UNIVERSITY OF SOUTHAMPTON

Models and estimation for repeated ordinal responses, with application to telecommunications experiments

by

Rory St.John Wolfe

Thesis submitted for the degree of Doctor of Philosophy

Faculty of Mathematical Studies

May 1996

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES

MATHEMATICS

Doctor of Philosophy

MODELS AND ESTIMATION FOR REPEATED ORDINAL RESPONSES, WITH APPLICATION TO TELECOMMUNICATIONS EXPERIMENTS

by Rory St.John Wolfe

Responses made on scales with ordered categories (ordinal responses) can be analysed using multinomial models which include 'threshold parameters'. These models have become established over the last 20 years, in which time research has focused on models with location terms which allow for a change in location of the threshold parameters. These location terms explain one type of difference between patterns of ordinal responses. Another type of difference can be explained by the inclusion of scaling terms in the model.

Where ordinal responses are observed repeatedly on the same subject the analyst has a challenge to explain the correlation between these intra-subject responses. This thesis presents a new approach to this challenge which involves fitting one of the multinomial models, the cumulative logit model, with subject-specific location and scaling terms. These terms may be fixed effects or random effects and both cases are investigated. The approach is motivated by data from telecommunications experiments, and when used to analyse these data, it is found that the model gives a good explanation of the correlation between intra-subject responses. A new piece of general-purpose software is introduced which allows the fitting, by maximum-likelihood, of cumulative link models with both location and scaling terms.

It is possible to fit the cumulative logit model by using generalized estimating equations (GEE). One particular type of GEE is discussed in detail and referred to as 'independent binomials'. An advantage to the use of this method of fitting the cumulative logit model is its straightforward implementation. Some theoretical comparisons are performed to compare the efficiency of independent-binomials estimation of model parameters with the efficiency of maximum-likelihood estimation. It is concluded that the loss of efficiency in independent-binomials cannot be considered too great to warrant its dismissal as a method of estimation.

Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

Sincere thanks to my supervisor, Dr David Firth, whose intelligence, intuition and constant encouragement made this work possible.

The thesis was completed while I was a research assistant on a joint project between the Mathematics Department at Southampton University and British Telecom. My thanks go to both institutions and in particular Dr Sue Lewis for her energy, dedication and academic excellence and Brian Surtees for his enthusiasm and insight.

Many members of staff in the Department have helped me enormously in my time here. A special thank you to Dr Phil Prescott, Dr Mac McDonald, Ray Brown, Yvonne Oliver and Prof Fred Smith.

Thanks also to all the students who have made life in the office pleasurable: Masoud, Denise W, Denise S, Christine, Parviz, Chris, Pang, Kath and Paula.

Finally, thank you to my family for their support.

Chapter 1

Introduction

This thesis examines issues involved with modelling ordinal response data. Ordinal responses are common features of research in psychology, medicine, agriculture, *etc.* A typical ordinal scale is labelled with the words "Good", "Fair", "Poor", "Bad". In a psychology experiment this scale might be used to record different subjects' opinions of a mood stimulus. In medicine the measurement of interest might be a patient's reaction to a prescribed drug and doctors might use the ordinal scale in the appraisal of the patients. In agriculture the scale might be used to classify plants according to a subjective quality such as hardiness.

The application of interest in this thesis is telecommunications research where the ordinal response is a subject's opinion of a telephone connection. Everyone who has experience of using telephonic communication knows that the quality of the connection can vary a great deal. At present, the general quality of connections on national telephone calls is good. However, the quality of the connection can be much worse if more modern telephones, *e.g.* mobile phones, are used. The quality of long-distance connections can also be quite poor. The quantification of the general public's opinion towards different connections and discovering which components of the connection are important in the formation of these opinions is important research for any telecommunications company.

A statistical topic of concern in this thesis is the problem posed by repeated observations being made on the same sampling unit. An assumption of much statistical modelling is the independence of observations (conditional on the effects included in the model). Where different observations are made on the same sampling unit it is quite likely that there will be correlation between these intra-unit observations. It is possible that the inclusion of suitable unit-specific explanatory effects in a model will allow for this intra-unit correlation. It may be necessary to model this correlation explicitly. The issue of repeated intra-unit observations has been vigorously addressed in recent literature. Analysis methods have been developed for all types of univariate repeated data, *i.e.* binary responses, continuous responses and both ordered and unordered categorical responses. The work in this thesis considers only the approach to the modelling of repeated ordinal responses where explanatory variables are included in the model to account for intra-unit correlation.

Historically, the most common method of modelling ordinal response data involved assigning increasing integer scores to the ordered categories and fitting a linear model to these scores. The parameters in the model were estimated by least-squares. This approach has two drawbacks. Firstly, the choice of scores is arbitrary and, secondly, it is necessary to assume that the data when scored are, conditionally on the covariates in the model, observations from a normal distribution.

Another method that has been used is to reduce the data to binomial responses by dichotomising the response scale at some arbitrary point. The dichotomised responses are then modelled using logistic regression for a binary response. This has the drawback that information is being discarded in the dichotomisation process. Also the choice of where to make the dichotomy is arbitrary and the results obtained from the logistic regression depend on this choice. Thus, different conclusions may be reached for different dichotomisations.

A more recent approach to modelling ordinal response data is to recognise explicitly that the responses are observations from a multinomial distribution. McCullagh (1980) proposed a family of models based on the cumulative probabilities of each category. Alternative models based on continuation-ratio probabilities are also discussed in that paper. These models are traditionally fitted by maximum-likelihood although recent work demonstrates how a Bayesian perspective of the model leads to the Gibbs sampler being used. It is also possible to define quasi-likelihood estimating equations to fit the model and an advantage of this is the straightforward extension to the analysis of data containing repeated ordinal responses on the same sampling unit.

The background to multinomial modelling of ordinal responses is presented in Chapter 2. A cohesive model framework is defined which encapsulates a wide variety of multinomial models for ordinal response data. This framework relates an underlying probability via a link function to a combination of explanatory variables. The three types of probability discussed in this thesis are cumulative probability, continuation-ratio probability and, more briefly, adjacent-category probability. The combination of explanatory variables is typically linear (and the corresponding parameters termed location parameters) although a particular instance of non-linearity, the model with a scaling term, is discussed in detail. A number of known equivalences between different multinomial models for ordinal responses are discussed and a new result demonstrating the asymptotical equivalence of the cumulative probit and adjacent-category probit models is introduced. Some terminological issues are discussed with a view to standardizing terminology. Also discussed briefly in Chapter 2 are some of the alternative methods that have been devised for modelling ordinal response data. These include the linear model with integer scores assigned to categories.

The telecommunications research motivating much of the work in this thesis was performed at the laboratories of British Telecom (BT) in Ipswich. The ordinal response is one of a variety of measurements recorded in designed experiments performed by BT. Chapter 3 describes the experimentation in detail. The experimental designs are introduced, the experimental procedure is described and the resultant data are discussed. Data from the BT experiments are used to make a comparison of the goodness of fit of two models. The first model is the linear model with integers assigned to categories. This model leads to the construction of an analysis of variance table to aid the testing of significance of explanatory variables. The second model is a continuation-ratio logit model which is an example of a multinomial model. Two criteria to compare the models, mean scores and fitted categories, are discussed and applied to the BT data.

The ordinal responses that are observed in the BT experiments are longitudinal in nature. A number of subjects respond to different sequences of stimuli. It is possible that the preceding stimulus or response will have an effect on the response given to the current stimulus. This type of effect is termed a carry-over effect. An approach to analysing the carry-over of response using the continuation-ratio logit model is presented in Chapter 3. This approach is applied to data from the BT experiments.

For the goodness of fit comparisons in Chapter 3 the intra-subject responses are assumed to be independent (conditional on the covariates in the model). This serves to demonstrate these comparisons but is not an intuitively appealing assumption. For the carry-over analysis it is assumed that responses are independent conditional on the immediately preceding responses (and other covariates in the model). It is possible to define generalized estimating equations (GEE) which allow for some general correlation structure between intra-subject responses. The use of GEE, assuming independence of intra-subject responses, to fit a multinomial model is demonstrated in this thesis. However, the use of GEE with a general correlation structure between intra-subject responses is not pursued. The reason for this is that informal analysis of the BT data suggests an alternative approach, which has as a key consideration the pattern of responses observed for each subject. In simple terms, this new approach allows for these patterns to differ in their location on the response scale and to differ in the degree to which they are spread across the response scale.

The vehicle for this new method is a particular multinomial model, the cumulative logit model, with location and scaling terms for each subject. Any of the other cumulative link models would also serve the purpose. A hierarchy of cumulative logit models can be fitted to test the different types of subject-specific effects. The application of this method to the BT data is the focus of Chapter 5. Results of fitting the model hierarchy are presented. The possibility of using this method when the subject effects are assumed to be random effects is also investigated in Chapter 5. Initially the cumulative logit model with a random subject location effect is fitted to the BT data and the results discussed. The necessary score equations are presented for maximum-likelihood estimation of the more complicated model where there is both a random subject scaling effect and a random subject location effect.

Chapter 4 gives detailed discussion of the different methods of fitting the cumulative logit model and software implementation of these methods. Maximum-likelihood estimation of the cumulative logit model with location effects only is possible in current widely-available software. A GEE method with an independence working correlation matrix, termed independent binomials, is an alternative to maximum likelihood. Under this working correlation the GEEs reduce to the score equations of a maximum-likelihood analysis of independent binomial responses and is thus implementable in any software with a routine for logistic regression of binary observations. Fitting the model with a scaling term is more difficult to implement. Two methods are considered in Chapter 4 to do this. Firstly, a set of GLIM4 macros is introduced to fit the model by maximum-likelihood. These macros are general purpose and have been included in the GLIM Macro Library. Secondly, the independent-binomials method is employed by succesive relaxations of the model.

Results from the use of both of these methods are given in Chapter 5 and differences discussed.

The method of independent binomials for fitting the cumulative logit model involves independence of the cumulative multinomial counts as a working assumption. This working assumption allows the method to be implemented in widely available software routines. However the method requires more justification than ease of implementation. The empirical and asymptotic relative efficiency comparisons made in Chapter 6 support the use of independent binomials as a method of estimation of the cumulative logit model. These comparisons are of estimates from various GEE methods of fitting the cumulative logit model (including independent binomials) with maximum-likelihood estimates. Further asymptotic relative efficiency comparisons are given for maximum-likelihood estimates and estimates obtained when the analysis is reduced to a binomial one by the dichotomisation of the response scale.

Finally, some concluding remarks are made in Chapter 7. The outcomes of the analyses in Chapter 5 and the conclusions to be drawn from them are discussed. The main results in the thesis are discussed in a wider context and avenues for future research are suggested. Comparisons are made between the approach taken in this thesis to modelling longitudinal ordinal responses and alternative approaches considered in the literature.

Chapter 2

Ordinal response modelling

2.1 Introduction

The use of ordered categorical response scales to collect data is widespread. In social surveys this has been a popular method of generating data for a number of years. An example comes from the 1972 U.S. General Social Survey of the National Data Program. Responses were sought on people's attitude toward abortion. The scale on which these responses were made has three categories: Generally disapprove, Middle position, Generally approve. In psychological research ordinal scales are frequently used as rating scales in experiments. An example which is examined in further detail in this thesis comes from Kijewski, Swensson and Judy (1989). They discuss an experiment in which a subject is required to respond to a range of visual stimuli, the responses being made on a 12-category ordinal scale with 6 grades of bright and 6 grades of dark. The statistical practice of grouping continuous data into ordered categories is another example of how ordinal responses originate.

A number of different methods of analysing data with an ordinal response have been developed. A simple approach involves assigning a score to each category and then analysing these scores using a standard linear model or non-parametric methods. These approaches do have the drawback that the scores assigned are arbitrary. One example of such an approach is the method that Agresti (1984, §8.3) calls the mean response model. This involves performing an analysis of variance on the scores assigned to the levels of the ordinal response. There are obvious advantages to using this model. The interpretation of the parameter estimates is in terms of means or adjusted means, both of which are very familiar concepts. Also, analysis of variance is a well known statistical technique and its application to ordinal response problems would remove the need for a different technique. However, this method does assume that, conditional on the covariates in the model, the integer scores assigned to categories are normally distributed.

Another potential drawback to the mean response model is that it does not necessarily give simple overall conclusions. For example, consider an ordinal scale which consists of four social class groupings and a variable such as income. If social class is considered as the response variable, successive integer scores (1 to 4) can be assigned to the ordered categories and then an analysis of variance can be performed on these scores, using income as an explanatory variable. From the fitted model the parameter estimate for the effect of income can easily be interpreted as the increase in fitted score for a unit increase in income. For a person of given income the model provides the following type of conclusion.

> With an income of 10,000 pounds a person will have on average a social class score of 2.38.

This conclusion requires further interpretation — what is a social class score of 2.38 and how does it relate to the initial categories? If the scores are changed do the conclusions change?

An approach to analysing ordinal responses which does not involve the assignment of arbitrary scores to categories is now motivated. If Y can be thought of as the ordered categorical manifestation of an underlying continuous random variable

Y^* then it is assumed that Y is observed in the following manner:

$$Y = j \qquad \text{if} \qquad \theta_{j-1} < Y^* \leq \theta_j \qquad (j = 1, \ldots, K)$$

where θ_j can be thought of as some unknown 'cut-point' on the underlying continuum. Now let Y^* take the linear form $Y_i^* = \mathbf{x}_i\beta + \epsilon$, where ϵ has a specified distribution F and $\mathrm{E}(\epsilon) = 0$. The design matrix here is $\mathbf{X}$ of which $\mathbf{x}_i$ is the appropriate row and the parameters to be estimated are contained in the vector β. Taking this form for Y^* gives:

$$\begin{aligned} \mathrm{P}(Y_i = j) &= \mathrm{P}(\theta_{j-1} < Y_i^* \leq \theta_j) \\ &= \mathrm{P}(\theta_{j-1} - \mathbf{x}_i\beta < \epsilon_i \leq \theta_j - \mathbf{x}_i\beta) \\ &= F(\theta_j - \mathbf{x}_i\beta) - F(\theta_{j-1} - \mathbf{x}_i\beta) \end{aligned}$$

which leads to

$$\mathrm{P}(Y_i \leq j) = F(\theta_j - \mathbf{x}_i\beta).$$

The θ_js are necessarily ordered, $\theta_1 < \theta_2 < \cdots < \theta_{K-1}$, and these are augmented by the definitions $\theta_0 = -\infty$ and $\theta_K = \infty$.

If we take the distribution of F to be the logistic distribution then we obtain the model

$$\log\left(\frac{\mathrm{P}(Y_i \leq j)}{1 - \mathrm{P}(Y_i \leq j)}\right) = \theta_j - \mathbf{x}_i\beta \tag{2.1}$$

which is the cumulative logit or proportional odds model for ordinal responses. This and related models are discussed in the influential paper by McCullagh (1980).

A model of this form (which is not based on arbitrary score-assignment) will give the following type of conclusion to the example above.

> With an income of 10,000 pounds a person has an 85% chance of being in social class group 2.

The tone of this conclusion may seem less definite than the conclusion from a score-based model. However there are no further interpretation problems since it is stated concisely in terms of the initial response scale.

This chapter contains an introduction to the cumulative logit model and other related models. Several connections between different models are noted and a new asymptotic equivalence is introduced. Terminological difficulties are discussed in detail for two models, the proportional odds or cumulative logit model and the proportional hazards or cumulative complementary-log-log model. Other relevant topics are touched on. These include the type of ordinal scale, the sampling mechanism employed to generate the data, and the different types of covariate effect that may be estimated via a model for an ordinal response. This latter topic includes the important considerations of models with scaling terms and models with random effects, both of which are examined in detail in other parts of the thesis.

2.2 A general model form

A general framework for a group of models for analysing an ordinal response Y, with K ordered categories, is

$$\mathrm{link}(\Pi_{ij}) = \theta_j - \mathbf{x}_i\beta. \tag{2.2}$$

The term Π_{ij} is a probability, which is some function of the individual category probabilities π_{ij}. An example is $\Pi_{ij} = \pi_{i1}+\cdots+\pi_{ij} = \gamma_{ij}$, the cumulative probability of category j. The term 'link' refers to a monotone function which maps $(0,1)$ onto $(-\infty,\infty)$, for example the logit link, $\log(\Pi/(1-\Pi))$. The threshold parameters, θ_j $(j = 1,\ldots,K-1)$, correspond to the boundaries between the $j-1$ and jth categories of Y. The row $\mathbf{x}_i$ of the design matrix, $\mathbf{X}$, contains explanatory variable values corresponding to the ith response. In the case of an ordinal explanatory variable the analyst may simply treat this variable as unordered categorical. Another possibility is is to apply an integer score to the categories and treat the variable as quantitative. The vector β contains parameters of interest which describe the covariate effects.

