underlying exchangeable multivariate normal distribution under correlation miss-specification ............................................................... 73

4.4 Stability plots of $\beta$ for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution .......................... 95

4.5 Consistency plots for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution .......................... 97
4.6 Consistency plots for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution under correlation miss-specification ........................................... 100

4.7 Histograms of the Likert and VAS observed values, for the 10 times .......... 106

4.8 Shadow plot of the first 5 VAS scores ............................................. 107

4.9 Shadow plot of the difference between the second to fifth VAS scores and the first VAS score ................................................................. 108

4.10 Correlation structures compared to sample correlations obtained from a fully unstructured GEEs estimation ............................................... 112

5.1 Consistency plots for the repetitive QEEs estimates of the correlation, mean and cut-off point parameters of the underlying multivariate normal distribution 132

5.2 Consistency plots for the repetitive QEEs estimates of the correlation, mean and cut-off point parameters of the underlying multivariate normal distribution when there is repetitive and repeated correlation miss-specification ...... 139
CHAPTER 1

Introduction

The analysis of repeated ordinal data is a broad statistical area, which is surprisingly lacking in methodology. In this thesis new methods will be formulated and investigated that will offer effective estimation of parameters. Initially we will look at established methods and work that has already been carried out in this area, so as to give an idea of the breadth of applications and also the need for further work into the subject. Repeated ordinal data occurs in a vast number of areas; agriculture, business, economics, medicine and sociology being just a few. It could be anything from two yes/no answers from individuals at different times to more complex situations with observations being made on numerous individuals over numerous occasions, with numerous possible responses.

'Repeated ordinal data' encompasses two themes, which are (obviously) 'repeated' and 'ordinal'. The term repeated implies that measurements are taken on a group of individuals at more than one occasion. As well as time, they could also be taken over space, such as the breaking strength of a cable being taken at various points along its length. However, for the purposes of this thesis the repeated nature of the data will be assumed (without loss of generality) to be time. In the repeated measures environment the observations taken on an individual will not be independent, so normal regression techniques can not be used. For example, if an individual starts off with a relatively high value, then it will tend to continue
to have high measurements at later times. If they were independent then there would be no reason for the statistician to consider the measurements as arising from individuals, since considering the entire set of data as separate measurements would give the same results. This correlation is an important feature of repeated measures that makes it both interesting and demanding. In the words of Laird, in the discussion of Zeger (1988) [84]

'...you were going to tell us the bad news and then the good news about longitudinal (repeated) data. I actually thought that the bad news was the good news. That is, you told us the bad news was that the data are dependent. But from certain perspectives, it seems to me, that is also the good news.'

Ordinal refers to the fact that the data are distinct categories which have some implied ordering, an example being the classification of a medical disease into none, little and severe. The ordering is another piece of information that we have about the data, so to ignore it would lead to a loss of 'information' and weaker results. Therefore we need to have methodology that takes this ordering into account. Although this thesis will focus on ordinal data, some of the preliminary work concerns categorical data (of which ordinal data is a special case). A categorical variable is one for which the possible values consist of a set of categories, for example yes/no answers (binary data). From this general definition there are sub-groups. If the categories have some natural implied ordering then the data is ordinal data (as already mentioned). If the categories do not have an ordering then it is nominal, such as gender. The ordering of an ordinal variable will have an effect on any statistical analysis carried out, whereas the ordering of a nominal variable will not. The last sub-grouping is interval data which consists of responses being banded. For instance, the income of people could be grouped into less than £10,000, between £10,000 and £20,000 and greater than £20,000. The type of the categorical data will determine what type of statistical analysis can be carried out on it. Methodology for nominal data can be used to analyse ordinal or interval data, and methods for ordinal data can be used for interval data. However, the opposites are not true.

1.1 Models for continuous and categorical data

There are several methods commonly used for both repeated continuous and categorical data. Although this thesis is based around repeated ordinal data analysis, we will look at continuous case as well since later work will be based on this. Repeated continuous data analysis is an established field with several powerful methods of parameter estimation. If we can extend the continuous case to the categorical case, then we may be able to extend this further to the ordinal case, which is the goal of this thesis.
1.1.1 Maximum Likelihood

Many of the results in repeated data analysis use maximum likelihood (ML) analysis. This is a general method that only requires the assumption of an underlying probability distribution for the data, a natural choice in the repeated measures environment being the multivariate normal. We maximise the total likelihood (or more usually the total log-likelihood) over the parameters to obtain estimates. Being such a general method, which can be applied to any estimation process rather than specifically repeated measures, it has several drawbacks. The main one concerns a trade off between bias and consistency. Consider a standard linear regression model in matrix notation, where $X$ represents the effects of a set of parameters, $\beta$ on the mean. If we have a nearly saturated model ($X$ has a large number of columns) then the estimates of the covariance structure will not be consistent. However, to obtain unbiased estimates of the mean parameters, $\beta$, we require there to be few parameters in the model (and therefore $X$ to have a small number of columns). This conflict brought about a new method being proposed by Patterson and Thompson (1971) [65], the method of restricted maximum likelihood estimation (REML). The use of REML to estimate the parameters counteracted these problems by linearly transforming the data, $y$, to be independent of the parameters. By doing so, the estimation process no longer depends directly on the parameters, so the size of the design matrix was not a problem. The multivariate normal case is analysed and in this case the method works well. However, when using categorical repeated measurements, the method has drawbacks. This is because it would be difficult to transform the model to be independent of the parameters, as required to carry out REML. Therefore the estimates will still be either biased or inconsistent depending on the number of columns in the design matrix. Although ML is an extremely effective technique for continuous data (due to the multivariate normal distribution), no flexible multidimensional categorical models exists, so this method is difficult to apply to repeated categorical data. Zeger (1988) [84] says

'The discrete longitudinal (repeated) data problem is also harder. There is not a multivariate distribution for discrete outcomes as flexible as the multivariate Gaussian, the basis of linear model theory. For discrete data, measures of location (mean) and of dependence (covariance) are not separable as they are in the Gaussian case. Discrete data models which attempt to account for dependence can therefore be more complicated to work with.'

This lack of obvious models for categorical and ordinal data means that specially formulated models need to be used which take into account the correlations between successive observations. Some of these models will be discussed later. Once we have a likelihood to work with, the application of ML becomes easy, with the simple numerical technique of Neill's
SIMPLEX algorithm (1971) [62] usually being sufficient. If the standard errors of parameter estimates are required, the more complicated multidimensional Newton-Raphson algorithm may be used. However, in general ML will be simple to implement.

1.1.2 Generalised Estimation Equations

Generalised estimation equations (GEEs) were proposed by Liang and Zeger (1986) [45] and further investigated by Zeger and Liang (1986) [87] and Zeger, Liang and Albert (1988) [88]. They are an extension of the generalised linear models of Nelder and Wedderburn (1972) [64] and are applied when the univariate distributions (marginals) are members of a generalised exponential family, with means which can be modelled as a function of linear combinations of explanatory variables, \( \mu = g(X \beta) \). In addition, a ‘working correlation’ matrix is needed which gives an indication of the relationship between the repeated measurements. It should be noted that this matrix is not the true correlation, but only a working approximation to it which is called the tetra-choric correlation. In the non-normal case the actual correlation matrix may depend on the mean parameters, but the working correlation matrix is assumed to be independent of \( \beta \). The estimation of the marginal parameters, \( \beta \), is carried out by solving the following estimation equations with respect to \( \beta \) where \( V \) is the working covariance matrix, defined by the working correlation matrix, \( R \) and the variances of the marginal distributions.

\[
\sum_{i=1}^{m} \left( \frac{\partial \mu}{\partial \beta} \right)^T V^{-1}(y)(y - \mu) = 0
\] (1.1)

They offer a framework for the analysis of both continuous and categorical repeated measurements, requiring few assumption concerning the data. Liang and Zeger (1986) [45] and Gourieroux, Monfort and Trognon (1984) [32] showed that if the correlation is estimated \( m^{1/2} \)-consistently, then the mean parameter estimates, \( \hat{\beta} \), are asymptotically unbiased and as efficient as if the correlation was correctly specified, meaning that the method is robust to slight miss-specification of the variance structure. That is to say that the loss of efficiency from an incorrect specification of the working correlation matrix, \( R \), will not be consequential when there is a large number of subjects. It is interesting to note that a choice of an independent working correlation matrix, which makes GEEs equivalent to fitting a regression that does not take into account the repeated nature of the data, actually leads to consistent estimates of \( \beta \). Standard methodology, such as simple ML using univariate distributions, which also does not take into account the correlated repeated measures leads to inconsistent estimates. Liang and Zeger (1986) [45] suggested the use of moments to estimate the correlation structure. However, this is a weak method which can provide inaccurate estimates. Prentice (1988) [66] considered the problem of moment estimation and extended GEEs by
considering a second set of estimation equations to simultaneously estimate the variance parameters along with the mean parameters. These are called GEEs2.

\[
\sum_{i=1}^{m} \left( \frac{\partial E(W_i)}{\partial \alpha} \right)^T H^{-1} [W_i - E(W_i)] = 0 \tag{1.2}
\]

where

\[
W_i = \begin{pmatrix}
Y_{i1} - \mu_{i1} \\
\sqrt{\mu_{i1}(1 - \mu_{i1})} \\
\vdots \\
Y_{in} - \mu_{in}
\end{pmatrix}
\]

\[
H_i = \text{diag} \left[ V \left( \frac{Y_{i1} - \mu_{i1}}{\sqrt{\mu_{i1}(1 - \mu_{i1})}} \right), \ldots, V \left( \frac{Y_{in} - \mu_{in}}{\sqrt{\mu_{in}(1 - \mu_{in})}} \right) \right]
\]

According to the particular model that is employed, \( H \) and \( W \) can be simplified. Prentice further showed that the solution to the second set of estimation equations gives asymptotically unbiased and consistent estimates of the correlation parameters and the mean parameters (\( \alpha \) and \( \beta \) respectively). Although this looks simple, in practice likelihood estimation of the correlation structure is difficult in the non-binary case. The expectations have no simple form and therefore the method becomes messy.

Once again, although the generalised linear model is quite general, there is little scope for categorical data. This is because categorical data tends to be method dependent (binary if you have only two possible outcomes, Poisson if you have count data etc.) and a rich class of models is not freely available. Continuous data, on the other hand, has the univariate normal distribution which offers a flexible set of modelling possibilities. Ordinal data has even fewer possibilities, with the binary case being the only commonly used ordinal distribution which is a member.

1.1.3 Random effects models

Another method for dealing with categorical and continuous repeated measures concerns the application of random effects models. It is assumed that data arise as independent observations from a generalised exponential family where the mean vector for any individual depends on both a set of common parameters and a vector of subject-specific parameters, i.e.

\[
h(\mu_{ij}|U_i) = X_{ij}\beta + d_{ij}U_i \tag{1.3}
\]

where \( h \) is a known link function. It is further assumed that these subject-specific parameters, \( U_i \), are mutually independent observations from a common underlying multivariate distribution. As an example, if \( d_{ij} = 1 \), then the random effect would represent a random intercept by which all measurements on that individual are changed relative to the population average.
This considers the natural heterogeneity amongst individuals from factors which have not been taken (or can not be taken) into account by assuming they are a subset of the parameters, such as intercepts, which can be represented by some known probability distribution. The correlation amongst repeated measurements is also accounted for by the distribution of $U_i$. Since we are dealing with individuals, rather than populations, this method leads to subject specific models, rather than population averaged models, Zeger, Liang and Albert (1988) [88]. The method is therefore used when the focus of the statistician is on individual subjects, rather than some population averaged effect. ML methods can be used to estimate the mean parameters and the correlations between the random effects.

1.2 Models for categorical data

Some methods are dedicated to modelling repeated categorical data. These were brought about because the previous methods left little scope for model flexibility when categorical data is used. The method that is used depends on the focus of the study. This can be split up into two basic subgroups, transitional and marginal. The former looks at an individual's category changes between each repeated time-step and the latter looks at how the marginal distributions changes over the different time periods. Put simply, the transitional models focus on how an individual changes over time and the marginal models focus on population averaged effects. There is much discussion concerning which of these two types of analysis should be used, which goes beyond considering the objective of the study. For instance, Ware, Lipsitz and Speizer (1988) [79] say

'...marginal models do not fully utilise the information in longitudinal data. In particular, they do not model individual changes over time or the effects of co-variates on individual changes.'

whereas Stram, Wei and Ware (1988) [74] comment that

'...simple models for the transition probabilities lead to complex models for marginal probabilities.'

'In addition, if the treatments of exposures under study affect the initial state as well as the transitions of study participants between states, models for transition probabilities may not capture this information.'
1.2.1 Markov models

For this type of model, it is assumed that

\[ H_{ij} = (y_{i1}, \ldots, y_{i,j-1}) \]
\[ \mu_{ij} = E(Y_{ij}|H_{ij}) \] (1.4)
\[ \nu_{ij} = \text{var}(Y_{ij}|H_{ij}) = \nu(\mu_{ij}) \phi \] (1.5)
\[ h(\mu_{ij}) = X_{ij}\beta + \sum_{r=1}^{s} f_{r}(H_{ij}; \alpha) \]

Equations 1.4 and 1.5 are the conditional mean and variance respectively of \( Y_{ij} \) given the past responses and the explanatory variables. The transitional functions, \( f_r \), are modelled with the transformed past outcomes, \( f_r(H_{ij}) \), being the explanatory variables. Several authors have considered different types of links, \( h \), to model different scenarios.

- **Linear Link**
  Tsay (1984) [77] considered a linear link, with \( h(\mu_{ij}) = \mu_{ij} \) and \( f_r = \alpha_r(Y_{ij-r} - X_{ij-r}\beta) \). This form gives a model of the type

\[ Y_{ij} = X_{ij}\beta + \sum_{r=1}^{s} \alpha_r(Y_{ij-r} - X_{ij-r}\beta) + Z_{ij} \]

where \( Z_{ij} \sim N(0, \sigma^2) \). The formulation of the link implies that the current observation depends on its mean and the deviations of previous measurements from their means.

- **Logit Link**
  Korn and Whittemore (1979) [41] and Zeger, Liang and Self (1985) [89] fit a logit link to binary data, to give

\[ \text{logit}[P(Y_{ij} = 1|H_{ij})] = X_{ij}\beta + \sum_{r=1}^{s} \alpha_r Y_{ij-r} \]

for the \( s \)th order Markov chain.

- **Log-linear Link**
  If there is count data a log-linear model may be assumed to fit the data. Zeger and Qaqish (1988) [90] discuss a 1st order Markov chain where \( f_1 = \alpha \{ \log[\text{max}(Y_{ij-1}, c)] - X_{ij-1}\beta \} \) with \( 0 < c < 1 \) to yield

\[ \mu_{ij} = e^{X_{ij}\beta} \left[ \frac{\text{max}(Y_{ij-1}, c)}{e^{X_{ij-1}\beta}} \right]^\alpha \] (1.6)
1.2.2 Multinomial models

Consider the simple 2 by 2 contingency table in table 1.1.

<table>
<thead>
<tr>
<th>Response at time 1</th>
<th>Response at time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p_{11}$ $p_{12}$ $p_{13}$ $p_{14}$ $\pi_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>$p_{21}$ $p_{22}$ $p_{23}$ $p_{24}$ $\pi_{12}$</td>
</tr>
<tr>
<td>3</td>
<td>$p_{31}$ $p_{32}$ $p_{33}$ $p_{34}$ $\pi_{13}$</td>
</tr>
<tr>
<td>4</td>
<td>$p_{41}$ $p_{42}$ $p_{43}$ $p_{44}$ $\pi_{14}$</td>
</tr>
</tbody>
</table>

Table 1.1: Contingency table showing probabilities of different joint outcomes

The marginals, $\pi_{jk}$, are obtained by summing the probabilities in that row/column, $\pi_{1j} = \sum_{k=1}^{4} p_{jk}$ and $\pi_{2j} = \sum_{k=1}^{4} p_{kj}$. Multinomial models assume that the underlying joint distribution of the response variable, $Y = (Y_1, Y_2)^T$, has a multinomial distribution for each different set of covariates, $X$. The approach taken by Grizzle, Starmer and Koch (1969) [33], Koch, Landis, Freeman, Lehnen (1977) [40] and Landis, Miller, Davis and Koch (1988) [43] is to estimate these multinomial cell probabilities for each level of $X$. To do this, the sample probabilities and a weighted least squares procedure to obtain ML estimates of the 16 parameters, $P(X) = [\tilde{p}_{11}(X), \ldots, \tilde{p}_{44}(X)]^T$ with the constraint $\sum_{j=1}^{4} \sum_{k=1}^{4} \tilde{p}_{jk} = 1$. These turn out to be the observed proportions, $\tilde{p}_{jk} = p_{jk}$. The variance of $\tilde{P}(X)$ can also be obtained as a standard result of multinomial models.

This method has several inherent problems. Firstly, the data have to be stratified into sub-populations which have the same set of covariates, $X$. Although this is easy if there are a few factors, if there are continuous covariates this gives rise to lots of small sub-samples and poor estimates. Secondly, by stratifying the sample we obtain several sets of estimates and have to find a method to join these together to find an overall 'population averaged' parameter such as weighted least squares as suggested by Landis, Miller, Davis and Koch (1988) [33].

1.2.3 Generalised Logit Model

Suppose we look at the individual probabilities of a categorical response at any time. If we choose a ‘baseline’ response then we can describe the odds of making each response relative to this and construct a model that describes how generalised logits depend on the repeated
measurements and a set of covariates. Suppose that $\phi_{ik}(j, X)$ represents the probability of response $k$ at time $j$ for individual $i$. If we choose the first category as the baseline then

$$L(t; X) = \log \frac{\phi_{ik}(j, X)}{\phi_{i1}(j, X)}$$ (1.7)

represents the generalised logit of individual $i$ at time $j$ making response $k$. It is no easy task to maximise this likelihood since it depends on marginal probabilities rather than individual probabilities. Instead an iterative routine is used to maximise the full likelihood subject to the marginal distributions satisfying the specified model. Weighted least squares is an example of an algorithm to fit such models as suggested in Agresti (1996) [1].

1.3 Methods for ordinal data

Within the field of categorical data there are special cases, which have already been described. We are focussing on repeated ordinal data, where there is a natural ordering to the data. This means there is an extra piece of information about the data, which improves methodology.

1.3.1 Proportional Odds

The basic framework of multinomial models has already been discussed. Mark and Gail (1994) [52] extended this methodology for the analysis of repeated ordinal data by looking at the logit of the cumulative probabilities at each occasion. Define $L_{jk}$ as the $k$th logit at time $j$

$$L_{jk}[\hat{P}(X)] = \log \frac{\hat{p}_{j1}(X) + \cdots + \hat{p}_{jk}(X)}{\hat{p}_{j,k+1}(X) + \cdots + \hat{p}_{jt}(X)}$$

We model these logits using a vector of known functions, $F(L_{jk})$. These lead to a weighted least squares estimate of the parameters, $\hat{\beta}$, of

$$\hat{\beta} = (X^TV_P^{-1}X)^{-1}X^TV_P^{-1}F(\hat{P})$$

which provides fully efficient unbiased estimates of the parameters. This simple model forms the basis for the 'proportional odds' model. If we have a categorical outcome, which is non-ordinal, one possible method of analysis is to use generalised logits. For ordinal data, rather than modelling odds relative to a 'baseline' category we could model cumulative logits, as suggested by McCullagh (1980) [56] and then estimate the parameters by ML.

$$L_k(t; X) = \log \frac{\phi_1(t; X) + \cdots + \phi_k(t; X)}{\phi_{k+1}(t; X) + \cdots + \phi_l(t; X)}$$ (1.8)
This gives the odds of getting a category of \( k \) or less compared to category \( k+1 \) or higher. This was shown to be a multivariate extension of the generalised linear models of Nelder and Wedderburn (1972) [64]. The estimation procedure based on ML is computationally heavy and McCullagh showed that even when there are non-linear models, the method of iterative weighted least squares converged to the ML estimates, which simplifies the calculations. Agresti and Lang (1993) [3] furthered the proportional odds model by considering a generalised Rasch model and then fitting proportional odds models to all possible binary collapsings of the response. They proposed the use of a Newton-Raphson algorithm to fit the model subject to a set of constraints, which produced consistent estimation of the subject-specific parameters.

1.3.2 GEE marginal methods

One member of the generalised exponential family is the multinomial distribution. Given a set of marginals, such as those in table 1.1, GEEs can be applied to model the mean according to a set of covariates. However, the working correlation has to be estimated and this causes problems. One way around this is to use the independence GEEs model (so no working correlation is necessary) to find estimates. This method works since the GEEs estimates are consistent even if the correlation is miss-specified (as it probably would be by assuming independent observations). Stram, Wei and Ware (1988) [74] recommend using the independent GEEs to find estimates for parameters which change over time ("treatment" effects). These different estimates are then joined using weighted-least squares, where the weights are derived from the estimate covariance matrix of the parameters. Mark and Gail (1994) [52] argued that the assumption of independence is naive and recommended using an empirical estimate of the variance/covariance matrix as an unstructured working correlation, with Lipsitz, Kim and Zhao (1994) [48] producing correlation estimates for different structures based on this sample variance/covariance matrix.

1.4 Latent variables

The method of latent variables is a useful and intuitive method to model the distribution of discrete ordinal data. In many areas of science we look at a variable that is not directly measurable. For instance, although we use the IQ scale as a measure of intelligence, there is no direct measurement that can be made. Another example, which will feature in this thesis, is the measuring of pain. Although an individual might give a response to how much pain they can feel (no pain, little pain, etc.) we can not measure directly the exact pain
that they are feeling. The main methods for latent variable analysis consider the correlations between a set of observed variables. This leads to finding a set of 'loadings' which describe the correlations in terms of some underlying latent and non-latent variables.

### 1.4.1 Factor analysis

Assume that we have \( p \) variables, \( x = (x_1, \ldots, x_p)^T \), which we wish to model as \( x = \Lambda d + u \), where \( u = (u_1, \ldots, u_p)^T \) are the residuals which are assumed to be uncorrelated with each other and independent of the latent variables. This means that the variance matrix of \( x \) is given by

\[
\Sigma = \Lambda \Phi \Lambda^T + \Psi
\]

where \( \Psi \) is a \( p \times p \) diagonal matrix containing the variances of the residuals, \( u \). Lawley (1940s) proposed estimation of the parameters by considering the observed covariance matrix, \( S = \sum_{i=1}^n x_i x_i^T/(n-1) \). From this a likelihood is obtained which is maximised, \( L = -[\log |\Sigma| + \text{trace}(S \Sigma^{-1})] \). Further work was carried out by Joreskog (1973) [38] and Wiley (1973) [83]. They developed a new model, for which the factor analysis model was a special case, called 'LISREL' model (LInear Structural RELationship). As well as having the independent latent variables they additionally assumed that some of them were dependent, \( \eta = B \eta + \Gamma \xi + \xi \), where \( \eta = (\eta_1, \ldots, \eta_p)^T \) are dependent latent variables and \( \xi = (\xi_1, \ldots, \xi_m)^T \) are the independent latent variables which are linked to the observed variables as \( y = \Lambda_y \eta + \epsilon \) and \( x = \Lambda_x \xi + \delta \) with the vectors \( \epsilon = (\epsilon_1, \ldots, \epsilon_q)^T \) and \( \delta = (\delta_1, \ldots, \delta_p)^T \) being the residuals. With this formulation they calculated the variance matrices to be

\[
\begin{align*}
\Sigma_{xx} &= \Lambda_x \Phi \Lambda_x^T + \theta_x \\
\Sigma_{yy} &= \Lambda_y (B^*)^{-1}(\Gamma \Phi \Gamma^T + \Psi) (B^*)^{-1} \Lambda_y^T + \theta_x \\
\Sigma_{xy} &= \Lambda_y (B^*)^{-1} \Gamma \Phi \Lambda_x^T 
\end{align*}
\]

with \( B^* = I - B, \, \Phi = E(\xi \xi^T), \, \theta_x = E(\delta \delta^T), \, \Psi = E(\xi \xi^T) \) and \( \theta_x = E(\epsilon \epsilon^T) \). To estimate the parameters a function of the sample variance matrix was optimised, usually the likelihood as in the previous section. Both of these models assume that the data are continuous, although some work has been carried out into the discrete case. Bock and Lieberman (1970) [8], Christofferson (1975) [13] and Muthen (1978) [60]

### 1.4.2 Latent trends

The work carried out into latent variables to date has considered the correlations between the measurements. In addition to considering correlations, we could assume a latent trend
exists, from which the data are 'manifestations'. This model is appealing from a practical viewpoint, for instance a toxicologist analysing an ordinal measurement is provided with a natural explanation of the biological mechanism that leads to an adverse effect, such as death. Work has not been carried out into this scenario since there is a 'loss of information' in the data from the premise. By assuming there is an underlying continuous variable, we are only able to 'see' (and measure) the discrete realisations of this. This leads to problems, such as the identifiability of parameters. However, no research has been carried out in this direction for repeated measurement designs, although the Probit model is well known in univariate statistics.

### 1.5 Outline of thesis

The basic methods used for the analysis will be modifications of maximum likelihood (ML) and generalised estimation equations (GEEs).

In Chapter 2 the application of ML and REML to continuous data to find mean parameters will be discussed. This work will later be modified to cope with ordinal data. As well as the estimation, different types of error structures will be considered in section 2.4 to allow for the correlations between individuals.

Chapter 3 will consider the formulation of the GEEs to estimate mean parameters (section 3.2) and Pearson residuals to estimate the correlation structure (section 3.3). Section 3.4 will present methodology for which the GEEs can be applied to estimate an underlying trend for repeated ordinal data, under the assumption that it arises as a discrete realisation of an underlying exponential function. Simulations will be used to examine the estimates obtained and the effect of miss-specification of the correlation structure.

Chapter 4 shows a way that maximum likelihood analysis could be applied to ordinal data to estimate the 'change points' at which the categories switch over, assuming that there is a defined underlying (latent) continuous probability function. A new estimation procedure, 'quasi-estimation equations' (QEEs), will be formulated in section 4.1, which estimates both the cut-off points and the underlying distribution, with simulation studies used to look at the performance of the method, particularly under correlation miss-specification. The asymptotics of these estimates will also be considered, both graphically and theoretically in section 4.2. As a comparison between current methodology and this new method, section 4.3 will apply QEEs and GEEs to the simple case when there is a binary outcome, since the mean parameter estimates in this case will be comparable. Finally in section 4.4, the QEEs will be applied to a real life situation concerning repeated measurements of pain from runners after
a race. Both ordinal and continuous measurements were recorded and methods to analyse them both are applied and the results compared.

Chapter 5 will look at the case when we have more than one response variable at each time (repetitive measurements), with the emphasis on making two measurements each time. The methodology of QEEs will be extended to analyse the situation with simulation studies examining the estimation procedure in section 5.2.

The thesis will finish with a chapter of conclusions concerning the content and possible extensions to this work.
CHAPTER 2

Likelihood Analysis

When estimating parameters for linear models, it is usual to use the method of maximum likelihood (ML). However a feature of repeated measurement data is that there is correlation between measurements taken on any one individual. Although ML copes with this, many statisticians use a method called 'restricted maximum likelihood' (REML) because ML does not estimate consistently under such circumstances.

2.1 Notation

Assume there are m 'subjects', which have had measurements taken at n common times, \( t_j, j = 1, \ldots, n \). Let \( y_{ij} \) be the \( j \)th observed measurement on the \( i \)th subject \((i = 1, \ldots, m, j = 1, \ldots, n)\). Expanding this into matrix notation gives a vector of observed measurements for each subject, \( i \).

\[
y_i = (y_{i1}, \ldots, y_{in})^T
\]

and a vector containing all \( N = mn \) measurements.

\[
y = (y_1, \ldots, y_m)^T
\]
Let $X_i$ be an $n \times p$ design matrix describing the relationship between the measurements on individual $i$ and the explanatory variables. Join these together to form one $N \times p$ data matrix, $X = (X_1^T, \ldots, X_N^T)^T$. Suppose that the variance matrix for individual $i$ is $\sigma^2 V_i$. These matrices can be combined to form a complete variance matrix for the data, $\sigma^2 V$ which is block diagonal, with non-zero blocks $\sigma^2 V_i$.

$$V = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_m \end{pmatrix}$$

It should be noted that the block-diagonal nature of the variance matrix says that the measurements taken from one person (within individuals) are correlated, but the measurements taken from two different people (between individuals) are independent no matter which time they were taken at, which is a feature of repeated measurement data.

We wish to estimate values for the $p$ parameters, $\beta = (\beta_1, \ldots, \beta_p)^T$. It is usual to assume the data has a multivariate normal distribution which is linear with respect to the parameter vector, $\beta$. That is to say

$$Y_i \sim N_n(X_i \beta, \sigma^2 V_i)$$

So, the complete set of data is normally distributed

$$Y \sim N_N(X \beta, \sigma^2 V)$$

These are assumed to be normally distributed since then there is a likelihood to work with, although other distributions can be used in a similar manner.

### 2.2 Maximum Likelihood estimation

Under the stated normal assumption, the likelihood for the observed data, $y$, is

$$Lik(\beta, \sigma^2, V) = \frac{\exp[-\frac{1}{2}(y - X \beta)^T(\sigma^2 V)^{-1}(y - X \beta)]}{(2\pi)^{\frac{1}{2}N}|\sigma^2 V|^{\frac{1}{2}}}$$

Therefore the log-likelihood is

$$L(\beta, \sigma^2, V) = \log \frac{\exp[-\frac{1}{2}(y - X \beta)^T(\sigma^2 V)^{-1}(y - X \beta)]}{(2\pi)^{\frac{1}{2}N}|\sigma^2 V|^{\frac{1}{2}}}$$

$$= -\frac{1}{2}(y - X \beta)^T(\sigma^2 V)^{-1}(y - X \beta) - \frac{1}{2}N \log(2\pi) - \frac{1}{2} \log |\sigma^2 V|$$

$$= -\frac{1}{2}\left[\sigma^{-2}(y - X \beta)^T V^{-1}(y - X \beta) + N \log(2\pi) + N \log \sigma^2 + \log |V| \right]$$

$$= -\frac{1}{2}\left[\sigma^{-2} \text{RSS} + N \log(2\pi) + N \log \sigma^2 + \sum_{i=1}^m \log |V_i| \right]$$

(2.3)
where the residual sum of squares (RSS) is defined to be

\[ \text{RSS}(V) = \text{RSS} = (y - X\beta)^T V^{-1} (y - X\beta) = \sum_{i=1}^{m} (y_i - X_i\beta)^T V_i^{-1} (y_i - X_i\beta) \] (2.4)

To find the maximum likelihood estimates, this log-likelihood is maximised with respect to the variables; \( \beta, \sigma^2 \) and \( V \). Holding \( \sigma^2 \) and \( V \) fixed, the only varying term in equation 2.3 is \(-\frac{1}{2} \sigma^2 \text{RSS}\), so this is maximised,

\[
\max \left[ -\frac{1}{2} \sigma^2 \text{RSS} \right] \equiv \max [-\text{RSS}] = \max \left[ -\sum_{i=1}^{m} (y_i - X_i\beta)^T V_i^{-1} (y_i - X_i\beta) \right]
\]

\[
\equiv \min \sum_{i=1}^{m} (y_i - X_i\beta)^T V_i^{-1} (y_i - X_i\beta)
\]

which is the method of generalised least-squares estimation. The solution can be shown to be

\[
\hat{\beta}(V) = (X^T V^{-1} X)^{-1} X^T V^{-1} y
\] (2.5)

Differentiating equation 2.3 with respect to \( \sigma^2 \) and equating to zero gives:

\[
\frac{d}{d\sigma^2} L(\hat{\beta}(V), \sigma^2, V) = -\frac{1}{2} \left\{ N \frac{\text{RSS}}{\sigma^2} - \sigma^2 \right\} = 0
\]

\[ \Rightarrow \hat{\sigma}^2(V) = \frac{\text{RSS}}{N} \] (2.6)

Substituting these two estimates into equation 2.3 gives

\[
L[\hat{\beta}, \hat{\sigma}^2, V] = -\frac{1}{2} \left[ \left( \frac{\text{RSS}}{N} \right)^{-1} \text{RSS} + N \log(2\pi) + N \log \left( \frac{\text{RSS}}{N} \right) + \sum_{i=1}^{m} \log |V_i| \right]
\]

\[
= -\frac{1}{2} \left[ N + N(\log(2\pi) + N \log \text{RSS} - N \log N + \sum_{i=1}^{m} \log |V_i| \right]
\]

Removing anything that does not depend on \( V \), gives the reduced log-likelihood function for \( V \).

\[
L_r(V) = -\frac{1}{2} \left[ N \log \text{RSS} + \sum_{i=1}^{m} \log |V_i| \right]
\] (2.7)

Maximising the reduced log-likelihood, equation 2.7, gives an estimate of the variance matrix, \( V \). Once this value is know, \( \hat{\beta}(V) \) and \( \hat{\sigma}^2(V) \) are found by substitution into equations 2.5 and 2.6 respectively. This maximisation will have to be done numerically, on a computer, because the calculations will usually be long and complicated.
2.3 Restricted Maximum Likelihood Estimation

The REML estimate is defined as the ML estimator based on the linearly transformed set of data \( Y^* = AY \), where \( A \) is chosen such that the distribution of \( Y^* \) does not depend on \( \beta \).

By creating a new variance matrix, \( H = \sigma^2 V \)

\[ Y \sim N_N (X\beta, H) \]

We define \( B \) (of dimension \( N \times N - p \)) as a matrix which satisfies the criterion \( B^T B = I_{N-p} \).

For fixed \( H \), the ML estimate of \( \beta \), is (as in equation 2.5)

\[ \hat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} Y \equiv GY \]

and the probability density functions (pdf's) of \( Y \) and \( \hat{\beta} \) are

\[ f(y) = (2\pi)^{-\frac{1}{2}N} |H|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (y - X\beta)^T H^{-1} (y - X\beta) \right] \]

\[ g(\hat{\beta}) = (2\pi)^{-\frac{1}{2}p} |X^T H^{-1} X|^{\frac{1}{2}} \exp\left[-\frac{1}{2} (\hat{\beta} - \beta)^T X^T H^{-1} X (\hat{\beta} - \beta) \right] \]

Transforming the \( Y \) vector to \( Z \) as follows

\[ Z = B^T Y \]

then it can be shown (Diggle, Liang and Zeger (1994) [20] pp. 65-66) that there exists a \( B \) such that

\[ E(Z) = 0 \]

\[ Z \text{ and } \hat{\beta} \text{ are independent.} \]

and the pdf of \( Z = B^T Y \) is proportional to

\[ \frac{f(y)}{g(\hat{\beta})} = (2\pi)^{-\frac{1}{2}(N-p)} |H|^{-\frac{1}{2}} |X^T H^{-1} X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (y - X\hat{\beta})^T H^{-1} (y - X\hat{\beta}) \right] \]

(2.8)

where the constant term is omitted as it does not depend on any model parameters. Equation 2.8 gives the REML estimator of \( H, \hat{H} \), as the estimate which maximises the log-likelihood function

\[ L'(H) = -\frac{1}{2} \log |H| - \frac{1}{2} \log |X^T H^{-1} X| - \frac{1}{2} (y - X\hat{\beta})^T H^{-1} (y - X\hat{\beta}) \]

(2.9)

compared to the ML estimate which excludes the middle term, \(-\frac{1}{2} \log |X^T H^{-1} X|\).

Rewriting to the original problem, for any given \( V \),

\[ \hat{\beta}(V) = (X^T V^{-1} X)^{-1} X^T V^{-1} y \]

(2.10)

\[ \text{RSS}(V) = \sum_{i=1}^{m} (y_i - X_i\hat{\beta})^T V^{-1} (y_i - X_i\hat{\beta}) \]

(2.11)
The unbiased REML estimator for $\sigma^2$, $\hat{\sigma}^2$, is defined to be
\[
\hat{\sigma}^2(V) = \frac{\text{RSS}}{N - p}
\] (2.12)
and the REML estimator for $V$ maximises the reduced log-likelihood function
\[
L^*(V) = -\frac{1}{2} \left[ N \log \text{RSS} + \sum_{i=1}^{m} \log |V_i| \right] - \frac{1}{2} \sum_{i=1}^{m} |X_i^T V_i^{-1} X_i|
\] (2.13)
Once again this maximisation problem will have to be conducted numerically. This estimate, $\hat{V}$, is substituted from this equation into equations 2.5 and 2.12 to find the REML estimates $\hat{\beta}(\hat{V})$ and $\hat{\sigma}^2(\hat{V})$.

It should be noted that the REML log-likelihood is the same as the ML log-likelihood with the addition of the term $-\frac{1}{2} \log |X^T V^{-1} X|$. This is a $p \times p$ matrix, compared to $L(V)$ which is of order $N$. As $N$ increases, this term decreases in size relative to the ML log-likelihood, which implies that for large $N$, REML($\hat{\beta}$) converges to ML($\hat{\beta}$) as the sample size increases. This also means that the ML and REML estimates will be asymptotically consistent or inconsistent together.

### 2.4 Models for the variance matrix

In the previous sections the problem of estimating the parameters for a longitudinal set of data was considered, using a general form for the covariance structure ($V_i$ was not specified). Now consider the case when the variance matrix is modelled by a set of parameters, which need to be estimated along with the parameters of the actual regression, $\beta$.

#### Error structure

When analysing repeated measures data, there will always be some sort of random variation. This tends to come from one of three sources.

- Measurement error
- Random Effects
- Serial Correlation

The first of these, measurement error, occurs due to how data is collected. In real life, there are very few things that are 'perfect'. When taking measurements, they are always subject
to some sort of error. For example a machine measuring the data may only be able to give so many digits of precision, so a rounded off value is recorded or the user might make a small mistake when taking the measurement, and so on.

Random effects are the most general of the errors considered. They describe how individuals' measurements differ due to random fluctuations specific to that individual rather than the entire population. As an example of this, consider accurately measured weights of identical twins who have exactly the same life style. The difference between their weights can be thought of as being caused by them having different random effects errors.

The last source of error is serial correlation. This is like a knock on effect of random effects, which can be summed up as 'High values breed high values and low values breed low values', i.e. if an individual starts off with a high value, they will tend to give higher measurements at later times, and vice-versa. This means that there will be some correlation between the measurements made on any one given individual.

These three random errors can be incorporated into the model in numerous ways, but for the following, they are assumed to be additive. Even if this is not the case, the data may sometimes be transformed so that this holds. For instance if the errors were multiplicative, then analysing the logs of the measurements would give an additive error structure.

The model being analysed is

$$Y \sim N_N[X\beta, V(\alpha)]$$

where $\alpha$ is a vector which parameterises the variance/covariance matrix using the three types of error discussed above. It is more useful to split this into a mean term and an error term as follows.

$$Y = X\beta + \epsilon$$

$$\epsilon \sim N_N[0, V(\alpha)]$$

The additive nature of the errors gives

$$\epsilon = \text{Random effects} + \text{Serial correlation} + \text{Measurement error} \quad (2.14)$$

Call the measurement error on the $j$th measurement of the $i$th unit $Z_{ij}$, where

$$Z_{ij} \sim N(0, \tau^2)$$

Measurement errors are assumed to be independent of the individual and time. This means that the variance matrix of the measurement error terms is

$$V(\text{Measurement error}) = \tau^2 I \quad (2.15)$$
Let the serial correlation, $W_i(t_{ij})$, be $m$ independent (stationary) normal distributions with zero mean, variance of $\sigma^2$ and correlation $\rho(u)$. This gives a variance matrix for the $i$th individual as the matrix of serial correlation errors, $\sigma^2 H_i$, where $H_i$ is the $n \times n$ square matrix containing correlations between $W_i(t_{ij})$ and $W_i(t_{ik})$.

$$H_i = \begin{pmatrix}
\rho(|t_{i,1} - t_{i,1}|) & \rho(|t_{i,1} - t_{i,2}|) & \cdots & \rho(|t_{i,1} - t_{i,n}|) \\
\rho(|t_{i,2} - t_{i,1}|) & \rho(|t_{i,2} - t_{i,2}|) & \cdots & \rho(|t_{i,2} - t_{i,n}|) \\
\vdots & \vdots & \ddots & \vdots \\
\rho(|t_{i,n} - t_{i,1}|) & \rho(|t_{i,n} - t_{i,2}|) & \cdots & \rho(|t_{i,n} - t_{i,n}|)
\end{pmatrix}$$ (2.16)

Combining the square matrices, $H_i$, into one block-diagonal matrix, $H$, with diagonal ‘blocks’ equal to the above matrices (in order) gives the variance matrix for the serial correlation error terms as

$$V(\text{Serial Correlation error}) = \sigma^2 H$$ (2.17)

The last term that needs to be considered is the random effects error. For any individual term, this error can be expressed as:

$$d_{ij}^T U_i$$

where

$$U_i \sim N_r(0, G)$$

are $m$ mutually independent normal processes and $d_{ij}$ are $r$-element vectors which contain the explanatory variables for each random effect. Calling $D_i$ the $n_i \times r$ matrix with $j$th row $d_{ij}$, then the variance matrix of the random effects term for individual $i$, is $D_i G D_i^T$. Combining these gives a block diagonal variance matrix of the random effects error.

$$V(\text{Random effects errors}) = DGD^T$$ (2.18)

Substituting the three sources of errors, equations 2.15, 2.17 and 2.18, into equation 2.14 gives an expression for the total error.

$$V(\varepsilon) = V(t, \alpha) = DGD^T + \sigma^2 H + \tau^2 I$$ (2.19)

This gives the full variance matrix if the correlation within the matrix is known (equation 2.16 is fully defined).
Random Intercept, Serial Correlation and Measurement Error

One special form of the error structure exists which gives an easy interpretation for the parameters. Assuming that the random effects error is given by $U \sim N(0, \nu^2)$ and $d_{ij} = 1$ then the value of $U$ represents a random intercept for each individual. That is to say, it is the amount by which all measurements on any given individual are changed relative to the population average. Using this model, the variance matrix, equation 2.19, becomes:

$$V(\epsilon) = \nu^2 J + \sigma^2 H + \tau^2 I$$ (2.20)

where $J$ is a $N \times N$ matrix with block diagonal matrices, of dimension $n_i \times n_i$, whose components are all equal to 1.

Using this model, it is easy to explain what each of the components represents. $\tau$ is a measure of the error between repeated measurements on any given individual. $\nu$ is a measure of the error between individuals. $\sigma$ is a measure of the overall error. This interpretation of the covariance parameters makes this model very attractive.

Correlation structures

One of the most popular choices for the correlation is the ‘exponential correlation model’, where

$$\rho(|t_j - t_k|) = e^{-\phi|t_j - t_k|} \quad \phi > 0$$ (2.21)

Using this form, the correlation between observations from one individual becomes smaller as the time difference increases. This is the usual scenario in real life, so it is a realistic form for the correlation function. However, since the rate at which the correlation decreases is not known, it is impossible to know if this is the best choice.

Another common choice is the ‘Gaussian correlation function’

$$\rho(|t_j - t_k|) = e^{-\phi(t_j - t_k)^2}$$ (2.22)

This function has similar properties to the exponential correlation model, but gives smaller correlations between observations at further time differences than the previous one. Once again, the correlation decreases as the time difference increases.

A less used structure is the ‘linear correlation function’

$$\rho(|t_j - t_k|) = \begin{cases} 1 - \phi |t_j - t_k| \quad & |t_j - t_k| \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$ (2.23)
which is in a different form from the previous two. It gives smaller correlations to further observations, but the decrease is linear (as implied in the name) rather than exponential (as it was in the previous two functions). This means that further observations will tend to have larger correlations than with the other two functions. However this could give unsatisfactory values for the correlations, with some values with modulus greater than 1 if the time difference was sufficiently large. This may happen if the model is being used to project forwards (or backwards) in time.

2.5 Application - ‘Milk data’

A data set, contained in Diggle, Liang and Zeger (1994) [20], provided by Ms. Alison Frensham was analysed. It is concerned with the analysis of protein content taken weekly from 79 Australian cows. The cows were split into three different groups according to their diet, as shown in table 2.1

<table>
<thead>
<tr>
<th>Group</th>
<th>Size</th>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>Barley only</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>Barley and Lupins</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>Lupins only</td>
</tr>
</tbody>
</table>

Table 2.1: Grouping of the cows to different diets

The objective of this study was to determine how diet affects the protein content of the cows milk. The three shadow plots in figure 2.1 show a random sample from the data for each of the different diets, with each line representing the repeated measurements for one cow. The ‘shadowing’ of the plot by removing some of the data allows the overall trends to be shown. Although these are still slightly messy there seems to be a slight difference in the profiles of the groups, although this is not marked. Also the overall shapes appear to be similar for each group, with a decrease in yield for the first few times and then a slow increase as time goes on.
Figure 2.1: Shadowed time series plots for the three groups of cows
Model choice

Three different correlation structures have been considered:

- Exponential correlation
- Exchangeable correlation
- Independent correlation

Exponential correlation model

This model has random intercept, serial correlation and measurement error with the exponential correlation function (as detailed in equations 2.20 and 2.21 respectively).

\[
V(\varepsilon) = \nu^2 J + \sigma^2 H + \tau^2 I
\]

\[
\rho|t_j - t_k| = e^{-\phi|t_j - t_k|} \quad \phi > 0
\]

This correlation was chosen arbitrarily, although there was no reason to suggest that it was not a sensible choice for the data. It would be expected that measurements would be highly correlated with other measurements taken at close times, and correlated to a lesser extent with measurements taken at more distant times and the exponential correlation structure describes this.

Uniform correlation model

Having taken the above model as the full model, some less parameterised models were considered. The first parameter which was removed was the serial correlation (by putting \( \sigma = 0 \)). Therefore each individual has a random intercept and measurement error. This means that the variance matrix for an individual, \( V_i \), has all its off diagonal elements equal and the diagonal elements equal, but taking a different value to the off-diagonal ones, i.e.

\[
V(\varepsilon) = \nu^2 J + \tau^2 I
\]

Independent correlation model

The simplest model to be considered was all measurements on an individual being independent. This would only leave the measurement errors, since \( \nu = 0 \) makes the individual
variance matrices, $V_i$, to become diagonal, completely eliminating correlations between different measurements on each individual.

$$V(\varepsilon) = \tau^2 I$$

Mean trend

The previous graphs exhibit a vague, but similar trend. They start with a linear trend and then switch over to a non-linear trend. To simplify the modelling, it will be assumed that when they start increasing non-linearly around time 3, they do so at a quadratic rate. This would lead to a 'typical cows' measurements being similar to graph 2.2. However, the focus of this study was the difference in protein yield according to the cow's diet. Therefore a different intercept will be given for each group of cows. This model leads to mean trends as in table 2.2 were used for the different groups, where $\mu_i$ is the mean for group $i$ ($i = 1, 2, 3$).

![Graph showing mean trend for a typical cow](image)

**Figure 2.2:** Typical mean trend for a typical cow

<table>
<thead>
<tr>
<th>Protein Content</th>
<th>Time</th>
</tr>
</thead>
</table>

$$\mu_1 = \begin{cases} \beta_1 + t\beta_4 & t \leq 3 \\ \beta_1 + 3\beta_4 + (t - 3)\beta_5 + (t - 3)^2\beta_6 & t > 3 \end{cases}$$

$$\mu_2 = \begin{cases} \beta_2 + t\beta_4 & t \leq 3 \\ \beta_2 + 3\beta_4 + (t - 3)\beta_5 + (t - 3)^2\beta_6 & t > 3 \end{cases}$$

$$\mu_3 = \begin{cases} \beta_3 + t\beta_4 & t \leq 3 \\ \beta_3 + 3\beta_4 + (t - 3)\beta_5 + (t - 3)^2\beta_6 & t > 3 \end{cases}$$

**Table 2.2:** Mean models for cows from different groups
Types of analysis

For the analysis both ML and REML were used so as to give a comparison between the two methods. Since there were missing data, different imputations methods were also considered to see what effect they had on the parameter estimates. Although this was not necessary for this data, when there are sparse measurements, imputation could become a powerful technique to improve the fit of a model.

No missing value imputation leaves the data as it is and does not impute the missing values. Therefore the likelihood estimation is carried out using partial records for cows that did not have all measurements taken on them.

Average value imputation assigns the mean value for the group at that time to the measurement. This does not consider different individuals since it assigns values just on the time they were meant to have been taken at.

Moving average imputation calculates the value of a missing point by taking the average of the nearest two measurements before and after. This does not work very well for long strings of missing values on any individual, since it will assign them all the same value.

The results from this analysis are in appendix A.

Analysis of results

From a practical point of view, this study does not show strong results. There is a difference in the means for the different groups, although this is not marked, and a t-test shows that it is not a significant difference. Interestingly, where all three errors were included in the model, $\tau$ has a small value. However, by removing the serial correlation to obtain the 'uniform correlation' model, the values increases. This suggests that the serial correlation and measurement error were related to each other.

By comparing these results conclusions about the different methods used can be drawn. The first being that REML estimation did not appear to be a significantly better method than ML estimation. Both give estimates and standard errors of the parameters which were virtually identical. The REML estimates were slightly larger, but only in the third or fourth figure.

Next the different correlation structures were compared. Before looking at the results, we know that the exponential correlation should be better than uniform correlation, which in
turn should be better than independent correlation, because the number of parameters in
the model decreases as the correlation structure was changes. Also, the different forms were
subsets due to the nested nature of the correlation modelling: The independence model is a
subset of the uniform model which in turn is a subset of the exponential model. Although
the extra parameters gave a better model, it increased the program run time dramatically.
The estimates for the parameters were similar for all three models, but there were major
differences between their standard errors. They were smaller for the uniform and independent
correlation models than the exponential model. This feature, correlation miss-specification
and its effects on repeated measurement estimates, will be considered in more depth later in
this thesis.

The different imputation methods were also compared (no imputation, time averaged missing
imputation and moving average imputation). The actual estimates for all three methods were
almost identical, with very similar standard errors.

To conclude, for this data set, it seems that a model calculated using maximum likelihood
estimation (ML) or restricted maximum likelihood estimation (REML) with exponential
correlation and no missing value imputation gave the 'best' results.
Two methods of estimating parameters in repeated data analysis have already been investigated. Another such method is 'generalised estimation equations', abbreviated to GEEs, which were formulated by Liang and Zeger (1986) [45], Zeger and Liang (1986) [87] and Zeger, Liang and Albert (1988) [88]. These methods give consistent estimates of the regression parameters and their variance, under a few assumptions about the time dependence of measurements. A specific application of GEEs to repeated ordinal data analysis will also be considered.

3.1 Theory

In the previous methods we assumed the data was multivariate normally distributed. GEEs are more general, in the sense that they can cope with non-normal data, as long as it is a member of an exponential family

\[ f(y_{ij}) = \exp \{ [y_{ij} - a(\theta_{ij}) + b(y_{ij})] \phi \} \]  

(3.1)

where \( \theta_{ij} = g(\eta_{ij}) \), \( \eta_{ij} = X_{ij}\beta \). This formulation only specifies the marginal distribution and not the complete multivariate distribution as previously. Although this offers more modelling
possibilities, the scope for ordinal data analysis is still somewhat limited, with only a few cases, such as binary data, being members of the exponential family.

When in this form two standard results of generalised linear modelling hold.

\[ E(y_{ij}) = \alpha'(\theta_{ij}) \tag{3.2} \]
\[ V(y_{ij}) = \frac{\alpha''(\theta_{ij})}{\phi} \tag{3.3} \]

Let \( R(\alpha) \) be a \( n \times n \) 'working correlation' matrix which is fully parameterised by the \( s \times 1 \) vector, \( \alpha \). This may, or may not be the same parameterisation as discussed in section 2.4. The elements must all be real, with modulus less than or equal to 1 and the matrix must be symmetric and non-singular. The following are defined

\[ A_i = \text{diag}[\alpha''(\theta_{i1}), \ldots, \alpha''(\theta_{in})] = \phi \times \text{diag}[V(y_{i1}), \ldots, V(y_{in})] \]
\[ \Delta_i = \text{diag}\left[ \frac{d\theta_{ij}}{d\eta_{ij}} \right] \]
\[ D_i = \frac{d[\alpha'(\theta_i)]}{d\beta} = A_i \Delta_i X_i \]
\[ V_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi \]
\[ S_i = y_i - \alpha'(\theta_i) \tag{3.4} \]

where \( X \) is a design matrix which specifies the relationship between the parameters and the mean (via the link function). If \( R(\alpha) \) is the real correlation matrix for the data then equation 3.4 would be exactly equal to the variance matrix, \( V(y_i) = V_i \), which is why it is called the working correlation matrix.

Using the above notation, the general estimating equations (GEEs) are defined as the root of the following equation.

\[ \sum_{i=1}^{m} D_i^T V_i^{-1} S_i = 0 = \sum_{i=1}^{m} U_i(\beta, \alpha) \tag{3.5} \]

Suppose the estimate of \( \alpha, \hat{\alpha}(y, \beta, \phi) \), is found when \( \beta \) and \( \phi \) are both known. Substituting this into equations 3.4 and 3.5, and then estimating \( \phi \) by a \( m^{1/2} \) consistent estimator, \( \hat{\phi}(y, \beta) \), transforms equation 3.5 into

\[ \sum_{i=1}^{m} U_i \{ \beta, \alpha[\beta, \hat{\phi}(\beta)] \} = 0 \tag{3.6} \]

The estimates of the 'mean parameters', \( \hat{\beta}_{GEE} \), are the solutions to the homogenous equations 3.6. The large sample distribution of \( \hat{\beta}_{GEE} \) can be found, subject to three conditions, which are:

\[ (i) \quad m^{1/2} |\hat{\alpha} - \alpha| = O_P(1) \text{ given } \beta \text{ and } \phi \tag{3.7} \]
Then asymptotically, where

\[ m^{1/2} |\hat{\beta}_{GEE} - \beta| \sim N_p(0, V_{GEE}) \]  

(3.10)

asymptotically, where

\[ V_{GEE} = \lim_{m \to \infty} \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right] \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} V(y_i) V_i^{-1} D_i \right]^{-1} \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right]^{-1} \]  

(3.11)

The full proof of this theorem is included as an appendix to this chapter in section 3.5.

3.2 Estimation using GEEs

The equations that need to be solved to find the GEE estimates have now been defined. If \( \alpha \) and \( \phi \) are known, then modified Fisher scoring could be used to find \( \hat{\beta}_{GEE} \). However, in general, the values of these parameters are not known, so they too have to be estimated. The method most commonly used is modified Fisher scoring to find \( \hat{\beta}_{GEE} \) and then (at each iteration) moment estimation for \( \alpha \) and \( \phi \). Starting off with two ‘variance’ estimates, \( \hat{\alpha} \) and \( \hat{\phi} \), Liang and Zeger (1986) [45] suggest the following iterative procedure for \( \beta \) to give one half of the estimation process.

\[ \hat{\beta}_{j+1} = \hat{\beta}_j - \left[ \sum_{i=1}^{m} D_i^T (\hat{\beta}_j) \hat{V}_i^{-1} (\hat{\beta}_j) D_i (\hat{\beta}_j) \right]^{-1} \left[ \sum_{i=1}^{m} D_i^T (\hat{\beta}_j) \hat{V}_i^{-1} (\hat{\beta}_j) S_i (\hat{\beta}_j) \right] \]  

(3.12)

where

\[ \hat{V}_i = V_i(\beta_j, \hat{\alpha}(\beta_j), \hat{\phi}(\beta_j)) \]

Once the new \( \beta_j \) has been estimated, the estimates of \( \alpha \) and \( \phi \) are updated using Pearson residuals

\[ \hat{r}_{ij} = \frac{y_{ij} - a^\prime(\hat{\theta}_{ij})}{\sqrt{a''(\hat{\theta}_{ij})}} \]  

(3.13)

where \( \hat{\theta}_{ij} \) depends on the current value of \( \beta \). \( \phi \) is estimated by:

\[ \hat{\phi}^{-1} = \sum_{i=1}^{m} \sum_{t=1}^{n} \frac{\hat{r}_{it}^2}{N - p} \]  

(3.14)

There is no ‘exact’ estimate of \( \alpha \), since no exact form has been specified, but the sample correlations,

\[ \hat{R}_{ij}^{-1} = \sum_{a=1}^{m} \frac{\hat{r}_{ia} \hat{r}_{aj}}{N - p} \]  

(3.15)

30
can be used to find $\alpha_j$. It should be noted for all of the above analysis to be valid, any missing data should be missing at random.

3.3 Different ‘working correlation’ matrices

As mentioned before, the correlation does not have to be specified exactly to obtain consistent and asymptotically normal estimates, $\hat{\beta}_{GEE}$. The only requirement is that $\alpha$ and $\phi$ are estimated consistently and that $V_{GEE}$ (when divided by $m$) converges to a fixed matrix. However, the closer the working correlation matrix is to the actual correlation matrix, the more efficient the estimates will be.

Some of the most common forms for $R(\alpha)$ are:

- Independence
- One dependence
- Exchangeable correlation
- Exponential correlation

Independence

Assuming that the observations are independent within individuals, then $R_i(\alpha) = I_n$ and is completely specified without any estimation.

One dependence

Consider a special case of a tri-diagonal correlation matrix, where the correlations next to the diagonals are equal. By averaging the off-diagonal sample correlations we obtain an estimate for this correlation.

$$\hat{\alpha} = \sum_{j=1}^{n-1} \frac{\hat{\alpha}_{ij}}{n-1} = \frac{\phi}{m} \sum_{i=1}^{m} \sum_{j=1}^{n-1} \frac{\hat{r}_{ij}\hat{r}_{i,j+1}}{(n-1)(m-p)} \quad (3.16)$$
Exchangeable correlation

Assume that the correlation matrix has all its off diagonal elements equal to \( \alpha \), so there is only one parameter to estimate.

\[
corr(y_{ij}, y_{ik}) = \begin{cases} 
  1 & j = k \\
  \alpha & j \neq k 
\end{cases}
\]

This form for the correlation structure explicitly accounts for the relationship with the previous measurement (as does 1-dependence) and also all other measurements that any individual makes. \( \alpha \) can be estimated by

\[
\hat{\alpha} = \frac{\sum_{i=1}^{m} \sum_{j>k} \hat{r}_{ij}\hat{r}_{ik}}{\frac{1}{2}mn(n - 1) - p}
\]

Exponential correlation

Instead of having a constant, \( \alpha \), on the off diagonal elements of the correlation matrix, we have exponentially decreasing correlations,

\[
corr(y_{ij}, y_{ik}) = \begin{cases} 
  1 & j = k \\
  \alpha^{|j-k|} & j \neq k 
\end{cases}
\]  
(3.17)

For this model

\[
E(\hat{r}_{ij}\hat{r}_{ik}) \approx \alpha^{|j-k|}
\]  
(3.18)

This means that by plotting the data on a graph with a y-axis log \( |(\hat{r}_{ij}\hat{r}_{ik})| \) and x-axis \( |j - k| \) the value of log \( \alpha \) can be estimated by the gradient of the regression line.

This formulation of the correlation matrix is more flexible than ML/REML. Firstly, since only the marginal distribution is defined and not the full multivariate distribution, this matrix does not have to be positive definite, whereas to have a fully defined likelihood for the multivariate normal distribution the correlation must be positive definite. This is taken into account by limiting the number of specifications for the correlation structure. It should be noted that some of the models are identical. Obviously, the independence models are the same, exchangeable is the same as ML/REML uniform correlation and the exponential model is identical to the ML/REML measurement error and serial correlation structure. However, since GEEs use marginal distributions, rather than the full multivariate distribution, the estimates will be different unless the correlation is independent.
3.4 An application of GEEs to latent trend analysis

This example assumes that the data arose as discrete observations from a continuous distribution (exponential for the purposes of the following work). Under such an assumption, the categorical data are members of the exponential families, so the GEEs can be applied to estimate the underlying distribution and the parameters associated with it. This is one of the few possibilities where GEEs can be applied to repeated ordinal data to estimate a latent trend. Suppose the $j$th measurement on individual $i$ ($i = 1, \ldots, m$, $j = 1, \ldots, n$) is the categorical response $y_{ij}$. Assume that this arises as the end-point of an interval between two successive integers of an exponential distribution with parameter $\lambda_{ij}$.

<table>
<thead>
<tr>
<th>Category</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$P(0 &lt; x \leq 1)$</td>
</tr>
<tr>
<td>2</td>
<td>$P(1 &lt; x \leq 2)$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k$</td>
<td>$P(k - 1 &lt; x \leq k)$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
</tbody>
</table>

This is shown diagrammatically in figure 3.1, where $x$ is the 'time' variable of the exponential distribution.

$$f(x, \lambda_{ij}) = \lambda_{ij} e^{-\lambda_{ij} x}$$

![Figure 3.1: Representation of discrete data arising as realisations of a continuous distribution](image)

For each individual there is a design matrix, $X_i$ which is used to linearly model the $p$ parameters, which are some function (the link) of $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{in})^T$. 

33
Method

Using the above notation, the pdf of the categorical data is given by

\[
P(y_{ij} = k) = \frac{1}{\lambda_{ij}} e^{-\frac{x_{ij}}{\lambda_{ij}}} dx
\]

\[
= -e^{-\frac{k}{\lambda_{ij}}} + e^{-\frac{k-1}{\lambda_{ij}}} = e^{-\frac{k}{\lambda_{ij}}} \left( e^{\frac{1}{\lambda_{ij}}} - 1 \right)
\]

\[
= e^{-\frac{k}{\lambda_{ij}}} \log \left( e^{\frac{1}{\lambda_{ij}}} - 1 \right) = e^{-\frac{k}{\lambda_{ij}}} + \log \left( e^{\frac{1}{\lambda_{ij}}} - 1 \right)
\]

(3.19)

Equation 3.19 is in the form of an exponential family,

\[
e^{[y \theta_{ij} - a(\theta_{ij}) + b(y_{ij})] \phi}
\]

(3.20)

with

\[
y = k
\]

\[
\theta_{ij} = -\frac{1}{\lambda_{ij}}
\]

\[
a(\theta_{ij}) = -\log(e^{-\theta_{ij}} - 1)
\]

Therefore the following results hold

\[
\mu(y_{ij}) = \frac{da}{d\theta_{ij}} = \frac{e^{-\theta_{ij}}}{e^{-\theta_{ij}} - 1}
\]

(3.21)

\[
V(y_{ij}) = \frac{a''(\theta_{ij})}{\phi} = \frac{e^{-\theta_{ij}}}{(e^{-\theta_{ij}} - 1)^2}
\]

(3.22)

This is then modelled using the generalised estimation equations of Liang and Zeger (1994) [45]. The only unspecified expression is the link, which (if we use the canonical link) is

\[
a'(\theta_{ij}) = \frac{e^{-\theta_{ij}}}{e^{-\theta_{ij}} - 1} = \mu_{ij}
\]

\[
\Rightarrow \theta_{ij} = -\log \left( \frac{|\mu_{ij} - 1|}{\mu_{ij}} \right) = g(\mu_{ij})
\]

(3.23)

Parameter estimates

A simple example was considered to see how well this method performed in a simulation study. 2 groups of 4 observations with 100 'subjects' in each set were randomly generated. These were chosen to be 1 plus the integer parts of independent random exponential observations with the following parameters:

\[
\text{Group 1: } \theta_1 = \beta_0 + t \times \beta_2
\]

(3.24)

\[
\text{Group 2: } \theta_2 = \beta_0 + \beta_1 + t \times \beta_2
\]

(3.25)
where \((\beta_0, \beta_1, \beta_2) = (-0.2, 0.2, -0.1)^T\). Four different structures were used to model the correlation between the repeated measurements; independence, one-dependence, exchangeable and exponential. These all took \(\alpha = 0.5\) as the correlation parameter. 100 simulations were run, yielding 100 sets of GEE estimates for the parameters. The mean and standard deviation of these estimates were calculated and are shown in table 3.1 according to the correlation structure imposed.