This framework includes the cumulative logit model discussed in the introduction to this chapter. The framework is a convenient way of describing, in one formula, a range of closely related models. It also highlights the similarities among a variety of these models. Models in the framework (2.2) may have some of the following properties.

1. *Stochastic ordering.* Strict stochastic ordering of two groups ($i = 1, 2$) is defined if either $\gamma_{1j} > \gamma_{2j}$ or $\gamma_{1j} < \gamma_{2j}$ for all j. From equation (2.2)

$$\text{link}(\Pi_{1j}) - \text{link}(\Pi_{2j}) = (\mathbf{x_2} - \mathbf{x_1})\beta$$

   and this implies strict stochastic ordering.

2. *Palindromic invariance.* If a model is invariant to a reversal of the category ordering, then it is said to be palindromic invariant (after McCullagh 1978). This invariance holds if the only change in the parameters after a reversal of category order is that β changes sign, and the θ_js reverse order.

3. *Invariance to contiguous-category collapsing.* If the form of the model and the parameters (β, θ_j) do not change when the responses from two neighbouring categories are merged, then that model is said to be invariant to contiguous-category collapsing. In fact the only change is that the relevant θ_j parameter is removed. Anderson & Philips (1981) give an example of this consideration.

2.2.1 Link functions

The distribution F (of which the link function in (2.2) is the inverse) can be taken as any monotone increasing function that maps $[0,1]$ onto $(-\infty,\infty)$. The logit link

$$F^{-1}(\Pi) = \log\left(\frac{\Pi}{1-\Pi}\right) \qquad (2.3)$$

has been introduced in this section. This link function is obtained when the logistic distribution is the assumed form for F. The function defined in (2.3) is symmetric about $\Pi = 0.5$, where $F^{-1}(\Pi) = 0$, in that $F^{-1}(\Pi) = -F^{-1}(1-\Pi)$.

A link function which behaves in a similar manner to the logit link is the probit link function. The probit link function is defined as

$$F^{-1}(\Pi) = \Phi^{-1}(\Pi)$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal random variable. This link function is also symmetric about $\Pi = 0.5$, where $F^{-1}(\Pi) = 0$, in that $F^{-1}(\Pi) = -F^{-1}(1-\Pi)$.

There is a pair of commonly-used non-symmetric link functions. These are the complementary log-log and log-log links. The complementary log-log link is defined as

$$F_{\mathrm{cll}}^{-1}(\Pi) = \log\{-\log(1-\Pi)\}$$

and is obtained when the assumed form of F is the standard extreme minimum value distribution. The log-log link is defined as

$$F_{\mathrm{ll}}^{-1}(\Pi) = \log\{-\log(\Pi)\}$$

and is obtained when the assumed form of F is the standard extreme maximum value distribution. It is related to the complementary log-log link by $F_{cll}^{-1}(\Pi) = F_{ll}^{-1}(1-\Pi)$. The complementary log-log link is more usually employed because it behaves in an almost identical fashion to the logit link in the region $\Pi < 0.5$.

For a more complete discussion of link functions refer to §4.3 in McCullagh & Nelder (1989).

2.2.2 Probabilities Π

Cumulative probability

The case of $\Pi_{ij} = \mathrm{P}(Y_i \leq j) = \gamma_{ij}$, the *cumulative* probability of category j, has been introduced already in this section. The cumulative logit model is defined in equation (2.1). In this thesis models based on the cumulative probability are referred to as cumulative link models. When discussing a particular member of the class of cumulative link models, the word link is replaced by the name of the particular link function being used, hence cumulative logit model.

Terminology here has become somewhat confused. McCullagh proposed the name 'proportional odds' for the cumulative logit model and 'proportional hazards' for the cumulative complementary log-log model. The second of these can be misleading as will be demonstrated in §2.5.4. In this thesis, the terminology that gives rise to the names cumulative logit and cumulative complementary log-log will be adhered to.

All models based on the cumulative probability imply stochastic ordering. The cumulative logit and cumulative probit models are palindromic invariant due to a combination of the symmetric nature of the link functions and the simple form of the cumulative probability. The cumulative complementary log-log model is not palindromic invariant but the same β parameters and the same θ_j parameters in reverse will be obtained if the categories are defined in reverse order and a cumulative log-log model is fitted. All the cumulative link models are invariant to contiguous category collapsing.

Continuation-ratio probability

Another probability, to be used in place of Π_{ij} in equation (2.2), that has been discussed in the literature (by *e.g.* Agresti, 1984) is the *continuation-ratio* probability

$$\Pi_{ij} \quad = \quad \mathrm{P}(Y_i = j | Y_i \geq j) \quad = \quad \frac{\pi_{ij}}{\pi_{ij} + \cdots + \pi_{iK}} \qquad (j = 1, \ldots, K).$$

Any appropriate link function may be used to form a model with the continuation-ratio probability. For example, what will be referred to as the continuation-ratio logit model:

$$\operatorname{logit}\left(\frac{\pi_{ij}}{\pi_{ij} + \cdots + \pi_{iK}}\right) = \theta_j - \mathbf{x}_i \beta$$

or equivalently

$$\log\left(\frac{\pi_{ij}}{1 - \gamma_{ij}}\right) = \theta_j - \mathbf{x}_i \beta.$$

Similarly there are the continuation-ratio complementary log-log and continuation-ratio probit models when the link function is the complementary log-log and probit respectively. The continuation-ratio link models do not have the same appeal to an underlying continuum as is the case with the cumulative link models. Note that the threshold parameters (θ_js) in a continuation-ratio link model are not, in general, ordered. All continuation-ratio link models define strict stochastic ordering but they are not, in general, invariant to the collapsing of contiguous categories.

It is straightforward to use general purpose binary regression software to fit the three continuation-ratio link models mentioned here. This is possible because the likelihood function for continuation-ratio link models can be split into $K - 1$ independent binomial likelihood functions. To demonstrate, consider the case of an ordinal response having four categories ($K = 4$). The likelihood for the ith observation is proportional to

$$\pi_{i1}^{y_{i1}} \pi_{i2}^{y_{i2}} \pi_{i3}^{y_{i3}} \pi_{i4}^{y_{i4}}$$

where $y_{ij} = 1$ if the ordinal response $y_i = j$ and $y_{ij} = 0$ otherwise. This may equivalently be written as

$$\begin{aligned}
&\pi_{i1}^{y_{i1}} (\pi_{i2} + \pi_{i3} + \pi_{i4})^{y_{i2}+y_{i3}+y_{i4}} \\
&\times \left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3} + \pi_{i4}} \right)^{y_{i2}} \left(\frac{\pi_{i3} + \pi_{i4}}{\pi_{i2} + \pi_{i3} + \pi_{i4}} \right)^{y_{i3}+y_{i4}} \\
&\times \left(\frac{\pi_{i3}}{\pi_{i3} + \pi_{i4}} \right)^{y_{i3}} \left(\frac{\pi_{i4}}{\pi_{i3} + \pi_{i4}} \right)^{y_{i4}}
\end{aligned}$$

But, since $\sum_j \pi_{ij} = 1$, this is also the joint likelihood for one observation from each of the three independent binomial distributions,

$$\begin{aligned}
&\mathrm{B}(1, \pi_{i1}) \\
&\mathrm{B}\left(y_{i2} + y_{i3} + y_{i4}, \frac{\pi_{i2}}{\pi_{i2} + \pi_{i3} + \pi_{i4}} \right) \\
&\mathrm{B}\left(y_{i3} + y_{i4}, \frac{\pi_{i3}}{\pi_{i3} + \pi_{i4}} \right)
\end{aligned}$$

A continuation-ratio link model is in this case

$$\begin{aligned}
\operatorname{link}(\pi_{i1}) &= \theta_1 - \mathbf{x}_i \beta \\
\operatorname{link}\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3} + \pi_{i4}} \right) &= \theta_2 - \mathbf{x}_i \beta \\
\operatorname{link}\left(\frac{\pi_{i3}}{\pi_{i3} + \pi_{i4}} \right) &= \theta_3 - \mathbf{x}_i \beta
\end{aligned}$$

So if the continuation-ratio link model is considered to have three levels then each level models one of the binomial probabilities in the expanded likelihood. This means that any software package that includes the facility for fitting logistic regression models to binomial data can be used to fit continuation-ratio link models. This is a practical advantage to using these models.

Adjacent-category probability

A third probability, to be used in place of Π_{ij} in equation (2.2), that has been given attention in the literature is the *adjacent-category* probability

$$\Pi_{ij} = \frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)}}.$$

This is the probability of an observation being in category j given that it is in category j or $j+1$. The adjacent-category link models imply stochastic ordering. One particular case, the logit version, is equivalent to a log-linear model for expected cell frequencies that employs integer scores for the ordered categories of the response — see Agresti (1984) §7.

2.2.3 Covariate effects

The general framework introduced in equation (2.2) allows for covariate effects in the linear form $\theta_j - \mathbf{x}_i\beta$. The parameters β will be referred to as location effect parameters. These are fixed effects. When dealing with cumulative link models these location effects can be interpreted as changes in the location of the underlying response distribution on the underlying continuum.

It also makes sense to consider changes in scale of the underlying response distribution. With this in mind McCullagh (1980) defines a general cumulative link model

$$\text{link}(\gamma_{ij}) = (\theta_j - \mathbf{x}_i\beta)/\exp(\mathbf{z}_i\tau) \tag{2.4}$$

where the parameters τ can be interpreted as changes in the scale or dispersion of the underlying distribution. In general, this model no longer defines a mode of strict stochastic ordering. Chapter 5 includes an analysis using scaling terms applied to the BT data described in Chapter 3. In Chapter 4 a new general purpose method of fitting this model (and cumulative link models with location terms only) is introduced. This general purpose method is implemented in GLIM4 and is published in the GLIM4 Macro Library. As scaling terms are heavily investigated in this thesis the next section will discuss the issue of scaling terms in a cumulative logit model in more detail.

Another covariate structure, more flexible than the scaling structure, is one that includes a full interaction between covariate effects and the threshold parameters.

In the model framework this is represented by,

$$\text{link}(\Pi_{ij}) = \theta_{ij} - \mathbf{x}_i\beta \tag{2.5}$$

where now the $\theta_{ij} = \theta_j(\mathbf{z}_i)$ are functions of what are termed 'category-specific' variables $\mathbf{z}_i$. Tutz (1991) discusses the covariate effects described in equation (2.5) for both cumulative models and continuation-ratio models. Examples of the use of models of this form are given in Chapter 3 and Chapter 5.

If the ordinal response data under consideration are in the form of repeated measures on the same sampling units then the inclusion of random effects in the model may be appropriate. The repeated measures may be clustered observations, *e.g* responses on different children where family is the sampling unit or responses from different eyes where the sampling units are individuals. The repeated measures may be longitudinal observations where responses are observed at different time points on the same sampling unit.

Hedeker & Gibbons (1994) have developed an appropriate methodology for the inclusion of random effects in cumulative link models. They propose a model for the continuous response underlying the repeated ordinal response as follows:

$$Y_{is}^* = \mathbf{w}_{is}\alpha_i + \mathbf{x}_{is}\beta + \epsilon_{is} \tag{2.6}$$

where there are $s = 1, \ldots, n_i$ observations on $i = 1, \ldots, n$ sampling units. The vector $\mathbf{w}_{is}$ is a design vector for the random effects α_i. The covariate design vector is given by $\mathbf{x}_{is}$ and the regression parameters are contained in β. Hedeker & Gibbons (1994) assume that the distribution of the random effects is multivariate normal, independent of the errors ϵ_{is} which are assumed to be independent and normally distributed. This implies the cumulative probit model. Assuming the logistic distribution for the errors would give the cumulative logit model.

The estimation of the parameters in the model may be performed by maximum likelihood but this requires the evaluation of an integral as the random effect distribution needs to be integrated out of the likelihood. This integration is performed by Gaussian quadrature in the software developed by Hedeker (1993). All models in the framework (2.2) have the same form of likelihood function, *i.e.* multinomial, and so the consideration of a random effect is similar for all such models. An alternative to numerical integration is to use a Gibbs sampling approach.

Jansen (1990) gives an example of the use of random effects in clustered ordinal data. The application that he discusses is an agricultural experiment where an ordinal response is observed for each strawberry plant within a cluster, the cluster being defined by the plot in which the strawberry plant was grown. A random effect for plots is included in the cumulative logit model and a normal distribution is assumed as the prior for this random effect. For parameter estimation Jansen (1990) uses Gaussian quadrature to perform the integration in the likelihood function. A similar application is that of Ezzet & Whitehead (1991) who analyse longitudinal data from a cross-over trial using a random effects model. There are two responses measured on each subject in the trial and a subject-specific random effect is considered in the same manner as the plot random effect in Jansen (1990). Ezzet & Whitehead (1991) integrate the likelihood by using numerical integration routines in FORTRAN.

Ten Have & Uttal (1994) deal with an example from a clinical psychology experiment where repeated ordinal observations are made on children. They use the continuation-ratio logit model and include random effects for children. To evaluate the resulting integral in the likelihood function Ten Have & Uttal (1994) employ the Gibbs sampler.

A special case, the cumulative complementary log-log model, is investigated by Crouchley (1995). He discusses this model with a distribution for the random effects from the Hougaard family (Hougaard, 1986). In this particular situation Crouchley (1995) shows how a closed form is obtained for the likelihood function and recourse to numerical integration or Gaussian quadrature is unnecessary.

2.3 The cumulative logit model with a scaling term

2.3.1 Introduction

The cumulative logit model with a scaling term τ proposed by McCullagh (1980) is

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = (\theta_j - \mathbf{x}_i\beta)/\exp(\mathbf{z}_i\tau). \tag{2.7}$$

As mentioned in the previous section the scaling term describes changes in the variance or scale (or dispersion) of the underlying response distribution.

A simple example of the cumulative logit model with a scaling term is now given. Imagine a set of data which can be divided into 2 groups according to the values of a binary characteristic. Consider

$$\left.\begin{aligned} &\operatorname{logit}(\gamma_{1j}) = (\theta_j - \beta_1)/\exp(\tau_1) \\ &\operatorname{logit}(\gamma_{2j}) = (\theta_j - \beta_2)/\exp(\tau_2) \\ &\beta_1 = 0 \qquad \tau_1 = 0 \end{aligned}\right\}, \tag{2.8}$$

a cumulative logit model with a scaling term. The binary explanatory variable has both a location and scale parameter associated with it. The responses are split into 2 groups according to the value of the binary covariate. The responses in these 2 groups are assumed to be observations from 2 underlying probability distributions. The difference between the underlying distributions is described via the cumulative logit model in two ways:

1. by the difference in their location, described by β_2
2. and the difference in their dispersion, described by τ_2.

The example given by model (2.8) demonstrates a very simple covariate structure in a cumulative logit model with a scaling term. With more complex covariate structures a more complicated model can be formulated, for example

$$\log\left(\frac{\gamma_{ghij}}{1-\gamma_{ghij}}\right) = (\theta_j - \mathbf{w}_g\alpha - \mathbf{x}_i\beta)/\exp(\mathbf{x}_i\tau + \mathbf{z}_h\upsilon). \tag{2.9}$$

In this model there are three distinct ways for a covariate to describe the underlying distribution:

1. Through a location parameter only, *e.g.* covariates $\mathbf{w}_g$ in model (2.9).

2. Through a scale parameter only, *e.g.* covariates $\mathbf{z}_h$ in model (2.9).

3. Via both location and scale parameters, *e.g.* covariates $\mathbf{x}_i$ in model (2.9).

McCullagh (1980) points out that the model with a scaling term no longer implies stochastic ordering. Where stochastic ordering holds, either $\gamma_{1j} > \gamma_{2j}$ for all j, or $\gamma_{2j} > \gamma_{1j}$ for all j. The model with a scaling term (2.8) can be interpreted as describing underlying distributions which are not, in general, stochastically ordered. However, given values of the covariates of type 2 and 3 above it can be said that the underlying response distributions are stochastically ordered regardless of covariates of type 1.

An alternative form of the cumulative logit model with a scaling term is

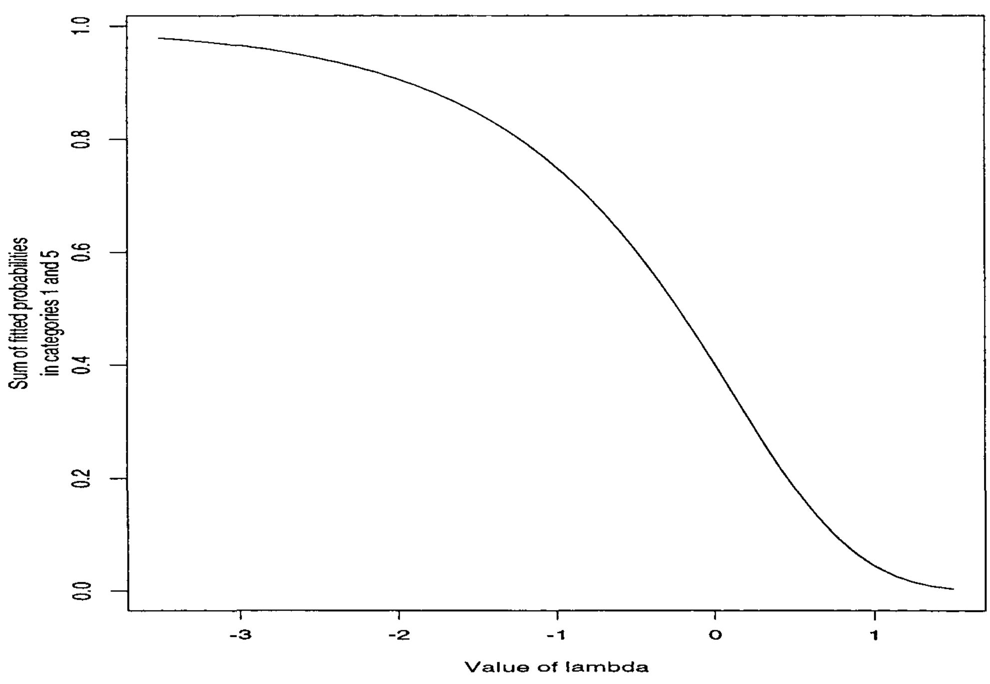
$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j \exp(\mathbf{z}_i\lambda) - \mathbf{x}_i\beta. \tag{2.10}$$

which is discussed by Kijewski et al (1989). The advantage of this parameterisation is that the scaling parameters λ can be interpreted more directly in terms of fitted category probabilities. The $\exp(\mathbf{z}_i\lambda)$ term has a multiplicative effect on the cut-point parameters θ_j. Consider a simple example of model (2.10) with one binary covariate:

$$\left.\begin{array}{l} \log\left(\frac{\gamma_{1j}}{1-\gamma_{1j}}\right) = \theta_j \exp(\lambda_1) \\ \log\left(\frac{\gamma_{2j}}{1-\gamma_{2j}}\right) = \theta_j \exp(\lambda_2) \\ \lambda_1 = 0 \end{array}\right\}. \tag{2.11}$$

The value of λ_2 has the effect of either clustering the cut-points on the underlying continuum (if $\lambda_2 < 0$) or spreading them out (if $\lambda_2 > 0$). The estimates of the cut-point parameters give fitted probabilities for level 1 of the covariate. If the cut-points are more spread out in level 2 of the covariate (by a value of $\lambda_2 > 0$) then the fitted probabilities in the extreme categories are smaller for level 2 than for level 1 of the covariate. Similarly if the cut-points are more clustered in level 2 of the covariate (by a value of $\lambda_2 < 0$) then the fitted probabilities in the extreme categories are larger for level 2 than for level 1 of the covariate.

Figure 2.1: Effect of scaling term (λ) on the fitted probabilities in the extreme categories of a 5 category response



The theme of interpreting a scaling term, λ, with respect to fitted probabilities can be continued further. Consider again the situation of a cumulative logit model with one binary covariate but now suppose that the ordinal response has 5 categories. Assume that the category probabilities for level 1 are known and equal ($\pi_{1j} = 0.2$

Table 2.1: Quality of right eye vision in men and women

	Vision quality				
	Highest	2	3	Lowest	Total
Men	1053	782	893	514	3242
Women	1976	2256	2456	789	7477

Cumulative proportions

Men	0.32	0.57	0.84	1.00
Women	0.26	0.57	0.89	1.00

for all j). Figure 2.1 illustrates how the extreme-category probabilities for level 2 vary with the value of the scaling term, λ. From Figure 2.1 it can be seen how increasing values of the scaling term result in a shift of the total probability from π_1 and π_5 to π_2, π_3 and π_4. Also note that when $\lambda = 0$ the fitted probabilities for level 2 equal the probabilities for level 1 and thus $\hat{\pi}_{21} + \hat{\pi}_{25} = 0.2 + 0.2 = 0.4$.

2.3.2 When is a model with a scaling term appropriate?

In this chapter the cumulative logit model is introduced in terms of an underlying continuum. The location and scaling terms are interpreted with respect to underlying probability density functions defined on this continuum. In these terms the model with a scaling term is appropriate when it is possible that the underlying cumulative probability functions associated with some combination of explanatory variables are not stochastically ordered.

The data in Table 2.1 are analysed by McCullagh (1980). The explanatory variable in this example is sex, a binary variable. If the cumulative proportions given in Table 2.1 are observations from two underlying cumulative probability distributions

F_1 and F_2, then it would seem that F_1 and F_2 are not stochastically ordered. Hence it is appropriate to consider a model with a scaling term (*e.g.* model 2.7) in this example.

The quality-of-vision data in Table 2.1 is one example of a lack of stochastic ordering in real data. Another example is the data from the BT experiments described in Chapter 3. In the BT data the lack of stochastic ordering (which is highlighted in §5.2.1) is due to differences between subjects in their use of the ordinal scale.

2.4 Sampling

The likelihood function that is used in estimation procedures for fitting the models in the framework (2.2) is the multinomial likelihood function. For a K-category ordinal response this is proportional to

$$L = \prod_{ij} \pi_{ij}^{y_{ij}}$$

where y_{ij} is the observation in the ijth cell. The n multinomial observations $y_i = (y_{i1}, y_{i2}, \ldots, y_{iK})$, where $\sum_j y_{ij} = m_i$, have expectation

$$E(Y_i) = (m_i\pi_{i1}, m_i\pi_{i2}, \ldots, m_i\pi_{iK}).$$

It is assumed, conditional on all effects in the model, that the multinomial variables Y_i are independent of each other and that the multinomial totals m_i are fixed.

If the multinomial observations y_i are made on a randomly-selected subset of a large population then it may be safe to assume that the variables Y_i are independent. Care must be taken however as it is not hard to imagine situations where the sampling procedure yields a set of observations which come from correlated Y_i variables.

The case of repeated sampling, *i.e.* several multinomial observations taken on the same sampling unit, gives rise to longitudinal data. It is likely that there will be correlation between the repeated observations on the same unit. In Chapter 4 a method of fitting the cumulative logit model using generalized estimating equations (GEE) is discussed. This fitting method involves a 'working' correlation matrix being formed. In the case of longitudinal data this working correlation matrix may be defined to reflect possible correlation between the intra-unit repeated observations. Whilst this thesis examines in detail the use of GEE as a method of fitting the cumulative logit model, the extension of GEE to longitudinal ordinal responses is not pursued. Recent work has been done in this direction by Clayton (1992), Lipsitz et al (1994) and Miller et al (1993).

2.5 Model relationships

2.5.1 The equivalence of the cumulative complementary-log-log and continuation-ratio complementary-log-log models

This relationship has long been recognised, for example see the comments of Pregibon in the discussion of McCullagh (1980). The relationship is formally specified by Läärä & Matthews (1985). If the cumulative complementary log-log model is defined as

$$\log(-\log(1-\gamma_{ij})) = \theta_j - \mathbf{x}_i\beta$$

and the continuation-ratio complementary log-log model is defined as

$$\log(-\log(1-\delta_{ij})) = \psi_j - \mathbf{x}_i\alpha$$

where $\delta_{ij} = \mathrm{P}(Y_i = j | Y_i \geq j)$ is the continuation-ratio probability, then the parameters α and β are equivalent and the threshold parameters θ_j and ψ_j of the two models are related by

$$\theta_1 = \psi_1 \qquad \theta_j = \log\left(e^{\psi_j} + e^{\psi_{j-1}}\right) \quad (j = 2, \ldots, K-1).$$

2.5.2 Approximate equivalence of cumulative probit and adjacent-category probit models

Further exploration of relationships like that in the previous section, between models from the framework in equation (2.2), led to the following: when the number of categories is large, a cumulative probit model with equally-spaced cutpoints is approximately equivalent to the corresponding adjacent-category probit model. This finding appears to be new and a heuristic derivation is now presented.

Consider the cumulative probit model

$$\Phi^{-1}(\gamma_{ij}) = \theta_j - \mathbf{x}_i\beta \tag{2.12}$$

with equally-spaced cutpoints, $\theta_j - \theta_{j-1} = h$ $(j = 2, \ldots, K-1)$. The 'corresponding' adjacent-category probit model is

$$\Phi^{-1}\left(\frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)}}\right) = \theta_j^* - \mathbf{x}_i\beta^* \qquad (j = 1, \ldots, K-1). \tag{2.13}$$

To see that model (2.13) is approximately equivalent to model (2.12) when K is large (and hence h small), note that for $2 \leq j \leq K-2$,

$$\begin{aligned} \phi(\theta_{j+1} - \mathbf{x}_i\beta) &= \phi(\theta_j + h - \mathbf{x}_i\beta) \\ &= \phi(\theta_j - \mathbf{x}_i\beta) + h\phi'(\theta_j - \mathbf{x}_i\beta) + O(h^2) \end{aligned}$$

by the Taylor expansion, where $\phi(z)$ denotes the normal density. Note also that for $2 \leq j \leq K-2$

$$\pi_{ij} = h\phi(\theta_j - \mathbf{x}_i\beta) - \frac{1}{2}h^2\phi'(\theta_j - \mathbf{x}_i\beta) + O(h^3),$$

and hence

$$\pi_{i(j+1)} = h\phi(\theta_j - \mathbf{x}_i\beta) + \frac{1}{2}h^2\phi'(\theta_j - \mathbf{x}_i\beta) + O(h^3).$$

Thus

$$\begin{aligned} \frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)}} &= \frac{h\phi(\theta_j - \mathbf{x}_i\beta) - \frac{1}{2}h^2\phi'(\theta_j - \mathbf{x}_i\beta) + O(h^3)}{2h\phi(\theta_j - \mathbf{x}_i\beta) + O(h^3)} \\ &= \frac{1}{2} - \frac{h}{4}\frac{\phi'(\theta_j - \mathbf{x}_i\beta)}{\phi(\theta_j - \mathbf{x}_i\beta)} + O(h^2). \end{aligned}$$

But $\phi'(z)/\phi(z) = -z$, so

$$\frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)}} = \frac{1}{2} + \frac{h}{4}(\theta_j - \mathbf{x}_i\beta) + O(h^2)$$

and thus, using the Taylor expansion on the right hand side,

$$\begin{aligned} \Phi^{-1}\left(\frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)}}\right) &= \Phi^{-1}(1/2) + \frac{h}{4}(\theta_j - \mathbf{x}_i\beta)\Phi^{-1'}(1/2) + O(h^2) \\ &= \frac{h\sqrt{2\pi}}{4}(\theta_j - \mathbf{x}_i\beta) + O(h^2). \end{aligned}$$

Thus model (2.13) is approximately equivalent to model (2.12) for $j = 2, \ldots, K-2$. Provided that the end categories can be neglected, which is not unreasonable when the number of categories is large, the cumulative probit model is approximately equivalent to the adjacent-category probit model. Note that the same is not true for other link functions: the key property of the probit in this regard is that $\phi'(z)/\phi(z)$ is linear in z.

No application for this result is immediately apparent. It has no consequences for the remainder of the thesis, and so is not pursued further.

2.5.3 The relationship between continuation-ratio link models and Cox's proportional hazards for survival data

The hazard function

$$\mathrm{h}(t; \mathbf{x}_i) = \lim_{dt \to 0} \frac{\mathrm{P}(t \le T_i \le t + dt | T_i > t)}{dt}$$

is used to define the well-known proportional hazards model for survival data in continuous time

$$\mathrm{h}(t; \mathbf{x}_i) = \psi(\mathbf{x}_i\beta)\mathrm{h}_0(t)$$

for some functions ψ and $\mathrm{h}_0(t)$. Consider the exponential function as a form for ψ and

$$H_0(t) = \int_0^t \mathrm{h}_0(t)\, dt.$$

Then the proportional hazards model is equivalent to the continuation-ratio complementary log-log model

$$\log(-\log(1-\delta_{ij})) = \theta_j - \mathbf{x}_i\beta$$

with

$$\theta_j = \log\left[H_0(t_j) - H_0(t_{j-1})\right].$$

An alternative version of the proportional hazards model for discrete time (Cox 1972) is

$$\frac{\mathrm{h}(t;\mathbf{x}_i)}{1-\mathrm{h}(t;\mathbf{x}_i)} = \psi(\mathbf{x}_i\beta)\frac{\mathrm{h}_0(t)}{1-\mathrm{h}_0(t)}$$

where now $\mathrm{h}(t;\mathbf{x}_i) = \mathrm{P}(T_i \leq t+1|T_i > t)$. The relational function ψ recommended by Cox & Oakes (1984) §7 is $\psi(\mathbf{x}_i\beta) = \exp(\mathbf{x}_i\beta)$. In this case the model becomes a logistic model for the hazard function in discrete time. It is then equivalent to the continuation-ratio logit model

$$\log\left(\frac{\delta_{ij}}{1-\delta_{ij}}\right) = \theta_j - \mathbf{x}_i\beta$$

with

$$\theta_j = \log\left(\frac{\mathrm{h}_0(t_j)}{1-\mathrm{h}_0(t_j)}\right).$$

2.5.4 Note on terminology

Having established the close relationship of both the continuation-ratio logit and continuation-ratio complementary log-log models to the proportional hazards model for survival data it becomes misleading to refer to the cumulative complementary log-log as the proportional hazards model.

Following the argument of McCullagh (1980) motivating the use of the name, it seems that a more appropriate name for him to have arrived at would have been the proportional log survivors model. This can be seen from the relationship between the continuation-ratio complementary log-log model and the proportional hazards model for survival data and the equivalence between the former and the cumulative complementary log-log model shown in §2.5.1. To make the distinction clear this is now shown explicitly for the cumulative complementary log-log model. In discrete time the survivor function $\mathrm{S}(t; \mathbf{x}_i) = \mathrm{P}(T_i > t)$ becomes $(1 - \gamma_{it})$. For two groups $(i = 1, 2)$ the cumulative complementary log-log model is specified as

$$\log(-\log(1 - \gamma_{1j})) = \theta_j - \mathbf{x}_1\beta \tag{2.14}$$

$$\log(-\log(1 - \gamma_{2j})) = \theta_j - \mathbf{x}_2\beta \tag{2.15}$$

and (equation 2.14) − (equation 2.15) gives

$$\frac{\log(1 - \gamma_{1j})}{\log(1 - \gamma_{2j})} = \exp(\mathbf{x}_2 - \mathbf{x}_1).$$

So the cumulative complementary log-log model defines proportional log survivor functions in the case of discrete time. McCullagh used the name proportional hazards because proportional log survivor functions are also defined (in the continuous case) by the proportional hazards model.

Given the different application of models to survival data and to ordinal response data, it seems appropriate that different names should be used for models in the two arenas. The use of the name cumulative complementary log-log model will prevent confusion arising firstly from its equivalence to the continuation-ratio complementary log-log model and secondly from the connections of both it and the continuation-ratio logit model to models for survival data.

It is also worth noting the motivation of the name 'proportional odds' model. The cumulative logit model is a proportional odds model for the event $Y \leq j$. However it is also true that the continuation-ratio logit model is a proportional odds model for the event $Y = j | Y \geq j$. Thus the name proportional odds model may equally well be applied to either. As a means of reference and as a means of description the name proportional odds is unhelpful. The name cumulative logit is descriptive of the model and, within the terminology framework suggested in this chapter, it is an appropriate reference.

2.6 Special scales

2.6.1 Ordinal scales, known cut-points

Terza (1985) gives an account of this special kind of response scale when the data are analysed by the cumulative probit model. Ordinal data often arises from grouping continuous data. Usually when this happens the cut-points of the continuous scale used to define the ordered groups are known, *e.g.* income being grouped into $< \$10,000$, $\$10,000 - \$20,000$ and $> \$20,000$. Terza demonstrates how this information may be used in a model of the form given in equation (2.5).

2.6.2 Partially ordered response scales

Consider a response scale, the categories of which can be grouped into two separate groups. For example, consider a 5-category scale separated into a group containing the ordered categories 1 to 3 and a second group containing the categories 4 and 5. Methods for dealing with this form of response scale have been considered by Tutz (1989) and Wang (1986). Wang introduces a parameterization of the multinomial likelihood equation which allows for partial ordering, with nominal and ordinal scales as extreme cases. He re-defines the likelihood in terms of $\lambda = (A')^{-1} \log \pi$. The matrix A is called the relational matrix and it contains all the ordering information for the categories. Wang recommends models based on the parameters λ. In the extreme case of an ordinal scale these parameters are given by $\lambda_{ij} = -\log(\pi_{i\,j-1} / \pi_{ij})$, *i.e.* the adjacent-category logits.

The approach of Tutz is different. He constructs a multiple-level ordinal response model that he refers to as a compound model. The motivation behind this can be illustrated by considering the example mentioned above. Consider the category allocation as a binary choice at the first level between the first three categories and the other two categories. If the allocation is to the former then there is an ordinal structure to model within that group. If allocation is to the latter then there is a further binary structure to model within that group. The modelling procedure at the second level is conditional on the first level.