In all cases, the estimates were close to the true values, \((-0.2, 0.2, -0.1)\), so the method seems like a reasonable estimation process for the estimation of the latent trend. The standard errors were also small in comparison to the estimates which means that the estimates of the underlying parameters were consistently estimated close to their true values. The second part of this simulation study was to examine the effectiveness of the correlation estimation. These results show that the correlation structure was not estimated well by this method, with a distinct downward bias away from the true value of 0.5, suggesting that the correlations were poorly estimated using moment estimation.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>(\alpha) Estimate</th>
<th>S.D.</th>
<th>(\beta) Estimate</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td>-0.2157</td>
<td>0.00526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2221</td>
<td>0.00462</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1121</td>
<td>0.00171</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.3648</td>
<td>0.03508</td>
<td>-0.2121</td>
<td>0.00617</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2218</td>
<td>0.00557</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1170</td>
<td>0.00208</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.4735</td>
<td>0.04266</td>
<td>-0.1968</td>
<td>0.00598</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2139</td>
<td>0.00544</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1244</td>
<td>0.00238</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.4190</td>
<td>0.05848</td>
<td>-0.2101</td>
<td>0.00645</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2211</td>
<td>0.00603</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1194</td>
<td>0.00204</td>
</tr>
</tbody>
</table>

Table 3.1: GEE estimates of the underlying exponential distribution

**Model miss-specification**

Suppose now that the variance structure is miss-specified. Simulations were used to estimate the parameters under such circumstances, with the results in tables 3.2, 3.3, 3.4, 3.5.
Table 3.2: GEE estimates of the independent underlying exponential distribution under correlation miss-specification

<table>
<thead>
<tr>
<th>Correlation</th>
<th>α Estimate</th>
<th>S.D.</th>
<th>β Estimate</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td>-0.2157</td>
<td>0.00526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2221</td>
<td>0.00462</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1121</td>
<td>0.00171</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.0010</td>
<td>0.04528</td>
<td>-0.2162</td>
<td>0.00541</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2222</td>
<td>0.00443</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1119</td>
<td>0.00173</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>-0.0018</td>
<td>0.03339</td>
<td>-0.2159</td>
<td>0.00542</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2224</td>
<td>0.00491</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1120</td>
<td>0.00183</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.1075</td>
<td>0.03530</td>
<td>-0.2150</td>
<td>0.00550</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2211</td>
<td>0.00464</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1122</td>
<td>0.00179</td>
</tr>
</tbody>
</table>

Table 3.3: GEE estimates of the 1-dependent underlying exponential distribution under correlation miss-specification

<table>
<thead>
<tr>
<th>Correlation</th>
<th>α Estimate</th>
<th>S.D.</th>
<th>β Estimate</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td>-0.2150</td>
<td>0.00634</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2230</td>
<td>0.00550</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1161</td>
<td>0.00211</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.3648</td>
<td>0.03508</td>
<td>-0.2121</td>
<td>0.00617</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2218</td>
<td>0.00557</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1170</td>
<td>0.00208</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.1734</td>
<td>0.04218</td>
<td>-0.2134</td>
<td>0.00668</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2206</td>
<td>0.00600</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1158</td>
<td>0.00217</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.2958</td>
<td>0.04735</td>
<td>-0.2135</td>
<td>0.00635</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2223</td>
<td>0.00562</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.1166</td>
<td>0.00212</td>
</tr>
</tbody>
</table>
These results show that the parameter estimates were close to the true values, even when there was correlation miss-specification. This means (as Liang and Zeger (1986) [45] suggest) that the estimation by GEEs is robust to miss-specification of the correlation. However, the
standard errors associated with the estimates varied when there was miss-specification. With the exception of the independence case, the standard errors were smallest (overall) when the parameters were estimated with the correct correlation structure. It should be noted that when the independence case is the real model, estimation of a different model should give better results, since the independence case will be a subset of the miss-specified model. In the study, the correlation estimates were close to zero in the miss-specified models, so the estimates became almost identical as expected. The correlation estimates show that the method works poorly even under no miss-specification. As mentioned before, this may be due to the moment estimation step of GEEs being poor.

3.5 Proof of the GEEs asymptotics

Theorem

\[
m^{\frac{1}{2}} |\hat{\beta}_{GEE} - \beta| \sim N_p(0, V_{GEE})
\]

asymptotically, where

\[
V_{GEE} = \lim_{m \to \infty} m \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right]^{-1} \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} V(y_i) V_i^{-1} D_i \right] \left[ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right]^{-1}
\]

Proof

Let \( \alpha^*(\beta) = \hat{\alpha}[\beta, \hat{\phi}(\beta)] \) and \( \tilde{V}_i = V_i(\alpha^*) \). The solution to the GEEs is found using an iterative scheme, which has been detailed already.

\[
\hat{\beta}_{j+1} = \hat{\beta}_j - \left\{ \sum_{i=1}^{m} D_i^T \tilde{V}_i^{-1} (\hat{\beta}_j) D_i (\hat{\beta}_j) \right\}^{-1} \left\{ \sum_{i=1}^{m} D_i^T \tilde{V}_i^{-1} (\hat{\beta}_j) S_i (\hat{\beta}_j) \right\}
\]

As the iteration process progresses, taking \( j \) sufficiently large to assume that the method has converged, \( \hat{\beta}_{j+1} \approx \beta \), leaving us with an estimate of \( \beta \), \( \hat{\beta}_j = \hat{\beta}_{GEE} \).

\[
\hat{\beta}_{GEE} - \beta = \left\{ \sum_{i=1}^{m} D_i^T (\hat{\beta}_{GEE}) \tilde{V}_i^{-1} (\hat{\beta}_{GEE}) D_i (\hat{\beta}_{GEE}) \right\}^{-1} \left\{ \sum_{i=1}^{m} D_i^T (\hat{\beta}_{GEE}) \tilde{V}_i^{-1} (\hat{\beta}_{GEE}) S_i (\hat{\beta}_{GEE}) \right\}
\]

\[
\approx \left\{ \sum_{i=1}^{m} \frac{\delta D_i^T (\hat{\beta}_{GEE}) \tilde{V}_i^{-1} (\hat{\beta}_{GEE}) S_i (\hat{\beta}_{GEE})}{\delta \beta^T} \right\}^{-1} \left\{ \sum_{i=1}^{m} D_i^T (\hat{\beta}_{GEE}) \tilde{V}_i^{-1} (\hat{\beta}_{GEE}) S_i (\hat{\beta}_{GEE}) \right\}
\]

\[
= \left\{ \sum_{i=1}^{m} \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta^T} \right\}^{-1} \left\{ \sum_{i=1}^{m} U_i[\beta, \alpha^*(\beta)] \right\}
\]
\[
\hat{\beta}_{\text{GEE}} - \beta \\
\approx \frac{1}{m} \left( \sum_{i=1}^{m} \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta} \right)^{-1} \left( \sum_{i=1}^{m} U_i[\beta, \alpha^*(\beta)] \right) \\
= \left\{ \frac{1}{m} \left( \sum_{i=1}^{m} \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta^T} \right) \times \frac{1}{m} \right\}^{-1} \left\{ \sum_{i=1}^{m} U_i[\beta, \alpha^*(\beta)] \times \frac{1}{m^2} \right\}
\]

(3.26)

In equation 3.26 the first half of the right hand side can be expanded to give

\[
\frac{1}{m} \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta^T} \approx \frac{1}{m} \frac{\partial U_i[\beta, \alpha^*(\beta)]}{\partial \beta^T} + \frac{1}{m} \frac{\partial U_i[\beta, \alpha^*(\beta)]}{\partial \alpha^T} \times \frac{\partial \alpha^*(\beta)}{\partial \beta^T} \frac{A_i}{B_i} \times C
\]

\[
\lim_{m \to \infty} \sum_{i=1}^{m} A_i = \lim_{m \to \infty} \sum_{i=1}^{m} \frac{1}{m} \times \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta^T} = \lim_{m \to \infty} \sum_{i=1}^{m} \frac{1}{m} \times \frac{\partial [D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)S_i(\beta)]}{\partial \beta^T}
\]

\[
= \lim_{m \to \infty} \sum_{i=1}^{m} \frac{D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)}{m} \times \frac{\partial [S_i(\beta)]}{\partial \beta^T} = \lim_{m \to \infty} \sum_{i=1}^{m} \frac{D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)}{m} \times \frac{\partial [y_i - a_i^T]}{\partial \beta^T}
\]

\[
= \lim_{m \to \infty} \sum_{i=1}^{m} \frac{D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)}{m} \times \frac{-\partial a_i^T}{\partial \beta^T} = - \lim_{m \to \infty} \sum_{i=1}^{m} \frac{D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)}{m} \times D_i(\beta)
\]

\[
= - \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} D_i^T(\beta)\hat{V}_{i}^{-1}(\beta) = \text{constant}
\]

\[
\lim_{m \to \infty} \sum_{i=1}^{m} B_i = \lim_{m \to \infty} \sum_{i=1}^{m} \frac{1}{m} \times \frac{\partial U_i[\beta, \alpha^*(\beta)]}{\partial \alpha^T} = \lim_{m \to \infty} \sum_{i=1}^{m} \frac{\partial [D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)S_i(\beta)]}{\partial \alpha^T}
\]

\[
= \lim_{m \to \infty} \sum_{i=1}^{m} S_i(\beta) \frac{\partial [D_i^T(\beta)\hat{V}_{i}^{-1}(\beta)]}{\partial \alpha^T} = \sum_{i=1}^{m} \frac{\partial \alpha^*(\beta)}{\partial \beta^T} \times \hat{V}_{i}^{-1}(\beta)
\]

\[
= \frac{1}{m} \times \frac{\partial \alpha^*(\beta)}{\partial \beta^T} \times \hat{V}_{i}^{-1}(\beta)
\]

\[
= \frac{1}{m} \times \frac{\partial \alpha^*(\beta)}{\partial \beta^T} \times \hat{V}_{i}^{-1}(\beta)
\]

\[
\lim_{m \to \infty} C = \lim_{m \to \infty} \frac{\partial \alpha^*(\beta)}{\partial \beta^T} \times \hat{V}_{i}^{-1}(\beta) = \text{constant}
\]

39
Now assume $\beta$ is fixed. Looking at the second half of the right-hand side of equation 3.26, Taylor expansion about $\alpha$ gives

$$m^{-\frac{1}{2}} \sum_{i=1}^{m} m^{-\frac{1}{2}} U_i[\beta, \alpha^*(\beta)]$$

$$= m^{-\frac{1}{2}} \sum_{i=1}^{m} U_i[\beta, \alpha] + m^{-\frac{1}{2}} \sum_{i=1}^{m} \frac{\partial U_i[\beta, \alpha^*(\beta)]}{\partial \alpha^T} \times m^{\frac{1}{2}} (\alpha^* - \alpha) + o_P(1) \quad (3.28)$$

$$= A^* + B^* \times C^* + o_P(1)$$

$$A^* = m^{-\frac{1}{2}} \sum_{i=1}^{m} U_i[\beta, \alpha] = m^{-\frac{1}{2}} \sum_{i=1}^{m} D_i^T V_i^{-1} S_i$$

$$E(A^*) = E \left[ m^{-\frac{1}{2}} \sum_{i=1}^{m} (D_i^T V_i^{-1} S_i) \right] = m^{-\frac{1}{2}} \sum_{i=1}^{m} E(D_i^T V_i^{-1} S_i)$$

$$= m^{-\frac{1}{2}} \sum_{i=1}^{m} D_i^T V_i^{-1} E(S_i) = m^{-\frac{1}{2}} \sum_{i=1}^{m} D_i^T V_i^{-1} E(y_i - E(y_i))$$

$$\to 0$$

$$V(A^*) = V \left[ m^{-\frac{1}{2}} \sum_{i=1}^{m} D_i^T V_i^{-1} S_i \right] = m^{-1} \sum_{i=1}^{m} V(D_i^T V_i^{-1} S_i)$$

$$= m^{-1} \sum_{i=1}^{m} (D_i^T V_i^{-1}) \times V(S_i) \times (D_i^T V_i^{-1})^T$$

$$= m^{-1} \sum_{i=1}^{m} D_i^T V_i^{-1} \times V[(y_i - E(y_i))] \times V_i^{-1} D_i$$

$$= m^{-1} \sum_{i=1}^{m} D_i^T V_i^{-1} \times V(y_i) \times V_i^{-1} D_i$$

So $A^*$ has the following (asymptotic) multivariate normal distribution due to the Central Limit theorem.

$$A^* \sim N_p \left[ 0, \lim_{m \to \infty} \left\{ \frac{1}{m} \sum_{i=1}^{m} D_i^T V_i^{-1} \times V(y_i) \times V_i^{-1} D_i \right\} \right] \quad (3.29)$$

$$B^* = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_i[\beta, \alpha^*(\beta)]}{\partial \alpha^T} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial [D_i^T V_i^{-1} S_i]}{\partial \alpha^T}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial [D_i^T V_i^{-1}]}{\partial \alpha^T} S_i = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial [D_i^T V_i^{-1}]}{\partial \alpha^T} [y_i - E(y_i)]$$

$$\to \frac{1}{m} o_P(m) = o_P(1) \quad (3.30)$$

Taylor's expansion of $\hat{\alpha}(\beta, \hat{\phi})$ about $\phi$ gives

$$\hat{\alpha}[\beta, \hat{\phi}] = \hat{\alpha}[\beta, \phi] + \frac{\partial^2 \hat{\alpha}[\beta, \phi]}{\partial \phi^2} (\hat{\phi} - \phi) \Rightarrow \hat{\alpha}[\beta, \phi] - \hat{\alpha}[\beta, \phi] = \frac{\partial \hat{\alpha}[\beta, \phi]}{\partial \phi} (\hat{\phi} - \phi)$$

40
\[ C^* = m^{\frac{1}{2}}(\alpha^* - \alpha) = m^{\frac{1}{2}}[\hat{X}(\beta, \phi) - \hat{X}(\beta, \phi) + \hat{X}(\beta, \phi) - \alpha] = m^{\frac{1}{2}} \left\{ \frac{\partial \hat{X}[\beta, \phi]}{\partial \phi} (\hat{\phi} - \phi) + \hat{X}[\beta, \phi] - \alpha \right\} \to O_P(1) + O_P(1) \to O_P(1) \text{ as } m \to \infty \]

By the theorem conditions, equations 3.7 and 3.8. Substituting equations 3.29, 3.30 and 3.31 into equation 3.28 shows that \( m^{-\frac{1}{2}} \sum U_i \) has the same distribution as \( A^* \) in the limit.

\[ m^{-\frac{1}{2}} \sum_{i=1}^{m} U_i[\beta, \alpha^*(\beta)] \sim N_p \left[ 0_p, \lim_{m \to \infty} \left\{ \frac{1}{m} \sum_{i=1}^{m} D_i^T V_i^{-1} \times V(y_i) \times V_i^{-1} D_i \right\} \right] \]

Going back to equation 3.26,

\[ m^{\frac{1}{2}}(\hat{\beta}_{GEE} - \beta) \approx \left\{ \sum_{i=1}^{m} \frac{\delta U_i[\beta, \alpha^*(\beta)]}{\delta \beta^T} \times \frac{1}{m} \right\}^{-1} \left\{ \sum_{i=1}^{m} U_i[\beta, \alpha^*(\beta)] \times \frac{1}{m} \right\}^{-1} \to \left\{ \frac{-m}{m} \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \times O_P(1) \right\}^{-1} \{A^* + O_P(1)\} \to -m \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} A^* \]

Therefore

\[ E[m^{\frac{1}{2}}(\hat{\beta}_{GEE} - \beta)] = -m \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \times E(A^*) = - \left\{ \frac{1}{m} \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \times 0_p \to 0_p \]

\[ \lim_{m \to \infty} V[m^{\frac{1}{2}}(\hat{\beta}_{GEE} - \beta)] = \lim_{m \to \infty} m^2 \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \times V(A^*) \times \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1}^T \]

\[ = \lim_{m \to \infty} m^2 \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \times \left\{ \frac{1}{m} \sum_{i=1}^{m} D_i^T V_i^{-1} V(y_i) V_i^{-1} D_i \right\} \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \]

\[ = \lim_{m \to \infty} m \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \times \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} V(y_i) V_i^{-1} D_i \right\} \left\{ \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right\}^{-1} \]

\[ = V_{GEE} \]

Since \( A^* \) is multivariate normal so is \( m^{\frac{1}{2}}(\hat{\beta}_{GEE} - \beta) \) since it is a linear product of a constant matrix and a multivariate normal vector.
ML and REML methods are easily applied to continuous data using the multivariate normal distribution, as shown in Chapter 2, to give a broad class of plausible models. However, there are no rich distributions for categorical data, so ML based methods cannot be applied. Chapter 3 introduced GEEs, which assumes the data arise from univariate generalised exponential families rather than a multivariate distribution. These are then 'merged' using a working correlation matrix to form an overall model for the data. However, although offering new categorical models, such as binary, binomial and Poisson, the scope of application to repeated ordinal data is still limited, with binary being the only obvious example. In this chapter, methodology will be developed and analysed which can estimate parameters for repeated ordinal data. A model based on the data arising as discrete realisations of multivariate distributions (latent trends) is considered. The mean parameters, variance and correlation structure will be estimated using a mixture of SIMPLEX maximum likelihood and Fisher's scoring. A set of quasi-estimation equations or QEEs will be developed, their name derived from the similarities to the formulation of the GEEs.
4.1 Quasi-estimation equations (QEEs)

The fundamental assumption for the formulation of QEEs is that there is an underlying \( n \)-variate distribution. Although this can be any distribution, we will concentrate on the multivariate normal distribution with an unknown mean, \( \mu_i \), and variance matrix, \( V_i \), for each individual. However, this work is easily applied to the case where the underlying distribution is not multivariate normal. Call the underlying variable from this distribution \( y_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{in}^*) \). Since we know the underlying distribution, we have a likelihood,

\[
L_i(y_i^*, \mu_i, V_i) = L_i = \frac{\exp\left[-\frac{1}{2}(y_i^* - \mu_i)^T V_i^{-1}(y_i^* - \mu_i)\right]}{(2\pi)^{\frac{1}{2}} |V_i|^{\frac{1}{2}}}
\]  

(4.1)

However the data are discrete realisations of this distribution, so the likelihood can not be maximised directly. Instead, since the data are realisations of equation 4.1, a set of cut-off points will be used to define the boundaries between different categories. The form of these cut-offs needs to be known, and will be considered later in this chapter. If there are \( l_1 \) categories for the first measurement, \( l_2 \) categories for the second etc., then the vectors of cut-offs for any individual are defined as

\[
c_{i1} = (c_{i1,1}, c_{i1,2}, \ldots, c_{i1,l_1-1})^T
\]

\[
c_{in} = (c_{i,n,1}, c_{i,n,2}, \ldots, c_{i,n,l_n-1})^T
\]

and a matrix containing all the cut-offs for that individual, \( c = (c_{i1}, \ldots, c_{in}) \). This gives \( \sum_{j=1}^n (l_j - 1) \) cut-offs to estimate. Now the boundaries between the different ordinal responses have been defined, we can calculate the probability of obtaining a specified set of data; the 'likelihood' of the set occurring.

\[
P_i = P_i(c_i, \mu_i, V_i, y_i) = P(c_{i1,y_{i1}-1} \leq y_{i1}^* < c_{i1,y_{i1}+1}, \ldots, c_{i,n,y_{in}-1} \leq y_{in}^* < c_{i,n,y_{in}})
\]

(4.2)

\[
P_i = P(y_i^* \text{ is in region } R_{y_i})
\]

\[
= \int_{c_{i1,y_{i1}-1}}^{c_{i1,y_{i1}+1}} \cdots \int_{c_{i,n,y_{in}-1}}^{c_{i,n,y_{in}+1}} \frac{\exp\left[-\frac{1}{2}(y_i^* - \mu_i)^T V_i^{-1}(y_i^* - \mu_i)\right]}{(2\pi)^{\frac{1}{2}} |V_i|^{\frac{1}{2}}} dy_{n}^* \cdots dy_1^*
\]

(4.3)

where

\[
c_{i1,0} = c_{i2,0} = c_{i3,0} = \cdots = -\infty
\]

\[
c_{i1,l_1} = c_{i2,l_2} = c_{i3,l_3} = \cdots = \infty
\]
and \( \mathcal{R}_{y_i} \) is the \( n \)-dimensional rectangular region defined by the \( n \) inequalities in equation 4.2.

Equation 4.3 is the likelihood for individual \( i \), so these \( m \) expressions are multiplied together to find the likelihood for the data, which depends on the individual’s means and variance matrix and the \( \sum_{j=1}^{n} (l_j - 1) \) cut-off points.

\[
L = \prod_{i=1}^{m} P_i(c_i, y_i; \mu_i, V_i) \quad (4.4)
\]

This likelihood will be very small, so the log of equation 4.4 is usually maximised

\[
\text{Lik} = \log L = \log \left( \prod_{i=1}^{m} P_i(c_i, y_i; \mu_i, V_i) \right)
= \sum_{i=1}^{m} \log P_i(c_i, y_i; \mu_i, V_i) \quad (4.5)
\]

Equation 4.5 is maximised with respect to \( c_i \) to find the estimates of the cut-off points. However, the likelihood also depends on the means and variance. A three stage maximisation process is proposed to separate out the estimation of the three sets of parameters. Each one is estimated by assuming that the other two are fixed and this is repeated until the estimates converge. It should be noted that the mean vector and variance matrix for an individual, \( \mu_i \) and \( V_i \) respectively, will usually be modelled.

\[
\begin{align*}
\mu_i &= X_i \beta \\
V_i &= V_i(\alpha)
\end{align*}
\]

This is similar to the work from chapter 2, meaning that the log-likelihood, equation 4.5, has to be maximised with respect to \( c, \alpha, \beta \).

### 4.1.1 Modelling cut-off points

So far the cut-off points have been not been modelled. However, given the number of cut-off points, it would logical to model them as we do for the means. Within the framework of this method, it is possible to does this and then go on to estimate the cut-off parameters. Assume the following (linear) modelling for the cut-off points.

\[
\begin{align*}
c_{i,1} &= Z_{i,1} \gamma \\
&\vdots \\
c_{i,n} &= Z_{i,n} \gamma
\end{align*}
\]

where \( \gamma \) is a \( q \times 1 \) vector of parameters and \( Z_{ij} \) is a \( (l_j - 1) \times q \) data matrix for the cut-offs at time \( j \). This could, depending on the choice of \( Z_{ij} \) reduce the number of parameters dramatically. Two examples of cut-off data matrices follow.
Equal cut-off points across time

Assume that there are an equal number of categories for each time, so \( t_1 = t_2 = \cdots = t \). By defining the \( Z_{ik} \) matrices to be identity matrices, \( I_t \),

\[
c_{i,1,1} = c_{i,2,1} = \cdots = c_{i,n,1} = \gamma_1
\]

\[
c_{i,1,2} = c_{i,2,2} = \cdots = c_{i,n,2} = \gamma_2
\]

\[
\vdots
\]

\[
c_{i,1,t} = c_{i,2,t} = \cdots = c_{i,n,t} = \gamma_t
\]

which means that the cut-off points do not change across time.

Equally spaced

Another method of parameterising the cut-off points is to consider them as being equally spaced. For this to occur, the data matrix for \( Z_{ik} \) is

\[
\begin{pmatrix}
1 & 0 \\
1 & 1 \\
\vdots & \vdots \\
1 & t-1
\end{pmatrix}
\]

which gives the cut-off points as

\[
c_{i,1,1} = c_{i,2,1} = \cdots = c_{i,n,1} = \gamma_1
\]

\[
c_{i,1,2} = c_{i,2,2} = \cdots = c_{i,n,2} = \gamma_1 + \gamma_2
\]

\[
\vdots
\]

\[
c_{i,1,t} = c_{i,2,t} = \cdots = c_{i,n,t} = \gamma_1 + (t-1)\gamma_2
\]

4.1.2 Quasi-estimation equations for \( \beta \)

Consider the problem of maximising the log-likelihood, equation 4.5, with respect to \( \beta \) whilst keeping \( \gamma \) and \( \alpha \) constant. To do this, the partial derivatives are equated to zero as follows

\[
\frac{\partial \text{Lik}}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{m} \log \int_{R_{ui}} L_i d\gamma_i^* = \sum_{i=1}^{m} \frac{\partial}{\partial \beta} \log \int_{R_{ui}} L_i d\gamma_i^* = \sum_{i=1}^{m} \int_{R_{ui}} \frac{\partial L_i}{\partial \beta} d\gamma_i^* = \sum_{i=1}^{m} \int_{R_{ui}} L_i d\gamma_i^* = 0_p
\]
To set up quasi-estimation equations, the second derivative of equation 4.5 (which is the derivative of equation 4.6) needs to be found.

\[
\frac{\partial^2 \text{Lik}}{\partial \beta^T \partial \beta} = \frac{\partial}{\partial \beta^T} \frac{\partial \text{Lik}}{\partial \beta} = \frac{\partial}{\partial \beta^T} \sum_{i=1}^{m} \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \int \int_{\mathcal{R}_v} L_i dy_i^* = \sum_{i=1}^{m} \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \int \int_{\mathcal{R}_v} L_i dy_i^* \\
= \sum_{i=1}^{m} \left[ \int \int_{\mathcal{R}_v} \frac{\partial^2 L_i}{\partial \beta^T \partial \beta} dy_i^* \int \int_{\mathcal{R}_v} L_i dy_i^* - \left( \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \right) \left( \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \right)^T \right] \\
= \sum_{i=1}^{m} \left[ \int \int_{\mathcal{R}_v} \frac{\partial^2 L_i}{\partial \beta^T \partial \beta} dy_i^* \int \int_{\mathcal{R}_v} L_i dy_i^* \right] - \left( \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \right) \left( \int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \right)^T \right] \\
\text{(4.7)}
\]

By defining

\[
A_{i1} = \frac{\int \int_{\mathcal{R}_v} \frac{\partial L_i}{\partial \beta} dy_i^* \int \int_{\mathcal{R}_v} L_i dy_i^*}{\int \int_{\mathcal{R}_v} L_i dy_i^*} \\
A_{i2} = \frac{\int \int_{\mathcal{R}_v} \frac{\partial^2 L_i}{\partial \beta^T \partial \beta} dy_i^*}{\int \int_{\mathcal{R}_v} L_i dy_i^*}
\]

then

\[
\frac{\partial \text{Lik}}{\partial \beta} = \sum_{i=1}^{m} A_{i1} \\
\frac{\partial^2 \text{Lik}}{\partial \beta^T \partial \beta} = \sum_{i=1}^{m} \left( A_{i2} - A_{i1} A_{i1}^T \right)
\]

The quasi-estimation equations can be solved iteratively using Fisher's scoring to give the iterative scheme

\[
\beta_{j+1} = \beta_j - \left[ \sum_{i=1}^{m} \left( A_{i2} - A_{i1} A_{i1}^T \right) \right]^{-1} \sum_{i=1}^{m} A_{i1}
\text{ (4.8)}
\]

This holds for any multivariate distribution, although for these equations to be fully defined, the two derivatives of the individual log-likelihoods, equation 4.1, need to be known. For the multivariate normal distribution these are

\[
L_i = \frac{\exp\left[-\frac{1}{2}(y_i^* - \mu_i)^T V_i^{-1}(y_i^* - \mu_i)\right]}{(2\pi)^{\frac{n}{2}} |V_i|^{\frac{1}{2}}}
\]

46
This gives

\[ A_{i1} = \frac{\int \cdots \int_{R_m} \frac{\partial L_i}{\partial \beta} \, dy_i^*}{\int \cdots \int_{R_m} L_i \, dy_i^*} \]

\[ = \frac{\int \cdots \int_{R_m} X_i^T V_i^{-1} (y_i^* - X_i \beta) L_i \, dy_i^*}{\int \cdots \int_{R_m} L_i \, dy_i^*} \]

\[ = -X_i^T V_i^{-1} X_i L_i + \frac{\int \cdots \int_{R_m} X_i^T V_i^{-1} (y_i^* - \mu_i)(y_i^* - \mu_i)^T V_i^{-1} X_i L_i \, dy_i^*}{\int \cdots \int_{R_m} L_i \, dy_i^*} \]

\[ A_{i2} = \frac{\int \cdots \int_{R_m} \frac{\partial^2 L_i}{\partial \beta \partial \beta^T} \, dy_i^*}{\int \cdots \int_{R_m} L_i \, dy_i^*} \]

\[ = -X_i^T V_i^{-1} X_i L_i + \frac{\int \cdots \int_{R_m} X_i^T V_i^{-1} (y_i^* - \mu_i)(y_i^* - \mu_i)^T V_i^{-1} X_i L_i \, dy_i^*}{\int \cdots \int_{R_m} L_i \, dy_i^*} \]

### 4.1.3 Quasi-estimation equations for \( \gamma \)

Using the proof outlined in appendix B the derivatives of the log-likelihood, 4.5 can be found.

\[ \frac{\partial \text{Lik}}{\partial \gamma} = \frac{\partial}{\partial \gamma} \sum_{i=1}^{m} \log \int \cdots \int_{R_m} L_i \, dy_i^* \]

\[ = \sum_{i=1}^{m} \frac{\partial}{\partial \gamma} \log \int \cdots \int_{R_m} L_i \, dy_i^* \]
\[
\frac{\partial^2 \text{Lik}}{\partial \gamma_j \partial \gamma_k} = \sum_{i=1}^{m} \left( \int_{c_{11}}^{c_{12}} \int_{c_{11}}^{c_{12}} \left( \int_{c_{21}}^{c_{22}} \ldots \int_{c_{n1}}^{c_{n2}} \text{Ldy}^* \right) \frac{\partial c_{12}}{\partial \gamma_j} - \int_{c_{21}}^{c_{22}} \left( \int_{c_{11}}^{c_{12}} \int_{c_{11}}^{c_{12}} \left( \int_{c_{21}}^{c_{22}} \ldots \int_{c_{n1}}^{c_{n2}} \text{Ldy}^* \right) \frac{\partial c_{11}}{\partial \gamma_j} \right) \right) \right)
\]

\[(4.11)\]
\[
\sum_{i=1}^{m} \left[ \int_{c_{11}}^{c_{12}} \int_{c_{n-2,1}}^{c_{n-2,2}} Ldy^* \right] \frac{\partial c_{n2}}{\partial \gamma_j} \frac{\partial c_{n-1,2}}{\partial \gamma_k} - \sum_{i=1}^{m} \left[ \int_{c_{11}}^{c_{12}} \int_{c_{n-2,1}}^{c_{n-2,2}} Ldy^* \right] \frac{\partial c_{n2}}{\partial \gamma_j} \frac{\partial c_{n-1,1}}{\partial \gamma_k}
\]

\[- \sum_{i=1}^{m} \left[ \int_{c_{12}}^{c_{n-1,1}} \int_{c_{11}}^{c_{n-1,1}} Ldy^* \right] \frac{\partial c_{n1}}{\partial \gamma_j} \frac{\partial c_{n2}}{\partial \gamma_k} + \sum_{i=1}^{m} \left[ \int_{c_{12}}^{c_{n-1,1}} \int_{c_{11}}^{c_{n-1,1}} Ldy^* \right] \frac{\partial c_{n1}}{\partial \gamma_j} \frac{\partial c_{n-1,1}}{\partial \gamma_k}
\]

\[
\left( \frac{\partial L_{ik}}{\partial \gamma_j} \right)^T \left( \frac{\partial L_{ik}}{\partial \gamma_j} \right)
\]

where the square brackets around integrals means that the likelihood, \( L \), is evaluated at the point(s) given in the constraint.

It is only left to define some of the terms within this expression. Since the cut-off points are linear combinations of the \( \gamma \),

\[
\sum_{i=1}^{m} \left[ \int_{c_{12}}^{c_{n-1,2}} \int_{c_{11}}^{c_{n-1,1}} Ldy^* \right] \frac{\partial^2 c_{n1}}{\partial \gamma_j \partial \gamma_k} = 0
\]

\[
\sum_{i=1}^{m} \left[ \int_{c_{12}}^{c_{n-1,2}} \int_{c_{11}}^{c_{n-1,1}} Ldy^* \right] \frac{\partial^2 c_{n2}}{\partial \gamma_j \partial \gamma_k} = 0
\]

\[
\vdots
\]

\[
\sum_{i=1}^{m} \left[ \int_{c_{12}}^{c_{n-1,2}} \int_{c_{11}}^{c_{n-1,1}} Ldy^* \right] \frac{\partial^2 c_{k2}}{\partial \gamma_j \partial \gamma_k} = 0
\]

\[
\frac{\partial c_{k2}}{\partial \gamma_j} = (Z_k)_{c_{k+1,j}}
\]
\[ \frac{\partial c_{k1}}{\partial \gamma_j} = (Z_k)_{c_{k1}j} \]

If \( y_j \) is fixed as \( c_{j2} \) then

\[
\frac{\partial L}{\partial \gamma_k} = \frac{\partial L}{\partial y_1} \frac{d y_1}{d \gamma_k} + \frac{\partial L}{\partial y_2} \frac{d y_2}{d \gamma_k} + \ldots + \frac{\partial L}{\partial y_n} \frac{d y_n}{d \gamma_k}
\]

\[ = \frac{\partial L}{\partial y_j} \times (Z_j)_{c_{j2,k}} \quad (4.13) \]

If \( y_j \) is fixed as \( c_{j1} \) then

\[
\frac{\partial L}{\partial \gamma_k} = \frac{\partial L}{\partial y_1} \frac{d y_1}{d \gamma_k} + \frac{\partial L}{\partial y_2} \frac{d y_2}{d \gamma_k} + \ldots + \frac{\partial L}{\partial y_n} \frac{d y_n}{d \gamma_k}
\]

\[ = \frac{\partial L}{\partial y_j} \times (Z_j)_{c_{j1,k}} \quad (4.14) \]

\[
L_i = \exp\left[-\frac{1}{2}(y_i^T V_i^{-1} y_i^* - y_i^T V_i^{-1} X_i \beta - \beta^T X_i^T V_i^{-1} y_i^* + \beta^T X_i^T V_i^{-1} X_i \beta]\right]
\]

\[ = \frac{1}{(2\pi)^{\frac{n}{2}} V_i^{\frac{1}{2}}} \]

\[
\frac{\partial L_i}{\partial y} = -\frac{1}{2}(-V_i^{-1} X_i \beta - V_i^{-1} X_i y_i^*) L_i
\]

\[ = V_i^{-1} (X_i \beta - y_i^*) L_i \]

\[ = \left( \frac{\partial L_i}{\partial y_1}, \ldots, \frac{\partial L_i}{\partial y_n} \right)^T \quad (4.15) \]

Using all of these equations 4.12 is fully defined.

The quasi-estimation equations which are used to estimate \( \gamma \) are defined as the roots of equation 4.11. i.e.

\[
\sum_{i=1}^{m} \frac{\partial}{\partial \gamma_j} \int \ldots \int_{R_{y_i}} L_i dy_i^* = 0 \quad (4.16)
\]

which are solved using an iterative multidimensional Newton-Raphson algorithm:

\[
\gamma_{j+1} = \gamma_j - \left[ \sum_{i=1}^{m} \int \ldots \int_{R_{y_i}} \frac{\partial^2 L_i}{\partial \gamma T \partial \gamma} \right]^{-1} \left[ \int \ldots \int_{R_{y_i}} \frac{\partial L_i}{\partial \gamma} dy_i^* \left( \int \ldots \int_{R_{y_i}} L_i dy_i^* \right) \right]
\]

\[ = \gamma_j - \sum_{i=1}^{m} \frac{\int \ldots \int_{R_{y_i}} \frac{\partial L_i}{\partial \gamma} dy_i^*}{\int \ldots \int_{R_{y_i}} L_i dy_i^*} \left( \int \ldots \int_{R_{y_i}} L_i dy_i^* \right) \left( \int \ldots \int_{R_{y_i}} L_i dy_i^* \right)^T \quad (4.17) \]
Using these two methods, it is possible to estimate the parameters, $\hat{\beta}_{QEE}$ and $\hat{\gamma}_{QEE}$. Although it is possible to use similar techniques to establish quasi-estimation equations for the $\alpha$ estimate, this could lead to problems with impossible correlation matrices being created, with elements greater than 1. Therefore ML will be used for their estimation, using a combination of methods discussed in previous chapters.

4.1.4 Non-linear links

In some cases it might be natural to assume that the underlying trend is some non-linear function of the parameters. For instance, consider the very simplified case where a cancerous growth is graded according to its size, which is dependent on the underlying number of cancerous cells within the growth. Although the number of cells would be discrete, the large number of them would make it approximately continuous. In this case, a log-link would make more sense than a linear link. That is to say, $y^*_i = \exp(X_i\beta)$ rather than $y^*_i = X_i\beta$ since the latent number of cells is, by necessity, positive. Also, the factors might be assumed to be multiplicative in nature, which makes this a natural link to use. In such a case, the above analysis would be invalid. Although it is possible to modify the QEEs and derive estimation equations for such analysis, the estimation of the cut-off points becomes extremely complicated. This is because they need to be modelled using the same non-linear link as the mean parameters, but this interferes with the multivariate calculus given in appendix B.

4.1.5 Computation of QEEs estimates

A very important aspect of this work is the computation of the QEE estimates. Although the equations have now been fully defined, the actual computation is extremely tricky. A computer program has been developed in FORTRAN to carry out QEEs estimation of the mean, cut-off and correlation parameters and is available from the author on request. It incorporates the NAG (Numerical Algorithms Group) libraries to evaluate some of the complicated functions, such as the multidimensional integrals, although most of the code has been built from scratch since no existing program could implement the estimation procedure at a sufficient speed.

4.1.6 Simulation studies

Simulations were used to examine the performance of this estimation method. Two groups of 50 repeated measurements (giving 100 individuals) were generated from a random tri-variate normal distribution (giving 3 time measurements per individual), with mean and variance
structure as in table 4.1. The variance/covariance values were taken as 2 for the variance and $\frac{1}{2}$ for the correlation term, $\alpha$. These continuous random numbers were then changed into categorical data using the cut-off points $(0, 2, 4, 6)$ for the three times, the model in section 4.1.1. To summarise, $\alpha = \frac{1}{2}$, $\beta = (1, 2)^T$ and $\gamma = (0, 2, 4, 6)^T$ were estimated using the 100 sets of 3 repeated measurements. 100 Monte Carlo simulations were carried out and the average and variances of the 100 estimates were calculated. The program used the SIMPLEX method to minimise the negative log-likelihood, equation 4.5, with respect to $\alpha$ and QEEs to estimate $\beta$ and $\gamma$. It should be noted that the variance value acted as a scale factor. That is to say that by multiplying the variance by a constant, $k$, would give estimates of $\beta$ and $\gamma$ as $k^{\frac{1}{2}}\beta$ and $k^{\frac{1}{2}}\gamma$ respectively. This was due to a lack of identifiability of the underlying latent trend's scale. The implication of this is that the variance had to be fixed in order to obtain comparable estimates from each simulation. The results are shown in table 4.2. The reported figures are the mean of the 100 estimates, the standard deviation of the estimates, a test statistic (which is below the parameter estimates) and its critical value (below the standard deviations) for a 95% significant level. There were two test statistics used, the standard $t$-test and its multivariate equivalent the Hotelling T-statistic, calculated as:

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$$

$$T = \frac{100 - p}{p} (\bar{x} - \mu)^T S^{-1} (\bar{x} - \mu) \sim F_{p, 100-p}$$

These are tests of the hypothesis $H_0 : x \sim N(\mu, \sigma^2)$ and $H_0 : x \sim N(\mu, \Sigma)$ respectively, depending if one or more parameters were estimated.
### Table 4.1: Specification of the simulation study for QEEs

<table>
<thead>
<tr>
<th>Correlation structure</th>
<th>Variance matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>$2 \begin{pmatrix} 1 &amp; 0 &amp; 0 \ 0 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>1-dependent</td>
<td>$2 \begin{pmatrix} 1 &amp; 1/2 &amp; 0 \ 1/2 &amp; 1 &amp; 1/2 \ 0 &amp; 1/2 &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>$2 \begin{pmatrix} 1 &amp; 1/2 &amp; 1/2 \ 1/2 &amp; 1 &amp; 1/2 \ 1/2 &amp; 1/2 &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$2 \begin{pmatrix} 1 &amp; 1/2 &amp; 1/2^2 \ 1/2 &amp; 1 &amp; 1/2 \ 1/2 &amp; 1/2 &amp; 1 \end{pmatrix}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Design Matrix / Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>$\mu = \begin{pmatrix} 0 \ 0 \ 0 \ 2 \ 4 \end{pmatrix}$</td>
</tr>
<tr>
<td>Group 2</td>
<td>$\mu = \begin{pmatrix} 1 \ 0 \ 1 \ 2 \ 3 \ 5 \end{pmatrix}$</td>
</tr>
</tbody>
</table>
These results show that the parameters were estimated well. That is to say that the average values were around the true values and the standard deviations were small enough to suggest that most estimates were close to their true value. The test statistics verify that the estimates were consistent with their true values for the correlation parameter (where present) and the mean parameters. However significant values were obtained for all of the cut-off point estimates, which suggests a problem. The $\gamma$ estimates were close to their true values and t-tests (conducted on each parameter individually) gave non-significant values, so individually they had constant means and variances. Therefore the problem concerns them having non-constant variance/covariance matrix over all the estimates.

Table 4.2: QEEs estimates of the underlying multivariate normal distribution, using ML for $\alpha$ and QEEs for $\beta$ and $\gamma$
Stability

As a prelude to the theoretical analysis of the asymptotic $\alpha$, $\beta$ and $\gamma$ estimates, a set of simulations were carried out to examine the behaviour of the estimates as the sample size (number of individuals) increased. Simulations were run using $10, 12, \ldots, 200$ individuals and the estimates were calculated and plotted. Due to the large number of possible graphs, only the exchangeable correlation structure was used since these results were typical of the others. The results in figure 4.1 show the estimated values of the correlation parameter, the mean parameters and the cut-off parameters. The dotted lines which have been added are the smoothed plots, using robust locally linearly fitted values. The bands, where shown, are $\pm 2$ smoothed standard deviations as calculated from the negative inverse of the second derivative matrix of the log-likelihood.
Figure 4.1(a) $\alpha$
Figure 4.1(b) $\beta$
Figure 4.1. Stability plots for the QREs estimates of the correlation, mean, and cut-off point parameters of the underlying exchangeable multivariate normal distribution.
The estimates seemed to settle to their true value, but the convergence was quite slow, with some estimates still being off when there was a sample size of 200. However, the standard error bands around the $\beta$ and $\gamma$ estimates decreased as the number of individuals increased, suggesting that the method is asymptotically unbiased, i.e.

$$
\lim_{m \to \infty} \hat{\alpha} = \alpha \\
\lim_{m \to \infty} \hat{\beta} = \beta \\
\lim_{m \to \infty} \hat{\gamma} = \gamma
$$

These are proved theoretically in section 4.2. In addition to these graphs, 'consistency plots' follow in figure 4.2. These show $m^{1/2}(\hat{\alpha} - \alpha)$, $m^{1/2}(\hat{\beta} - \beta)$ and $m^{1/2}(\hat{\gamma} - \gamma)$, with the ±2 standard deviations where possible. The values are not deviating from zero as the sample size increases, so they seem to converge at a rate close to $m^{1/2}$. The standard errors of the estimates also seem to converge to a fixed value (although it is quite slow) at a rate of $m^{1/2}$. This would suggest that the estimates are are asymptotically $m^{1/2}$ consistent. That is to say that

$$
\lim_{m \to \infty} m^{1/2}(\hat{\alpha} - \alpha) = O_P(1) \\
\lim_{m \to \infty} m^{1/2}(\hat{\beta} - \beta) = O_p + O_P(1) \\
\lim_{m \to \infty} m^{1/2} \sqrt{V(\hat{\beta} - \beta)} = \text{constant} \\
\lim_{m \to \infty} m^{1/2}(\hat{\gamma} - \gamma) = O_q + O_P(1) \\
\lim_{m \to \infty} m^{1/2} \sqrt{V(\hat{\gamma} - \gamma)} = \text{constant}
$$

This will also be investigated in section 4.2.
Figure 4.2(a) $m^\frac{1}{4}(\hat{a} - a)$
Figure 4.2(b) $m_1^\dagger (\hat{\beta} - \beta)$
Figure 4.2(c) $m^{\frac{1}{2}}sd(\beta)$
Figure 4.2(d) \( m^{\frac{1}{2}} (\gamma - \gamma) \)
Figure 4.2: Consistency plots for the QEEs estimates of the correlation, mean and cut-off point parameters of the underlying exchangeable multivariate normal distribution.
Correlation Miss-specification

The correlation structure in a repeated measurement experiment is important. In fact, many of the modern latent variable models are only concerned with the correlation (such as factor analysis and the 'LISREL' model). It is difficult to specify exactly what form this takes, but choosing the incorrect structure may lead to poor results. QEEs rely on there being an underlying latent trend from which the ordinal data are discrete realisations, so there is a need to know the correlation structure.

A set of simulations was run to test for incorrect correlation structures. The results are in tables 4.3, 4.4, 4.5 and 4.6. In all cases, the results show that the parameters were estimated almost identically, although with the miss-specified models there tended to be larger standard deviations. This suggests that the estimation process (QEEs) produces unbiased estimates when there is correlation miss-specification, although the true value will not be as consistently estimated. This is further reflected in the t-test and T-test values. Table 4.3 is redundant since there is no real miss-specification. When the underlying correlation is independent, the other correlation structures should model this well with correlation parameters close to zero. This occurred in these simulations, with t-statistics which are not significant, giving little evidence to reject the null hypothesis that they were zero. However a slight difference was shown in the $\beta$ and $\gamma$ parameters. Table 4.4 illustrates the effects of correlation miss-specification. Firstly, the t-test values are significant for the correlation parameter, except when the true correlation structure is used. Therefore the correlation parameter is extremely poorly estimated when there is miss-specification. For the $\beta$ and $\gamma$ estimates, the Hotelling T-values are smallest when the true correlation structure is used. This suggests that if the incorrect correlation structure is used, the resulting mean and cut-off parameters estimates will be less reliable than when there is no miss-specification. Tables 4.5 and 4.6 show similar trends, with significant values for the miss-specified correlation parameters, and smaller T-values when the true correlation structures are used.
<table>
<thead>
<tr>
<th>Correlation</th>
<th>α</th>
<th>S.D.</th>
<th>β</th>
<th>S.D.</th>
<th>γ</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9704</td>
<td>0.1638</td>
<td>1.9705</td>
<td>0.1903</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.9707</td>
<td>0.1574</td>
<td>3.9545</td>
<td>0.2524</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.0884</td>
<td>3.0892</td>
<td>5.9803</td>
<td>0.3397</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.7553</td>
<td>2.4665</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>-0.010</td>
<td>0.0768</td>
<td>1.0052</td>
<td>0.1581</td>
<td>-0.051</td>
<td>0.1660</td>
</tr>
<tr>
<td></td>
<td>1.2888</td>
<td>1.9842</td>
<td>1.9936</td>
<td>0.1348</td>
<td>1.9886</td>
<td>0.2029</td>
</tr>
<tr>
<td></td>
<td>0.1684</td>
<td>3.0892</td>
<td>0.1638</td>
<td>6.0351</td>
<td>0.3435</td>
<td>2.4665</td>
</tr>
<tr>
<td></td>
<td>3.3716</td>
<td>2.4665</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.0115</td>
<td>0.0589</td>
<td>0.9846</td>
<td>0.1784</td>
<td>-0.052</td>
<td>0.1629</td>
</tr>
<tr>
<td></td>
<td>1.9506</td>
<td>1.9842</td>
<td>1.9751</td>
<td>0.1458</td>
<td>1.9572</td>
<td>0.1768</td>
</tr>
<tr>
<td></td>
<td>1.6554</td>
<td>3.0892</td>
<td>0.1684</td>
<td>3.9562</td>
<td>0.2724</td>
<td>3.4094</td>
</tr>
<tr>
<td>Exponential</td>
<td>-0.008</td>
<td>0.0896</td>
<td>0.9835</td>
<td>0.1989</td>
<td>-0.062</td>
<td>0.1721</td>
</tr>
<tr>
<td></td>
<td>0.8822</td>
<td>1.9842</td>
<td>2.0071</td>
<td>0.1506</td>
<td>1.9823</td>
<td>0.2118</td>
</tr>
<tr>
<td></td>
<td>0.5627</td>
<td>3.0892</td>
<td>0.5627</td>
<td>3.9994</td>
<td>0.3000</td>
<td>5.8284</td>
</tr>
</tbody>
</table>

Table 4.3: QEEs estimates of the underlying independent multivariate normal distribution under correlation miss-specification
<table>
<thead>
<tr>
<th>Correlation</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
<th>$\gamma$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>0.9936</td>
<td>0.2387</td>
<td>1.9733</td>
<td>0.2603</td>
<td>-0.053</td>
<td>0.2036</td>
</tr>
<tr>
<td></td>
<td>2.0131</td>
<td>0.1710</td>
<td>4.0283</td>
<td>0.3483</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3953</td>
<td>3.0892</td>
<td>6.0934</td>
<td>0.4541</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.1275</td>
<td>2.4665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.5067</td>
<td>0.0467</td>
<td>0.9936</td>
<td>0.2279</td>
<td>-0.054</td>
<td>0.1923</td>
</tr>
<tr>
<td></td>
<td>1.4381</td>
<td>1.9842</td>
<td>2.0070</td>
<td>0.1516</td>
<td>2.0016</td>
<td>0.2230</td>
</tr>
<tr>
<td></td>
<td>0.1604</td>
<td>3.0892</td>
<td>3.9881</td>
<td>0.2990</td>
<td>3.9981</td>
<td>0.2990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.9632</td>
<td>2.4665</td>
<td>3.9632</td>
<td>2.4665</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.3656</td>
<td>0.0743</td>
<td>0.9702</td>
<td>0.2477</td>
<td>-0.051</td>
<td>0.1890</td>
</tr>
<tr>
<td></td>
<td>18.094</td>
<td>1.9842</td>
<td>1.9981</td>
<td>0.1456</td>
<td>2.0039</td>
<td>0.2439</td>
</tr>
<tr>
<td></td>
<td>0.7214</td>
<td>3.0892</td>
<td>4.0101</td>
<td>0.3181</td>
<td>4.0101</td>
<td>0.3181</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.2256</td>
<td>2.4665</td>
<td>4.2256</td>
<td>2.4665</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.4807</td>
<td>0.0573</td>
<td>0.9719</td>
<td>0.2177</td>
<td>-0.055</td>
<td>0.1891</td>
</tr>
<tr>
<td></td>
<td>3.3650</td>
<td>1.9842</td>
<td>2.0365</td>
<td>0.1391</td>
<td>1.9960</td>
<td>0.2121</td>
</tr>
<tr>
<td></td>
<td>5.2781</td>
<td>3.0892</td>
<td>4.0441</td>
<td>0.3072</td>
<td>4.0441</td>
<td>0.3072</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.0622</td>
<td>0.3781</td>
<td>6.0622</td>
<td>0.3781</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.2692</td>
<td>2.4665</td>
<td>4.2692</td>
<td>2.4665</td>
</tr>
</tbody>
</table>

Table 4.4: QEEs estimates of the underlying 1-dependent multivariate normal distribution under correlation miss-specification
<table>
<thead>
<tr>
<th>Correlation</th>
<th>$\alpha$ S.D.</th>
<th>$\beta$ S.D.</th>
<th>$\gamma$ S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>1.0037 0.2378</td>
<td>4.0490 0.3025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0376 0.1426</td>
<td>6.0869 0.4284</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.4748 3.0892</td>
<td>6.5328 2.4665</td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.3727 0.0671</td>
<td>0.9718 0.2416</td>
<td>-0.071 0.1788</td>
</tr>
<tr>
<td></td>
<td>18.972 1.9842</td>
<td>2.0482 0.1409</td>
<td>1.9941 0.2227</td>
</tr>
<tr>
<td></td>
<td>7.8086 3.0892</td>
<td>6.1252 0.4289</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.6002 2.4665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.5016 0.0793</td>
<td>0.9744 0.2601</td>
<td>-0.061 0.1818</td>
</tr>
<tr>
<td></td>
<td>0.2039 1.9842</td>
<td>2.0097 0.1355</td>
<td>2.0004 0.2486</td>
</tr>
<tr>
<td></td>
<td>0.9251 3.0892</td>
<td>6.0316 0.4170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.3814 2.4665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>0.5203 0.0764</td>
<td>0.9301 0.2664</td>
<td>-0.071 0.1947</td>
</tr>
<tr>
<td></td>
<td>2.6535 1.9842</td>
<td>1.9938 0.1338</td>
<td>3.9600 0.3466</td>
</tr>
<tr>
<td></td>
<td>3.4348 3.0892</td>
<td>5.9618 0.4506</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.3960 2.4665</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: QEEs estimates of the underlying exchangeable multivariate normal distribution under correlation miss-specification
<table>
<thead>
<tr>
<th>Correlation</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
<th>$\gamma$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9705</td>
<td>0.2689</td>
<td>2.0270</td>
<td>0.1533</td>
<td>-0.068</td>
<td>0.1996</td>
</tr>
<tr>
<td></td>
<td>3.5491</td>
<td>3.0892</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.6962</td>
<td>2.4665</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.4276</td>
<td>0.0586</td>
<td>2.0140</td>
<td>0.1417</td>
<td>-0.090</td>
<td>0.2239</td>
</tr>
<tr>
<td></td>
<td>12.353</td>
<td>1.9842</td>
<td>1.5027</td>
<td>3.0892</td>
<td>1.9714</td>
<td>0.2463</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.4384</td>
<td>0.0674</td>
<td>1.0292</td>
<td>0.2256</td>
<td>-0.042</td>
<td>0.2220</td>
</tr>
<tr>
<td></td>
<td>9.1414</td>
<td>1.9842</td>
<td>1.9766</td>
<td>0.1436</td>
<td>1.9885</td>
<td>0.2294</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>0.5121</td>
<td>0.0672</td>
<td>0.9867</td>
<td>0.2239</td>
<td>-0.536</td>
<td>0.1757</td>
</tr>
<tr>
<td></td>
<td>1.7932</td>
<td>1.9842</td>
<td>1.9965</td>
<td>0.1659</td>
<td>1.9928</td>
<td>0.2352</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.6:** QEEs estimates of the underlying exponential multivariate normal distribution under correlation miss-specification

The consistency of the estimation processes under correlation miss-specification was also considered. Again, due to the number of graphs, only the case with an exchangeable underlying correlation (which is typical of the other correlations) is shown in figure 4.3. The estimation process remained unbiased even under miss-specification, with the process converging at a rate of $m^{\frac{1}{2}}$. However the parameter standard errors were poorly estimated under miss-specification. 'Bands' were formed with the miss-specified model estimates being displaced from the true model estimates, with the $\beta$ estimates being more distinct than the $\gamma$ estimates. However, they all seemed to converge to a constant, suggesting that QEEs offer $m^{\frac{1}{2}}$ consistent estimates of the parameters even under correlation miss-specification.
Figure 4.3(a) $m^\dagger (\beta - \beta)$
Figure 4.3(b) \( m^{1/2} sd(\hat{\beta}) \)
Figure 4.3(c) $m^{\frac{1}{2}} \gamma (\gamma - \gamma)$
Figure 4.3(d) $m^{\frac{1}{2}}sd(\hat{\gamma})$

**Figure 4.3:** Consistency plots for the QEEs estimates of the mean and cut-off point parameters of the underlying exchangeable multivariate normal distribution under correlation miss-specification
4.2 Asymptotics of the QEE estimates

In the previous work we looked at graphical evidence of the asymptotics of the QEEs estimates. Following are proofs which show, given some conditions,

\[ m^\frac{1}{2} (\hat{\alpha}_{ML} - \alpha) \overset{D}{\rightarrow} N(0, I_\alpha^{-1}) \quad (4.18) \]
\[ m^\frac{1}{2} (\hat{\beta}_{QEE} - \beta) \overset{D}{\rightarrow} N_p(0_p, I_\beta^{-1}) \quad (4.19) \]
\[ m^\frac{1}{2} (\hat{\gamma}_{QEE} - \gamma) \overset{D}{\rightarrow} N_q(0_q, I_\gamma^{-1}) \quad (4.20) \]

where the variance matrices are the corresponding information matrices for that parameter. This agrees with the graphical evidence already shown. It should be noted that the conditions of the proofs require that there is no scale invariability, so the variances must be fixed for the theorems to hold. Although these theorems say that for a large enough sample the estimates will converge to their true values, in practice 'large enough' is difficult to define. From the graphical analysis, we have seen that 200 observations gave the suggestion of convergence, although a large sample size was still needed for full convergence.
4.2.1 $\beta$ estimates

Theorem

\[ m^{\frac{1}{2}}|\hat{\beta}_{QBE} - \beta| \overset{D}{\rightarrow} N_p(0, (I_\beta)^{-1}) \]  

(4.21)

asymptotically, where

\[ I_\beta = E_\beta \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right)^T \right] \]

Assumptions

1. The likelihood and all derivatives exist and are finite.
   - (a) $\int \cdots \int_{\mathcal{R}_y} L_i dy_i$
   - (b) $\frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_y} L_i dy_i$
   - (c) $\frac{\partial}{\partial \beta \partial \beta^T} \int \cdots \int_{\mathcal{R}_y} L_i dy_i$

2. $\alpha$ and $\gamma$ have been estimated (by ML and quasi-estimation equations respectively) and converged to $\hat{\alpha}$ and $\hat{\gamma}$, which means that the following exist and are finite.
   - (a) $\frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{R}_y} L_i dy_i$
   - (b) $\frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_y} L_i dy_i$

3. The following information matrices exist and are finite.
   - (a) $I_\beta = E_\beta \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right)^T \right]$
   - (b) $I_{\beta,\alpha} = E_\beta \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right) \right]$
   - (c) $I_{\beta,\gamma} = E_\beta \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_y} L_i dy_i \right)^T \right]$

4. The derivatives of the estimates exist and are finite.
   - (a) $\frac{\partial \hat{\alpha}}{\partial \beta} \not\rightarrow \infty$
   - (b) $\frac{\partial \hat{\gamma}}{\partial \beta} \not\rightarrow \infty$
Proof

Using the iterative scheme detailed in section 4.1.2 (equation 4.8), we can approximate \( \hat{\beta}_{QEE} - \beta \) by

\[
m^{1/2}(\hat{\beta}_{QEE} - \beta) \approx m^{1/2} \left( \sum_{i=1}^{m} \frac{\delta U_{ii}^T}{\delta \beta} \right)^{-1} \left( \sum_{i=1}^{m} U_{ii} \right)
\]

\[
= \left( \sum_{i=1}^{m} \frac{1}{m} \frac{\delta U_{ii}^T}{\delta \beta} \right)^{-1} \left( \sum_{i=1}^{m} \frac{1}{m^2} U_{ii} \right) \tag{4.22}
\]

where

\[
U_{ii} = \frac{\partial}{\partial \beta} \int \ldots \int r_{v_i} L_i dy_i^* = U_{ii}(\beta, \alpha, \gamma)
\]

\[
\hat{\alpha} = \hat{\alpha}(\hat{\beta}, \hat{\gamma}) \text{ is the ML estimate of } \alpha \text{ when } \beta, \gamma \text{ are constant}
\]

\[
\hat{\gamma} = \hat{\gamma}(\beta, \hat{\alpha}) \text{ is the ML estimate of } \gamma \text{ when } \beta, \alpha \text{ are constant}
\]

Using calculus to expand the matrix in equation 4.22 gives

\[
\frac{\delta U_{ii}^T}{\delta \beta} = \frac{\partial U_{ii}^T}{\partial \beta} + \frac{\partial U_{ii}^T}{\partial \alpha} \frac{\partial \alpha}{\partial \beta} + \frac{\partial U_{ii}^T}{\partial \gamma} \frac{\partial \gamma}{\partial \beta} \tag{4.23}
\]

\[
E_\beta(A_{ii})
\]

\[
= E_\beta \left[ \frac{\partial}{\partial \beta} \left( \frac{\partial}{\partial \beta} \int \ldots \int r_{v_i} L_i dy_i^* \right)^T \right]
\]

\[
= E_\beta \left[ \frac{\partial^2}{\partial \beta \partial \beta^T} \int \ldots \int r_{v_i} L_i dy_i^* - \left( \frac{\partial}{\partial \beta} \int \ldots \int r_{v_i} L_i dy_i^* \right) \left( \frac{\partial}{\partial \beta} \int \ldots \int r_{v_i} L_i dy_i^* \right)^T \right]
\]

\[
= \int \ldots \int r_{v_i} \left[ \frac{\partial^2}{\partial \beta \partial \beta^T} \int \ldots \int r_{v_i} L_i dy_i^* - \int \ldots \int r_{v_i} L_i dy_i^* \right] \int \ldots \int r_{v_i} L_i dy_i^* d\beta - I_\beta
\]

\[
= \int \ldots \int r_{v_i} \frac{\partial^2}{\partial \beta \partial \beta^T} \int \ldots \int r_{v_i} L_i dy_i^* d\beta - I_\beta
\]

\[
= \frac{\partial^2}{\partial \beta \partial \beta^T} \int \ldots \int r_{v_i} L_i dy_i^* d\beta - I_\beta
\]

\[
= \frac{\partial^2}{\partial \beta \partial \beta^T} 1 - I_\beta = -I_\beta
\]
Therefore, by the Khintchine Weak Law of Large Numbers, we have

\[
\frac{1}{m} \sum_{i=1}^{m} A_{ii} \xrightarrow{p} -I_{\beta} \quad (4.24)
\]

\[
\frac{1}{m} \sum_{i=1}^{m} B_{ii} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{ii}^{T}}{\partial \alpha} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \left( \frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right)^{T} \\
= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right)^{T} \\
= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \beta} \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \\
= \frac{1}{m} \frac{\partial}{\partial \beta} \left( \frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right) = 0_{1,p} \quad (4.25)
\]

since $\alpha$ is the ML estimate, which means that the derivative of the log-likelihood is zero, i.e.

\[
\frac{\partial}{\partial \alpha} \log L_{ik} = \frac{m}{\partial \alpha} \sum_{i=1}^{m} \log L_{i}dy_{i}^{*} = \sum_{i=1}^{m} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \\
= 0
\]

\[
\frac{1}{m} \sum_{i=1}^{m} C_{ii} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{ii}^{T}}{\partial \gamma} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \gamma} \left( \frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right)^{T} \\
= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \gamma} \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right)^{T} \\
= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \beta} \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \\
= \frac{1}{m} \frac{\partial}{\partial \beta} \left( \frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} \right) = 0_{q,p} \quad (4.26)
\]

since $\gamma$ is the QEE estimate which solves the following equation (by assumption, this estimation process has already converged, so the root exists).

\[
\frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_{\nu_{i}}} L_{i}dy_{i}^{*} = 0_{q}
\]
From the assumptions we know that

\[ D_1 = \frac{\partial \hat{\alpha}}{\partial \beta} \to \infty \quad (4.27) \]

\[ E_1 = \frac{\partial \hat{\gamma}^T}{\partial \beta} \to \infty \quad (4.28) \]

By substituting equations 4.24, 4.25, 4.26, 4.27 and 4.28 into equation 4.23 we have

\[ -\frac{1}{m} \frac{\delta U^T_{i1}}{\delta \beta} + \mathbf{I}_\beta \quad (4.29) \]

which is the first component of the right hand side of equation 4.22. Now consider the vector part (second component) of this equation. Using multidimensional Taylor's expansion

\[ U_{i1}(\beta, \hat{\alpha}, \hat{\gamma}) = U_{i1} + \frac{\partial U_{i1}}{\partial \alpha} (\hat{\alpha} - \alpha) + \frac{\partial U_{i1}}{\partial \gamma} (\hat{\gamma} - \gamma) + o_p(1) \quad (4.30) \]

\[ A_{i1} + B_{i1} D_{i1} + C_{i1} E_{i1} + o_p(1) \]

where \( U_{i1} = U_{i1}(\beta, \alpha, \gamma) \).

\[ E_{\beta}(A_{i1}) = E_{\beta} \left[ \frac{\partial}{\partial \beta} \left[ \cdots \int_{R_1} \cdots L_i dy^*_i \right] \right] = \int \cdots \int_{R_\beta} \frac{\partial}{\partial \beta} \left[ \cdots \int_{R_1} \cdots L_i dy^*_i \right] \cdots \int_{R_\beta} \cdots \int_{R_1} \cdots L_i dy^*_i d\beta = \frac{\partial}{\partial \beta} I_{i1} = o_p \quad (4.31) \]

\[ V_{\beta}(A_{i1}) = V_{\beta} \left[ \frac{\partial}{\partial \beta} \left[ \cdots \int_{R_1} \cdots L_i dy^*_i \right] \right] = E_{\beta} \left[ \left( \frac{\partial}{\partial \beta} \left[ \cdots \int_{R_1} \cdots L_i dy^*_i \right] \right) \left( \frac{\partial}{\partial \beta} \left[ \cdots \int_{R_1} \cdots L_i dy^*_i \right] \right)^T \right] = I_{i1} \quad (4.32) \]

By combining equations 4.31 and 4.32 and using the central limit theorem, we have

\[ \frac{1}{m^2} \sum_{i=1}^{m} A_{i1} \overset{d}{\to} N_p [0_p, I_{i1}] \quad (4.33) \]
Therefore

\[
\frac{1}{m} \sum_{i=1}^{m} B_{1i}^* = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{1i}}{\partial \alpha} P_{\beta} - I_{\beta,\alpha}
\]  
(4.34)

\[
E_\beta(C_{1i}^*)
\]

\[
E_\beta \left[ \frac{\partial}{\partial \gamma^T} \int \cdots \int_{R_\gamma} L_i dy_i^* \right]
\]

\[
E_\beta \left[ \frac{\partial}{\partial \gamma^T} \int \cdots \int_{R_\gamma} L_i dy_i^* \right] - \left( \frac{\partial}{\partial \gamma} \int \cdots \int_{R_\gamma} L_i dy_i^* \right) \left( \frac{\partial}{\partial \gamma} \int \cdots \int_{R_\gamma} L_i dy_i^* \right)^T
\]

\[
E_\beta \left[ \frac{\partial}{\partial \gamma^T} \int \cdots \int_{R_\gamma} L_i dy_i^* \right] - I_{\beta,\gamma}
\]

\[
\int \cdots \int_{R_\gamma} \frac{\partial}{\partial \gamma^T} \int \cdots \int_{R_\gamma} L_i dy_i^* d\beta - I_{\beta,\gamma}
\]

79
\[
\begin{align*}
\int \cdots \int_{R_0} \frac{\partial}{\partial \beta} \frac{\partial}{\partial \gamma} \int \cdots \int_{R_1} L \, du_i \, dy_i \, d\beta - I_{\beta, \gamma} \\
\frac{\partial}{\partial \beta} \frac{\partial}{\partial \gamma} \int \cdots \int_{R_0} \int \cdots \int_{R_1} L \, du_i \, dy_i \, d\beta - I_{\beta, \gamma} \\
= \frac{\partial}{\partial \beta} \frac{\partial}{\partial \gamma} 1 - I_{\beta, \gamma} = -I_{\beta, \gamma}
\end{align*}
\]

Therefore

\[
\frac{1}{m} \sum_{i=1}^{m} G_{ii}^* = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{ii}}{\partial \gamma^T} \xrightarrow{p} -I_{\beta, \gamma} \tag{4.35}
\]

From the assumption that the ML estimate of \(\alpha\) is convergent, we have from equation 4.43

\[
m^{\frac{1}{2}} (\alpha_{ML} - \alpha) \xrightarrow{p} N(0, I_{\alpha}^{-1})
\]

which means that

\[
m^{\frac{1}{2}} D_i^* = m^{\frac{1}{2}} [\hat{\alpha}(\beta, \hat{\gamma}) - \alpha] \xrightarrow{p} 0 \tag{4.36}
\]

Since the estimation of \(\gamma\) by QEEs is assumed to be convergent, equation 4.42 shows that

\[
m^{\frac{1}{2}} (\gamma_{QEE} - \gamma) \xrightarrow{p} N_q(0_q, I_{\gamma}^{-1})
\]

which means that

\[
m^{\frac{1}{2}} E_i^* = m^{\frac{1}{2}} [\hat{\gamma}(\beta, \hat{\alpha}) - \gamma] \xrightarrow{p} 0_q \tag{4.37}
\]

Therefore the vector part of equation 4.22 is asymptotically identical to \(\frac{1}{m} \sum_{i=1}^{m} A_i^*\)

\[
\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} U_{ii}(\beta, \hat{\alpha}, \hat{\gamma}) \xrightarrow{p} \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} A_i^* \\
\xrightarrow{p} N_p(0_p, I_{\beta}) \tag{4.38}
\]

since

\[
\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} B_{ii}^* D_i^* \xrightarrow{p} 0_p \\
\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} C_{ii}^* E_i^* \xrightarrow{p} 0_p
\]

Combining the results from equations 4.29 and 4.38 gives us (asymptotically)

\[
m^{\frac{1}{2}} (\hat{\gamma}_{QEE} - \beta) = -I_{\beta}^{-1} A_i^*
\]

80
\[ E_{\beta}[m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta)] = -I_{\beta}^{-1}E_{\beta}[A_1^*] = 0_p \]  \hfill (4.39)

\[ V_{\beta}[m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta)] = I_{\beta}^{-1}V_{\beta}(A_1^*)I_{\beta}^{-1} = I_{\beta}^{-1}I_{\beta}I_{\beta}^{-1} \]
\[ = I_{\beta}^{-1} \]  \hfill (4.40)

Since \( A_1^* \) is normally distributed, so will \( m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \), meaning that

\[ m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \xrightarrow{D} N_p [0_p, I_{\beta}^{-1}] \]  \hfill (4.41)
4.2.2 $\gamma$ estimates

Theorem

$$m^{\frac{1}{2}}[\hat{\gamma}_{QE} - \gamma] \xrightarrow{D} N_{q}(0, (I_{\gamma})^{-1})$$

asymptotically, where

$$I_{\gamma} = E_{\gamma} \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right)^T \right]$$

Assumptions

1. The likelihood and all derivatives exist and are finite.
   (a) $\int \cdots \int_{\mathcal{Y}} L_i \, d\gamma$
   (b) $\frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma$
   (c) $\frac{\partial}{\partial \gamma} \frac{\partial}{\partial \gamma^T} \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma$

2. $\beta$ and $\alpha$ have been estimated (by quasi-estimation equations and ML respectively) and converged to $\hat{\beta}$ and $\hat{\alpha}$, which means that the following exist and are finite
   (a) $\frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma$
   (b) $\frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma$

3. The following information matrices exist and are finite.
   (a) $I_{\gamma} = E_{\gamma} \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right)^T \right]$
   (b) $I_{\gamma,\beta} = E_{\gamma} \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right)^T \right]$
   (c) $I_{\gamma,\alpha} = E_{\gamma} \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{Y}} L_i \, d\gamma \right)^T \right]$

4. The derivatives of the estimates exist and are finite.
   (a) $\frac{\partial \hat{\beta}}{\partial \gamma} \not\to \infty$
   (b) $\frac{\partial \hat{\alpha}}{\partial \gamma} \not\to \infty$

Proof

The proof of this theorem is omitted, since it can be shown using the same steps as in the above theorem, by interchanging $\beta$ and $\gamma$.  