In the example mentioned above the two approaches yield models based on the following transformations of the category probabilities π_j. Wang's method is to model $\log(\boldsymbol{\Lambda})$ where

$$\boldsymbol{\Lambda} = \left(\pi_1, \frac{\pi_2}{\pi_1}, \frac{\pi_3}{\pi_2}, \pi_4, \frac{\pi_5}{\pi_4}\right)'$$

but with the constraint $\sum \pi_j = 1$ one of these is redundant. Tutz's method, employing continuation-ratio logit models at each of the two levels of the response scale, models $\log(\boldsymbol{\Lambda}^*)$ where,

$$\boldsymbol{\Lambda}^* = \left(\frac{\pi_1 + \pi_2 + \pi_3}{\pi_4 + \pi_5}, \frac{\pi_1}{\pi_2 + \pi_3}, \frac{\pi_2}{\pi_3}, \frac{\pi_4}{\pi_5}\right)'.$$

2.7 Alternative model approaches

There is a brief discussion in the introduction to this chapter of the modelling approaches of assigning arbitrary scores to categories and analysing the resultant data via a log-linear model or an analysis of variance. The extension of log-linear modelling to ordinal data has been widely investigated and is well documented by Agresti (1984 §5). This method is useful when investigating association between variables (if no response variable can be identified).

As mentioned before, the use of any score-based model carries with it the arbitrariness of the scoring mechanism but, as Agresti (1984 §8) points out, interpretation may be easier in such a model than in a multinomial model. A generalization of the score-based modelling approaches is to make the score an estimable parameter of the model (Agresti 1984 §8). Thus the score giving the best fit is chosen. While this may lead to a set of scores which give a good fit to the data, interpretation is not necessarily any easier than in a model with a simple (*e.g.* integer) set of scores.

One final alternative approach which has not yet been mentioned is the *stereotype* model suggested by Anderson (1984). Consider the logistic regression model for categorical data ,

$$\mathrm{P}(Y_i = j) = \frac{\exp(\beta_{0j} - \mathbf{x}_i\beta_j)}{\sum_{t=1}^{K} \exp(\beta_{0t} - \mathbf{x}_i\beta_t)}.$$

Anderson (1984) proposed a modification to this formulation which replaces $\mathbf{x}_i\beta_j$ with $\phi_j\mathbf{x}_i\beta$ (*i.e.* taking the β_j to be parallel). Considering the ϕ_j parameters to be monotone decreasing leads to the stereotype model:

$$\mathrm{P}(Y_i = j) = \frac{\exp(\beta_{0j} - \phi_j\mathbf{x}_i\beta)}{\sum_{t=1}^{K} \exp(\beta_{0t} - \phi_t\mathbf{x}_i\beta)}.$$

The fitting of this model does not involve prior constraint of the ϕ_j parameters, Anderson regards *a posteriori* investigation of the ordering of these parameters as the key to determining the ordinality of the response scale. A concept which is used in this context is indistinguishability. If an explanatory variable is not predictive between two contiguous categories then those categories are said to be indistinguishable. This could be seen from the parameter estimates if two categories had equal values of ϕ_j. Anderson's advice is to then combine these categories and repeat the analysis.

Chapter 3

Introduction to the British Telecom data

3.1 Introduction

When we use the telephone we frequently experience 'good lines' and 'bad lines'. This is due to a whole range of factors such as the type of telephone we are using, the distance between these telephones, the type of electrical connection used to relay the conversation, the amount of background noise where we are making the call, and our own previous experience of telephones. Most of these factors are quantitative, *e.g.* electrical features such as signal loss from one end of a connection to the other.

British Telecom (BT) is involved in ongoing experimentation into all aspects of telephony. Their experiments are run at British Telecom Laboratories (BTL) in Martlesham. One group at BTL is concerned with measuring the speech-transmission performance of complete telephone connections, as perceived by the human users. It does this by eliciting the opinion of people on different telephone connections. Much of the work in this thesis arose through a research contract on this topic between BT and Southampton University.

Table 3.1: BT Experiments

Conversation	
Limited duration:	E198, E200, E212, E216.
Unlimited duration:	E136, E139, E140, E199, E211, E213.
Listening	
2 subjects per row:	E247, E264.
1 subject per row:	F064, F065.
Special design:	F100.

The data from a number of BT experiments are analysed in this thesis. A full list of the experiments is contained in Table 3.1 from which it can be seen that the experiments fall into two distinct types: conversation experiments and listening experiments. Conversation experiments involve pairs of subjects conversing on the telephone, whereas listening experiments involve individual subjects listening to pre-recorded speech over the telephone.

In both types of experiment the aim is to obtain subject responses in order to compare different transmission conditions. The transmission conditions are combinations of settings of particular electrical features of telephone connections such as loudness, signal loss, artificial interference and echo, which is the name given to the delayed audible return of one's own voice from the distant end of the connection. The method of eliciting peoples' opinion has developed over the many years that BT have been investigating the performance of different transmission conditions. In the data from BT experiments analysed in this thesis, opinion has been recorded on a 5-category ordered scale (see Figures 3.1 and 3.2 for example).

In this chapter a preliminary description of each type of BT experiment: conversation and listening, is given. Terminology for experimental design is introduced and the types of design used for the BT experiments are outlined. It is possible that a subject's response is influenced by exposure to earlier periods in the experiment. For example if the condition in the preceding period was bad then a subject may respond favourably in the current period to a moderate transmission condition. If the condition in the preceding period was very good then the subject may respond unfavourably to a moderate condition in the current period. The effect of features of preceding periods of an experiment on the response in the current period are called carry-over effects, and these are discussed in this chapter. Two separate modelling issues are introduced, both motivated by data from the BT experiments. The first issue is how the goodness of fit of the linear model underlying analysis of variance (ANOVA) compares with the continuation-ratio logit model. Both of these models were introduced in Chapter 2. Secondly a new method of analysing carry-over effects using the continuation-ratio logit model is introduced. Since the motivation for this method of carry-over analysis comes from the BT data, application of the method to these data is presented.

3.1.1 Listening experiments

The typical procedure for a listening experiment can be described as follows. After receiving initial instructions a subject sits in a soundproofed cabinet with a telephone set and the instructions for the experimental procedure on a table in front of them. During the experiment, an experimental controller (located in a control room) instructs the subject via a loud-speaker. The telephone rings, the subject picks it up and listens as some sentences are read out. When these have finished a signal is given for the subject to respond with their opinion of the transmission condition. Typically this is done by choosing a category from a five-point scale, graded descriptively from 'No meaning understood with any feasible effort' to 'Complete

Figure 3.1: Typical response scale for listening experiments.

Effort required to understand the meaning of sentences
4. Complete relaxation possible; no effort required
3. Attention necessary; no appreciable effort required
2. Moderate effort required
1. Considerable effort required
0. No meaning understood with any feasible effort

relaxation possible; no effort required' (see Figure 3.1). The subject listens to a variety of sentences spoken by different voices under different transmission conditions, and at every prompt records an opinion.

The order in which a subject hears the conditions is determined by an experimental design. This design can be set out as a two-way layout in which each row corresponds to a subject and each column to a period. There is a well-developed literature on the design of experiments, *e.g.* Montgomery (1984), and on row-column layouts in particular, *e.g.* Jones & Kenward (1989) and Street & Street (1987). At each intersection of a row and a column any number of attributes (or factors) may be allocated. The allocation employed affects the usefulness of the information gained from the experiment.

In the BT listening experiments facilities are usually available to perform the experiment in two different cabinets simultaneously. In this case there are two subjects corresponding to each row of the design. For a given row-column intersection, values of three factors (transmission condition, sentence list and talker) are specified. The design is complicated by the fact that each period contains a number of sub-periods (typically 5). The transmission condition, sentence list and talker are all allocated to the period and are thus the same for each sub-period within that period. However, the listening level is different in each sub-period. The listening level affects the loudness at which the subject hears the talker over the telephone. The number of sets of sentences in the sentence list equals the number of listening levels (which is also equal to the number of sub-periods) and in each sub-period a different set of sentences from the list is used. Although the listening level is treated as a factor at the experimental design stage, the different levels correspond to signal level measurements and hence listening level may be treated as either a continuous variable or a factor at the analysis stage.

An example of the values of the factors and the data obtained from one period of a listening experiment is

Col-umn	Row	Cond-ition	Voice	Sentence list	Listening level	Opinion score Cabinet 1	Opinion score Cabinet 2
					4	3	2
					2	1	2
3	10	2	1	8	1	0	0
					5	3	4
					3	2	2

Note that the column, row, condition, voice and sentence list apply to the entire period and hence apply to each sub-period. In this example there are 5 sub-periods. There are also two subjects in the row (one in cabinet 1 and the other in cabinet 2). The opinion responses are made using the 5-point scale and numerically coded as in Figure 3.1.

3.1.2 Conversation experiments

There are three main differences between a conversation and a listening experiment.

1. A conversation experiment consists of a number of pairs of subjects and each pair engages in conversation over the telephone; this is in contrast to a listening experiment where single subjects listen to pre-recorded speech. The two subjects in each pair sit in different cabinets which will be referred to as ends, implying the two ends of the telephone connection. As in the listening experiments, each subject sits at a table on which there are a telephone set and a sheet of instructions describing the experimental procedure. One of the instructions given to the subjects is to avoid discussing their opinion of the transmission condition. This is an attempt to remove one possible violation of independence of the opinion scores of the two subjects. Similarly to listening experiments, there is an experimental controller who instructs the subjects on procedure.

2. Another factor is introduced in order to stimulate conversation but which is not of primary interest in the analysis of the experiments. This stimulus takes the form of a set of picture-cards (a picture-set) for which subjects are asked to complete a task. For example, the pair of subjects will have 8 cards each of which only 6 are common to both. The subjects have to discover which cards they have in common and agree on an order of preference for them. It is important that subjects receive fresh stimuli in each period regardless of the task being performed. This helps to avoid fatigue in the subjects.

3. An important practical difference between listening and conversation experiments is the amount of time it takes to arrive at the subjects' opinion response. In a listening experiment a subject only has to listen to a few sentences before recording an opinion. In a conversation experiment a subject engages in conversation and then gives an opinion. In some conversation experiments the duration of this conversation is determined by the subjects. They talk for as long as is necessary to complete the task that they have been set. In other

conversation experiments the duration of the conversation is constrained (*e.g.* to 3 minutes). In the time-constrained experiments, the subjects are usually asked to select and discuss the picture-sets rather than perform a particular task with them. The conversation is halted by the experimental controller at the appropriate time. Time-constrained experiments are open to the criticism that they are not representative of real telephone conversations. They arise from practical concerns about the potential length of time an experiment will take if the subjects are free to converse for as long as they wish. Conversation experiments where the duration of the conversation is at the subjects' discretion can occasionally result in extremely long conversations.

Figure 3.2: Typical response scale for conversation experiments.

Opinion of the telephone connection
4. Excellent
3. Good
2. Fair
1. Poor
0. Bad

One period in a conversation experiment is as follows. One subject rings the other and they hold a conversation either to accomplish the task involving the picture-set or to converse for the specified time. When the conversation is finished the subjects hang-up and are prompted by the experiment controller to give an opinion of the transmission condition. This opinion is typically given on a five point scale graded descriptively from 'Bad' to 'Excellent' (see Figure 3.2). The subjects also give a binary response to a question on difficulty in hearing over the connection. This completes one period of the experiment. Unknown to the subjects, various objective measurements on the speech signals are also made. Only the ordinal response is investigated in this thesis, so the other response variables will not be mentioned further.

The experimental design for conversation experiments has a row-column structure, but unlike listening experiments they have no nested sub-periods within the periods. In addition to the transmission conditions the picture-sets must be allocated to the row-column intersections.

3.1.3 Carry-over considerations

In any given period of an experiment a subject's response may be influenced by the various factors in the experimental procedure. If a subject's response is influenced by the value of a factor in the previous period then that subject is said to experience a carry-over effect due to that factor. For example in the BT conversation experiments, a subject's response might be determined not just by the picture-set and condition experienced in a period but also by the condition that was experienced in the previous period. It is possible to imagine some sort of comparative process going on in the subject's mind. Does the condition in the current period seem better or worse than the condition in the previous condition?

This concern over the possible effect on the response in the present period of features of the previous period is a common one in experimentation where subjects receive a sequence of treatments. One of the first authors on this subject was Williams (1949, 1950) who was concerned with applications in agricultural experiments. A complete introduction to experimentation in the presence of carry-over is given by Jones & Kenward (1989) where the emphasis is on medical applications. Other applications where carry-over effects cause concern include psychology and human-factors engineering.

The carry-over effects need not be limited to the immediately preceding period.

There may be carry-over effects from two periods ago or earlier. The terminology usually used is to call carry-over from the immediately preceding period, *first-order carry-over*, and carry-over from the last-but-one period, *second-order carry-over*, and so on.

3.2 Design of the BT experiments

The purpose of experimental designs in general is to allocate experimental factors, usually in such a way as to maximise the information extracted from the experiment. Not all of the factors are necessarily of direct interest, those that are not are called *nuisance* factors. Each factor has a number of levels, *i.e.* values (quantitative or qualitative) which the factor may take. Two factors are said to be *orthogonal* if every level of one of the factors occurs in the design the same number of times with every level of the other factor. A level of a factor may be *aliased* with certain levels of other factors. When this occurs it is impossible to estimate the effect of the level of the factor separately of those levels of other factors with which it is aliased (see McCullagh & Nelder, 1989 §3.5).

Figure 3.3: A 4×4 Latin square

	Column			
	1	2	3	4
Row				
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

A Latin square design is a row-column arrangement for 1 treatment factor in which every level of the treatment factor occurs exactly once in every row and once in every column. Thus treatments are orthogonal to both rows and columns. An example of a Latin square is given in Figure 3.3. In this example the rows can be taken to represent subjects and columns correspond to periods of the experiment. The letters A to D represent the 4 levels of the treatment factor. A Latin square design is optimal in the sense that when the responses from the experiment are analysed by a linear model, the variance of the treatment contrasts is minimised. This optimality only holds if there is no carry-over effect of treatment.

A Graeco-Latin square is a row-column design for two treatment factors. At every intersection of row and column the levels of two factors are specified. These levels are specified in such a way that every level of one treatment occurs with every level of the other treatment exactly once across the design. Thus the design gives orthogonality between the two treatment factors. Further there is pairwise orthogonality between the treatment factors, rows and columns. A Graeco-Latin square can be constructed by superimposing two orthogonal Latin squares.

It is possible to superimpose more than two mutually-orthogonal Latin squares. Thus row-column designs for more than two factors can be constructed. These are called Hyper-Graeco-Latin square designs and in any one of them there is pairwise orthogonality between the treatment factors, rows and columns.

A Williams Latin square design (Williams 1949) uses a particular form of Latin square and has been widely used for experiments where carry-over effects are anticipated. If the columns of the design denote periods, then the specification of treatment levels across columns gives the allocation order of the treatment for each row. A treatment is said to be *balanced* if every level of the treatment is preceded by every other level of that factor the same number of times in the design. The treatment factor in a Williams Latin square design is balanced across columns. In Figure 3.4, a Williams Latin square design for a treatment with 4 levels, it can be

Figure 3.4: A 4×4 Williams Latin square

	Column			
	1	2	3	4
Row				
1	A	B	C	D
2	B	D	A	C
3	C	A	D	B
4	D	C	B	A

seen that every level of treatment is followed by every other level exactly once in the design.

3.2.1 The design of conversation experiment E199

In experiment E199 there were 8 transmission conditions of interest. To ensure that each subject received each condition once, 8 periods were included in the experiment. This required the use of 8 different picture-sets. For this experiment, the carry-over of condition was a consideration at the design stage. The design is constructed as follows. An 8×8 Williams Latin square for conditions (rows 1 to 8 and columns 1 to 8 of the condition matrix in Table 3.2) is repeated but with the rows permutated in a specific order. This results in a rectangle of size 16 rows by 8 columns for the allocation of conditions. An 8×8 Latin square (rows 1 to 8 and columns 1 to 8 of the picture-set matrix in Table 3.2) is repeated without permutation of the rows to give a rectangle for the allocation of picture-sets. Note that the initial Latin square for picture-sets is simply constructed by starting with row 1 containing the integers from 1 to 8 inclusive and cycling this row by adding 1, modulo 8, to give the other 7 rows.