82
4.2.3 $\alpha$ estimate

Theorem

\[ m^{\frac{1}{2}}|\hat{\alpha}_{ML} - \alpha| \xrightarrow{P} \mathcal{N}(0, I^{-1}_\alpha) \]  \hspace{1cm} (4.43)

asymptotically, where

\[ I_\alpha = E_\alpha \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \right)^2 \right] \]

Assumptions

1. The likelihood and all derivatives exist and are finite.

   (a) \( \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \)

   (b) \( \frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \)

   (c) \( \frac{\partial^2}{\partial \alpha^2} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \)

2. As \( \delta \to 0 \) we have

\[ E_\alpha \left[ \sup_{h:||h|| \leq \delta} \left| \frac{\partial^2}{\partial \alpha^2} \left( \int \cdots \int_{\mathcal{R}_u} L_i(\alpha + h, \hat{\beta}, \hat{\gamma})d\mathbf{y}_i^* - \int \cdots \int_{\mathcal{R}_u} L_i(\alpha, \hat{\beta}, \hat{\gamma})d\mathbf{y}_i^* \right) \right| \right] = \psi_\delta \to 0 \]

\[ (4.44) \]

3. \( \beta \) and \( \gamma \) have been estimated by quasi-estimation equations and converged to \( \hat{\beta} \) and \( \hat{\gamma} \), which means that the following hold

   (a) \( \sum_{i=1}^{m} \frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* = 0_p \)

   (b) \( \sum_{i=1}^{m} \frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* = 0_q \)

   (c) \( \frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \) exists and is finite

   (d) \( \frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \) exists and is finite

4. The following information matrices exist and are finite.

   (a) \( I_\alpha = E_\alpha \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_u} L_i d\mathbf{y}_i^* \right)^2 \right] \)
(b) \( I_{\alpha, \beta} = E_{\alpha} \left[ \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \beta, \gamma) dy_i^* \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_{\beta}} L_i(\alpha, \beta, \gamma) dy_i^* \right] \)

(c) \( I_{\alpha, \gamma} = E_{\alpha} \left[ \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \beta, \gamma) \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_{\beta}} L_i(\alpha, \beta, \gamma) dy_i^* \right] \)

Proof

When maximising the log-likelihood we estimate \( \beta \) and \( \gamma \) by \( \hat{\beta} \) and \( \hat{\gamma} \) respectively and then estimate \( \alpha \). Ideally, we wish to carry out the estimation using the true values, but these are unknown. Therefore we are maximising the log-likelihood

\[
Lik(\alpha, \hat{\beta}, \hat{\gamma}) = \sum_{i=1}^{m} \log \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^*
\]

and not the true likelihood,

\[
Lik(\alpha, \beta, \gamma) = \sum_{i=1}^{m} \log \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \beta, \gamma) dy_i^*
\]

We can expand this likelihood with respect to the two estimated parameters to give

\[
Lik(\alpha, \hat{\beta}, \hat{\gamma}) = Lik(\alpha, \beta, \gamma) + \frac{1}{2} (\hat{\beta} - \beta) \frac{\partial}{\partial \beta} Lik(\alpha, \beta, \gamma) + \frac{1}{2} (\hat{\gamma} - \gamma) \frac{\partial}{\partial \gamma} Lik(\alpha, \beta, \gamma)
\]

\[
= Lik(\alpha, \beta, \gamma) + m \frac{1}{2} (\hat{\beta} - \beta) m^{-\frac{1}{2}} \frac{\partial}{\partial \beta} Lik(\alpha, \beta, \gamma) + m \frac{1}{2} (\hat{\gamma} - \gamma) m^{-\frac{1}{2}} \frac{\partial}{\partial \gamma} Lik(\alpha, \beta, \gamma)
\]

\[
\rightarrow Lik(\alpha, \beta, \gamma) + o_p(1) + o_p(1)
\]

\[
= Lik(\alpha, \beta, \gamma) + o_p(1)
\]

from equations 4.21 and 4.42 and the assumption that the QEEs estimates of \( \beta \) and \( \gamma \) have converged. Therefore, by maximising the likelihood \( Lik(\alpha, \hat{\beta}, \hat{\gamma}) \), we will obtain estimates that will be asymptotically identical to the ML estimate of \( \alpha \) by maximising the true likelihood, \( Lik(\alpha, \beta, \gamma) \).

Two standard results of generalised linear modelling are needed.

\[
E_{\alpha} \left[ \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* \right] = \int_{\mathcal{R}_{\alpha}} \frac{\partial}{\partial \alpha} \left[ \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* \right] \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* d\alpha
\]

\[
= \int_{\mathcal{R}_{\alpha}} \frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* d\alpha = \frac{\partial}{\partial \alpha} \int_{\mathcal{R}_{\alpha}} \int \cdots \int_{\mathcal{R}_{\alpha}} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* d\alpha
\]

\[
= \frac{\partial}{\partial \alpha} 1 = 0
\]

(4.45)
If we now define

\[ U_3 = \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \log \int_{R} L_i(\alpha, \beta, \gamma) dy_i^* \]

\[ V_3 = \sum_{i=1}^{m} \frac{\partial^2}{\partial \alpha^2} \log \int_{R} L_i(\alpha, \beta, \gamma) dy_i^* \]

then by the central limit theorem and Khintchine Weak Law of Numbers equations 4.45 and 4.46 give

\[ m^{-\frac{1}{2}} U_3 \xrightarrow{D} N(0, I_\alpha) \]

\[ m^{-1} V_3 \xrightarrow{p} -I_\alpha \]

Consider a number, u which is such that \(|u| \leq K\) for \(0 < K < \infty\). Taylor expansion about \(\alpha\) gives

\[ \sum_{i=1}^{m} \left[ \log \int_{R} L_i(\alpha + m^{-\frac{1}{2}} u, \beta, \gamma) dy_i^* - \log \int_{R} L_i(\alpha, \beta, \gamma) dy_i^* \right] \]

(4.47)
\[ \frac{u}{m^{\frac{3}{2}}} U_3 + \frac{u^2}{2m} V_3 + \frac{u^2}{2} Z(u) \]

where

\[ \alpha^* \in (\alpha, \alpha + m^{-\frac{1}{2}} u) \]

\[ Z(u) = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial^2}{\partial \alpha^2} \left( \log \int \cdots \int_{\mathbb{R}_+} L_i(\alpha^*, \hat{\beta}, \hat{\gamma}) dy_i^* - \log \int \cdots \int_{\mathbb{R}_+} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* \right) \quad (4.48) \]

Since \( u \) is an arbitrary number, we can choose it such that for every \( \delta > 0 \) there exists a number, \( m_0 \), such that \( m^{-\frac{1}{2}} |u| \leq m^{-\frac{1}{2}} K < \delta \) for all \( m > m_0 \). Therefore, for sufficiently large \( m \), we have

\[ |Z(u)| \leq \frac{1}{m} \sum_{i=1}^{m} \sup_{|h| \leq m^{-\frac{1}{2}} |u|} \left| \frac{\partial^2}{\partial \alpha^2} \left( \log \int \cdots \int_{\mathbb{R}_+} L_i(\alpha + h, \hat{\beta}, \hat{\gamma}) dy_i^* - \log \int \cdots \int_{\mathbb{R}_+} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* \right) \right| \]

\[ \overset{a.s.}{\Rightarrow} \psi_\delta \to 0 \quad (4.49) \]

by the Khintchine Weak Law of Numbers and the assumptions. Therefore, asymptotically, we have

\[ \lambda(u) = \frac{u}{m^{\frac{3}{2}}} U_3 - \frac{u^2}{2} I_\alpha + o_p(1) \]

We wish to maximise \( \lambda \) with respect to \( u \), so we find the first derivative and equate to zero to give

\[ \frac{d\lambda}{du} = \frac{U_3}{m^{\frac{3}{2}}} - u I_\alpha + o_p(1) = 0 \]

\[ \Rightarrow \hat{u} = \frac{U_3}{m^{\frac{3}{2}} I_\alpha} + o_p(1) \]

We know, from the definition of \( u \) and \( \lambda \), that \( \alpha + m^{-\frac{1}{2}} \hat{u} \) will be close to the value which maximises \( \int \cdots \int_{\mathbb{R}_+} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i^* \), which occurs at the maximum likelihood estimate of \( \alpha, \hat{\alpha} \).

\[ \hat{\alpha}_{ML} = \alpha + m^{-\frac{1}{2}} \hat{u} + o_p(m^{-\frac{1}{2}}) \]

\[ = \alpha + m^{-1} \left[ \frac{U_3}{I_\alpha} + o_p(m^{\frac{1}{2}}) \right] \]

\[ m^{\frac{3}{2}} (\hat{\alpha}_{ML} - \alpha) = \frac{U_3}{m^{\frac{3}{2}} I_\alpha} + o_p(1) \]

\[ \overset{d}{\Rightarrow} N(0, I_\alpha^{-1}) \quad (4.50) \]
4.3 Comparison with GEEs

In order to assess the effectiveness of QEEs as a method of parameter estimation, a series of simulations were carried out, to compare QEEs and the GEEs of Liang and Zeger (1986) [45]. In these simulations, a binary response was assumed to have been recorded, which arose as a discrete realisation from a correlated normal distribution. If the continuous value was negative, then the observation was category 1 (failure) and if it was positive, then it was category 2 (success). The aim of the simulations was to estimate the underlying normal distribution. It should be noted that a simple binary distribution was used, with a probit link, so as to ensure the methods were trying to estimate the same mean parameters. This simple example is one of the few cases for which GEEs are able to estimate underlying latent variables. However, the correlation estimate obtained from the GEEs will be the tetrachoric correlation, rather than the marginal correlation estimated by the QEEs.

The underlying distribution was multivariate normal with mean and correlation as in table 4.1. Once again 100 Monte-Carlo simulations were run on the two groups of 50 individuals, with the results in table 4.7, where the estimated parameters were $\mathbf{\beta} = (-2, 1, 2)^T$ and $\alpha_1 = 0.5$ and $\gamma = 0$ was treated as being fixed. These results show two things. Firstly, the averages of the $\mathbf{\beta}$ estimates were similar for both methods, as were the standard deviations attached to the estimates. The Hotelling T-test statistics also indicate that QEEs and GEEs were equally good estimation method. The correlation parameter estimated by GEEs is the marginal tetrachoric correlation, so direct comparison with the multivariate correlation estimate of QEEs is meaningless. Suffice it to say that the underlying correlation was well estimated by QEEs.
<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
<th>α Estimate</th>
<th>S.D.</th>
<th>β Estimate</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>QEEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.028</td>
<td>0.2682</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0233</td>
<td>0.2852</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0155</td>
<td>0.1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3899</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td></td>
<td></td>
<td>-1.967</td>
<td>0.2262</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9678</td>
<td>0.2339</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9916</td>
<td>0.1847</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0307</td>
<td>2.6984</td>
</tr>
<tr>
<td>1-dependent</td>
<td>QEEs</td>
<td>0.5053</td>
<td>0.3336</td>
<td>0.1589</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.0341</td>
<td>3.7953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2051</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1188</td>
<td>75.500</td>
<td>0.0505</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.056</td>
<td>3.9639</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2695</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>QEEs</td>
<td>0.5146</td>
<td>0.7294</td>
<td>0.2000</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.030</td>
<td>1.8120</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2550</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.0869</td>
<td>130.46</td>
<td>0.0317</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.005</td>
<td>3.3991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2482</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exponential</td>
<td>QEEs</td>
<td>0.5395</td>
<td>1.9869</td>
<td>0.1991</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.053</td>
<td>2.5634</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2773</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1266</td>
<td>77.690</td>
<td>0.0481</td>
<td>1.9842</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.012</td>
<td>1.5423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2537</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2371</td>
<td>1.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2871</td>
<td>2.6984</td>
</tr>
</tbody>
</table>

Table 4.7: QEEs and GEEs estimates using binary realisations from an underlying multivariate normal distribution
4.3.1 Correlation Miss-specification

Following are the results when the underlying correlation structure is miss-specified, similar to the work carried out before. Tables, 4.8, 4.9, 4.10 and 4.11, show several features of the estimation processes. Firstly the estimates under miss-specification were close to their true values, suggesting unbiasedness. Secondly, the standard errors associated with these estimates tended to be larger in the miss-specified models (the independent case excluded because of previously mentioned reasons concerning there being no true miss-specification). This difference was more noticeable in the QEEs estimates than the GEEs estimates. This is because the GEEs were estimating the tetrachoric correlation with values close to zero in most cases. Therefore, if the structure of the correlation was miss-specified, there would be little change in the model that is being used. However QEEs picked up changes in the marginal correlation although the t-values attached to the estimates were inflated when there was correlation miss-specification, so the estimates were not close to their true value of 0.5. This reiterates what we have previously shown both experimentally and theoretically. The main interest of these tables concerns the mean parameter estimates. We know that the T-values should be (and actually were) lowest for the true models in the case of QEEs and the GEEs. In addition to this, the T-values attached to the estimates were similar for both methods, which means that they were providing equally good estimates under correlation miss-specification. However many of the test values were significant which implies that the estimates had non-constant variance/covariance matrices since t-test applied to each parameter individually gave non-significant values.
<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>QEEs</td>
<td></td>
<td></td>
<td>-2.028</td>
<td>0.2682</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0233</td>
<td>0.2852</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0155</td>
<td>0.1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3899</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td></td>
<td></td>
<td>-1.967</td>
<td>0.2262</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9678</td>
<td>0.2339</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9916</td>
<td>0.1847</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0307</td>
<td>2.6984</td>
</tr>
<tr>
<td>1-dependent</td>
<td>QEEs</td>
<td>-0.024</td>
<td>0.1836</td>
<td>-2.027</td>
<td>0.2801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2897</td>
<td>1.9842</td>
<td>1.0071</td>
<td>0.2852</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0283</td>
<td>0.1683</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9691</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>-0.001</td>
<td>0.0781</td>
<td>-2.031</td>
<td>0.2088</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0048</td>
<td>0.1870</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0370</td>
<td>0.1868</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3525</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>QEEs</td>
<td>0.0253</td>
<td>0.2512</td>
<td>-1.976</td>
<td>0.1407</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0051</td>
<td>1.9842</td>
<td>0.9920</td>
<td>0.1356</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9921</td>
<td>0.1225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.2399</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>-0.006</td>
<td>0.0508</td>
<td>-2.022</td>
<td>0.2530</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0219</td>
<td>0.2395</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0018</td>
<td>0.1930</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4745</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exponential</td>
<td>QEEs</td>
<td>0.0025</td>
<td>0.1658</td>
<td>-2.058</td>
<td>0.3067</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1509</td>
<td>1.9842</td>
<td>1.0526</td>
<td>0.3188</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0490</td>
<td>0.2178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0685</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.0023</td>
<td>0.0766</td>
<td>-2.030</td>
<td>0.2489</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0066</td>
<td>0.1959</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0375</td>
<td>0.2184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0113</td>
<td>2.6984</td>
</tr>
</tbody>
</table>

Table 4.8: QEEs and GEEs estimates using binary realisations from an underlying independent multivariate normal distribution
<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>QEEs</td>
<td></td>
<td></td>
<td>-2.080</td>
<td>0.2721</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0604</td>
<td>0.3028</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0368</td>
<td>0.1852</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.8161</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td></td>
<td></td>
<td>-2.054</td>
<td>0.2615</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0121</td>
<td>0.2448</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0475</td>
<td>0.1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3167</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>QEEs</td>
<td>0.5053</td>
<td>0.1589</td>
<td>-1.995</td>
<td>0.2738</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3336</td>
<td>1.9842</td>
<td>3.7953</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0341</td>
<td>0.2051</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1188</td>
<td>0.0505</td>
<td>-2.056</td>
<td>0.2695</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0153</td>
<td>0.2314</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0553</td>
<td>0.1875</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.9639</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchangeable</td>
<td>QEEs</td>
<td>0.5445</td>
<td>0.2469</td>
<td>-2.129</td>
<td>0.2928</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.8027</td>
<td>1.9842</td>
<td>6.7960</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0718</td>
<td>0.2178</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.0702</td>
<td>0.0448</td>
<td>-2.084</td>
<td>0.2589</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0541</td>
<td>0.2323</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0752</td>
<td>0.1881</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.7315</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>QEEs</td>
<td>0.5407</td>
<td>0.2318</td>
<td>-2.111</td>
<td>0.2913</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7557</td>
<td>1.9842</td>
<td>7.2180</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0971</td>
<td>0.2111</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1209</td>
<td>0.0452</td>
<td>-2.041</td>
<td>0.2477</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0021</td>
<td>0.2445</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0260</td>
<td>0.1851</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5570</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: QEEs and GEEs estimates using binary realisations from a one-dependent underlying multivariate normal distribution.
<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>QEEs</td>
<td>-2.056</td>
<td>0.2852</td>
<td>1.0454</td>
<td>0.3290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0541</td>
<td>0.1769</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.6089</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>-2.037</td>
<td>0.2519</td>
<td>1.0152</td>
<td>0.2155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0389</td>
<td>0.1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2497</td>
<td>2.6984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-dependent</td>
<td>QEEs</td>
<td>0.4734</td>
<td>0.1497</td>
<td>-2.073</td>
<td>0.2918</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7746</td>
<td>1.9842</td>
<td>1.0244</td>
<td>0.3215</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0441</td>
<td>0.1915</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.5200</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1134</td>
<td>0.0440</td>
<td>-2.071</td>
<td>0.2544</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0190</td>
<td>0.2393</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0713</td>
<td>0.1786</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.3688</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>QEEs</td>
<td>0.5146</td>
<td>0.2000</td>
<td>-2.030</td>
<td>0.2550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7294</td>
<td>1.9842</td>
<td>1.0098</td>
<td>0.2811</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0440</td>
<td>0.1978</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.8120</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.0869</td>
<td>0.0317</td>
<td>-2.005</td>
<td>0.2482</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9543</td>
<td>0.2688</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0009</td>
<td>0.1796</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.3991</td>
<td>2.6984</td>
</tr>
<tr>
<td>Exponential</td>
<td>QEEs</td>
<td>0.5567</td>
<td>0.2236</td>
<td>-2.052</td>
<td>0.2760</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5377</td>
<td>1.9842</td>
<td>1.0825</td>
<td>0.2964</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0467</td>
<td>0.1934</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.9302</td>
<td>2.6984</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1180</td>
<td>0.0513</td>
<td>-2.024</td>
<td>0.2224</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9876</td>
<td>0.2345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0168</td>
<td>0.1821</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3743</td>
<td>2.6984</td>
</tr>
</tbody>
</table>

Table 4.10: QEEs and GEEs estimates using binary realisations from an exchangeable underlying multivariate normal distribution
<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>QEEs</td>
<td>-2.030</td>
<td>0.2813</td>
<td>2.0365</td>
<td>0.1817</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>-2.015</td>
<td>0.2345</td>
<td>0.9862</td>
<td>0.2783</td>
</tr>
<tr>
<td>1-dependent</td>
<td>QEEs</td>
<td>0.9862</td>
<td>0.2783</td>
<td>2.0295</td>
<td>0.1656</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.9677</td>
<td>0.2964</td>
<td>0.2813</td>
<td>0.1817</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>QEEs</td>
<td>0.5053</td>
<td>0.1589</td>
<td>0.1589</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1243</td>
<td>0.0403</td>
<td>0.0403</td>
<td>0.0403</td>
</tr>
<tr>
<td>Exponential</td>
<td>QEEs</td>
<td>0.5117</td>
<td>0.2027</td>
<td>0.2027</td>
<td>0.0202</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.0805</td>
<td>0.0352</td>
<td>0.0352</td>
<td>0.0352</td>
</tr>
<tr>
<td></td>
<td>QEEs</td>
<td>0.5395</td>
<td>0.1991</td>
<td>0.1991</td>
<td>0.0199</td>
</tr>
<tr>
<td></td>
<td>GEEs</td>
<td>0.1266</td>
<td>0.0481</td>
<td>0.0481</td>
<td>0.0481</td>
</tr>
</tbody>
</table>

Table 4.11: QEEs and GEEs estimates using binary realisations from an exponential underlying multivariate normal distribution
4.3.2 Stability

As before the stability of the estimates was considered so a comparison could be made between their possible asymptotic properties. The graphs in figure 4.4 show these stability plots for the exchangeable case. Both methods estimate the $\beta$ values well. They both converged to their true values as the theory suggests. Figure 4.5 shows the consistency of the estimates. The $m^{\frac{1}{2}}(\hat{\beta} - \beta)$ graphs oscillated around zero in both cases and did not diverge, which would suggest that they are converging at a rate of $m^{\frac{1}{2}}$ to zero. $m^{\frac{1}{2}}sd(\hat{\beta})$ showed marked differences between the methods. The QEEs estimates had lower values than the GEEs, meaning that the GEEs were giving inconsistent results. It is important to note that QEEs offer a superior method of estimation for estimating this type of binary latent trend, with fewer individuals being required to give 'good' parameter estimates. They converged after around 50-60 individuals, but GEEs did not converge until around 150 individuals. This practical aspect is important, since data collection on many individuals (so as to give good results) is costly both in terms of money and time, so a lower number of observations necessary for the stability/convergence is extremely desirable.
Figure 4.4: Stability plots of $\beta$ for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution
Figure 4.5(a) $m^t(\beta - \beta)$
Figure 4.5(b) $m^{1/2} \text{sd}(\hat{\beta})$

Figure 4.5: Consistency plots for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution
The final comparison between the methods was the stability of the estimates under correlation miss-specification. Figure 4.6 has the same properties as previous work. Both methods produced estimates which converged to their true values at a rate of $m^{\frac{1}{2}}$, with standard errors that also decreased at a rate of $m^{\frac{1}{2}}$. This implies that they are both $m^{\frac{1}{2}}$ consistent estimators. Also for QEEs, miss-specified models produced incorrect standard errors (shown by the 'bands' of values) so they would give an inefficient estimation method. The GEEs did not show a similar property, since the correlation that they estimated were all close to zero, so correlation miss-specification had little effect on the estimation procedure.
Figure 4.6(a) $m^1 (\hat{\beta} - \beta)$
Figure 4.6: Consistency plots for the QEEs and GEEs estimates using binary realisations from an underlying exchangeable multivariate normal distribution under correlation miss-specification
4.4 Runners example

Andrew Vickers from 'The Research Council for Complementary Medicine' provided a data set concerning a group of individuals that ran a race of varying lengths and durations. Every 12 hours after the race (starting at 9pm in the evening of the race) the runners recorded measurements of the pain that they experienced in their legs. This was repeated over a time period of 5 days, giving a total of 10 repeated measurements per runner. Two measurements were taken, one approximately continuous (visual analogue scale or VAS) and one ordinal (Likert scale). The former consisted of the runners marking on a scale of total length 100 the pain they felt (0 corresponds to no pain) and the latter consisted of ticking a box corresponding to their pain. Table 4.12 gives the possible Likert responses. It should be noted that this is naturally ordinal data. There are distinct categories and an implicit ordering (0 is less than 1, which is in turn less than 2 etc.).

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>A complete absence of soreness</td>
</tr>
<tr>
<td>1</td>
<td>A light pain felt only when touched / a vague ache</td>
</tr>
<tr>
<td>2</td>
<td>A moderate pain felt only when touched / a slight persistent pain</td>
</tr>
<tr>
<td>3</td>
<td>A light pain when walking up or down stairs</td>
</tr>
<tr>
<td>4</td>
<td>A light pain when walking on a flat surface / painful</td>
</tr>
<tr>
<td>5</td>
<td>A moderate pain, stiffness or weakness when walking / very painful</td>
</tr>
<tr>
<td>6</td>
<td>A severe pain that limits my ability to move</td>
</tr>
</tbody>
</table>

Table 4.12: Likert Scale of Muscle Soreness

In addition to the response, several covariates were measured, which were

**Group** is a binary covariate which shows if a runner was in the placebo or active group. Subjects took pills starting the evening before the race and continuing until 9 doses had been taken. 0 corresponds to the active group and 1 corresponds to the placebo group. The main aim of the study concerns working out the effect of this drug on the pain the runners feel in their legs.

**Age** of the runner

**Sex** of the runner. 0 indicates a male and 1 indicates a female.

**Training** miles per week was also recorded so as to give an indication of the runners' ability to cope with the pain of a race. Intuitively, the more training they do, the less likely
they are to feel pain in their legs after the race since their bodies are more prepared for the actual race.

**Race length** gives the length of the race in miles. As before, we would expect a positive correlation with pain.

**Race time** gives the time in minutes to complete the race. Again, a high race time would suggest a high pain measurement. There will also be a high correlation between the race time and race length, which may lead to one of these variables being superfluous.

**Injury** is another binary variable which indicates if the runner suffered an injury during the race. 0 means that they had no injury and 1 means that the runner experienced an injury during the race. No details were given as to if the injury occurred mid-race or at the finish. For the purposes of this work we will assume that the presence of an injury meant that the runner stopped the race and the race time/length recorded was the total time/length run until the injury occurred.

It should be noted that several of these parameters give a method for comparison between different factors. For instance, the coefficient on the ‘sex’ parameter has the interpretation of the difference between an average male and female response. ‘Group’ and ‘Injury’ have a similar interpretation. The aim of this study was to examine if the drug had any effect on the pain that runners suffered after their race, which is equivalent to saying that the coefficient of ‘group’ has a significant effect on the pain measurement. Methods previously discussed were applied to this data set. For the continuous VAS measurements, maximum likelihood (ML), restricted maximum likelihood (REML) and generalised estimation equations (GEEs) were applied. To analyse the ordinal Likert scores, quasi estimation equations (QEEs) were applied to estimate the underlying latent trend, pain, that a runner feels. It should be noted that these two types of distribution (continuous and underlying continuous) were not expected to be identical, one reason being that the variance of the measurements in the underlying case acts as a scaling parameter. However, if the variance of the underlying distribution was close to the variance of the VAS scores, we would expect some similarities. Although the study was aimed at finding if the drug had an effect on pain, we will be comparing and contrasting the methods, particularly making comparisons between the underlying continuous distribution estimate by QEEs and the actual continuous distribution estimated by the other methods.

### 4.4.1 Graphical analysis

Histograms of the continuous (VAS) and ordinal (Likert) responses were plotted so as to obtain an idea of the data set. These are in figure 4.7. As expected, these show that the pain
felt by the runners decreases over time. One immediate problem with this data set concerns the VAS scores. As the VAS score decreases, there is a natural lower limit of 0. After around time 5, the number of VAS scores close to zero becomes extremely large compared to the number of non-zero VAS scores. This 'skewing' of the data could cause problems for standard methodology, so we will only analyse the measurements taken between times 1 and 5 (inclusively). It should be noted that although the Likert scores are also skewed at later times, this would not cause a problem for QEEs, since the lowest category, 0, is assumed to have arisen as a continuous measurement between minus infinity and the first cut-off point. This is a major advantage of QEEs.

![Figure 4.7(a) Time=1](image)

![Figure 4.7(b) Time=2](image)
Figure 4.1(c) Time=3

Figure 4.1(d) Time=4

Figure 4.1(e) Time=5
Figure 4.7(f) Time=6

Figure 4.7(g) Time=7

Figure 4.7(h) Time=8
The second graphical analysis was to examine how the VAS scores behave over time in addition to their decreasing nature. A sample of runners were randomly picked and the time series plots of their measurements are given in figure 4.8. In addition to this, figure 4.9 shows the VAS scores of these individuals minus their initial VAS score, so we can see the trend once some of the individual random effects have been removed. Both plots show some features of the VAS data. Firstly, it is generally decreasing, which is intuitive since after the race has finished, we would expect runners to feel less and less pain in their legs. However, the 'average' rate of decrease does not seem to be constant, therefore time can be not included as a linear variable. Also, there does not seem to be any pattern to the decrease, so it seems that we should model a shift specific to each time to account for the longitudinal trend of the data.
Figure 4.8: Shadow plot of the first 5 VAS scores
Figure 4.9: Shadow plot of the difference between the second to fifth VAS scores and the first VAS score
One major problem specific to this data set is that there are individuals which have increasing VAS scores. That is to say, that the pain they experience increased after the race. This is a major problem in the analysis, so it was decided to exclude any individuals with VAS exhibiting this phenomenon since the aim of this study was to compare the QEEs with other repeated measurement techniques. In practice, if the statistician did not wish to remove these individuals, then the model would need to include extra factors to try and account for this.