Table 3.2: Design for Experiment E199

Column	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
Row	Condition matrix									Picture-set matrix								
1	8	8	1	7	2	6	3	5	4	0	1	2	3	4	5	6	7	8
2	1	1	2	8	3	7	4	6	5	0	8	1	2	3	4	5	6	7
3	2	2	3	1	4	8	5	7	6	0	7	8	1	2	3	4	5	6
4	3	3	4	2	5	1	6	8	7	0	6	7	8	1	2	3	4	5
5	4	4	5	3	6	2	7	1	8	0	5	6	7	8	1	2	3	4
6	5	5	6	4	7	3	8	2	1	0	4	5	6	7	8	1	2	3
7	6	6	7	5	8	4	1	3	2	0	3	4	5	6	7	8	1	2
8	7	7	8	6	1	5	2	4	3	0	2	3	4	5	6	7	8	1
9	4	4	5	3	6	2	7	1	8	0	1	2	3	4	5	6	7	8
10	5	5	6	4	7	3	8	2	1	0	8	1	2	3	4	5	6	7
11	6	6	7	5	8	4	1	3	2	0	7	8	1	2	3	4	5	6
12	7	7	8	6	1	5	2	4	3	0	6	7	8	1	2	3	4	5
13	8	8	1	7	2	6	3	5	4	0	5	6	7	8	1	2	3	4
14	1	1	2	8	3	7	4	6	5	0	4	5	6	7	8	1	2	3
15	2	2	3	1	4	8	5	7	6	0	3	4	5	6	7	8	1	2
16	3	3	4	2	5	1	6	8	7	0	2	3	4	5	6	7	8	1

Table 3.2 shows the two rectangles that are superimposed to produce the Graeco-Latin square design for experiment E199. This design gives pairwise orthogonality between conditions, picture-sets, rows and columns. In fact each condition appears in each row once, in each column twice and with each picture-set twice. Similarly each picture-set appears in each row once and in each column twice. Conditions are balanced in the design because a Williams Latin square was initially used for them.

A warm-up period, labelled column 0 in the design, is included in the experiment to familiarise subjects with the experimental procedure. The data from this warm-up period are not included in the analysis of the experiment. A picture-set which is not used in the rest of the experiment (labelled picture-set 0) is given to all subjects in the warm-up period. The conditions allocated to column 1 are duplicated in the warm-up period (column 0). The use of a warm-up period means that there is a preceding condition associated with the responses in column 1. Thus for the analysis, every condition is preceded by every other condition, *and itself*, the same number of times. This property will be referred to as *complete balance.*

The design in columns 1 to 8 does not give orthogonality between picture-sets and condition carry-over; every picture-set occurs with 6 carry-over conditions twice and 2 carry-over conditions once. However, if the information on condition carry-over from the warm-up period is used in analysis then picture-sets and condition carry-over are orthogonal.

3.2.2 The design of listening experiment E247

In experiment E247 there are 12 conditions and hence 12 periods are used. Of these conditions, 3 are synthesised conditions. This means that synthetic speech is used as opposed to a human voice. In this case condition and voice are not completely separate factors. The position in the design of each synthesiser condition fixes the position of the corresponding synthesiser voice in the design. The other 9 conditions occur with speech from 3 different human male talkers. The design ensures that each human voice occurs in every row 3 times and in every column 3 times. Also every voice occurs with every one of the 9 non-synthesised conditions the same number of times.

Each row has a pair of subjects assigned to it, one subject in each of two cabinets. There are 12 different sentence lists, each list containing 5 sets of sentences. These lists are used with both synthesiser and other conditions. The design is constructed by superimposing two Latin squares and one modified Latin square. The Latin square for conditions is given in Table 3.3, the Latin square for sentence lists is given in Table 3.4 and finally the square for voices is given in Table 3.5. Note that in Table 3.5 the numbers 1,2,3 represent human voices and the numbers 10,11,12 represent synthetic speech. These last three numbers correspond directly to conditions 2,4,6 respectively in Table 3.3 and hence voice 10 is aliased with condition 2, *etc.*

The three squares are superimposed to produce the design but note that this is not a Hyper-Graeco-Latin square design because the squares for conditions and voices are not orthogonal. This is because of the restriction relating to synthesised speech. The square for sentence-lists is orthogonal to both of the other squares. This design gives pairwise orthogonality between conditions, sentence-lists, voices, rows and columns except for the pairing conditions with voices.

The design was obtained by performing an exhaustive search of 12×12 Latin squares listed by Fisher & Yates (1963) to find combinations such that if three symbols represented three synthetic voices and the other nine symbols represented three human talkers, each condition and each sentence list would still occur equally often with each other and with each talker.

There were 5 listening levels used in 5 sub-periods in each period of the experiment. One listening level was allocated to each sub-period at random. A different set of sentences from the sentence list was used in each sub-period.

Note that the design was augmented by a warm-up period, labelled column 0 in Tables 3.3, 3.4 and 3.5. A different condition from those used in the rest of the experiment was used in this warm-up period and labelled condition 0. Similarly sentence-list 0 and voice 0 were used in the warm-up period. The listening levels used in the warm-up period were the same as in the design in columns 1-12 and were allocated at random to the 5 sub-periods in the warm-up period.

Table 3.3: Design matrix for conditions in experiment E247.

Column	0	1	2	3	4	5	6	7	8	9	10	11	12
Row													
1	0	8	6	2	9	11	1	5	10	4	7	3	12
2	0	3	7	9	4	6	8	12	5	11	2	10	1
3	0	12	4	6	7	9	5	3	8	2	11	1	10
4	0	10	2	4	11	7	3	1	12	6	9	5	8
5	0	4	8	10	5	1	9	7	6	12	3	11	2
6	0	5	9	11	6	2	10	8	1	7	4	12	3
7	0	1	11	7	2	4	12	10	3	9	6	8	5
8	0	9	1	3	10	12	2	6	11	5	8	4	7
9	0	6	10	12	1	3	11	9	2	8	5	7	4
10	0	7	5	1	8	10	6	4	9	3	12	2	11
11	0	11	3	5	12	8	4	2	7	1	10	6	9
12	0	2	12	8	3	5	7	11	4	10	1	9	6

Table 3.4: Design matrix for sentence lists in experiment E247

Column	0	1	2	3	4	5	6	7	8	9	10	11	12
Row													
1	0	7	5	1	8	10	6	4	9	3	12	2	11
2	0	1	11	7	2	4	12	10	3	9	6	8	5
3	0	12	4	6	7	9	5	3	8	2	11	1	10
4	0	3	7	9	4	6	8	12	5	11	2	10	1
5	0	6	10	12	1	3	11	9	2	8	5	7	4
6	0	2	12	8	3	5	7	11	4	10	1	9	6
7	0	11	3	5	12	8	4	2	7	1	10	6	9
8	0	4	8	10	5	1	9	7	6	12	3	11	2
9	0	9	1	3	10	12	2	6	11	5	8	4	7
10	0	8	6	2	9	11	1	5	10	4	7	3	12
11	0	5	9	11	6	2	10	8	1	7	4	12	3
12	0	10	2	4	11	7	3	1	12	6	9	5	8

Table 3.5: Design matrix for voices in experiment E247.

Column	0	1	2	3	4	5	6	7	8	9	10	11	12
Row													
1	0	2	12	10	3	2	1	2	1	11	1	3	3
2	0	3	1	3	11	12	2	3	2	2	10	1	1
3	0	1	11	12	2	1	3	1	3	10	3	2	2
4	0	3	10	11	1	3	2	3	2	12	2	1	1
5	0	11	1	3	1	3	2	3	12	2	2	1	10
6	0	1	2	1	12	10	3	1	3	3	11	2	2
7	0	2	3	2	10	11	1	2	1	1	12	3	3
8	0	3	1	3	1	3	10	12	2	2	2	11	1
9	0	12	2	1	2	1	3	1	10	3	3	2	11
10	0	2	3	2	3	2	12	11	1	1	1	10	3
11	0	1	2	1	2	1	11	10	3	3	3	12	2
12	0	10	3	2	3	2	1	2	11	1	1	3	12

3.3 Comparing the goodness of fit of analysis of variance and the continuation-ratio logit model

3.3.1 Introduction

Table 3.6: Important explanatory variables for the BT Experiments

Conversation	
Limited duration:	Cabinet+Row+Cabinet.Row+Column+Condition
Unlimited duration:	Cabinet+Row+Cabinet.Row+Column+Condition +Picture set
Listening	
2 subjects per row:	Cabinet+Row+Cabinet.Row+Column+Condition +Voice+Listening level+Sentence list
1 subject per row:	Subject+Column+Condition +Voice+Listening level+Sentence list

The British Telecom data described in detail in this chapter give rise to an ordinal response problem. The scores given on the opinion rating scale in Figure 3.1 or Figure 3.2 constitute a 5-level ordinal response variable. Factors that may be included as explanatory variables in a model for the data from these experiments are outlined in Table 3.6. Note that in this table Cabinet.Row means the interaction between Cabinet and Row. Also note that the combination Cabinet+Row+Cabinet.Row can be replaced with a single factor for subject.

The traditional British Telecom method of analysis for opinion score responses has been to perform an analysis of variance (ANOVA) on the numerical scores (0 to 4) assigned to the categories (see Richards 1973). One of the assumptions underlying the ANOVA procedure is that the errors in the response variable follow a normal distribution. The opinion score is constrained to one of five values, and so is a discrete rather than a continuous response. Approximating a discrete response with five values by a normal curve is rather a crude approach. Also the use of integers from 0 to 4 is arbitrary. The use of any linear combination of this set of integers would give the exact same results when used in ANOVA. But any set of scores that are not a linear combination of these integers will give different results. The first of these considerations has the consequence of inefficient estimation of parameters. The second has the consequence that the results may change with different scores and, thus, the conclusions from the model may change with different scores. This is not a desirable property.

The types of designs used in these experiments, *e.g.* those discussed in this chapter based on Hyper-Graeco-Latin squares and Williams Latin squares, give orthogonality between most factors of concern. However, not all possible combinations of factors are included, *e.g* if in a listening experiment, conditions, voices and sentence-lists are pairwise orthogonal, not every possible combination of the levels of these three factors will necessarily appear in the design. There is no replication of the design for any of the experiments. Replication would involve repeating the experimental design for further subjects. Because of the lack of replication and because the experiments do not exhaustively cover every possible combination of levels of factors the data arising from these experiments are sparse.

If the data are regarded as a contingency table cross-classified by the ordinal response and the categorical covariates, then the characteristic of sparse data is that many cells contain zero counts. This has repercussions on the model fitting procedures and is a problem well documented for logistic regression (*e.g.* Collett, §3 1991). However, it is a problem that has not received much attention in the case of ordinal-response regression. Anderson gives voice to the concern raised by this problem in the discussion of McCullagh (1980), as do Fienberg, Atkinson and the author himself. A practical effect of sparse data on some models is to make parameter estimates infinite as the model predicts perfectly in some cells of the table.

British Telecom desires to bring their analysis techniques up to date. They are concerned about the problem of inefficiency of the ANOVA model and the arbitrariness of the integers chosen. Their software of choice for analyses is GLIM (Francis et al 1993). A probability link model of the type described in Chapter 2 would be theoretically more appealing for ordinal response analysis than ANOVA. At the time that this work was done the particular models that could most straightforwardly be fitted in GLIM were the continuation-ratio link models. This situation has changed with the introduction of the GLIM4 macros (Wolfe 1996), which allow general purpose fitting of cumulative link models. However, the GLIM4 macros are not flexible enough to enable the fitting of interaction terms involving the threshold parameters, of which the carry-over parameter matrix introduced in this chapter is an example. This carry-over parameter matrix is introduced as a useful analysis where other carry-over analyses are not possible. To maintain consistency of the work the continuation-ratio models are used throughout this chapter.

In order to adopt a single continuation-ratio link model from the range of possibilities, some comparative work was done on data from a number of experiments. The same selection of explanatory variables was used in all models. For both the logit and probit links convergence was quite fast (5-6 cycles). The fit (in terms of deviance) was very similar for these two models. It is not surprising that these should give such similar fits, as the two link functions are similar. One difference between these two models is that the model algebra is simpler for the continuation-ratio logit model than for the continuation-ratio probit model. The other link examined was the complementary log-log link. Note that the continuation-ratio complementary log-log model is equivalent to the cumulative complementary log-log model, see §2.5.1. When all of the required explanatory variables are included in this model, trouble is encountered during the fitting process as the linear predictor tends to plus or minus infinity for some data points. When this occurs GLIM sets the relevant fitted value to zero or one, meaning that the model fits perfectly in these cells of the table. This problem is due to the sparseness of the data. A subset of the explanatory variables was chosen and the deviance values for the complementary log-log logit and probit link models were compared. It was found that the continuation-ratio complementary log-log model gave a worse fit to the data than the other two[1]. As a result of this comparative work the continuation-ratio logit model was chosen as the model for further work.

The model to be used for comparison with the ANOVA results is the continuation-ratio logit model which is defined in §2.2.2 as

$$\log\left(\frac{\pi_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \mathbf{x}_i\beta \tag{3.1}$$

where the vector $\mathbf{x}_i$ consists of indicators for the levels of the explanatory factors, (*e.g.* condition, row, column *etc.*). A fitting procedure gives parameter estimates $\hat{\beta}$ and from these, estimates of the cell probabilities $\hat{\pi}_{ij}$ may be calculated.

One potential advantage of ANOVA over the continuation-ratio logit model is the ability to construct a unique ANOVA table for the variables in the model if all the variables are mutually orthogonal, and from this table to draw direct conclusions about the significance of the variables. It is possible to construct a corresponding table when fitting the continuation-ratio logit model, the so-called analysis-of-deviance table (McCullagh & Nelder, §2.3.2 1989). It is formed by fitting the model one explanatory variable at a time and attributing deviance figures to each successive variable by inspection of the residual deviance at each stage. Unlike the ANOVA table, however, the analysis-of-deviance (ANODE) table is arbitrary in that the

[1]For example in experiment E198 the deviances for the logit, probit and complementary log-log link are 557.8, 557.6 and 584.2 respectively on 746 degrees of freedom. This is for a model containing Cabinet and Condition as explanatory variables.

precise deviance attributed to each variable depends on which other variables have already been included in the model, regardless of orthogonality.

However this apparent advantage of the ANOVA table is illusory. Firstly its uniqueness depends on orthogonality between all the variables in the model. Also, as the table is based on sums of squares obtained from the scores assigned to each category, this implies that the ANOVA table depends on the choice of scores. A different assignment of scores to categories will produce a different ANOVA table. Thus the ANOVA table is arbitrary with an infinite number of variations. The arbitrariness of the ANODE table is perhaps less serious, in that the number of variations is limited by the number of explanatory variables in the model.

3.3.2 Methods and results of comparison

A thorough comparison of the continuation-ratio logit model and the analysis-of-variance (ANOVA) model is now outlined. The residual sum of squares is the quantity that is minimised when fitting ANOVA. The continuation-ratio logit model is fitted by maximising the multinomial likelihood function for the model. A contrived residual sum of squares (RSS_{CR}), calculated from the fitted probabilities of the continuation-ratio logit model, is proposed. This contrived residual sum of squares for the continuation-ratio logit model can be directly compared with the residual sum of squares as minimised in ANOVA. The proposed sum of squares for the continuation-ratio logit model is calculated as follows:

$$RSS_{CR} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

with $\hat{y}_i$ given by

$$\hat{y}_i = \sum_{j=0}^{k-1} j \ \hat{\pi}_{ij}.$$

The fitted value $\hat{y}_i$ is calculated by multiplying the fitted probabilities from the continuation-ratio model $\hat{\pi}_{ij}$ by the scores $j = 0$ to 4 as used in the ANOVA model for this data, giving a fitted mean score for the continuation-ratio logit model.

This method of comparison has been applied to the results from experiments E198, E199, E247 and E264. Table 3.7 summarises the comparison. For all experiments we note that the value of the contrived sum of squares (RSS_{CR}) for the continuation-ratio logit model is smaller than the ANOVA sum of squares (RSS_A). Thus the continuation-ratio logit model is outperforming ANOVA in the criterion which has actually been minimised in order to fit the latter model. This implies that the continuation-ratio logit model is fitting the data better than the linear model underlying ANOVA.

Table 3.7: Sums of squares comparison

Experiment	RSS_A	RSS_A^*	RSS_{CR}
E198	76.6	76.0	75.6
E199	52.9	51.9	50.8
E247	598.1	573.6	566.5
E264	642.1	619.3	607.5

There is one flaw in this comparison which might explain why RSS_{CR} is consistently smaller than RSS_A. For ANOVA the range of fitted values is

$$-\infty < \hat{y} < \infty$$

whereas for the continuation-ratio logit model the range of fitted values is

$$0 < \hat{y} < 4.$$

This means that the differences $y - \hat{y}$ may be larger for the analysis of variance, not because it fits the data worse than the continuation-ratio logit model, but because the fitted values are not constrained to lie in a narrow interval of the real line. To adjust the comparison method to allow for this difference in fitted-value range, the following is proposed: crudely constrain the ANOVA fitted values by letting

$$\hat{y}^* = 0 \quad \text{if} \quad \hat{y} < 0$$

$$\hat{y}^* = 4 \quad \text{if} \quad \hat{y} > 4.$$

The new constrained sum of squares will be called RSS_A^*. Table 3.7 also gives values of RSS_A^* for the four experiments and it can be seen that the values of RSS_{CR} are smaller than these constrained ANOVA sums of squares. The conclusion is the same as before, the continuation-ratio model is fitting the data in these experiments better than the ANOVA model.