4.4.2 ANOVA analysis

In an initial analysis of the data set (on the remaining 103 individuals), ANOVA was carried out in two scenarios. Firstly, by looking at the first VAS score alone to consider how the factors affect the first response independent of any follow up measurements. The second scenario was under the assumption that the repeated measurements were independent, which, given the nature of the data, is invalid. However, this will form a suitable first approximation to give an idea of the effect of the variables on the response.

For the first case, tables 4.13 and 4.14 give the full model and then a reduced model eliminating variables, one at a time, with small F-values until the remaining variables are significant. The tables show that three of the parameters are important but there is some doubt concerning ‘Race Time’, ‘Training’ and ‘Sex’. Although ‘Group’ was not significant, it was still included in the model since the effect of the drug was the focus of this study. Examination of the rejected parameters gives an interesting insight into the pain measurement in the runners legs. Race time not being significant is because of the high degree of correlation between the race time and the race length, since a long race would tend to take a long time. Also, if a runner has undertaken long training runs, then he/she is likely to have run longer races, giving rise (again) to a high degree of correlation with the race length and causing training to be unimportant. The most interesting is the sex parameter. This is not important which tells us that males and females have an equal level of pain at the end of their races. Finally, since group does not produce a significant F-value, the drug does not have an immediate effect on the pain the runners feel.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-Value</th>
<th>Prob(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>99.268</td>
<td>0.2215</td>
<td>0.6390</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>27.866</td>
<td>0.0622</td>
<td>0.8036</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>16.460</td>
<td>0.0367</td>
<td>0.8484</td>
</tr>
<tr>
<td>Training</td>
<td>1</td>
<td>104.37</td>
<td>0.2329</td>
<td>0.6305</td>
</tr>
<tr>
<td>Race Length</td>
<td>1</td>
<td>6598.7</td>
<td>14.725</td>
<td>0.0002</td>
</tr>
<tr>
<td>Race Time</td>
<td>1</td>
<td>267.18</td>
<td>0.5962</td>
<td>0.4419</td>
</tr>
<tr>
<td>Injury</td>
<td>1</td>
<td>1273.2</td>
<td>2.8412</td>
<td>0.0952</td>
</tr>
<tr>
<td>Residuals</td>
<td>95</td>
<td>42571</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.13: ANOVA table for the 1st VAS score with the full model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-Value</th>
<th>Prob(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>99.268</td>
<td>0.2261</td>
<td>0.6354</td>
</tr>
<tr>
<td>Race Length</td>
<td>1</td>
<td>5629.1</td>
<td>12.823</td>
<td>0.0005</td>
</tr>
<tr>
<td>Injury</td>
<td>1</td>
<td>1769.6</td>
<td>4.0310</td>
<td>0.0474</td>
</tr>
<tr>
<td>Residuals</td>
<td>99</td>
<td>43460</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.14: ANOVA table for the 1st VAS score with the reduced model

Assuming that the VAS scores are independent between and within individuals we can join all the data as one long vector of responses. Tables 4.15 and 4.16 give the full model and then a reduced model for this independent case, where time is accounted for using four different levels and treating the effect at time 1 being the baseline. These results agreed with the previous analysis, suggesting that the Group, Race Length and Injury parameters were important for modelling the VAS scores. Also modelling the time effect on the VAS scores using factors seems reasonable, as shown by the highly significant F-test value.
### Table 4.15: ANOVA table for the VAS score (assuming independence) with the full model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-Value</th>
<th>Prob(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>767.36</td>
<td>2.0568</td>
<td>0.1522</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>12.150</td>
<td>0.0326</td>
<td>0.8569</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>0.9212</td>
<td>0.0025</td>
<td>0.9603</td>
</tr>
<tr>
<td>Training</td>
<td>1</td>
<td>23.852</td>
<td>0.0639</td>
<td>0.8005</td>
</tr>
<tr>
<td>Race Length</td>
<td>1</td>
<td>22358</td>
<td>59.927</td>
<td>0.0000</td>
</tr>
<tr>
<td>Race Time</td>
<td>1</td>
<td>2078.5</td>
<td>5.5711</td>
<td>0.0186</td>
</tr>
<tr>
<td>Injury</td>
<td>1</td>
<td>8331.2</td>
<td>22.330</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time</td>
<td>4</td>
<td>22722</td>
<td>60.902</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residuals</td>
<td>503</td>
<td>187665</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.16: ANOVA table for the VAS scores (assuming independence) with the reduced model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-Value</th>
<th>Prob(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>767.4</td>
<td>2.0431</td>
<td>0.1535</td>
</tr>
<tr>
<td>Race Length</td>
<td>1</td>
<td>19966</td>
<td>53.159</td>
<td>0.0000</td>
</tr>
<tr>
<td>Injury</td>
<td>1</td>
<td>10075</td>
<td>26.825</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time</td>
<td>4</td>
<td>22721</td>
<td>60.495</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residuals</td>
<td>507</td>
<td>190428</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.4.3 Correlation analysis

An alternative approach to finding which parameters for the variables are important to the mean model, it is possible to select a correlation model, and then use an estimation method to find out which have significant t-test values. Using GEEs (with a totally unspecified correlation structure), the following correlation matrix is estimated for the 5 repeated measurements when all of the parameters were included linearly.

\[
\begin{pmatrix}
1.0000 & 0.8924 & 0.8078 & 0.6588 & 0.5066 \\
0.8924 & 1.0000 & 0.9252 & 0.7861 & 0.6577 \\
0.8078 & 0.9252 & 1.0000 & 0.8544 & 0.7569 \\
0.6588 & 0.7861 & 0.8544 & 1.0000 & 0.8121 \\
0.5066 & 0.6577 & 0.7569 & 0.8121 & 1.0000 \\
\end{pmatrix}
\] (4.51)

Figure 4.10 shows these points, along with the correlation values according to the different structures. This shows that the exponential structure was the best fit when all parameters...
were all included in the model.

![Graph showing correlation structures compared to sample correlations obtained from a fully unstructured GEEs estimation.](image)

**Figure 4.10:** Correlation structures compared to sample correlations obtained from a fully unstructured GEEs estimation

We now use GEEs to reduce the model. Table 4.17 shows this reduction, where the least significant (lowest value/standard error value) coefficient was removed between models, again leaving the 'Group' factor since this was the parameter of interest. The remaining factors were all significant and therefore important to the model. This agrees with the ANOVA analysis, in so much as it finds that the Intercept, Group, Race Length, Injury and Time variables were all needed in the model for a 'good fit'.
Therefore, in the following workings, we will assume that the model for the VAS scores and the trend underlying the Likert scores is

\[ \mu_i = X_i \beta \]

\[
\begin{pmatrix}
\mu_{i1} \\
\mu_{i2} \\
\mu_{i3} \\
\mu_{i4} \\
\mu_{i5}
\end{pmatrix} =
\begin{pmatrix}
1 & \text{Group}_i & \text{Length}_i & \text{Injury}_i & 0 & 0 & 0 \\
1 & \text{Group}_i & \text{Length}_i & \text{Injury}_i & 1 & 0 & 0 \\
1 & \text{Group}_i & \text{Length}_i & \text{Injury}_i & 0 & 1 & 0 \\
1 & \text{Group}_i & \text{Length}_i & \text{Injury}_i & 0 & 0 & 1 \\
1 & \text{Group}_i & \text{Length}_i & \text{Injury}_i & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\beta_7 \\
\beta_8
\end{pmatrix}
\] (4.52)

where \( \mu_i \) is the mean for the VAS scores measured on the \( i \)th individual, \( \text{Group}_i \), \( \text{Length}_i \) and \( \text{Injury}_i \) are the values of the group, race length and injury status measured on individual \( i \). The parameters \( \beta_5, \beta_6, \beta_7 \) and \( \beta_8 \) give a measure of the difference in the mean trend between the VAS scores at time 2, 3, 4 and 5 respectively compared to the VAS score at time 1.
4.4.4 ML/REML analysis

Two of the most common methods used to analyse repeated data are ML and REML, which were discussed in depth in chapter 2. The correlation matrix can be described in numerous ways, but for this analysis we will use structures which were described for GEEs in section 3.3. It should be noted that the 1-dependent structure is not possible using the parameterisation of the variance/covariance matrix describe in section 2.4, since ML and REML are constrained to having a positive definite matrix in order to have a properly defined multivariate normal distribution. These results are in tables 4.18 and 4.19 for ML and REML respectively.

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>Exchangeable</th>
<th>Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>29.764(4.6141)</td>
<td>29.764(8.7001)</td>
<td>31.371(5.8018)</td>
</tr>
<tr>
<td>Group</td>
<td>2.4945(1.7219)</td>
<td>2.4945(3.4753)</td>
<td>2.3870(3.0945)</td>
</tr>
<tr>
<td>Length</td>
<td>0.7017(0.1161)</td>
<td>0.7017(0.2343)</td>
<td>0.6571(0.2083)</td>
</tr>
<tr>
<td>Injury</td>
<td>11.957(2.2904)</td>
<td>11.957(4.6228)</td>
<td>12.787(4.1178)</td>
</tr>
<tr>
<td>Time2</td>
<td>-11.23(2.6793)</td>
<td>-11.23(1.2879)</td>
<td>-11.29(1.5916)</td>
</tr>
<tr>
<td>Time3</td>
<td>-21.78(2.6793)</td>
<td>-21.78(1.2879)</td>
<td>-21.84(2.3503)</td>
</tr>
<tr>
<td>Time4</td>
<td>-28.89(2.6793)</td>
<td>-28.89(1.2879)</td>
<td>-28.97(2.5078)</td>
</tr>
<tr>
<td>Time5</td>
<td>-37.98(2.6793)</td>
<td>-37.98(1.2879)</td>
<td>-38.05(1.5961)</td>
</tr>
<tr>
<td>Variance</td>
<td>369.69</td>
<td>370.00</td>
<td>352.64</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.7690</td>
<td>0.7875</td>
<td></td>
</tr>
<tr>
<td>Likelihood</td>
<td>-2253.3</td>
<td>-2023.9</td>
<td>-1881.2</td>
</tr>
</tbody>
</table>

*Table 4.18: Parameter estimates using ML*
Since the data are naturally repeated, we would expect the independence model to perform badly for ML and REML. This was reflected in the associated likelihoods which were less negative than those for the models which took into account the correlation between the repeated measures. Also, although the parameter estimates do not change, their standard errors were different between the correlation models. This suggests, as previously shown, that correlation miss-specification leads to estimates with incorrect standard deviations. The difference between ML and REML is also shown in these tables. The estimates remained almost unchanged, as did their standard errors. However, there were increases in the estimated variance under REML. This is because REML takes into account the uncertainty when using estimated parameter values, and not their true values. The likelihood for REML estimates were slightly larger, but not drastically so. The likelihoods from these estimates show that the ML or REML model with an exponential correlation structure provided the best fit to the data.

### 4.4.5 GEEs analysis

For the GEEs we assume a normal model for the marginal distributions was fitted, with different correlations as discussed in chapter 3. By fitting the parameters linearly, we obtained the results for the different correlation specifications shown in table 4.20. The likelihood stated is calculated post-estimation assuming the multivariate normal distribution defined by the estimates given.
Again, if we assumed that the data were independent observations on individuals (in which case we would expect correlation miss specification) then the estimates remain unbiased although their associated standard errors were different. Also, the likelihood was larger, indicating that the independence model was a poor fit for the data. The other models had similar values for the parameters, although the standard errors varied slightly. Overall, the exponential model seemed to be the best fit due to its having the smallest likelihood.

ML/REML and GEEs are methods which estimate the parameters in different ways. ML and REML (in this scenario) assumed that the data were multivariate normal and maximised the resulting likelihood. GEEs did not make this assumption, only that the marginal errors were normally distributed (again in this scenario). Therefore, they will only give the same estimates for the independence models, when marginal normal errors are identical to multivariate normal errors. However, all three methods suggest that the exponential structure is the best correlation structure for the data, with the smallest likelihoods.

### 4.4.6 QEEs analysis

Now we will consider the more complicated scenario of parameter estimation assuming that we do not know the continuous VAS, but only the ordinal Likert score. In this example, it would be natural to assume that there is an underlying latent variable which gives a measure of the underlying pain. In fact, the underlying latent variable should be 'equivalent' to the actual continuous variable measured, which we will examine later. The first problem we have
to consider is how to model the cut-off points. Since nothing changes between successive measurements (apart from the underlying pain) we have no reason to suspect that the cut-off point between the ordinal values would change over time. Therefore we will consider equal time cut-off points, as detailed in section 4.1.1. With this model, and by fitting the parameters linearly, we obtained the estimates in table 4.21 for the usual four correlation types. Once again there is a lack of identifiability due to the scaling parameter, the variance term. This was fixed at an arbitrary value of 500, although other values would have led to multiples (by the square root of the ratio of the variances) of the $\beta$ and $\gamma$ estimates being obtained. The correlation and likelihood are unaffected by this scale invariance problem.

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>1-Dependent</th>
<th>Exchangeable</th>
<th>Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>6.0047(2.0187)</td>
<td>5.4247(2.6092)</td>
<td>5.6186(3.3036)</td>
<td>4.9930(3.7598)</td>
</tr>
<tr>
<td>Length</td>
<td>1.2275(0.0855)</td>
<td>1.3126(0.0962)</td>
<td>1.2319(0.2400)</td>
<td>1.1663(0.1134)</td>
</tr>
<tr>
<td>Injury</td>
<td>15.543(2.7921)</td>
<td>17.913(3.6187)</td>
<td>15.348(4.9848)</td>
<td>16.686(5.2337)</td>
</tr>
<tr>
<td>Time2</td>
<td>-12.83(3.0139)</td>
<td>-13.18(3.1221)</td>
<td>-11.98(1.6057)</td>
<td>-12.10(1.5530)</td>
</tr>
<tr>
<td>Time3</td>
<td>-23.86(3.0094)</td>
<td>-25.78(3.1087)</td>
<td>-23.03(1.5847)</td>
<td>-23.36(1.9049)</td>
</tr>
<tr>
<td>Time4</td>
<td>-32.68(3.0127)</td>
<td>-35.06(3.1120)</td>
<td>-31.76(1.5862)</td>
<td>-31.76(2.1539)</td>
</tr>
<tr>
<td>Time5</td>
<td>-42.69(3.0185)</td>
<td>-46.22(3.1243)</td>
<td>-41.84(1.5915)</td>
<td>-42.18(2.3604)</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.5354</td>
<td>0.7192</td>
<td>0.8431</td>
<td></td>
</tr>
<tr>
<td>Cut-off 1</td>
<td>-40.51(2.6839)</td>
<td>-44.31(3.0530)</td>
<td>-38.68(2.9929)</td>
<td>-39.96(3.0605)</td>
</tr>
<tr>
<td>Cut-off 2</td>
<td>-12.79(1.5853)</td>
<td>-14.92(1.8903)</td>
<td>-11.79(2.2505)</td>
<td>-13.90(2.2766)</td>
</tr>
<tr>
<td>Cut-off 3</td>
<td>-4.048(1.4572)</td>
<td>-5.201(1.7639)</td>
<td>-3.183(2.1768)</td>
<td>-5.038(2.1894)</td>
</tr>
<tr>
<td>Cut-off 4</td>
<td>13.734(1.3671)</td>
<td>14.611(1.6819)</td>
<td>14.844(2.1267)</td>
<td>12.832(2.1361)</td>
</tr>
<tr>
<td>Cut-off 5</td>
<td>30.787(1.4861)</td>
<td>33.425(1.8037)</td>
<td>31.608(2.1822)</td>
<td>29.738(2.2052)</td>
</tr>
<tr>
<td>Cut-off 6</td>
<td>62.064(2.5014)</td>
<td>66.823(2.8397)</td>
<td>60.879(2.7746)</td>
<td>59.620(2.9293)</td>
</tr>
<tr>
<td>Likelihood</td>
<td>-782.97</td>
<td>-663.20</td>
<td>-635.01</td>
<td>-590.40</td>
</tr>
</tbody>
</table>

Table 4.21: Parameter estimates using QEEs

For QEEs we arrive at the same conclusion as for the continuous methods, that the exponential structure seemed to be the best model for the data, indicated by the smallest associated likelihood. Given the repeated nature of the data, we would expect the correlated models to be more 'reasonable' than the independence model. This has the implication that the independence model performed poorly, with the standard errors of the estimates being vastly different from the more reasonable correlated models. The 1-dependence model was also poor, but this may be because the correlation is constrained to giving a positive definite model which will be discussed later.
4.4.7 Model comparisons

There are two possible avenues for comparing the continuous and ordinal methods, either to convert the continuous data to categories and then compare results, or to simply look at the continuous trends that have been estimated (the actual trend or the underlying trend for VAS and Likert scores respectively). The latter is a more natural comparison method, although there are problems with the invariant properties of the QEEs. The results in table 4.21 are invariant with respect to the variance and the addition of an intercept. Therefore they were transformed so that the first and last cut-off points ($\gamma_1$ and $\gamma_6$) were equal to the observed cutoff points. In order to do this, the VAS and ordinal data were sorted into increasing order and then the cut-off point was taken as the average of the VAS scores associated with the Likert score changing from one category to another. Table 4.22 gives the cut-off VAS scores for individual times and one overall VAS score which represents the cut-off points being constant across time.

<table>
<thead>
<tr>
<th>Cut-off</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
<th>Time 5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>18</td>
<td>18.5</td>
<td>18</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>25.5</td>
<td>25</td>
<td>25.5</td>
<td>25</td>
<td>26.5</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>48.5</td>
<td>46</td>
<td>48.5</td>
<td>48.5</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>64.5</td>
<td>64</td>
<td>64</td>
<td>67</td>
<td>65.5</td>
<td>66</td>
</tr>
<tr>
<td>6</td>
<td>85</td>
<td>86</td>
<td>85</td>
<td>88.5</td>
<td>86.5</td>
<td>86.5</td>
</tr>
</tbody>
</table>

Table 4.22: VAS score cut-off points when compared to the Likert scores

This table shows that the cut-off points do not seem to change across time, which is what we assumed in the QEEs analysis. Therefore, for the purpose of further analysis, we will assume that the VAS score can represent the underlying pain that an individual feels and the vector of cut-off points, $\gamma = (4, 19, 26, 49, 66, 86.5)^T$, are the points which show the underlying pain giving rise to different ordinal categories. The results for the scaled QEEs estimates are in table 4.23. The likelihood reported is the likelihood obtained if the underlying trend estimated by the QEEs using equation 4.52 was the trend for the VAS scores.
Table 4.23: Parameter estimates using QEEs when transformed to fit the first and last VAS cut-off value

These results compare well with the estimates of ML, REML and GEEs in tables 4.18, 4.19 and 4.20 respectively. The estimates themselves are quite similar, as are the standard deviations. The only case where this is not true is the 1-dependent model. The reason for this is that the variance matrix is constrained to be positive definite, although the data suggest that this is not that case. Therefore we would expect poorer results.

As mentioned, the variance matrix is constrained to be positive definite which might effect the results. However, algebra can show that the only correlation structure which can be negative definite is the 1-dependence. The variances for the QEE estimates are similar to their counterparts as are the correlations (except for the 1-dependence model).

The likelihoods are more negative for the QEEs than with ML, REML and GEEs, particularly for the independent and 1-dependent models, although we would expect this. We are trying to find the underlying continuous distribution from discrete realisations. Therefore we loose ‘information’ which will give poorer estimates. However, QEEs still suggest that the exponential is the best correlation model, so using an underlying continuous trend for the
ordinal data gives the same conclusion as knowing the true continuous data.

4.4.8 Model conclusions

As mentioned, the QEEs and the continuous methods give similar results, and therefore have similar interpretations. In all cases, the ‘Group’ variable has a positive, but non-significant value, which is surprising. Since it was positive, this means that the placebo had higher values (corresponding to higher pain levels), but the drug which was administered to the runners before and after the race had no significant effect on the pain the runners felt in their legs. The other conclusions drawn from the parameter estimates agree with general reasoning. The positive significant values for ‘Length’ and ‘Injury’ mean that more pain is felt for a runner who ran a long race and suffered an injury mid-race. The time coefficients are all negative, so as time goes on they feel significantly less pain. Also, they are decreasing, so the pain is therefore progressively decreasing. The last conclusion concerns the correlation parameters. All the methods suggested an exponential structure which means that previous pain a runner felt will affect their present pain, although it will have progressively less impact as time goes on.
CHAPTER 5

Multivariate responses

Within the statistical analysis of repeated measures, very little work has been carried out on the case when there is more than one response variable. In the simple case where there is only one measurement per time, methods such as ML, REML, GEEs and QEEs offer a variety of modelling possibilities. When there is more than one response variable at any time there is difficulty. It is necessary to include the correlations between the different measurements taken at any time on any individual (the repetitive measurements) as well as the correlations between the different measurements at each time (repeated measurements).

In this chapter we formulate the modified model for the general case of many repetitive measurements per individual and simulation studies are used to examine the properties of the resulting estimates. The model we focus on will be the simplest case when there are two ordered categorical measurements taken at each time on every individual, the bivariate ordinal case, although this is easily extended. Also, the runners data is re-analysed and the estimation of the VAS scores and Likert scores is carried out jointly using QEEs.
5.1 Methodology

Much of the notation will be a generalisation of chapter 4. We consider \(d\) measurements taken at each time, \(y_{ij1}, \ldots, y_{ijd}\), although we focus on the latent variables, \(y_{ij1}', \ldots, y_{ijd}'\) which follow some multivariate continuous distribution with their repeated dependence described by the matrices \(R_{i1}, \ldots, R_{id}\) for repetitive measurements \(1, \ldots, d\) respectively. Associated with each latent trend is a set of covariates, \(X_{ij1}, \ldots, X_{ijd}\) which describe factors that effect the mean. The ordinal data arise from these distributions as discrete realisations using a set of cut-off points \(c_{ij1}, \ldots, c_{ijd}\).

So far this notation describes \(d\) separate cases modelled by the methodology of chapter 4. If the repetitive measurements were unrelated (uncorrelated), we could apply the method of QEEs directly to obtain estimates for the mean, cut-off and correlation parameters, \(\beta, \gamma\) and \(\alpha\) respectively. However this is not a logical assumption since measurements arose from the same individual, so there would be some correlation expected. Therefore a modification of the QEEs model will be made to allow for this correlation.

5.1.1 Correlation structure

We require two sets of correlations, one between the repetitive measurements and one between the repeated measures. Assume we know the \(d\) repeated correlations as in chapter 4, and that their correlation matrices are \(R_{i1}, \ldots, R_{id}\). We join up these \(d\) correlation matrices by putting the actual matrices on the diagonal and the element-by-element products (denoted by \(A \otimes B\) for matrices \(A\) and \(B\), which must have the same dimensions) of the matrices on the off-diagonals to give

\[
R_i^a = \begin{pmatrix}
R_{i1} & R_{i1} \otimes R_{i2} & \cdots & R_{i1} \otimes R_{id} \\
R_{i2} \otimes R_{i1} & R_{i2} & \cdots & R_{i2} \otimes R_{id} \\
\vdots & \vdots & \ddots & \vdots \\
R_{id} \otimes R_{i1} & R_{id} \otimes R_{i2} & \cdots & R_{id}
\end{pmatrix}
\]

The second correlation matrix we need, \(R_i^b\), describes the correlation between the repetitive measurements. If we assume that the correlation is given by a \(d \times d\) matrix, \(R_i^b\) then

\[
R_i^b = R_i^c \otimes 1_{dd}
\]

where \(1_{dd}\) is a \(d \times d\) matrix of ones and \(\otimes\) is the Kronecker product. By combining \(R_i^c\) and \(R_i^b\) with the \(\otimes\) product, we obtain a correlation matrix, \(R_i^* = R_i^c \otimes R_i^b\), giving the relationship between all measurements.
Independent repetitive measurements

Suppose there is no relationship between the repetitive measurements. This means that the repetitive correlation is $I_d$ and the collective correlation matrix for the data can be represented as

$$R^i = R^i_d \otimes R^i_t = R^i_t \otimes (I_d \otimes 1_{dd})$$

$$= \begin{pmatrix} R_{t1} & \cdots & 0_{tt} \\ \vdots & \ddots & \vdots \\ 0_{tt} & \cdots & R_{td} \end{pmatrix} \tag{5.1}$$

With this type of correlation structure we do not need to consider them as repetitive measurements since we could carry out an analysis with $md$ items where each of the $m$ individuals has been split up into its $d$ different repetitive measurements and use QEEs to analyse the resulting data which will be far more efficient.

It is interesting to note that if the repeated measurements are independent ($R_{t1} = \cdots = R_{td} = I_n$) and the repetitive measurements are not ($R^i_t \neq I_d$) then we can turn the problem on its head. That is to say that we can analyse the data as independent repetitive measurements, with repeated correlations of $R^i_t$.

5.1.2 Model formulation

The above work gives rise to different structures for the correlations amongst the repeated and the repetitive measurements. We combine these correlations along with the means and cut-off points to create an overall likelihood which takes into account the repeated and repetitive nature of the data. Firstly we combine each of the $d$ repeated ordinal measurements into a set of vectors and the corresponding data matrices into one long data matrix.

$$Y_{ik} = (y_{ilk}, \ldots, y_{ink})^T$$

$$X_{ik} = (X_{tik}^T, \ldots, X_{ink}^T)^T$$

From the data matrix, $X_{ik}$, we have the mean for each repetitive latent trend.

$$\mu_{ik} = X_{ik} \beta$$

where $\beta$ is a $p \times 1$ vector of mean parameters. The correlation structure for the entire data has already been defined as $R^i$. From the assumption, the data arises as discrete realisations of a latent multivariate distribution, $L_i = L_i(\mu_i, R_i, y^*_i)$ where $\mu_i = (\mu^T_{i1}, \ldots, \mu^T_{id})^T = \mu_i(\beta)$, $R_i = R_i(\alpha)$ and $y^*_i = (y^*_tik, \ldots, y^*_tink)^T$. However we only know the ordinal measurements,
so we obtain a total log-likelihood of
\[
Lik = \sum_{i=1}^{m} \log \int_{c_{1,y_{11}}}^{c_{1,y_{11}-1}} \cdots \int_{c_{n,y_{1n}}}^{c_{n,y_{1n}-1}} \int_{c_{1,y_{12}}}^{c_{1,y_{12}-1}} \cdots \int_{c_{n,y_{nd}}}^{c_{n,y_{nd}-1}} L_i(\mu, R, y^*) dy_{nd}^* \cdots dy_{12}^* dy_{1n}^* \cdots dy_{11}^*
\]
(5.2)

5.2 Simulation studies

As before we explore this method using known data so we can assess how well the methodology works. The data was assumed to have come from an underlying multivariate normal distribution with cut-off points \( \gamma = (0, 2, 4, 6)^T \). Two sets of three times repeated measures (giving 6 observations per individual) were generated assuming that they arose as two different measurements taken on two different groups, with 25 individuals in the first group and 25 in the second, at three different times. The mean model for this is given in table 5.1 where the parameters were assumed to be \( \beta = (1, 2, 1)^T \).

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>( \beta_2(t - 1) )</td>
<td>( \beta_1 + \beta_2(t - 1) )</td>
</tr>
<tr>
<td>2nd</td>
<td>( \beta_2(t - 1) + \beta_3 )</td>
<td>( \beta_1 + \beta_2(t - 1) + \beta_3 )</td>
</tr>
</tbody>
</table>

Table 5.1: Specification of the simulation study for repetitive measurements

Different correlation structures were considered both for the repeated and the repetitive measurements, which have both been discussed before. When applicable, the repeated correlation was \( \alpha^a = 0.5 \) and the repetitive correlation was \( \alpha^b = 0.5 \). One interpretation of this model could be two doctors examining a patient and giving their subjective assessment of the pain (for instance) the patients are feeling using a set of five categories. In this case, \( \beta_1 \) gives the difference between groups, \( \beta_2 \) represents the increase of latent pain at different times and \( \beta_3 \) is the difference of perception of pain that doctor 2 has compared to doctor 1. The two different correlations can also be interpreted in this scenario. The repeated correlations could be thought of as each doctor having a look at their previous notes on the patient, which would sway their opinion and the repetitive correlations could be thought of as the doctors conferring with each other to give an interaction between their measurements. If the doctors were adamant about their decisions, then there would be no repetitive correlation (\( \alpha^b = 0 \)) but if they were influenced then there would be a correlation (\( \alpha^b \neq 0 \)). Once again, t-tests or Hotelling T-tests are reported under each estimate, and the critical values for these tests are tabulated under the standard deviations. The results are in tables 5.2 and 5.3 for the independent and correlated repetitive correlation structures respectively. These tables show
that in all cases the correlation parameter (where applicable), the mean parameters and the cut-off parameters were well estimated. Their average values were close to the true values and the associated standard deviations were not large in comparison to the values, meaning that the estimates were closely grouped around their true values. In addition to this, the test statistics were not significant (apart from the exponential repeated measurements and independent repetitive measurements model), therefore we accept the null hypothesis that the parameter estimates were multivariate normally distributed with the correct mean vectors and constant variance/covariance matrix. This exception could be due to the randomness of the simulation study, although the actual estimates and standard errors are comparable to the other models.

<table>
<thead>
<tr>
<th>Repeated Correlation</th>
<th>( \alpha )</th>
<th>S.D.</th>
<th>( \beta )</th>
<th>S.D.</th>
<th>( \gamma )</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>-</td>
<td>-</td>
<td>1.0159</td>
<td>0.1573</td>
<td>0.0064</td>
<td>0.1474</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>2.0104</td>
<td>0.1324</td>
<td>2.0031</td>
<td>0.1426</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>1.0000</td>
<td>0.0789</td>
<td>4.0170</td>
<td>0.1960</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4180</td>
<td>2.8024</td>
<td>6.0317</td>
<td>0.2627</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2645</td>
<td>2.5740</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.5078</td>
<td>0.0381</td>
<td>0.9769</td>
<td>0.1738</td>
<td>-0.048</td>
<td>0.1849</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>2.0014</td>
<td>0.1011</td>
<td>1.9708</td>
<td>0.1804</td>
</tr>
<tr>
<td></td>
<td>1.4426</td>
<td>2.0096</td>
<td>0.9887</td>
<td>0.1513</td>
<td>3.9823</td>
<td>0.2028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4041</td>
<td>2.8024</td>
<td>5.9820</td>
<td>0.2431</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8962</td>
<td>2.5740</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.4984</td>
<td>0.0533</td>
<td>1.0153</td>
<td>0.1701</td>
<td>0.0166</td>
<td>0.1735</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>2.0051</td>
<td>0.1154</td>
<td>2.0129</td>
<td>0.1965</td>
</tr>
<tr>
<td></td>
<td>0.2094</td>
<td>2.0096</td>
<td>1.0033</td>
<td>0.1540</td>
<td>4.0144</td>
<td>0.2585</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1554</td>
<td>2.8024</td>
<td>6.0381</td>
<td>0.3367</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3906</td>
<td>2.5740</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.4933</td>
<td>0.0499</td>
<td>1.0442</td>
<td>0.1681</td>
<td>0.0399</td>
<td>0.1670</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>2.0306</td>
<td>0.0978</td>
<td>2.0787</td>
<td>0.1715</td>
</tr>
<tr>
<td></td>
<td>0.9457</td>
<td>2.0096</td>
<td>1.0106</td>
<td>0.1823</td>
<td>4.0791</td>
<td>0.2351</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.3790</td>
<td>2.8024</td>
<td>6.1060</td>
<td>0.3118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.6053</td>
<td>2.5740</td>
</tr>
</tbody>
</table>

Table 5.2: Simulation parameter estimates of the underlying distribution, using an independent (uncorrelated) repetitive structure
<table>
<thead>
<tr>
<th>Repeated Correlation</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
<th>$\gamma$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>0.4966</td>
<td>0.0586</td>
<td>0.0061</td>
<td>0.1698</td>
<td>0.029</td>
<td>0.1511</td>
</tr>
<tr>
<td></td>
<td>0.4146</td>
<td>2.0096</td>
<td>0.9732</td>
<td>0.0913</td>
<td>2.8024</td>
<td>0.8535</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.4991</td>
<td>0.0386</td>
<td>0.9814</td>
<td>0.2328</td>
<td>1.9898</td>
<td>0.1979</td>
</tr>
<tr>
<td></td>
<td>0.4834</td>
<td>0.0546</td>
<td>2.0015</td>
<td>0.1532</td>
<td>0.1698</td>
<td>0.2575</td>
</tr>
<tr>
<td></td>
<td>2.3095</td>
<td>3.1907</td>
<td>1.0371</td>
<td>0.1171</td>
<td>6.0310</td>
<td>0.3814</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.4987</td>
<td>0.0519</td>
<td>1.0313</td>
<td>0.2294</td>
<td>0.1171</td>
<td>0.2117</td>
</tr>
<tr>
<td></td>
<td>0.4988</td>
<td>0.0585</td>
<td>1.9944</td>
<td>0.1115</td>
<td>4.0257</td>
<td>0.2364</td>
</tr>
<tr>
<td></td>
<td>0.0327</td>
<td>3.1907</td>
<td>1.0001</td>
<td>0.1464</td>
<td>6.0403</td>
<td>0.2990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4316</td>
<td>2.8024</td>
<td>0.6403</td>
<td>0.3814</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.5022</td>
<td>0.0518</td>
<td>1.0307</td>
<td>0.1971</td>
<td>1.9983</td>
<td>0.2117</td>
</tr>
<tr>
<td></td>
<td>0.5009</td>
<td>0.0638</td>
<td>1.9976</td>
<td>0.1244</td>
<td>1.9883</td>
<td>0.2367</td>
</tr>
<tr>
<td></td>
<td>0.6076</td>
<td>3.1907</td>
<td>1.0023</td>
<td>0.1181</td>
<td>1.9883</td>
<td>0.3324</td>
</tr>
</tbody>
</table>

Table 5.3: Simulation parameter estimates of the underlying distribution, using a correlated repetitive structure
Stability

As in chapter 4, graphical studies were carried out on the QEE estimation process. The results for the repetitive correlation structure are in figure 5.1. These show that the estimates did not diverge at a rate of \( m^{\frac{1}{4}} \), and their standard deviations seemed to converge to fixed values. That is to say that

\[
\begin{align*}
 m^{\frac{1}{2}}(\hat{\alpha} - \alpha) &\to 0_r + O_P(1) \\
m^{\frac{1}{2}}(\hat{\beta} - \beta) &\to 0_p + O_P(1) \\
m^{\frac{1}{2}}(\hat{\gamma} - \gamma) &\to 0_q + O_P(1) \\
m^{\frac{3}{2}}\sqrt{V(\hat{\beta})} &\to \text{constant} \\
m^{\frac{3}{2}}\sqrt{V(\hat{\gamma})} &\to \text{constant}
\end{align*}
\]

This suggests that the estimates are \( m^{\frac{3}{2}} \) consistent.
Figure 5.1(a) $m^\frac{1}{2}(\hat{\alpha} - \alpha)$
Figure 5.1(b) $m^{\frac{1}{2}}(\hat{\beta} - \beta)$
Figure 5.1(c) $m \frac{1}{2} \text{sd}(\hat{\beta})$
Figure 5.1(d) $m^3 (\gamma \hat{\gamma})$
Figure 5.1: Consistency plots for the repetitive QEEs estimates of the correlation, mean and cut-off point parameters of the underlying multivariate normal distribution.
5.2.1 Correlation Miss-specification

Suppose that both types of correlation may be miss-specified. There are lots of possible combinations, but consider an underlying process which has correlated repetitive correlations and an exchangeable repeated correlation structure since these results are typical of the others. This specification of 'true model' will be used throughout this section. The Monte-Carlo estimates for the different possibilities for this case are in tables 5.4 and 5.5. As expected, the results were best when the underlying distribution was not miss-specified, with smaller standard deviations and test statistics. The \( \beta \) and \( \gamma \) estimates remained close to their true values for all correlation structures, suggesting unbiasedness under correlation misspecification. However, the \( \alpha \) estimates were more effected by the misspecification, with larger test statistics and estimates which were further from their true values.

<table>
<thead>
<tr>
<th>Repeated Correlation</th>
<th>( \alpha ) S.D.</th>
<th>( \beta ) S.D.</th>
<th>( \gamma ) S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>1.0208 0.2360</td>
<td>-0.020 0.2529</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>2.0330 0.1877</td>
<td>1.9859 0.2680</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>1.0247 0.1655</td>
<td>4.0592 0.2846</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>0.9188 2.8024</td>
<td>6.0387 0.3408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.7529 2.5740</td>
</tr>
<tr>
<td><strong>1-dependent</strong></td>
<td>0.3764 0.0526</td>
<td>1.0199 0.1853</td>
<td>-0.043 0.1739</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>2.0537 0.1226</td>
<td>2.0019 0.2272</td>
</tr>
<tr>
<td>16.622 2.0096</td>
<td>-</td>
<td>0.9995 0.1552</td>
<td>4.0876 0.2702</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>3.8937 2.8024</td>
<td>6.1408 0.3379</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.7921 2.5740</td>
</tr>
<tr>
<td><strong>Exchangeable</strong></td>
<td>0.5027 0.0508</td>
<td>0.9976 0.2187</td>
<td>-0.053 0.1705</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>1.9801 0.1063</td>
<td>1.9669 0.2052</td>
</tr>
<tr>
<td>0.3808 2.0096</td>
<td>-</td>
<td>0.9916 0.1192</td>
<td>3.9679 0.2368</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.6803 2.8024</td>
<td>5.9698 0.3275</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.1808 2.5740</td>
</tr>
<tr>
<td><strong>Exponential</strong></td>
<td>0.5414 0.0511</td>
<td>0.9967 0.2081</td>
<td>-0.024 0.1524</td>
</tr>
<tr>
<td>- -</td>
<td>-</td>
<td>2.0000 0.0999</td>
<td>1.9996 0.1682</td>
</tr>
<tr>
<td>5.7293 2.0096</td>
<td>-</td>
<td>0.9852 0.1355</td>
<td>4.0099 0.2206</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.2096 2.8024</td>
<td>6.0062 0.3054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6274 2.5740</td>
</tr>
</tbody>
</table>

Table 5.4: Simulation parameter estimates of the underlying distribution, using an independent (uncorrelated) repetitive structure and exchangeable repeated correlation
<table>
<thead>
<tr>
<th>Repeated Correlation</th>
<th>$\alpha$</th>
<th>S.D.</th>
<th>$\beta$</th>
<th>S.D.</th>
<th>$\gamma$</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>0.4832</td>
<td>0.0638</td>
<td>1.0044</td>
<td>0.2154</td>
<td>-0.012</td>
<td>0.1649</td>
</tr>
<tr>
<td></td>
<td>1.8640</td>
<td>2.0096</td>
<td>2.0130</td>
<td>0.1078</td>
<td>4.0206</td>
<td>0.2129</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0073</td>
<td>0.1360</td>
<td>6.0556</td>
<td>0.3091</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2390</td>
<td>2.8024</td>
<td>1.8281</td>
<td>2.5740</td>
</tr>
<tr>
<td>1-dependent</td>
<td>0.3778</td>
<td>0.0431</td>
<td>1.0430</td>
<td>0.2100</td>
<td>0.0565</td>
<td>0.2138</td>
</tr>
<tr>
<td></td>
<td>0.5045</td>
<td>0.0567</td>
<td>2.0210</td>
<td>0.1137</td>
<td>4.1369</td>
<td>0.2744</td>
</tr>
<tr>
<td></td>
<td>197.16</td>
<td>3.1907</td>
<td>1.0298</td>
<td>0.1339</td>
<td>6.2019</td>
<td>0.3357</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.0933</td>
<td>2.8024</td>
<td>4.7718</td>
<td>2.5740</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>0.4987</td>
<td>0.0519</td>
<td>1.0294</td>
<td>0.1628</td>
<td>0.0169</td>
<td>0.1507</td>
</tr>
<tr>
<td></td>
<td>0.4988</td>
<td>0.0585</td>
<td>2.0038</td>
<td>0.0835</td>
<td>2.0052</td>
<td>0.1651</td>
</tr>
<tr>
<td></td>
<td>0.0327</td>
<td>3.1907</td>
<td>1.0113</td>
<td>0.0983</td>
<td>4.0315</td>
<td>0.1798</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6568</td>
<td>2.8024</td>
<td>6.0367</td>
<td>0.2284</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0423</td>
<td>2.5740</td>
<td>1.0423</td>
<td>2.5740</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.5333</td>
<td>0.0592</td>
<td>0.9427</td>
<td>0.2655</td>
<td>-0.076</td>
<td>0.1785</td>
</tr>
<tr>
<td></td>
<td>0.4897</td>
<td>0.0546</td>
<td>1.9868</td>
<td>0.1053</td>
<td>1.9127</td>
<td>0.2259</td>
</tr>
<tr>
<td></td>
<td>10.839</td>
<td>3.1907</td>
<td>0.9692</td>
<td>0.1142</td>
<td>3.8848</td>
<td>0.2588</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.7300</td>
<td>2.8024</td>
<td>5.9100</td>
<td>0.3184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.5068</td>
<td>2.5740</td>
<td>3.5068</td>
<td>2.5740</td>
</tr>
</tbody>
</table>

Table 5.5: Simulation parameter estimates of the underlying distribution, using a correlated repetitive structure and exchangeable repeated correlation

**Stability**

Figure 5.2 shows consistency and stability plots when there was (possibly) double correlation miss-specification. As before, the estimates did not diverge and their standard errors seemed to converge at a rate of $m^\frac{1}{2}$. However, distinct bands were formed suggesting that under correlation miss-specification QEEs give unbiased estimates with incorrect standard error estimates.
Figure 5.2(a) $m^2(\bar{\alpha} - \alpha)$
Figure 5.2(b) $m^\frac{1}{2} (\hat{\beta} - \beta)$
Figure 5.2(c) $m^{\frac{1}{2}}sd(\beta)$
Figure 5.2(d) $m^\frac{1}{2}(\gamma - \gamma)$
Figure 5.2(e) \( m^3 \sigma d(\hat{\gamma}) \)

Figure 5.2: Consistency plots for the repetitive QEEs estimates of the correlation, mean and cut-off point parameters of the underlying multivariate normal distribution when there is repetitive and repeated correlation miss-specification.
5.3 Asymptotics

We have looked at graphical evidence for the asymptotics properties of the QEEs estimates when there are multivariate repeated ordinal responses. Following are proofs which show, given some conditions,

\[ m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha) \overset{D}{\to} N_r(0_r, I_\alpha^{-1}) \]  
\[ m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \overset{D}{\to} N_p(0_p, I_\beta^{-1}) \]  
\[ m^{\frac{1}{2}}(\hat{\gamma}_{QEE} - \gamma) \overset{D}{\to} N_q(0_q, I_\gamma^{-1}) \]

(5.3) (5.4) (5.5)

where the variance matrices are the corresponding information matrices for that parameter. It should be noted that the conditions of the proofs require that there is no scale invariability, so \( \alpha \) is the repetitive and repeated correlations and not the variance term(s).
5.3.1 \( \beta \) estimates

Theorem

\[
m^2 |\hat{\beta}_{QEE} - \beta| \overset{P}{\to} N_p[0_p, (I_\beta)^{-1}]
\]  

(5.6) asymptotically, where

\[
I_\beta = E_\beta \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right)^T \right]
\]

Assumptions

1. The likelihood and all derivatives exist and are finite.

   (a) \( \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \)

   (b) \( \frac{\partial}{\partial \beta} \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \)

   (c) \( \frac{\partial}{\partial \beta \partial \beta^T} \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \)

2. \( \alpha \) and \( \gamma \) have been estimated (by ML and quasi-estimation equations respectively) and converged to \( \hat{\alpha} \) and \( \hat{\gamma} \), which means that the following exist and are finite

   (a) \( \frac{\partial}{\partial \alpha} \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \)

   (b) \( \frac{\partial}{\partial \gamma} \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \)

3. The following information matrices exist and are finite.

   (a) \( I_\beta = E_\beta \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right)^T \right] \)

   (b) \( I_{\beta, \alpha} = E_\beta \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right)^T \right] \)

   (c) \( I_{\beta, \gamma} = E_\beta \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right) \left( \frac{\partial}{\partial \beta} \log \int \cdots \int \mathbb{R}_v L_i d\mathbf{y}_i^* \right)^T \right] \)

4. The derivatives of the estimates exist and are finite.

   (a) \( \frac{\partial \hat{\alpha}^T}{\partial \beta} \not\to \infty \)

   (b) \( \frac{\partial \hat{\gamma}^T}{\partial \beta} \not\to \infty \)
Proof

Using the iterative scheme detailed in section 4.1.2 (equation 4.8), we can approximate $\hat{\beta}_{QEE} - \beta$ by

$$m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \approx m^{\frac{1}{2}} \left\{ \sum_{i=1}^{m} \frac{\delta U_{ii}^T}{\delta \beta} \right\}^{-1} \left\{ \sum_{i=1}^{m} U_{ii} \right\}$$

$$= \left\{ \sum_{i=1}^{m} \frac{1}{m} \frac{\delta U_{ii}^T}{\delta \beta} \right\}^{-1} \left\{ \sum_{i=1}^{m} \frac{1}{m^2} U_{ii} \right\} \quad (5.7)$$

where

$$U_{ii} = \frac{\partial}{\partial \beta} \int \cdots \int_{\mathbb{R}^m} \mathcal{L}_i d\mathbf{y}_i^* = U_{1i}(\beta, \alpha, \gamma)$$

$$\hat{\alpha} = \hat{\alpha}(\hat{\beta}, \hat{\gamma})$$

is the ML estimate of $\alpha$ when $\beta, \gamma$ are constant

$$\hat{\gamma} = \hat{\gamma}(\hat{\beta}, \hat{\alpha})$$

is the ML estimate of $\gamma$ when $\beta, \alpha$ are constant

Using calculus to expand the matrix in equation 5.7 gives

$$\frac{\delta U_{ii}^T}{\delta \beta} = \frac{\partial U_{ii}^T}{\partial \beta} + \frac{\partial U_{ii}^T}{\partial \alpha} \frac{\partial \alpha}{\partial \beta} + \frac{\partial U_{ii}^T}{\partial \gamma} \frac{\partial \gamma}{\partial \beta} \quad (5.8)$$

As before,

$$\frac{1}{m} \sum_{i=1}^{m} A_{ii} \rightarrow -I_p \quad (5.9)$$

$$\frac{1}{m} \sum_{i=1}^{m} C_{1i} \rightarrow 0_{q,p} \quad (5.10)$$

$$\frac{1}{m} \sum_{i=1}^{m} B_{1i} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{ii}^T}{\partial \alpha} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \left( \frac{\partial}{\partial \beta} \int \cdots \int_{\mathbb{R}^m} \mathcal{L}_i d\mathbf{y}_i^* \right)^T$$

$$= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathbb{R}^m} \mathcal{L}_i d\mathbf{y}_i^* \right)^T$$

$$= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \beta^T} \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathbb{R}^m} \mathcal{L}_i d\mathbf{y}_i^*$$

$$= \frac{1}{m} \frac{\partial}{\partial \beta^T} \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \int \cdots \int_{\mathbb{R}^m} \mathcal{L}_i d\mathbf{y}_i^* = 0_{r,p} \quad (5.11)$$
since $\hat{\alpha}$ is the ML estimate, which means that the derivative of the log-likelihood is zero, i.e.

$$
\frac{\partial}{\partial \hat{\alpha}} \text{Lik} = \frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{m} \log \int \cdots \int \mathbb{R}_{\nu_i} L_i dy_i^* = \sum_{i=1}^{m} \frac{\partial}{\partial \hat{\alpha}} \int \cdots \int \mathbb{R}_{\nu_i} L_i dy_i^*
$$

$$
= 0_r
$$

From the assumptions we know that

$$
D_1 = \frac{\partial \hat{\alpha}^T}{\partial \beta} \neq \infty
$$

$$
E_1 = \frac{\partial \hat{\gamma}^T}{\partial \beta} \neq \infty
$$

By substituting equations 5.9, 5.11, 5.10, 5.12 and 5.13 into equation 5.8 we have

$$
\frac{1}{m} \frac{\delta U_i}{\delta \beta} \overset{p}{\to} -I_{\beta}
$$

which is the first component of the right hand side of equation 5.7. Now consider the vector part of this equation (the second component of the right hand side of equation 5.7). Using multidimensional Taylor’s expansion

$$
U_{1i}(\beta, \hat{\alpha}, \hat{\gamma}) = U_{i1} + \frac{\partial U_{i1}}{\partial \hat{\alpha}} (\hat{\alpha} - \alpha) + \frac{\partial U_{i1}}{\partial \hat{\gamma}} (\hat{\gamma} - \gamma) + o_P(1)
$$

$$
A_{i1} + B_{i1} D_{i1} + C_{i1} E_{i1} + o_P(1)
$$

where $U_{i1} = U_{1i}(\beta, \alpha, \gamma)$.

As in the proof in chapter 4, we have

$$
\frac{1}{m} \sum_{i=1}^{m} A_{ii}^* \overset{p}{\to} N_p[0, I_{\beta}]
$$

$$
\frac{1}{m} \sum_{i=1}^{m} C_{ii}^* E_{i1}^* = \frac{1}{m} \sum_{i=1}^{m} C_{ii}^* m^\frac{1}{2} E_{i1}^* \overset{p}{\to} 0_p
$$
Therefore

\[
\frac{1}{m} \sum_{i=1}^{m} B_{li}^* = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial U_{li}}{\partial \alpha^T} \beta - I_{\beta,\alpha} \quad (5.18)
\]

From the assumption that the ML estimate of \( \alpha \) is convergent, we have from equation 5.24

\[ m^{\frac{1}{2}}(\alpha_{ML} - \alpha) \xrightarrow{p} N_r(0_r, I_{\alpha}^{-1}) \]

which means that

\[ m^{\frac{1}{2}}D_1^* = m^{\frac{1}{2}}[\hat{\alpha}(\beta, \gamma) - \alpha] \xrightarrow{p} 0_r \quad (5.19) \]

and

\[ \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} B_{li}^* \cdot D_1^* = \frac{1}{m} \sum_{i=1}^{m} B_{li}^* \cdot m^{\frac{1}{2}}D_1^* \xrightarrow{p} 0_p \quad (5.20) \]

Therefore the vector part of equation 5.7 is asymptotically identical to

\[ \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} A_1^* \]

\[ \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} U_{li}(\beta, \hat{\alpha}, \gamma) \xrightarrow{p} \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} A_1^* \xrightarrow{p} N_p [0_p, I_\beta] \quad (5.21) \]

Combining the results from equations 5.14 and 5.21 gives us (asymptotically)

\[ m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) = -I_\beta^{-1}A_1^* \]

and since \( A_1^* \) is normally distributed, so will \( m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \). So

\[ m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \xrightarrow{p} N_p [0_p, I_\beta^{-1}] \quad (5.22) \]
5.3.2 \( \gamma \) estimates

Theorem

\[
m^{1/2} |\hat{\gamma}_{QEE} - \gamma| \overset{P}{\to} N_0(0, (I_\gamma)^{-1})
\]

asymptotically, where

\[
I_\gamma = E_\gamma \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right)^T \right]
\]

Assumptions

1. The likelihood and all derivatives exist and are finite.
   (a) \( \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \)
   (b) \( \frac{\partial}{\partial \gamma} \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \)
   (c) \( \frac{\partial}{\partial \gamma} \frac{1}{\partial \gamma^T} \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \)

2. \( \beta \) and \( \alpha \) have been estimated (by quasi-estimation equations and ML respectively) and
   converged to \( \hat{\beta} \) and \( \hat{\alpha} \), which means that the following exist and are finite
   (a) \( \frac{\partial}{\partial \beta} \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \)
   (b) \( \frac{\partial}{\partial \alpha} \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \)

3. The following information matrices exist and are finite.
   (a) \( I_\gamma = E_\gamma \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right)^T \right] \)
   (b) \( I_{\gamma, \beta} = E_\gamma \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right)^T \right] \)
   (c) \( I_{\gamma, \alpha} = E_\gamma \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right) \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{\mathcal{R}_m} L_i dy_i^* \right)^T \right] \)

4. The derivatives of the estimates exist and are finite.
   (a) \( \frac{\partial \hat{\beta}^T}{\partial \gamma} \not\to \infty \)
   (b) \( \frac{\partial \hat{\alpha}^T}{\partial \gamma} \not\to \infty \)
Proof

The proof of this theorem is omitted, since it can be shown using the same steps as in the above theorem, by interchanging $\beta$ and $\gamma$. 
5.3.3 $\alpha$ estimates

Theorem

$$m^{\frac{1}{2}} |\hat{\alpha}_{ML} - \alpha| \overset{D}{\rightarrow} N_{n_0}[0_{n_0}, (I_\alpha)^{-1}] \tag{5.24}$$

asymptotically, where

$$I_\alpha = E_\alpha \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int r_{y_i} L_i dy_i^* \right) \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int r_{y_i} L_i dy_i^* \right)^T \right]$$

Assumptions

1. The likelihood and all derivatives exist and are finite.
   (a) $\int \cdots \int r_{y_i} L_i dy_i^*$
   (b) $\frac{\partial}{\partial \alpha} \int \cdots \int r_{y_i} L_i dy_i^*$
   (c) $\frac{\partial^2}{\partial \alpha \partial \alpha^T} \int \cdots \int r_{y_i} L_i dy_i^*$

2. As $\delta \to 0$ we have
   $$E_\alpha \left[ \sup_{h:|h| \leq \delta} \left| \frac{\partial^2}{\partial \alpha \partial \alpha^T} \int \cdots \int r_{y_i} L_i (\alpha + h, \hat{\beta}, \hat{\gamma}) dy_i^* \right| \right] = 0_{\delta} \tag{5.25}$$

3. $\beta$ and $\gamma$ have been estimated by quasi-estimation equations so the following hold.
   (a) $\sum_{i=1}^{m} \frac{\partial}{\partial \beta} \int \cdots \int r_{y_i} L_i dy_i^* = 0_p$
   (b) $\sum_{i=1}^{m} \frac{\partial}{\partial \gamma} \int \cdots \int r_{y_i} L_i dy_i^* = 0_q$
   (c) $\frac{\partial}{\partial \beta} \int \cdots \int r_{y_i} L_i dy_i^*$ exists and is finite
   (d) $\frac{\partial}{\partial \gamma} \int \cdots \int r_{y_i} L_i dy_i^*$ exists and is finite

4. The following information matrices exist and are finite.
   (a) $I_\alpha = E_\alpha \left[ \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int r_{y_i} L_i dy_i^* \right) \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int r_{y_i} L_i dy_i^* \right)^T \right]$
(b) \( I_{\alpha, \beta} = E_\alpha \left[ \left( \frac{\partial}{\partial \beta} \log \int \cdots \int_{R_v} L_i dy_i \right) \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{R_v} L_i dy_i \right)^T \right] \)

(c) \( I_{\alpha, \gamma} = E_\alpha \left[ \left( \frac{\partial}{\partial \gamma} \log \int \cdots \int_{R_v} L_i dy_i \right) \left( \frac{\partial}{\partial \alpha} \log \int \cdots \int_{R_v} L_i dy_i \right)^T \right] \)

Proof

As before, maximising the log-likelihood \( \text{Lik}(\alpha, \beta, \gamma) = \sum_{i=1}^{m} \log \int \cdots \int_{R_v} L_i(\alpha, \beta, \gamma) dy_i \) gives (asymptotically) the same \( \alpha \) estimate as when the true likelihood, \( \text{Lik}(\alpha, \beta, \gamma) \) is maximised.

Consider a vector, \( \mathbf{u} \) which is such that \( |\mathbf{u}| \leq K \) for \( 0 < K < \infty \). By defining \( \lambda \) and then expanding about \( \alpha \) gives

\[
\lambda = \sum_{i=1}^{m} \left[ \log \int \cdots \int_{R_v} L_i(\alpha + m^{-\frac{1}{2}} \mathbf{u}, \hat{\beta}, \hat{\gamma}) dy_i - \log \int \cdots \int_{R_v} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i \right] \tag{5.26}
\]

\[
= \frac{1}{m^2} \left[ \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} \log \int \cdots \int_{R_v} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i \right]^T \mathbf{u} + \frac{1}{2m} \mathbf{u}^T \left[ \sum_{i=1}^{m} \frac{\partial^2}{\partial \alpha^2} \log \int \cdots \int_{R_v} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i \right] \mathbf{u} \]

\[
= \frac{1}{m^2} \left[ \sum_{i=1}^{m} U_{3i}(\alpha, \hat{\beta}, \hat{\gamma}) \right]^T \mathbf{u} + \frac{1}{2m} \mathbf{u}^T \left[ \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha^*, \hat{\beta}, \hat{\gamma}) \right] \mathbf{u} \tag{5.27}
\]

where

\[
U_{3i}(\alpha, \hat{\beta}, \hat{\gamma}) = \frac{\partial}{\partial \alpha} \log \int \cdots \int_{R_v} L_i(\alpha, \hat{\beta}, \hat{\gamma}) dy_i
\]

\( \alpha^* \in (\alpha, \alpha + m^{-\frac{1}{2}} \mathbf{u}) \)

By defining \( Z(\mathbf{u}) \) as

\[
\frac{1}{m} \sum_{i=1}^{m} \left[ \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha^*, \hat{\beta}, \hat{\gamma}) - \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right]
\]

then

\[
\lambda = \frac{1}{m^2} \left[ \sum_{i=1}^{m} U_{3i}(\alpha, \hat{\beta}, \hat{\gamma}) \right]^T \mathbf{u} + \frac{1}{2m} \mathbf{u}^T Z(\mathbf{u}) \mathbf{u} \tag{5.28}
\]

Since \( \mathbf{u} \) is arbitrary vector, we can choose it such that for every \( \delta > 0 \) there exists a number, \( m_0 \), such that \( m^{-\frac{1}{2}} |\mathbf{u}| \leq m^{-\frac{1}{2}} K < \delta \) for all \( m > m_0 \). Therefore, for sufficiently large \( m \), we
have

\[ |Z(u)| \leq \frac{1}{m} \sum_{i=1}^{m} h_i |u_i| \leq m^{-\frac{1}{2}} |u| \left| \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha + h, \hat{\beta}, \hat{\gamma}) - \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right| \]

\[ \leq \frac{1}{m} \sum_{i=1}^{m} h_i |u_i| \leq m^{-\frac{1}{2}} \left| \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha + h, \hat{\beta}, \hat{\gamma}) - \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right| \]

\[ \sup_{u: |u| \leq K} \left[ \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha + h, \hat{\beta}, \hat{\gamma}) - \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right] \]

\[ \Rightarrow E_\alpha \left[ \sup_{u: |u| \leq K} \left[ \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha + h, \hat{\beta}, \hat{\gamma}) - \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right] \right] = \psi_\delta \rightarrow 0 \]

(5.29)

by the Khintchine Weak Law of Numbers and the assumptions. Therefore

\[ \sup_{u: |u| \leq K} |Z(u)| \lesssim 0 \]

(5.30)

Putting this into equation 5.28 yields

\[ \lambda(u) = \frac{1}{m^2} \left[ \sum_{i=1}^{m} U_{3i}(\alpha, \hat{\beta}, \hat{\gamma}) \right]^T u + \frac{1}{2m} u^T \left[ \sum_{i=1}^{m} \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right] u + o_P(1) \]

But

\[ E_\alpha \left[ \frac{\partial}{\partial \alpha} U_{3i}^T(\alpha, \hat{\beta}, \hat{\gamma}) \right] \]

\[ = E_\alpha \left[ \frac{\partial}{\partial \alpha} \left( \int_{\mathbb{R}_v} L_i dy_i^* \right)^T \right] \]

\[ = E_\alpha \left[ \frac{\partial^2}{\partial \alpha \partial \alpha^T} \int_{\mathbb{R}_v} L_i dy_i^* \right] - \left( \int_{\mathbb{R}_v} L_i dy_i^* \right) \left( \int_{\mathbb{R}_v} L_i dy_i^* \right)^T \]

\[ = \int \int \int h_\alpha \left[ \frac{\partial^2}{\partial \alpha \partial \alpha^T} \int_{\mathbb{R}_v} L_i dy_i^* \right] \int_{\mathbb{R}_v} L_i dy_i^* \alpha - I_\alpha \]

\[ = \int \int \int h_\alpha \frac{\partial^2}{\partial \alpha \partial \alpha^T} \int_{\mathbb{R}_v} L_i dy_i^* \alpha - I_\alpha \]

\[ = \frac{\partial^2}{\partial \alpha \partial \alpha^T} \int_{\mathbb{R}_v} \int_{\mathbb{R}_v} L_i dy_i^* \alpha - I_\alpha \]

\[ = \frac{\partial^2}{\partial \alpha \partial \alpha^T} 1 - I_\alpha = -I_\alpha \]

So

\[ \lambda(u) = \frac{1}{m^2} \left[ \sum_{i=1}^{m} U_{3i}(\alpha, \hat{\beta}, \hat{\gamma}) \right]^T u + \frac{1}{2} u^T I_\alpha u + o_P(1) \]

(5.31)
If we maximise equation 5.31 with respect to \( u \) we get the equation

\[
\frac{\partial \lambda}{\partial u} = m^{-\frac{1}{2}} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) - \frac{1}{2} I_{\alpha} u = 0_r
\]

\[
\Rightarrow \quad \hat{u} = m^{-\frac{1}{2}} I_{\alpha}^{-1} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma)
\]

From the definition of \( u \) we know that \( u \) is close to the value which maximises

\[\sum_{i=1}^{m} U_{3i}(\alpha + m^{-\frac{1}{2}} u, \beta, \gamma)dy_i^*, \]

which occurs at the maximum likelihood estimate of \( \alpha, \hat{\alpha} \).