The comparison of sum-of-squares values is a comparison which is unfair to the continuation-ratio logit model because the residual sum of squares is precisely the criterion which is minimised when fitting the ANOVA model. Thus it may be stressed that the continuation-ratio logit model is outperforming the linear model underlying ANOVA, giving a better fit to the data. Another type of comparison is considered which does not have any obvious favouritism. This second method of comparison also uses fitted values, but rather than fitted mean scores the comparison uses fitted categories.

For ANOVA the fitted category for each observation is calculated by

$$\text{Predicted category} = \begin{cases} 0 & \text{if} \quad \hat{y} \leq 0.5 \\ 1 & \text{if} \quad 0.5 < \hat{y} \leq 1.5 \\ 2 & \text{if} \quad 1.5 < \hat{y} \leq 2.5 \\ 3 & \text{if} \quad 2.5 < \hat{y} \leq 3.5 \\ 4 & \text{if} \quad 3.5 < \hat{y}. \end{cases}$$

Note that this is an arbitrary choice but one that ties in with the arbitrary integers 0 to 4 that are used in the ANOVA models for the BT data under consideration.

For the continuation-ratio logit model fitted categories may be calculated from the fitted probabilities of each category for each observation. There are several methods to calculate fitted categories. Anderson & Phillips (1981) suggest two. Firstly, they consider

$$\text{predicted category for } i\text{th response} = a \quad \text{where} \quad \hat{\pi}_{ia} = \max(\hat{\pi}_{ij} : \quad j = 0, 1, 2, 3, 4)$$

i.e. taking the predicted category to be the category with largest (modal) $\hat{\pi}_{ij}$. Secondly, Anderson & Phillips (1981) consider

$$\text{predicted category for } i\text{th response} = a \quad \text{where} \quad \hat{\theta}_{a-1} \leq \sum \mathbf{x}_i \hat{\beta} \leq \hat{\theta}_a.$$

However, they use this predictor in the context of a cumulative logit model, which has a strong appeal to an underlying continuum. Indeed for all cumulative link models the threshold parameters (θ_js) are ordered. This is not necessarily true for continuation-ratio link models. Hence, this approach would not be appropriate to use here. A third possiblity in keeping with the theme of comparison with ANOVA is to use the fitted scores from the continuation-ratio logit model that are calculated to obtain RSS_{CR} and use the same allocation rule as for fitted values from ANOVA. This is referred to as the mid-score allocation rule.

To choose one of these methods, either the method of modal $\hat{\pi}_{ij}$ or the mid-score allocation rule, cross-tabulations are made of observed against fitted categories. This was done for the 4 experiments in Table 3.7. It was found that the modal $\hat{\pi}_{ij}$ method gave much better results for all experiments. This method gave more correctly fitting categories, fewer fits that were incorrect by more than one category, and also gave a closer agreement of distribution of fitted categories to the distribution of obseved categories. Thus the comparison of the continuation-ratio logit model with ANOVA was done using fitted values obtained from the method of modal $\hat{\pi}_{ij}$.

The cross-tabulation of observed category by fitted category from ANOVA for experiment E264 is given in Table 3.8. The cross-tabulation of observed category by fitted category (by method of modal $\hat{\pi}_{ij}$) from the continuation-ratio logit model for experiment E264 is given in Table 3.9. From these 2 tables the following observations

Table 3.8: Observed by fitted categories from ANOVA (E264)

		Fitted category					
		0	1	2	3	4	Total
	0	114	97	12	0	0	223
Observed	1	28	177	103	7	0	315
category	2	1	53	209	64	3	330
	3	1	5	107	177	35	325
	4	0	0	10	113	124	247
Total		144	332	441	361	162	1440

are made. The continuation-ratio logit model gives more fitted categories equal to the observed categories (the total of the numbers on the diagonals, 840 and 801). For ANOVA, fewer fitted categories are more than one category away from the observed category (the numbers in the top-right and bottom-left regions of each table, totalling 39 for ANOVA and 40 for continuation-ratio logit). Finally, in the margins of the table it can be seen that the distribution of fitted categories from the continuation-ratio model is closer to the observed distribution than the distribution of fitted categories from the ANOVA model.

The difference in marginal distribution of fitted categories highlights a problem with this method of comparison. The fitted categories for the ANOVA are obtained by using arbitrary cut-points of the real line. Each cut-point is the mean of the scores allocated to the categories on either side of that cut-point (*e.g.* 2.5 is the cut-point between categories scored 2 and 3). However, it is clear from the ANOVA results in Table 3.8 that these cut-points produce a distribution of fitted categories that does not correspond very closely to the distribution of observed categories. If we wished to obtain a table of fitted categories for the ANOVA without making any comparison to fitted categories by another method, then the cut-point should

Table 3.9: Observed by fitted categories from continuation-ratio logit model (E264)

	Fitted category					
	0	1	2	3	4	Total
0	141	73	9	0	0	223
Observed 1	46	173	83	13	0	315
category 2	4	60	173	87	6	330
3	1	5	77	177	65	325
4	0	0	2	69	176	247
Total	192	311	344	346	247	1440

be chosen such that the marginal total of fitted categories is equal to the marginal total of observed categories.

Fixing the marginal totals for fitted categories to equal the marginal totals for observed categories is not suitable when we wish to make a comparison between the ANOVA results and the continuation-ratio logit results. For this comparison it is more appropriate to fix the marginal totals for fitted categories by the ANOVA model to equal the marginal totals for fitted categories (method of modal $\hat{\pi}_{ij}$) by the continuation-ratio logit model. For experiment E264 the appropriate cross-classification is given in Table 3.10. From this table it is observed that the marginal total of fitted categories is equal to the margin in Table 3.9 for the continuation-ratio logit model. The number of correct predictions is now 823 and the number of predictions more than one category out is 36. These numbers compare directly with 840 correct predictions and 40 predictions more than one category out for the continuation-ratio logit model shown in Table 3.9. There are more correctly fitted categories by the continuation-ratio logit model but this model also gives more predicted categories more than one category wrong. The comparison is inconclusive.

Table 3.10: ANOVA — Fitted margin fixed to equal continuation-ratio fitted margin

	Fitted category					
	0	1	2	3	4	Total
0	139	76	8	0	0	223
Observed 1	48	167	91	9	0	315
category 2	4	60	167	95	4	330
3	1	8	75	174	67	325
4	0	0	3	68	176	247
Total	192	315	330	325	247	1440

3.4 Modelling carry-over effects in the BT data

3.4.1 Introduction

In this chapter a model from the probability link class defined in Chapter 2 is chosen for modelling the data from BT experiments. The model is the continuation-ratio logit model given in equation (3.1) and the reasons for this choice are outlined in §3.3.1. In the previous section comparisons are made between this model and the traditional BT method of analysis, ANOVA. The continuation-ratio logit model is a more sophisticated analysis tool for ordinal data than the linear model underlying ANOVA. The conclusion from the comparisons in the previous section is that the continuation-ratio logit model fits the BT data better than the linear model that underlies ANOVA.

In §3.1.3 there is a discussion of carry-over effects. Researchers at British Telecom have become concerned about the possible existence of carry-over effects in their experiments. Unfortunately, due to the sparse nature of the data, it is difficult to examine interactions (principally the interaction between condition and preceding condition) which might shed light on the nature of any carry-over effects. With this in mind an experiment (F100) was designed and implemented in late 1995. To investigate the large number of other experiments for carry-over, an approach to modelling carry-over that involves a carry-over parameter matrix is discussed in this section.

The approach to modelling carry-over that is now introduced involves investigating an interaction between current response and previous response. This may be done by altering the formulation of model (3.1) to

$$\log\left(\frac{\pi_{ijk}}{1-\gamma_{ijk}}\right) = \theta_j + \theta_{jk} - \mathbf{x}_i\beta \tag{3.2}$$

where now k denotes the previous response. Fitting this model will give a matrix of parameters with elements θ_{jk} and of dimension $(K-1) \times K$ where K is the number of categories. Note that with the threshold parameters (θ_js) also in the model, there will be $K-1$ redundant parameters among the θ_{jk}s. Thus the number of degrees of freedom associated with θ_{jk} in this model is $(K-1) \times (K-1)$.

This matrix of carry-over parameters and its interpretation are discussed in §3.4.3. Beforehand residuals from model (3.1) are examined to detect if there is any pattern in the residuals which might further motivate the use of model (3.2).

3.4.2 Residual analysis

The continuation-ratio logit model (3.1) with explanatory variables as listed in Table 3.6 is fitted to data from the experiments in Table 3.1. For each experiment a tabulation of observed responses categorised by previous response is made. Also a tabulation summing the fitted probabilities from model (3.1) across previous response is made for each experiment. These two tables are given for a range of BT experiments in Appendix A.

Figure 3.5: Residual analysis : Group 1

This highlights the main pattern in the residuals given in Appendix A. The residuals for 9 experiments are given in this group, arranged in the order;

E198	E199	E200
E211	E212	E213
E136	E139	E140

Previous Response

Response	1	2	3	4	5
1	−′ −′ +′ +′ − −′ − + +′		− − −′ − − − −′ − −′		+ + + +′ + + + + +′
2					
3					
4			+ + + +′ + −′ +′ −′ +		− −′ − − − − − − −
5	+ + +′ +′ + + +′ + +		− −′ − − − +′ − +′ −		+ +′ +′ + + + + + +

Notes

1. These patterns are for residuals as defined in equation (3.3).
2. A plus represents a positive residual (underfitting) and a minus represents a negative residual (overfitting).
3. Residuals which satisfy $|r_k| < 0.6$ (*i.e.* are small in value) are marked with a prime.
4. Blank boxes are left where there are few large residuals or where the directions of the residuals are not consistent across the experiments. For example cell (1,1) could have been left blank. It is included for comparative purposes with the corresponding cell in Figure 3.6.

Figure 3.6: Residual analysis : Group 2

This highlights the second pattern in the residuals given in Appendix A. The residuals for 4 experiments are given in this group, arranged in the order;

E216 E247
E264 F064

Response \ Previous Response	1	2	3	4	5
1	+ + + +		− − − −		−′ − + +′
2					
3					
4			+ + +′ +		−′ + +′ +
5	+ + −′ +		− − −′ −		+ −′ − −′

Notes

1. These patterns are for residuals as defined in equation (3.3).
2. A plus represents a positive residual (underfitting) and a minus represents a negative residual (overfitting).
3. Residuals which satisfy $|r_k| < 0.6$ (*i.e.* are small in value) are marked with a prime.
4. Blank boxes are left where there are few large residuals or where the directions of the residuals are not consistent across the experiments. For example cell (5,5) could have been left blank. It is included for comparitive purposes with the corresponding cell in Figure 3.5.
5. The residuals from experiment F065 do not correspond either to the pattern in group 1 (Figure 3.5) or to the pattern in group 2

To compare these two tables for a particular experiment, a residual is calculated. This residual is defined for the kth cell by

$$r_k = \left(\frac{O_k - E_k}{\sqrt{E_k}} \right) \tag{3.3}$$

where O_k and E_k are the observed and expected (or fitted) values in the kth cell. Where the value of the residual is close to 0 the model fits the data well. Negative residuals show where the model is over-fitting and positive residuals show where the model is under-fitting. In Figure 3.5 the experiments are grouped according to the overall pattern of their values of r_k. For group 1 the model consistently underfits in three corners of the table. There is consistent over-fitting in cells (1,3) and (5,3). Thus more occasions are observed of one extreme response being followed by another extreme response (except for category 1 followed by itself) than the model predicts. Also fewer occasions are observed of an extreme response following the central response (category 3) than the model predicts. The other cells which have consistent residual values across these experiments are (4,3) and (4,5).

For group 2 as identified in Figure 3.6, a different overall pattern in the residuals is noted, albeit with similarities to the pattern for group 1. The differences between the pattern for group 1 and that for group 2 is in the corners and in cell (4,5). For group 2 there is no consistent value for the residuals in the 2 corners (5,1) and (5,5) as opposed to the consistent positive values noted for group 1. However in group 2 the residuals in cell (1,1), are consistently positive. In cell (4,5) for group 2 there is no consistent value for the residuals.

When group 1 in Figure 3.5 and group 2 in Figure 3.6 are compared with the grouping of types of experiment in Table 3.1 it can be seen that all the experiments in group 1 are conversation experiments. Three out of four of the experiments in group 2 are listening experiments.

3.4.3 Interpretation of carry-over parameter matrix

To understand the parameter estimates in the carry-over parameter matrix it is instructive to consider a simple example first. Suppose that the model

$$\log\left(\frac{\pi_{jk}}{1-\gamma_{jk}}\right) = \theta_j + \theta_{jk} \tag{3.4}$$

is fitted to an ordinal response with $K = 5$ categories, and that the parameter estimates from the fitting process are those given in Table 3.11. Now using these

Table 3.11: Parameter estimates for simple example

		Previous response k					
	$\hat{\theta}_{jk}$	1	2	3	4	5	
	1	0	-1	-3	-1	0	$\hat{\theta}_1=-2$
j	2	0	0	0	0	0	$\hat{\theta}_2=-1$
	3	0	0	0	0	0	$\hat{\theta}_3=\ 0$
	4	0	1	3	2	0	$\hat{\theta}_4=\ 1$

parameter estimates we may calculate fitted cell probabilities ($\hat{\pi}_{jk}$) for each combination of response and previous response according to the relationships defined by the continuation-ratio logit model:

$$\hat{\pi}_{1k} = \frac{\exp(\hat{\theta}_1 + \hat{\theta}_{1k})}{1+\exp(\hat{\theta}_1 + \hat{\theta}_{1k})} \qquad \text{and so forth to} \qquad \hat{\pi}_{5k} = 1 - \hat{\pi}_{1k} - \hat{\pi}_{2k} - \hat{\pi}_{3k} - \hat{\pi}_{4k}.$$

The fitted probabilities for each value of previous response k are given in Table 3.12. Note that the fitted probabilities in column 1 of Table 3.12 are calculated from the fitted values of the threshold parameters ($\hat{\theta}_j$s) alone as $\hat{\theta}_{j1} = 0$ for all j. If the model is a good fit to the original data then the fitted probabilities will be close in value to the observed probabilities. Assuming that this is the case here, the fitted probabilities in Table 3.12 show the structure in a set of observed data that can

Table 3.12: Fitted probabilities for simple example

	Previous response k				
	1	2	3	4	5
$\hat{\pi}_{1k}$	0.12	0.05	0.01	0.05	0.12
$\hat{\pi}_{2k}$	0.24	0.26	0.27	0.26	0.24
$\hat{\pi}_{3k}$	0.32	0.35	0.36	0.35	0.32
$\hat{\pi}_{4k}$	0.24	0.31	0.36	0.33	0.24
$\hat{\pi}_{5k}$	0.09	0.04	0.01	0.02	0.09

be represented by the cumulative logit model (3.4) with parameter estimates as in Table 3.11.

In order to interpret the carry-over parameter matrix, first note that each row in the matrix must have one parameter set to zero (typically the parameter in the first column) and that all parameters in a row are relative to this. Thus $\theta_{13} = -3$ is not an absolute parameter value for the effect of a previous score of 3 on category 1, it is relative to the effect of a previous score of 1 on category 1 being equal to zero.

From model (3.4) the linear predictor, $\theta_j + \theta_{jk}$ is defined in terms of an underlying continuation-ratio ρ_{jk}, where

$$\rho_{jk} = \frac{\pi_{jk}}{\pi_{jk} + \pi_{(j+1)k} + \cdots + \pi_{Kk}}.$$

So the effect of a previous score is to change the values of the underlying continuation-ratios. For $j = 1$, $\theta_1 + \theta_{1k}$ is defined solely in terms of π_{1k}. So a negative parameter estimate in the first row of the carry-over parameter matrix will reduce the fitted probability in category 1. From Table 3.12 it can be seen that the fitted probability for category 1 is smaller for a previous score of 3 than for a previous score of 1. If a parameter estimate in the first row of the carry-over parameter matrix is smaller than any other parameter estimate in that row then that indicates for which previous response the probability for category 1 is smallest. For $j = 4$, $\theta_4 + \theta_{4k}$ is defined in terms of π_{4k}/π_{5k}, the ratio of the probabilities for category 4 and category 5. Thus a positive parameter estimate in the fourth row of the carry-over parameter matrix will increase the ratio of the fitted probabilities for category 4 and category 5. From Table 3.12 it can be seen that the ratio of the fitted probabilities for category 4 and category 5 is larger for a previous score of 3 than for a previous score of 1. If a parameter estimate in the fourth row of the carry-over parameter matrix is larger than any other parameter estimate in that row then that indicates for which previous response the ratio of the probabilities in category 4 and category 5 is largest.

3.4.4 Carry-over matrices in analysis of BT data

The illustration in the previous section is tailored to reflect the results from the BT data sets. In Figures 3.5 and 3.6 patterns in the residuals from model (3.1) have been noted, and these patterns suggest consistent misfitting by the model. This consistent misfitting is in categories 1, 4 and 5 of the opinion score response. For a response in category 1 or 5, the model underfits if the previous response is in category 1 or 5, and overfits if the previous response is in category 3. The reverse is true for a current response in category 4; the model overfits if the previous response is in category 1 or 5, and underfits if the previous response is in category 3. So when fitting model (3.2) to data from these experiments significant parameter estimates in the first and fourth rows of the carry-over matrix of parameters are to be expected. For group 1 as identified in Figure 3.5, the relative values of the estimates in rows 1 and 4 should be similar to those in the example in §3.4.3 (given that column 1 estimates are set equal to zero). In row 1 decreasing estimates from column 1 to the central column 3 and increasing estimates from column 3 to column 5 are expected, with a positive estimate expected for column 5. In row 4 a pattern of increasing estimates from column 1 to column 3 and decreasing estimates from column 3 to column 5 with a negative estimate for column 5 are expected.

The carry-over parameter estimates for the BT experiments are given in Table 3.13. If there is a cell in the table of observed values with a zero count then the parameter estimate corresponding to that cell will be infinite. If this occurs in column 1 then all estimates in that row will be infinite (as they are evaluated relative to the estimate in column 1). In this case values for the parameter estimates in the columns other than column 1 may be obtained by changing the column of parameters that is set to zero. The deviance and degrees of freedom corresponding to the matrix are given in Table 3.13 and these values may be tested against the χ^2 distribution for significance of the parameter matrix (note that the 5% value of χ^2_{16} is 26.3). Thus the matrix is significant at a 5% level of significance in all experiments bar E200. For experiments in group 1 the parameters in rows 1 and 4 do appear to follow the values that we would expect.

Table 3.13: Carry-over parameter matrices

E198 deviance: 52.6

		Previous response				
		1	2	3	4	5
	1	0	1.49	-0.53	-0.26	1.72
Cut-	2	0	-1.51	-0.49	-0.45	1.32
Point	3	0	0.81	1.27	-0.05	0.17
	4	0	1.86	4.97	2.98	-0.18

E199 deviance: 26.8

	Previous response				
	1	2	3	4	5
1	0	-1.00	0.72	1.90	4.64
2	0	-1.01	1.12	1.54	0.91
3	0	0.96	-0.29	1.10	1.14
4	0	1.03	0.79	0.75	0.70

E200 deviance: 16.4

		Previous response				
		1	2	3	4	5
	1	0	-0.79	-0.36	-0.56	2.18
Cut-	2	0	-0.35	-0.24	0.51	0.55
Point	3	0	1.51	0.16	0.58	0.87
	4	0	1.08	1.35	0.44	-0.02

E211 deviance: 35.8

	Previous response				
	1	2	3	4	5
1	0		∞		
2	0	-1.67	1.20	1.08	2.56
3	0	-0.62	-1.31	-2.31	-1.14
4	0	1.50	1.97	0.00	-1.85

E212 deviance: 28.2

		Previous response				
		1	2	3	4	5
	1	0	0.63	-0.99	0.25	0.83
Cut-	2	0	-2.08	-1.75	-1.13	-1.95
Point	3	0	0.37	-0.39	-0.36	0.24
	4	0	2.48	3.63	0.21	-0.08

E213 deviance: 37.4

	Previous response				
	1	2	3	4	5
1	0	1.26	∞	-0.94	1.83
2	0	-0.26	-4.00	-2.22	-2.63
3	0	-0.40	-0.86	-1.59	-0.32
4	0	1.25	0.79	1.54	-0.16

E216 deviance: 48.6

		Previous response				
		1	2	3	4	5
	1	0	-0.37	-1.02	-0.93	-0.90
Cut-	2	0	-1.17	-0.86	-0.64	1.61
Point	3	0	0.92	-0.19	0.04	-0.95
	4	0	1.42	3.12	1.45	0.45

E136 deviance: 27.0

	Previous response				
	1	2	3	4	5
1	0		∞		
2	0	0.34	-0.26	-1.07	-0.28
3	0	-0.81	-0.37	-0.64	-0.98
4	0	1.87	1.98	1.37	-0.39

Table 3.13: Carry-over parameter matrices *continued*

E139 deviance: 42.0

		Previous response				
		1	2	3	4	5
	1	0	-1.37	-3.02	0.24	1.21
Cut-	2	0	-0.44	-0.40	-0.04	-1.88
Point	3	0	-0.98	-0.73	-0.99	-1.42
	4	0	0.56	0.91	4.42	-1.40

E140 deviance: 40.9

	Previous response				
	1	2	3	4	5
1	0	-0.50	-0.63	0.23	1.15
2	0	-0.00	1.93	0.73	4.53
3	0	2.46	1.96	2.79	2.62
4	0	1.15	2.65	2.29	0.35

E247 deviance: 37.7

		Previous response				
		1	2	3	4	5
	1	0	-0.79	-0.98	-1.09	-1.08
Cut-	2	0	-0.26	-0.04	0.12	-0.17
Point	3	0	0.76	0.74	0.83	0.80
	4	0	1.58	2.46	1.79	2.22

E264 deviance: 70.4

	Previous response				
	1	2	3	4	5
1	0	-0.87	-1.70	-1.08	0.35
2	0	-1.18	-1.25	-1.32	-0.59
3	0	0.66	0.57	-0.10	0.72
4	0	-0.75	0.07	-0.36	0.14

F064 deviance: 37.7

		Previous response				
		1	2	3	4	5
	1	0	-1.13	-1.54	-0.86	-0.19
Cut-	2	0	0.21	-0.19	-0.72	-0.42
Point	3	0	0.23	-0.16	0.21	-1.53
	4	0	0.76	0.94	0.05	0.76

F065 deviance: 27.3

	Previous response				
	1	2	3	4	5
1	0	-0.06	-0.08	1.28	2.23
2	0	0.93	0.69	1.13	1.45
3	0	1.28	-0.22	-1.01	0.46
4	0	-2.00	-0.71	-1.48	-1.76

3.4.5 Conclusions

For the experiments considered (in Table 3.1), there does appear to be a significant carry-over effect of response which can be modelled by the carry-over parameter matrices and which is consistent across experiments in 2 different patterns. However a note of caution must be introduced. The two distinct groups of experiments which have been noted in Figures 3.5 and 3.6 correspond to the grouping of experiments that is made from the informal analysis of score frequencies presented in Chapter 5. The conclusion from that analysis is that it is necessary to allow fully for subject variability in the model by including a subject scaling term *before* investigating for carry-over effects of any nature.

This conclusion is arrived at because of differences between subjects in their use of the extreme categories of the scale, a feature which is highlighted in the exploratory analysis of score frequencies in Chapter 5. The carry-over parameter matrices discussed in this chapter contain consistent patterns in the parameters corresponding to categories 1, 4 and 5. So it is entirely possible that what has been attributed to carry-over of response could in fact be a subject-specific effect. In other words, the subject scaling effect that is considered in Chapter 5 and the effect modelled by the carry-over parameter matrix may be confounded. It makes sense to analyse these experiments allowing fully for subject variability by inclusion of a scaling term before further analysing for carry-over.

Chapter 4

Methods of fitting the cumulative logit model

4.1 Introduction

In this chapter an account is given of several methods of fitting the cumulative logit model. Software facilities for implementing these methods are discussed. Two distinct cases of the cumulative logit model are considered: the model with location terms only, and the model with location and scaling terms. For the cumulative logit model with location terms only, two distinct approaches to parameter estimation are discussed. These approaches are maximum likelihood and generalized estimating equations. A set of general purpose GLIM macros is introduced for fitting cumulative logit and other cumulative link models in GLIM4 by maximum likelihood. The approach of generalized estimating equations (GEE) leads to a number of alternative methods based on using different working correlation matrices. GEE with the true correlation matrix for multinomial observations is equivalent to maximum likelihood estimation and this equivalence is demonstrated. Two other possible working correlation matrices are Clayton's (1992) correlation matrix and a matrix similar to Clayton's but based on equal marginal probabilities. A method employing the identity matrix as a working correlation is discussed in this chapter in detail. This particular GEE method is referred to as the method of independent binomials.

For the cumulative logit model with location and scaling terms, the general purpose GLIM4 macros introduced in this chapter may be used for maximum likelihood fitting of the model. Two forms of the model with a scaling term are discussed and both can be fitted using the GLIM4 macros. At present there is no widely-available routine for general-purpose fitting of these models in any other software. A successive relaxation approach to fitting the model by independent binomials is also discussed. This approach involves iterations between estimating the location effects holding the scaling effects constant and estimating the scaling effects holding the location effects constant.

Figure 4.1 displays current software options for fitting the cumulative logit model. It also serves as a guide to the organisation of this chapter. In §4.2 discussion is of fitting methods for the cumulative logit model with location effects only. In §4.3 fitting methods for the cumulative logit model with location and scaling effects are discussed. Figure 4.1 is meant as a quick reference guide. Details and discussion of each fitting method are contained in this chapter.

Figure 4.1: Software for implementing methods of fitting the cumulative logit model

Fitting Method	Cumulative logit model: Location effects only $\theta_j - \mathbf{x_i}\beta$	Cumulative logit model: Location and scale effects $(\theta_j - \mathbf{x_i}\beta)/\exp(\mathbf{z_i}\tau)$ or $\theta_j \exp(\mathbf{z_i}\lambda) - \mathbf{x_i}\beta$
Maximum Likelihood	SAS, SPlus, BMDP, LIMDEP, GLIM4 *etc.*	GLIM4 Specific SAS program
GEE - working correlation matrix:		
1) true	As for maximum-likelihood	As for maximum-likelihood
2) Clayton's	SPlus, Specific SAS macros	-
3) Equal-margins	SPlus, Specific SAS macros	-
Independent Binomials	SPlus, GLIM4, SAS, LIMDEP, BMDP, *etc*	Specific GLIM4 program

4.2 Fitting the cumulative logit model with location terms only

4.2.1 The cumulative logit model with location terms only

The cumulative logit model with location terms only (for a K-category ordinal response) is

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \mathrm{x}_i\beta \qquad (j = 1, \ldots, K-1) \tag{4.1}$$

as defined in Chapter 2. In this section maximum likelihood and generalized estimating equation approaches to fitting this model will be discussed. The latter approach includes as a special case a method which we call the method of independent binomials. This particular method has a practical appeal because it is straightforward to fit in widely-available logistic regression software.

4.2.2 Maximum likelihood

Introduction

Model (4.1) can be fitted using maximum likelihood. Note that this model is not a generalized linear model as defined in McCullagh & Nelder (1989) and this will now be shown with their notation suitably adapted. A vector of observations $\mathbf{y}$, with length $n(K-1)$, is assumed to be a realisation of a random variable $\mathbf{Y}$ whose components are independently distributed (the distribution being a member of an exponential family) with means $E(\mathbf{Y}) = \mu$. A systematic component η is defined as a linear combination of explanatory variables. The systematic component η is related to μ by

$$\eta_i = g(\mu_i) \qquad (i = 1, \ldots, n)$$

where $g(.)$ is termed a link function and can be any monotonic, differentiable function.

For the cumulative logit model we assume that the data are realisations of the multinomial distribution with

$$E(Y_{ij}) = m_i \pi_{ij}$$

where π_{ij} are the underlying cell probabilities and the row sums m_i are taken as fixed. Model (4.1) is defined in terms of the cumulative cell probabilities, γ_{ij}. The relationship between $\mu_{ij} = E(Y_{ij})$ and $\eta_{ij} = \theta_j - X_i\beta$ is

$$\begin{aligned}
\mu_{i1}/m_i &= (1 + e^{-\eta_{i1}})^{-1} \\
\mu_{i2}/m_i &= (1 + e^{-\eta_{i2}})^{-1} - (1 + e^{-\eta_{i1}})^{-1} \\
\mu_{i3}/m_i &= (1 + e^{-\eta_{i3}})^{-1} - (1 + e^{-\eta_{i2}})^{-1} \\
&\vdots \\
\mu_{iK}/m_i &= 1 - (1 + e^{-\eta_{i(K-1)}})^{-1}.
\end{aligned}$$

There is no single link function $g(.)$ that can be applied to the left-hand side of these equations to give the right-hand side. Thus the model does not fit into the generalized linear model framework. Note however that this relationship between μ_{ij} and η_{ij} for the cumulative logit model does satisfy the definition of a composite link as introduced by Thompson & Baker (1981). A composite link function is one that allows μ_{ij} to depend on more than one η_{ij}.

The important results for fitting the cumulative logit model by maximum likelihood are now given. The model may be written in the form

$$\text{logit}(\gamma_{ij}) = \sum_r x^*_{ijr} \beta^*_r$$

for notational simplicity where β^* is a vector containing the parameters (θ, β). The vector $\mathbf{x}^*_{ij}$ is a row from the expanded design matrix $\mathbf{X}^*$ which includes zero/one indicators for the cut-point parameters as well as the vectors $\mathbf{x}_i$. The log likelihood for the multinomial observations $y_i = (y_{i1}, \ldots, y_{iK})$ is proportional to

$$l(\pi; y) = \sum_{ij} y_{ij} \log \pi_{ij}. \tag{4.2}$$

For estimation the partial derivatives

$$\begin{aligned} \frac{\partial l}{\partial \beta_r^*} &= \sum_{ij} \frac{\partial l}{\partial \gamma_{ij}} \frac{\partial \gamma_{ij}}{\partial \beta_r^*} \\ &= \sum_{ij} x_{ijr}^* \gamma_{ij}(1-\gamma_{ij}) \frac{\partial l}{\partial \gamma_{ij}} \end{aligned} \tag{4.3}$$

are required and here

$$\frac{\partial l}{\partial \gamma_{ij}} = \frac{y_{ij}}{\pi_{ij}} - \frac{y_{i(j+1)}}{\pi_{i(j+1)}}$$

for $1 \leq j < K$. Note that this corrects a mistake on pages 172 and 173 of McCullagh and Nelder (1989).

The capacity to fit model (4.1) by maximum likelihood exists in some standard software packages. For example in SAS (SAS Institute, Inc. 1987) the procedure `PROC LOGISTIC` can be used. In GENSTAT (Payne *et al*, 1993) the command `ORDINALLOGISTIC` is available.

GLIM4 macros

Hutchinson (1985) was the first to propose a method for fitting model (4.1) in GLIM3.77. He uses the procedure for composite-link functions described by Thompson & Baker (1981). An important development in GLIM subsequent to Hutchinson's work was the production of a set of macros for fitting general non-linear models, which were originally published in Ekholm, Green & Palmgren (1986). Ekholm & Palmgren (1988) demonstrated the technique of using these macros to fit the cumulative logit model. The macros for fitting non-linear models have been updated for use in GLIM4 (Francis et al, 1993) by Ekholm & Green (1995).

The macros presented in Appendix B enable general purpose fitting of cumulative link models. They are much more than just an update of the work by Ekholm & Palmgren (1988). Two features of GLIM4 which are not available in GLIM3.77 are used in the macros. These are the ability to print to macros (page 143, Francis et al, 1993) and the method for fitting non-linear models described at *pp* 201–203 of Francis et al, 1993. An initialization macro very similar to that used by Ekholm & Green (1995) is employed. Most of the work which Ekholm & Palmgren (1988) require the user to do (*i.e.* specifying the form for `%eta`) is performed internally. The user is left with minimal work to do. The user is given a choice of link functions to use, enabling the fitting of cumulative probit and cumulative complementary-log-log models in addition to the cumultive logit model. The use by Ekholm & Palmgren (1988) of general numerical derivatives to update the model matrix is replaced with the use of the analytical derivatives for these cumulative link models. This greatly improves the time efficiency of the model fitting process. For a model fitted to data from BT experiment E198, for example, the time to fit the model in GLIM4 using the numerical derivatives was 3 minutes and 48 seconds whereas using the analytical derivatives the time taken was 25 seconds. This comparison was done on a SPARC station 2.

These macros have been included in the GLIM Macro Library under the name `ORDINAL`. The submission includes detailed instructions for their use which are printed in the Macro Library Guide. Further description may be found in the GLIM Newsletter article, Wolfe (1996). Confirmation that the output from these macros agrees with results obtained by other authors can be found in this article. The macros, the instructions on how to use them and the GLIM Newsletter article are all available on the Internet at URL

`http://www.maths.soton.ac.uk/rw/r_wolfe.html`

Further disussion of the macros can be found in §4.3.2.