Therefore

\[
\hat{\alpha}_{ML} = \alpha + m^{-\frac{1}{2}} \hat{u} + o_P(m^{-\frac{1}{2}})
\]

\[
= \alpha + m^{-1} \left[I_{\alpha}^{-1} \sum_{i=1}^{m} U_{3i}(\beta, \gamma) + o_P(m^{\frac{1}{2}})\right]
\]

\[
\hat{\alpha}_{ML} - \alpha = m^{-1} \left[I_{\alpha}^{-1} \sum_{i=1}^{m} U_{3i}(\beta, \gamma) + o_P(m^{\frac{1}{2}})\right]
\]

\[
m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha) = m^{-\frac{1}{2}} I_{\alpha}^{-1} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) + o_P(1) \tag{5.32}
\]

and

\[
E_{\alpha} \left[U_{3i}(\alpha, \beta, \gamma)\right]
\]

\[
= E_{\alpha} \left[ \frac{\partial}{\partial \alpha} \int_{R_{\alpha}} \cdots \int_{R_{\alpha}} L_i dy_i^* \right]
\]

\[
= \int_{R_{\alpha}} \frac{\partial}{\partial \alpha} \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right] \int_{R_{\alpha}} \cdots \int_{R_{\alpha}} L_i dy_i^* d\alpha
\]

\[
= \int_{R_{\alpha}} \frac{\partial}{\partial \alpha} \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right] \frac{\partial}{\partial \alpha} \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right] d\alpha
\]

\[
= \frac{\partial}{\partial \alpha} 1_r = 0_r \tag{5.33}
\]

and

\[
V_{\alpha} \left[U_{3i}(\alpha, \beta, \gamma)\right]
\]

\[
= V_{\alpha} \left[ \frac{\partial}{\partial \alpha} \int_{R_{\alpha}} \cdots \int_{R_{\alpha}} L_i dy_i^* \right]
\]

\[
= \left[ \frac{\partial}{\partial \alpha} \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right] \right]^T
\]

\[
= I_{\alpha} \tag{5.34}
\]

\[
= \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right] \left[ \cdots \int_{R_{\alpha}} L_i dy_i^* \right]^T
\]

\[
= I_{\alpha} \tag{5.35}
\]

150
By combining equations 5.33 and 5.35 and using the central limit theorem, we have

\[
\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) \xrightarrow{D} N_r[0_r, I_\alpha]
\]  
(5.36)

\[
E_\alpha[m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha)] = I_\alpha^{-1} E_\alpha \left[ m^{\frac{-1}{2}} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) \right]
\]

= 0_r 
(5.37)

\[
V_\alpha[m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha)] = I_\alpha^{-1} V_\alpha \left[ m^{\frac{-1}{2}} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) \right] I_\alpha^{-1}
\]

= I_\alpha^{-1} I_\alpha I_\alpha^{-1}

= I_\alpha^{-1} 
(5.38)

Since \( m^{\frac{-1}{2}} \sum_{i=1}^{m} U_{3i}(\alpha, \beta, \gamma) \) is normally distributed (equation 5.36), so will equation 5.32, i.e.

\[
m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha) \xrightarrow{D} N_r(0_r, I_\alpha^{-1})
\]  
(5.39)

5.4 Runners example

In section 4.4 a data set was introduced consisting of repeated ordinal and continuous measurements of pain felt by runners after they had run a long race. Although this is not multivariate repeated ordinal data, it is multivariate repeated data of which one of the variables is ordinal. Previous work considered these two measurements as separate entities and then compared the results. The QEEs in this chapter offer a framework for this type of analysis, although not explicitly. Going back to the original problem, if we assume there is a 5-variate normal distribution, from which the VAS scores (continuous measurements) arose and a continuous latent 5-variate normal distribution from which the Likert scores (ordinal measurements) were discrete realisations. We joined these two continuous distributions together to find a 10-variate normal likelihood, although this needed to be integrated over the 5 sets of limits (one for each ordinal measurement) to find a total likelihood for the actual data. This was then maximised with respect to \( \alpha, \beta \) and \( \gamma \) until convergence was achieved. The correlation structure for the repeated measurements was assumed to be exponential as this proved to be the best fit for the ordinal and continuous measurements. The mean structure was assumed to be as before, with the variables for the Likert and VAS scores being treated as the same. The estimates are in table 5.6.
### Table 5.6: Parameter estimates treating the ordinal and continuous data as having the same factor effects

<table>
<thead>
<tr>
<th>Latent Trend</th>
<th>VAS score</th>
<th>Latent Trend</th>
<th>VAS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>33.687(6.7846)</td>
<td>35.456(7.2387)</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>3.9612(3.6229)</td>
<td>3.7670(3.8876)</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>1.0511(0.2759)</td>
<td>0.9756(0.2671)</td>
<td></td>
</tr>
<tr>
<td>Injury</td>
<td>13.120(4.8864)</td>
<td>12.615(5.0198)</td>
<td></td>
</tr>
<tr>
<td>Time2</td>
<td>-11.10(1.4132)</td>
<td>-12.07(1.3287)</td>
<td></td>
</tr>
<tr>
<td>Time3</td>
<td>-21.51(1.7946)</td>
<td>-21.91(1.9987)</td>
<td></td>
</tr>
<tr>
<td>Time4</td>
<td>-28.73(1.8393)</td>
<td>-32.75(2.1098)</td>
<td></td>
</tr>
<tr>
<td>Time5</td>
<td>-37.90(2.2988)</td>
<td>-41.71(2.3058)</td>
<td></td>
</tr>
<tr>
<td>Repeated correlation</td>
<td>0.8456</td>
<td>0.8843</td>
<td>0.9350</td>
</tr>
<tr>
<td>Repetitive correlation</td>
<td>-</td>
<td>-</td>
<td>0.9441</td>
</tr>
<tr>
<td>Variance</td>
<td>408.55</td>
<td>374.52</td>
<td></td>
</tr>
<tr>
<td>Cut-off 1</td>
<td>-35.69(2.4551)</td>
<td>-37.12(2.2874)</td>
<td></td>
</tr>
<tr>
<td>Cut-off 2</td>
<td>-12.48(1.8274)</td>
<td>-12.44(1.9897)</td>
<td></td>
</tr>
<tr>
<td>Cut-off 3</td>
<td>-4.511(1.7759)</td>
<td>-5.127(1.7764)</td>
<td></td>
</tr>
<tr>
<td>Cut-off 4</td>
<td>12.504(1.7774)</td>
<td>13.536(1.7135)</td>
<td></td>
</tr>
<tr>
<td>Cut-off 5</td>
<td>30.639(1.8973)</td>
<td>31.330(1.9781)</td>
<td></td>
</tr>
<tr>
<td>Cut-off 6</td>
<td>59.542(2.7433)</td>
<td>64.823(2.6782)</td>
<td></td>
</tr>
<tr>
<td>Likelihood</td>
<td>-2546.99</td>
<td>-2264.85</td>
<td></td>
</tr>
</tbody>
</table>

If the VAS scores and Likert scores are treated as being independent (although related through having the same means) then the results are somewhat similar to the parameters for the analysis when carried out separately. However, by including a correlation between the repetitive measurements, the likelihood becomes less negative than when no correlation is included. This suggests, as logic does, that the VAS scores and the latent trend responsible for the Likert scores are strongly related and to ignore this relationship would lead to a poorer model.
CHAPTER 6

Conclusions and Further Work

This aim of this thesis was to create and investigate new methodology for the analysis of repeated ordinal data. This has been achieved and a set of estimation equations, QEEs, have been formulated to estimate the mean trend and the cut-off points between successive ordinal categories from a latent trend.

6.1 Chapter 3

In addition to showing the methodology of GEEs, a new scenario where GEEs can be applied to estimate a latent trend for ordinal data was formulated and investigated. This is when the data arise as discrete realisations of a latent exponential distribution. Under this assumption, simulation studies were used to examine the mean estimates and the results suggested that the estimates were unbiased. Also, it was found that under correlation miss-specification the mean estimates appeared to remain unbiased, although there was an increase in their variances, giving rise to inefficient estimates. It should be noted that there were severe model restrictions in order to apply GEEs to a latent trend, suggesting that new methodology is needed for the analysis of latent trends.
6.2 Chapter 4

Most of the new work contained within this thesis was presented in this chapter. We considered the data as being discrete realisations of an underlying multivariate distribution (normal for the majority of the analysis). Therefore we need to have a set of boundary (or cut-off) points which define the change from one ordinal category to another. Methodology, resulting in QEEs, was presented which estimated the mean of the underlying distribution (parameterised by $\alpha$), the cut-off points (parameterised by $\gamma$) and any correlation from the distribution ($\rho$). The variance could also be estimated as a nuisance parameter although it acts as a scale factor in most cases and can therefore be fixed at a specific value. The estimation process was examined using simulation studies and several features of the estimates were found. Firstly, they are unbiased and as before remain unbiased under correlation miss-specification. Secondly, they have small variance (which is incorrect under correlation miss-specification) which converges at order $m^{\frac{1}{2}}$ to a constant. Graphical evidence suggested that

$$
\lim_{m \to \infty} m^{\frac{1}{2}}(\hat{\alpha} - \alpha) = O_P(1)
$$

$$
\lim_{m \to \infty} m^{\frac{1}{2}}(\hat{\beta} - \beta) = 0 + O_P(1)
$$

$$
\lim_{m \to \infty} m^{\frac{1}{2}}\sqrt{V(\hat{\beta} - \beta)} = \text{constant}
$$

$$
\lim_{m \to \infty} m^{\frac{1}{2}}(\hat{\gamma} - \gamma) = 0 + O_P(1)
$$

$$
\lim_{m \to \infty} m^{\frac{1}{2}}\sqrt{V(\hat{\gamma} - \gamma)} = \text{constant}
$$

where $m$ is the sample size. A theoretical analysis was also carried out and it was proved that under a mild set of assumptions (one being that there is no correlation miss-specification)

$$
m^{\frac{1}{2}}(\hat{\alpha}_{ML} - \alpha) \overset{D}{\to} N(0, I^{-1}_\alpha) \quad (6.1)
$$

$$
m^{\frac{1}{2}}(\hat{\beta}_{QEE} - \beta) \overset{D}{\to} N_p(0_p, I^{-1}_\beta) \quad (6.2)
$$

$$
m^{\frac{1}{2}}(\hat{\gamma}_{QEE} - \gamma) \overset{D}{\to} N_q(0_q, I^{-1}_\gamma) \quad (6.3)
$$

which means that the estimates are asymptotically consistent.

The method of QEEs was then compared to GEEs in the simple case when we have binary data, with the GEEs using a probit link so that both methods were trying to estimate the same trend. It was found that QEEs and GEEs preformed equally, giving $m^{\frac{1}{2}}$ consistent estimates of the parameters (as shown by the equations above for QEEs). However, the GEEs can only estimate the latent trend in this case since the data are binary. If there were several ordinal categories, then there exists no 'suitable' marginal distribution, so GEEs could not be applied. On the other hand, QEEs can cope with many possible ordinal responses and offer a more flexible technique to model latent trends.
The final aim of this chapter was to analyse a real set of data consisting of repeated continuous and ordinal measurements of pain runners feel in their legs at several times after they have run a long race. ML/REML and GEEs were applied to the continuous data and QEEs to the ordinal data. From a practical point of view, it was found that the analysis of both types of data by any of the mentioned methods led to the same conclusions concerning which independent variables were important and what form the correlation structure took. From a more theoretical point of view, it was found that by estimating the parameters as if the data arose as discrete realisations of an underlying continuous distribution (QEEs) gave rise to a model similar to that found by using the true continuous data (ML/REML and GEEs). Therefore, this shows that the idea of a latent trend is applicable to this data set, and QEEs worked effectively even on the limited amount of information supplied by just having a set of ordinal categories.

6.3 Chapter 5

This chapter extended the QEEs to the case where there is more than one 'repetitive measurement' taken at several occasions. This introduced a natural 'repetitive correlation' which gives an indication of how one measurement is affected by another. Simulation studies indicated that ignoring this correlation led to poorer estimates for the mean trend and cut-off points. Again, graphical evidence was presented which suggested that

\[
\lim_{m \to \infty} m^\frac{1}{2}(\alpha - \alpha) = 0_r + O_P(1)
\]

\[
\lim_{m \to \infty} m^\frac{1}{2}(\beta - \beta) = 0_p + O_P(1)
\]

\[
\lim_{m \to \infty} m^\frac{1}{2}V(\beta - \beta) = \text{constant}
\]

\[
\lim_{m \to \infty} m^\frac{1}{2}(\gamma - \gamma) = 0_q + O_P(1)
\]

\[
\lim_{m \to \infty} m^\frac{1}{2}V(\gamma - \gamma) = \text{constant}
\]

and the theory presented in chapter 4 was extended to show that

\[
m^\frac{1}{2}(\hat{\alpha}_{ML} - \alpha) \overset{D}{\to} N_r(0_r, I_{\alpha}^{-1})
\] (6.4)

\[
m^\frac{1}{2}(\hat{\beta}_{QEE} - \beta) \overset{D}{\to} N_p(0_p, I_{\beta}^{-1})
\] (6.5)

\[
m^\frac{1}{2}(\hat{\gamma}_{QEE} - \gamma) \overset{D}{\to} N_q(0_q, I_{\gamma}^{-1})
\] (6.6)

where \(\alpha\) encompasses all repeated correlation and the repetitive correlation parameters. Again, the assumptions ensure that the correlation structures are not miss-specified.

This methodology was applied to the running data set, assuming that the two measurements were repetitive in nature (which was the case). It was found that by including this extra
correlation, a significant improvement was made to the fit of the model, although the conclusions reached were still the same. Therefore, to assume that measurements which are repetitive as being unrelated causes a detriment to the model, which may be enough to cause incorrect conclusions being drawn from data.

6.4 Further work

More work is needed in the field of repeated ordinal data analysis. At present, there is only one well established method, proportional odds, and other methodology is needed to provide flexibility in the modelling stage of the analysis. QEEs help bridge this gap by introducing a set of models when the existence of a latent trend is plausible.

This work can be extended in several directions. As it stands, the method of QEEs is computationally intensive, requiring numerous evaluations of the integral of the multivariate distribution. This is a major problem, especially when the number of occasions on which measurements are taken is large. It was found that above 10-15 occasions becomes impractical, since the integrals take too long computationally and the achieved accuracy becomes unsatisfactory. In such circumstances, several possible methods could be employed to get round this. For instance, it has been shown that minor correlation miss-specification produces unbiased (but inconsistent) results. If we assume that the correlation type is independence, then the multivariate integral can be evaluated as the product of the univariate normal integrals. By optimising the likelihood this way, a set of estimates will be obtained which are unbiased. Using these, it may be possible to ‘fine-tune’ the estimates by using methods similar to those suggested in Mark and Gail (1994) [52]. This would involve stratifying the population into sub-populations and then joining up the separate parameter estimates from each population, to achieve an overall population averaged estimate. This may lead to better estimates of the parameters, whereas the independence estimates would not be. Another possible method would be to ‘de-correlate’ the likelihood before estimation, by multiplying through by the matrix square-root of the correlation matrix. However, this causes problems since the hyper-rectangular cut-off point limits would be altered to be non-rectangular, making it difficult to evaluate the univariate integrals.

Another possible extension this work would be to work on a goodness-of-fit test for the latent trend. Since the trend itself is not ‘visible’, standard methods can not be used to test the model. If such a test could be formulated, then it might be applicable to test for the presence of a latent trend as well as the fit, which is a very important practical aspect. As the work stands, we only assume that a latent trend exists and it would be left to the statistician.
to decide if such an assumption is valid. Tests at a preliminary stage would eliminate this
decision which is fundamental to the estimation procedure.
APPENDIX A

Results from the ‘milk’ data

These are the results from the example in section 2.5.
A.1 No missing value imputation

Exponential correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St. Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4148E+01</td>
<td>0.5389E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4047E+01</td>
<td>0.5292E-01</td>
<td>0.530E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3936E+01</td>
<td>0.5293E-01</td>
<td>0.529E+00 0.538E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2286E+00</td>
<td>0.1564E-01</td>
<td>-0.609E+00 0.620E+00 0.619E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.7946E-02</td>
<td>0.8075E-02</td>
<td>-0.747E-01 -0.766E-01 -0.779E-01 -0.318E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.5951E-03</td>
<td>0.5069E-03</td>
<td>0.191E-01 0.209E-01 0.235E-01 0.227E+00 -0.927E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- tau 0.1542E+00
- nu 0.8486E-02
- sigmasq 0.7201E-01
- phi 0.1521E+00

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St. Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4148E+01</td>
<td>0.5449E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4047E+01</td>
<td>0.5350E-01</td>
<td>0.521E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3936E+01</td>
<td>0.5361E-01</td>
<td>0.520E+00 0.530E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2286E+00</td>
<td>0.1566E-01</td>
<td>-0.602E+00 -0.613E+00 -0.613E+00 0.100E+01</td>
</tr>
<tr>
<td>0.8049E-02</td>
<td>0.8114E-02</td>
<td>-0.770E-01 -0.790E-01 -0.803E-01 -0.315E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.6086E-03</td>
<td>0.5092E-03</td>
<td>0.196E-01 0.215E-01 0.241E-01 0.225E+00 -0.926E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- tau 0.1542E+00
- nu 0.8643E-02
- sigmasq 0.7399E-01
- phi 0.1479E+00
Uniform correlation

**ML estimation**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4143E+01</td>
<td>0.5325E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4047E+01</td>
<td>0.5236E-01</td>
<td>0.569E+00</td>
</tr>
<tr>
<td>0.3939E+01</td>
<td>0.5235E-01</td>
<td>0.569E+00</td>
</tr>
<tr>
<td>-0.2301E+00</td>
<td>0.1652E-01</td>
<td>-0.715E+00</td>
</tr>
<tr>
<td>0.1788E-02</td>
<td>0.5467E-02</td>
<td>0.219E+00</td>
</tr>
<tr>
<td>0.3346E-03</td>
<td>0.3527E-03</td>
<td>-0.179E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- \( \tau \) 0.2544E+00
- \( \nu \) 0.1684E+00

**REML estimation**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4143E+01</td>
<td>0.5325E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4047E+01</td>
<td>0.5236E-01</td>
<td>0.569E+00</td>
</tr>
<tr>
<td>0.3939E+01</td>
<td>0.5235E-01</td>
<td>0.569E+00</td>
</tr>
<tr>
<td>-0.2301E+00</td>
<td>0.1652E-01</td>
<td>-0.715E+00</td>
</tr>
<tr>
<td>0.1788E-02</td>
<td>0.5467E-02</td>
<td>0.219E+00</td>
</tr>
<tr>
<td>0.3346E-03</td>
<td>0.3527E-03</td>
<td>-0.179E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- \( \tau \) 0.2472E+00
- \( \nu \) 0.1671E+00
Independent correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4152E+01</td>
<td>0.4982E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4050E+01</td>
<td>0.4963E-01</td>
<td>0.918E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3932E+01</td>
<td>0.4962E-01</td>
<td>0.917E+00 0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2307E+00</td>
<td>0.1975E-01</td>
<td>-0.914E+00 -0.917E+00 -0.916E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1655E-02</td>
<td>0.6506E-02</td>
<td>0.280E+00 0.280E+00 0.280E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3232E-03</td>
<td>0.4164E-03</td>
<td>-0.229E+00 -0.229E+00 -0.228E+00 0.432E+00 -0.958E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

\[ \tau \approx 0.2993E+00 \]

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4152E+01</td>
<td>0.4993E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4050E+01</td>
<td>0.4974E-01</td>
<td>0.918E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3932E+01</td>
<td>0.4973E-01</td>
<td>0.917E+00 0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2307E+00</td>
<td>0.1979E-01</td>
<td>-0.914E+00 -0.917E+00 -0.916E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1655E-02</td>
<td>0.6521E-02</td>
<td>0.280E+00 0.280E+00 0.280E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3232E-03</td>
<td>0.4173E-03</td>
<td>-0.229E+00 -0.229E+00 -0.228E+00 0.432E+00 -0.958E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

\[ \tau \approx 0.3000E+00 \]
A.2 Time averaged imputation

Exponential correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.4163E+01</td>
<td>0.4912E-01</td>
</tr>
<tr>
<td></td>
<td>0.4048E+01</td>
<td>0.4840E-01</td>
</tr>
<tr>
<td></td>
<td>0.3916E+01</td>
<td>0.4840E-01</td>
</tr>
<tr>
<td></td>
<td>-0.2282E+00</td>
<td>0.1552E-01</td>
</tr>
<tr>
<td></td>
<td>0.3095E-02</td>
<td>0.7399E-02</td>
</tr>
<tr>
<td></td>
<td>0.1374E-03</td>
<td>0.4373E-03</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- tau: 0.1521E+00
- nu: 0.4965E-01
- sigmasq: 0.5597E-01
- phi: 0.2086E+00

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.4163E+01</td>
<td>0.4957E-01</td>
</tr>
<tr>
<td></td>
<td>0.4048E+01</td>
<td>0.4882E-01</td>
</tr>
<tr>
<td></td>
<td>0.3916E+01</td>
<td>0.4882E-01</td>
</tr>
<tr>
<td></td>
<td>-0.2281E+00</td>
<td>0.1554E-01</td>
</tr>
<tr>
<td></td>
<td>0.3088E-02</td>
<td>0.7428E-02</td>
</tr>
<tr>
<td></td>
<td>0.1370E-03</td>
<td>0.4388E-03</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:
- tau: 0.1520E+00
- nu: 0.5466E-01
- sigmasq: 0.5680E-01
- phi: 0.2062E+00
Uniform correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4161E+01</td>
<td>0.4961E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4054E+01</td>
<td>0.4690E-01</td>
<td>0.625E+00</td>
</tr>
<tr>
<td>0.3925E+01</td>
<td>0.4692E-01</td>
<td>0.626E+00</td>
</tr>
<tr>
<td>-.2323E+00</td>
<td>0.1608E-01</td>
<td>-0.750E+00</td>
</tr>
<tr>
<td>0.3035E-02</td>
<td>0.5050E-02</td>
<td>0.229E+00</td>
</tr>
<tr>
<td>0.2381E-03</td>
<td>0.3071E-03</td>
<td>-0.190E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

tau 0.2424E+00
nu 0.1414E+00

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4161E+01</td>
<td>0.5002E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4054E+01</td>
<td>0.4929E-01</td>
<td>0.617E+00</td>
</tr>
<tr>
<td>0.3925E+01</td>
<td>0.4929E-01</td>
<td>0.617E+00</td>
</tr>
<tr>
<td>-.2323E+00</td>
<td>0.1610E-01</td>
<td>-0.744E+00</td>
</tr>
<tr>
<td>0.3035E-02</td>
<td>0.5055E-02</td>
<td>0.227E+00</td>
</tr>
<tr>
<td>0.2381E-03</td>
<td>0.3074E-03</td>
<td>-0.189E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

tau 0.2449E+00
nu 0.1459E+00
### Independent correlation

#### ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4161E+01</td>
<td>0.4694E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4054E+01</td>
<td>0.4680E-01</td>
<td>0.926E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3925E+01</td>
<td>0.4680E-01</td>
<td>0.926E+00 0.928E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2323E+00</td>
<td>0.1861E-01</td>
<td>-0.917E+00 -0.920E+00 -0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3035E-02</td>
<td>0.5846E-02</td>
<td>0.280E+00 0.281E+00 0.281E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>0.2381E-03</td>
<td>0.3555E-03</td>
<td>-0.232E+00 -0.232E+00 -0.233E+00 -0.233E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

- \( \tau = 0.2837E+00 \)

#### REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4161E+01</td>
<td>0.4703E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4054E+01</td>
<td>0.4690E-01</td>
<td>0.926E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3925E+01</td>
<td>0.4690E-01</td>
<td>0.926E+00 0.928E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2323E+00</td>
<td>0.1855E-01</td>
<td>-0.917E+00 -0.920E+00 -0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3035E-02</td>
<td>0.5858E-02</td>
<td>0.280E+00 0.281E+00 0.281E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>0.2381E-03</td>
<td>0.3563E-03</td>
<td>-0.232E+00 -0.232E+00 -0.233E+00 -0.233E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

- \( \tau = 0.2843E+00 \)
A.3 Moving average imputation

Exponential correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4144E+01</td>
<td>0.5461E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4046E+01</td>
<td>0.5355E-01</td>
<td>0.489E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3945E+01</td>
<td>0.5355E-01</td>
<td>0.489E+00 0.499E+00 0.100E+01</td>
</tr>
<tr>
<td>-.2294E+00</td>
<td>0.1495E-01</td>
<td>-0.567E+00 -0.579E+00 -0.579E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1296E-01</td>
<td>0.7629E-02</td>
<td>-0.102E+00 -0.104E+00 -0.104E+00 -0.294E+00 0.100E+01</td>
</tr>
<tr>
<td>-.1231E-02</td>
<td>0.4451E-03</td>
<td>0.353E-01 0.360E-01 0.360E-01 0.210E+00 -0.931E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

- tau 0.1418E+00
- nu 0.8901E-02
- sigmasq 0.7920E-01
- phi 0.1285E+00

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4144E+01</td>
<td>0.5524E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4046E+01</td>
<td>0.5414E-01</td>
<td>0.481E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3945E+01</td>
<td>0.5414E-01</td>
<td>0.481E+00 0.491E+00 0.100E+01</td>
</tr>
<tr>
<td>-.2294E+00</td>
<td>0.1487E-01</td>
<td>-0.561E+00 -0.573E+00 -0.573E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1293E-01</td>
<td>0.7662E-02</td>
<td>-0.104E+00 -0.106E+00 -0.106E+00 -0.291E+00 0.100E+01</td>
</tr>
<tr>
<td>-.1236E-02</td>
<td>0.4466E-03</td>
<td>0.355E-01 0.362E-01 0.362E-01 0.208E+00 -0.930E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

- tau 0.1419E+00
- nu 0.9018E-02
- sigmasq 0.8128E-01
- phi 0.1285E+00
Uniform correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4156E+01</td>
<td>0.5446E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4057E+01</td>
<td>0.5359E-01</td>
<td>0.579E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3956E+01</td>
<td>0.5359E-01</td>
<td>0.579E+00 0.588E+00 0.100E+01</td>
</tr>
<tr>
<td>-.2404E+00</td>
<td>0.1696E-01</td>
<td>-0.720E+00 -0.732E+00 -0.732E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1834E+01</td>
<td>0.327E-02</td>
<td>0.220E+00 0.223E+00 0.223E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>-.1481E+02</td>
<td>0.324E-03</td>
<td>-0.182E+00 -0.185E+00 -0.185E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

\[
\begin{align*}
\text{tau} & \quad 0.2633E+00 \\
\text{nu} & \quad 0.1717E+00 
\end{align*}
\]

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St.Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4156E+01</td>
<td>0.5496E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4057E+01</td>
<td>0.5405E-01</td>
<td>0.570E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3956E+01</td>
<td>0.5405E-01</td>
<td>0.570E+00 0.579E+00 0.100E+01</td>
</tr>
<tr>
<td>-.2404E+00</td>
<td>0.1696E-01</td>
<td>-0.715E+00 -0.727E+00 -0.727E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1834E+01</td>
<td>0.3333E-02</td>
<td>0.218E+00 0.222E+00 0.222E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>-.1481E-02</td>
<td>0.3243E-03</td>
<td>-0.184E+00 -0.184E+00 -0.184E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

\[
\begin{align*}
\text{tau} & \quad 0.2588E+00 \\
\text{nu} & \quad 0.1723E+00 
\end{align*}
\]
Independent correlation

ML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St. Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4156E+01</td>
<td>0.5106E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4057E+01</td>
<td>0.5092E-01</td>
<td>0.926E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3956E+01</td>
<td>0.5092E-01</td>
<td>0.926E+00 0.928E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2404E+00</td>
<td>0.2025E-01</td>
<td>-0.917E+00 -0.920E+00 -0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1834E-01</td>
<td>0.6360E-02</td>
<td>0.280E+00 0.281E+00 0.281E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.1481E-02</td>
<td>0.3868E-03</td>
<td>-0.232E+00 -0.233E+00 -0.233E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

tau 0.3086E+00

REML estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>St. Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4156E+01</td>
<td>0.5117E-01</td>
<td>0.100E+01</td>
</tr>
<tr>
<td>0.4057E+01</td>
<td>0.5102E-01</td>
<td>0.926E+00 0.100E+01</td>
</tr>
<tr>
<td>0.3956E+01</td>
<td>0.5102E-01</td>
<td>0.926E+00 0.928E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.2404E+00</td>
<td>0.2029E-01</td>
<td>-0.917E+00 -0.920E+00 -0.920E+00 0.100E+01</td>
</tr>
<tr>
<td>0.1834E-01</td>
<td>0.6373E-02</td>
<td>0.280E+00 0.281E+00 0.281E+00 -0.529E+00 0.100E+01</td>
</tr>
<tr>
<td>-0.1481E-02</td>
<td>0.3876E-03</td>
<td>-0.232E+00 -0.233E+00 -0.233E+00 0.439E+00 -0.964E+00</td>
</tr>
</tbody>
</table>

Parameters in variance matrix:

tau 0.3190E+00
Multi-dimensional calculus results

Two theorems were used in the derivation of the quasi-likelihood equations. The proof of them follows.

**Theorem B.1 (Multi-dimensional Leibnitz)**

\[
\frac{d}{dt} \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x, \ldots, y, t) dy \cdots dx = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial f(x, \ldots, y, t)}{\partial t} dy \cdots dx \quad \text{ (B.1)}
\]

provided

\[
f(x, \ldots, y, t) \text{ is continuous} \quad \text{ (B.2)}
\]

\[
\frac{\partial f(x, \ldots, y, t)}{\partial t} \text{ is continuous} \quad \text{ (B.3)}
\]

**Proof**

Define

\[
g(t) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial f(x, \ldots, y, t)}{\partial t} dy \cdots dx
\]

for \( t_1 \leq t \leq t_2 \). By the condition in equation B.3, \( g(t) \) is also continuous within this interval.

Choose another time point, \( t_3 \) which satisfy \( t_1 \leq t_3 \leq t_2 \). Then

\[
\int_{t_1}^{t_3} g(t) dt
\]
= \int_{t_1}^{t_2} \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x, \ldots, y, t) \, dy \cdots dx dt
\]
\[
= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial}{\partial t} f(x, \ldots, y, t) \, dy \cdots dx
\]
\[
= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \left[ f(x, \ldots, y, t_3) - f(x, \ldots, y, t_1) \right] \, dy \cdots dx
\]
\[
= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x, \ldots, y, t_3) \, dy \cdots dx - \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x, \ldots, y, t_1) \, dy \cdots dx
\]
\[
= F(t_3) - F(t_1)
\]

Since \( t_3 \) was just an arbitrary time value, call it \( t \), which gives
\[
\int_{t_1}^{t} g(u) \, du = F(t) - F(t_1)
\]

Differentiate this with respect to \( t \)
\[
\frac{dF}{dt} = \frac{d}{dt} \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x, \ldots, y, t) \, dy \cdots dx
\]
\[
= g(t)
\]
\[
= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial}{\partial t} f(x, \ldots, y, t) \, dy \cdots dx \tag{B.4}
\]
as required.

**Theorem B.2**

\[
\frac{d}{dt} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(y_1, \ldots, y_n, t) \, dy_1 \cdots dy_n
\]
\[
= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial}{\partial t} f(y_1, \ldots, y_n, t) \, dy_1 \cdots dy_n
\]
\[
= \cdots
\]
\[
= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(y_1, \ldots, y_n-1, y_n, t) \, dy_1 \cdots dy_{n-1} \cdots dy_n
\]
\[
= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(y_1, \ldots, y_n-1, a_n, t) \, dy_1 \cdots dy_{n-1} \cdots dy_n
\]

provided
\[
f(x, \ldots, y, t) \text{ is continuous} \tag{B.5}
\]
\[
\frac{\partial f(x, \ldots, y, t)}{\partial t} \text{ is continuous} \tag{B.6}
\]
\[
(F.7)
\]

169
Proof

Define

\[ u_1 = a_1(t), \ldots, u_n = a_n(t) \]
\[ v_1 = b_1(t), \ldots, v_n = b_n(t) \]

\[ w = t \]

Then

\[ F(t) = \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_n, w) dy_n \cdots dy_1 \]
\[ = G(u_1, \ldots, u_n, v_1, \ldots, v_n, w) \]

and

\[
\frac{dF}{dt} = \frac{\partial G}{\partial u_1} \frac{du_1}{dt} + \cdots + \frac{\partial G}{\partial u_n} \frac{du_n}{dt} + \frac{\partial G}{\partial v_1} \frac{dv_1}{dt} + \cdots + \frac{\partial G}{\partial v_n} \frac{dv_n}{dt} + \frac{\partial G}{\partial w} \frac{dw}{dt}
\]

(B.8)

The terms in equation B.8 are almost identical to those that we are trying to prove. Only the partial derivatives need be derived to finish the proof.

\[
\frac{\partial G}{\partial u_k} = \frac{\partial}{\partial u_k} \int_{u_1}^{v_1} \cdots \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) dy_n \cdots dy_1
\]
\[ = \frac{\partial}{\partial u_k} \int_{u_1}^{v_1} \cdots \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) dy_k dy_n \cdots dy_1
\]
\[ = \sum_{u_1}^{v_1} \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \int_{u_1}^{v_1} \int_{u_n}^{v_n} \left[ \frac{\partial}{\partial u_k} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) \right] dy_k dy_n \cdots dy_1
\]

(B.9)

\[
\frac{\partial G}{\partial v_k} = \frac{\partial}{\partial v_k} \int_{u_1}^{v_1} \cdots \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) dy_n \cdots dy_1
\]
\[ = -\frac{\partial}{\partial v_k} \int_{u_1}^{v_1} \cdots \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) dy_n \cdots dy_1
\]
\[ = -\sum_{u_1}^{v_1} \int_{u_k-1}^{v_k-1} \int_{u_k+1}^{v_k+1} \int_{u_1}^{v_1} \int_{u_n}^{v_n} f(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_n, w) dy_n \cdots dy_1
\]

(B.10)

\[
\frac{\partial G}{\partial w} = \frac{\partial}{\partial w} \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} f(y_1, \ldots, y_n, w) dy_n \cdots dy_1
\]
\[ = \int_{u_1}^{v_1} \int_{u_n}^{v_n} \frac{\partial f(y_1, \ldots, y_n, w)}{\partial w} dy_n \cdots dy_1
\]

(B.11)

Substituting these 2k + 1 equations, equations B.9, B.10 and B.11, into B.8 gives

\[
\frac{d}{dt} \int_{a_1(t)}^{b_1(t)} \cdots \int_{a_n(t)}^{b_n(t)} f(y_1, \ldots, y_n, t) dy_n \cdots dy_1
\]

170
\begin{align*}
&= \int_{a_2(t)}^{b_2(t)} \cdots \int_{a_n(t)}^{b_n(t)} f(b_1, \ldots, y_n, t) \, dy_n \cdots dy_2 \\
&\quad + \cdots + \\
&\quad \int_{a_1(t)}^{b_1(t)} \cdots \int_{a_{n-1}(t)}^{b_{n-1}(t)} f(y_1, \ldots, y_{n-1}, b_n, t) \, dy_{n-1} \cdots dy_1 \\
&\quad - \int_{a_2(t)}^{b_2(t)} \cdots \int_{a_n(t)}^{b_n(t)} f(b_1, \ldots, y_n, t) \, dy_n \cdots dy_2 \\
&\quad - \cdots - \\
&\quad \int_{a_1(t)}^{b_1(t)} \cdots \int_{a_{n-1}(t)}^{b_{n-1}(t)} f(y_1, \ldots, y_{n-1}, a_n, t) \, dy_{n-1} \cdots dy_1 \\
&\quad + \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{\partial f(y_1, \ldots, y_n, t)}{\partial t} \, dy_n \cdots dy_1 \quad \text{as required.}
\end{align*}


[38] Joreskog, K.G. (1973) 'A general method for estimating a linear structural equation system'. Structural Equation Models in the Social Sciences


174


[65] Patterson, H.D., Thompson, R. (1971) 'Recovery of inter-block information when block sizes are unequal'. *Biometrika*, 58, 545-554.