4.2.3 Generalized estimating equations

An alternative approach to maximum likelihood for fitting model (4.1) is discussed by Clayton (1992). He suggests the use of generalized estimating equations (GEE, see Liang & Zeger, 1986) to estimate the parameters of the model. As a method of estimation GEE is an example of quasi-likelihood estimation (see Wedderburn 1974, McCullagh 1983) specifically tailored with longitudinal data analysis in mind. Longitudinal data are characterised by repeated observations on the same sampling unit. The use of GEE to fit model (4.1) makes the extension to analysis of longitudinal ordinal responses straightforward. The GEE notation introduced by Liang & Zeger (1986) is tailored in a particular manner for fitting the cumulative logit model. This tailoring is now described.

To apply GEE to fitting the cumulative logit model (Clayton, 1992) each response y_i on the K-category ordinal scale is converted into $(K-1)$ binary responses $y^{\dagger}_{i1}, \ldots, y^{\dagger}_{i(K-1)}$ by the rule

$$y^{\dagger}_{ij} = \begin{cases} 1 & y_i \leq j \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

These binary responses $y^{\dagger}_{ij}$ take the role of the repeated observations within a sampling unit.

For units $i = 1, \ldots, n$ we observe responses $y^{\dagger}_{ij}$ (repeated observations within i are at $j = 1, \ldots, K-1$). The response vector $y^{\dagger}_i = (y^{\dagger}_{i1}, \ldots, y^{\dagger}_{iK-1})$ is a realisation of the random variable $Y^{\dagger}_i$ which has expectation $E(Y^{\dagger}_i) = \gamma_i(\beta)$. The generalized estimating equations are defined as

$$\sum_{i=1}^{n} D_i^T V_i^{-1} \left(y^{\dagger}_i - \gamma_i \right) = 0 \tag{4.5}$$

where the matrix D_i contains the derivatives $\partial\gamma_i/\partial\beta_r$ and

$$V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} \tag{4.6}$$

with $A_i(\gamma_i)$ being a diagonal matrix of variance functions. Note that $V_i = \text{cov}(Y_i^\dagger)$ only if $R(\alpha)$ is the true correlation matrix for $Y_i^\dagger$. Otherwise $R(\alpha)$ is referred to as a 'working' correlation matrix which is characterised by the parameters α. The simplest form that $R(\alpha)$ can take is the identity matrix in which case the GEE reduce to the score equations of a likelihood analysis of binomial observations in which those observations are assumed to be independent. More complicated forms of $R(\alpha)$ can be used, often based on an educated guess which borrows strength across subjects to estimate the true underlying correlation structure. The benefit of an improved working correlation matrix is in improved efficiency in estimating the parameters β.

The estimated standard errors of the estimates $\hat{\beta}$ are only given by the square-roots of the diagonal elements of

$$\widehat{\text{cov}}_M(\hat{\beta}) = (\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} \tag{4.7}$$

if the covariance matrix $\text{cov}(Y^\dagger) = V$ is correctly specified. When a working correlation matrix based on a guess at the true correlation structure is being used this covariance matrix is not correct. The estimated standard errors of the estimates $\hat{\beta}$ are then given by the square-roots of the diagonal elements of the 'information sandwich'

$$\widehat{\text{cov}}_R(\hat{\beta}) = (\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} \hat{D}^T \hat{V}^{-1} (y^\dagger - \hat{\gamma})(y^\dagger - \hat{\gamma})^T \hat{V}^{-1} \hat{D} (\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} \tag{4.8}$$

and these are referred to as robust standard errors. These robust standard errors are frequently presented with the corresponding elements of $\widehat{\text{cov}}_M(\hat{\beta})$ which are then referred to as 'naive' standard errors.

Returning to the observed response y_i on the K-category ordinal scale; if it is assumed that y_i are independent of each other then the correlation matrix for $y_{ij}^\dagger$ will be block diagonal. The known correlation between the binary responses

$$\text{corr}(Y_{ij}^\dagger, Y_{ik}^\dagger) = \sqrt{\frac{\gamma_{ij}(1-\gamma_{ik})}{\gamma_{ik}(1-\gamma_{ij})}} \qquad j < k, \tag{4.9}$$

where $\gamma_{ij} = E(Y_{ij}^{\dagger})$, can be used in GEE removing the need to form a working correlation matrix. In this case the analysis is equivalent to maximum-likelihood but note that this equivalence depends on the assumption of independence of the y_is. This equivalence is shown explicitly for a particular case in §6.3.4.

Clayton (1992) suggests using the observed marginal cumulative proportions in place of γ_{ij} in equation (4.9) to obtain a working correlation matrix that is constant for all i. This method of fitting the model may be implemented in S-Plus (StatSci, 1993) using the function `gee()` (written by Vincent Carey) which is available via Statlib on the Internet at URL `http://unix.hensa.ac.uk`.

An alternative working correlation matrix (also constant for all i) can be constructed by assuming that the marginal cell probabilities are equally likely. This method can also be implemented in S-plus using the `gee()` function. The equal-margins GEE method is similar in spirit to Clayton's GEE method in that both use working correlation matrices which are constant for each multinomial observation. Clayton's GEE method uses the observed marginal cell proportions to calculate the working correlation. Equal-margins GEE takes the marginal cell probabilities to be equal. Thus the working correlation matrix is data-dependent in Clayton's GEE but not in equal-margins GEE. Because of this, equal-margins GEE is used in some theoretical comparisons presented in Chapter 6. The results for equal-margins GEE should give some indication of how Clayton's GEE would perform.

4.2.4 The method of independent binomials

The method of independent binomials to fit model (4.1) involves the use of generalized estimating equations (discussed in the previous section) with an independence working correlation matrix. This implies as a working assumption that the collapsed binomial observations $y_{ij}^{\dagger}$ are independent of each other. The important practical advantage of using GEE with an independence working correlation matrix is that it is possible to implement the method using widely-available routines which perform logistic regression on independent binary responses. When used to fit model (4.1) a logistic regression routine will produce estimated naive standard errors for the parameter estimates. Estimated robust standard errors can be calculated from the information sandwich formula (equation 4.8) using the true correlations given in equation (4.9) with γ_{ij} replaced by the estimated value $\hat{\gamma}_{ij}$ from the routine. An investigation of the efficiency of parameter estimation by this method relative to estimation by maximum likelihood can be found in Chapter 6.

A logistic regression routine estimates the parameters of model (4.1) by maximising the log-likelihood function for independent binary observations

$$l_B(\pi; y) = \sum_i \sum_j^{K-1} y_{ij}^{\dagger} \log \gamma_{ij} + (1 - y_{ij}^{\dagger}) \log(1 - \gamma_{ij}). \tag{4.10}$$

Note that a maximum-likelihood routine for fitting model (4.1) maximises the multinomial log-likelihood function

$$l_M(\pi; y) = \sum_{ij} y_{ij} \log \pi_{ij} \tag{4.11}$$

where $y_{ij} = 1$ if $y_i = j$. The deviance is related in general to a log-likelihood function by the relation

$$D(\pi; y) = 2l(y; y) - 2l(\pi; y) \tag{4.12}$$

where $l(y; y)$ is the log-likelihood function for the saturated model (*i.e.*, where the model fits the data perfectly). For both $l_M(\pi; y)$ and $l_B(\pi; y)$ the value of the log-likelihood in the saturated model is zero.

A logistic regression routine will give a deviance value which is calculated from the log-likelihood function for independent binary observations. This deviance will be labelled $D_B(y; \hat{\pi}_{IB})$. Using the fitted values $\hat{\pi}_{IB}$ from the logistic regression routine a value of the multinomial deviance function may be calculated. This deviance will be labelled $D_M(y; \hat{\pi}_{IB})$. It makes sense to compare both this deviance and

$D_B(y; \hat{\pi}_{IB})$ with the minimised value of the multinomial deviance function obtained when fitting the model by maximum-likelihood, $D_M(y; \hat{\pi}_{ML})$ say. Obviously

$$D_M(y; \hat{\pi}_{IB}) \geq D_M(y; \hat{\pi}_{ML}). \tag{4.13}$$

For the BT experiments analysed in Chapter 5, it is found that

$$D_M(y; \hat{\pi}_{IB}) \leq D_B(y; \hat{\pi}_{IB}). \tag{4.14}$$

Whether this inequality is true in general is now investigated. For one multinomial observation equation (4.10) gives

$$\begin{aligned} l_B(y_i; \hat{\pi}_{ij}) &= \log \hat{\pi}_{i1} + \cdots + \log(\hat{\pi}_{i1} + \cdots + \hat{\pi}_{ij}) \\ &\quad + \log(1 - \hat{\pi}_{i1} - \cdots - \hat{\pi}_{ij+1}) + \cdots + \log \hat{\pi}_{iK} \end{aligned}$$

and equation (4.11) gives

$$l_M(y_i; \hat{\pi}_{ij}) = \log \hat{\pi}_{ij}.$$

For inequality (4.14) to hold it is required that

$$\begin{aligned} l_M(y_i; \hat{\pi}_{ij}) &\geq l_B(y_i; \hat{\pi}_{ij}) \\ i.e. \quad \hat{\pi}_{ij} &\geq \hat{\pi}_{i1} \times \cdots \times (\hat{\pi}_{i1} + \cdots + \hat{\pi}_{ij}) \\ &\quad \times (1 - \hat{\pi}_{i1} - \cdots - \hat{\pi}_{ij+1}) \times \cdots \times \hat{\pi}_{iK} \end{aligned}$$

which is obviously true for $\hat{\pi}_{i1}$ and $\hat{\pi}_{iK}$. However, this does not necessarily hold in the intermediate categories and thus inequality (4.14) does not hold in general.

It is important to consider the deviance function as it is the quantity on which tests of significance are often based in generalized linear modelling. When the cumulative logit model is being fitted using independent binomials it is appropriate to use the multinomial deviance function calculated with the fitted values obtained from the logistic regression routine, *i.e.* $D_M(y; \hat{\pi}_{IB})$, and not the deviance $D_B(y; \hat{\pi}_{IB})$ that would routinely be reported by a binomial regression program.

4.3 Fitting the cumulative logit model with a scaling term

4.3.1 The cumulative logit model with location and scaling terms

Two forms of the cumulative logit model with location and scaling terms are considered in this chapter. The first of these two forms is

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = (\theta_j - \mathbf{x}_i\beta)/\exp(\mathbf{z}_i\tau). \tag{4.15}$$

where $\mathbf{z}_i$ is a row from the design matrix $\mathbf{Z}$ for the scaling terms and τ is the vector of scaling effect parameters. This model was introduced in Chapter 2. It is helpful to think of a continuum underlying the ordinal response and probability distributions being defined on this continuum. With this in mind the scaling effect parameters, τ, describe changes in the variance or scale (or dispersion) of the probability distribution defined on the underlying continuum.

The second form of the model is

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j \exp(\mathbf{z}_i\lambda) - \mathbf{x}_i\beta. \tag{4.16}$$

where the vector of scaling effect parameters, λ, can be interpreted in terms of fitted category probabilities. Both of the models, (4.15) and (4.16), are discussed in detail in §2.3.

4.3.2 Maximum likelihood using GLIM4 macros

The macros introduced in §4.2.2 can also be used to fit the cumulative logit model with a scaling term. As far as the internal workings of the macros are concerned the scaling term is just one more part of the non-linearity of the model. The macros allow the fitting of the two forms of the cumulative logit model with a scaling term given by (4.15) and (4.16).

The use of either the logit, probit or complementary-log-log link functions is possible when fitting either form of the model using the macros. Cox (1995) fits the cumulative logit model with a scaling term (model 4.15) using a special program written in SAS. However, the GLIM4 macros presented here are the only currently-available general-purpose routine for fitting cumulative link models with scaling terms in widely-available software. A demonstration of the use of the GLIM4 macros for fitting the cumulative logit model with a scaling term is now given.

Table 4.1: Data from rating visual stimuli

		Contrast level				
Category	Judgement	-2	-1	0	+1	+2
1	6 dark	19	3	0	0	0
2	5 dark	14	4	0	0	0
3	4 dark	3	11	0	0	0
4	3 dark	2	5	3	1	0
5	2 dark	0	13	15	4	0
6	1 dark	0	0	14	1	0
7	1 bright	0	2	4	8	0
8	2 bright	0	0	3	8	1
9	3 bright	0	0	0	11	4
10	4 bright	0	0	1	7	12
11	5 bright	0	0	0	0	8
12	6 bright	0	0	0	0	13
Total		38	38	40	40	38
Class		1	2	3	4	5

Consider the data in Table 4.1 from Kijewski, Swensson & Judy (1989). The data arise from one subject rating 196 visual stimuli. The ratings are made on a 12-category ordinal scale and the stimuli are drawn from 5 classes. These 5 classes are defined by the contrast level of the visual stimuli. The third class is taken as the reference class.

We wish to fit the cumulative logit model with a scaling term,

$$\mathrm{logit}(\gamma_{ij}) = \frac{\theta_j - \beta_i}{\exp(\tau_i)}, \tag{4.17}$$

where β_i denotes the location effect of class and τ_i denotes the scaling effect of class. Using GLIM4 the macros are inputted and then the following commands are used:

```
$units 60$
$data obs$
$read
19 3 0 0 0
14 4 0 0 0
3 11 0 0 0
2 5 3 1 0
0 13 15 4 0
0 0 14 1 0
0 2 4 8 0
0 0 3 8 1
0 0 0 11 4
0 0 1 7 12
0 0 0 0 8
0 0 0 0 13
$ca cats=%gl(12,5) : class=%gl(5,1) : rows=class$
$factor class 5 (3)$
$list model=class$
```

```
$list s_model=class$
$cycle 20 0.00001$
$use ORDINAL obs rows cats$
$d e$
$pri 'Chi-square, sum(o-e)^2/e : ' %X2$
$return
```

The location term is specified by `$list model=class$` and the scaling term by `$list s_model=class$`. The link function is set to logit by default. The convergence criterion is strengthened by `$cycle 20 0.00001$`. Finally the internally calculated χ^2 statistic is printed out. The output from the analysis is as follows.

```
[o] Checks completed, the model will now be fitted.
[o]
[o] scaled deviance =  27.648 at cycle 11
[o]     residual df =  36
[o]
[o]
[o] Cut-point 1 =      CUT_[1]                =   -7.244
[o] Cut-point 2 = exp(CUT_[2]) + cut-point 1 =   -5.562
[o] Cut-point 3 = exp(CUT_[3]) + cut-point 2 =   -3.960
[o] Cut-point 4 = exp(CUT_[4]) + cut-point 3 =   -2.755
[o] Cut-point 5 = exp(CUT_[5]) + cut-point 4 =  -0.0540
[o] Cut-point 6 = exp(CUT_[6]) + cut-point 5 =    1.076
[o] Cut-point 7 = exp(CUT_[7]) + cut-point 6 =    2.391
[o] Cut-point 8 = exp(CUT_[8]) + cut-point 7 =    3.685
[o] Cut-point 9 = exp(CUT_[9]) + cut-point 8 =    5.545
[o] Cut-point 10 = exp(CUT_[10]) + cut-point 9 =    8.967
[o] Cut-point 11 = exp(CUT_[11]) + cut-point 10 =    10.46
```

```
[o]
[o]          estimate        s.e.     parameter
[o]      1     -7.244       1.810      CUT_[1]
[o]      2     0.5198      0.4433      CUT_[2]
[o]      3     0.4711      0.3886      CUT_[3]
[o]      4     0.1870      0.3859      CUT_[4]
[o]      5     0.9935      0.2127      CUT_[5]
[o]      6     0.1221      0.2436      CUT_[6]
[o]      7     0.2739      0.2891      CUT_[7]
[o]      8     0.2574      0.3587      CUT_[8]
[o]      9     0.6210      0.3712      CUT_[9]
[o]     10      1.230      0.4398      CUT_[10]
[o]     11     0.4040      0.5381      CUT_[11]
[o]     12     -7.258       1.836      CLASS__[1]
[o]     13     -3.599      0.8680      CLASS__[2]
[o]     14      0.000     aliased      CLASS__[3]
[o]     15      3.318      0.7811      CLASS__[4]
[o]     16      9.272       2.578      CLASS__[5]
[o]     17   -0.04848      0.4469      S_CLASS_[1]
[o]     18     0.3449      0.2865      S_CLASS_[2]
[o]     19      0.000     aliased      S_CLASS_[3]
[o]     20     0.3538      0.2750      S_CLASS_[4]
[o]     21     0.5625      0.4374      S_CLASS_[5]
[o] scale parameter 1.000
[o]
[o] Chi-square, sum(o-e)^2/e :    25.73
```

The estimates of the location effect parameters (β_1, β_2, *etc.*) are given by CLASS__[1], CLASS__[2], *etc.* The estimates of the scaling effect parameters (τ_1,

τ_2, *etc.*) are given by S_CLASS_[1], S_CLASS_[2], *etc.* The cut-point values indicate the position of the cut-points on the underlying probability distribution for class 3. The location effect parameters quantify the shift in this distribution for the other classes. The scaling effect parameters quantify the change in dispersion of this distribution for the other classes. As an example of the interpretation of the scaling effect parameters consider $\tau_5 = 0.56$. This value is with reference to $\tau_3 = 0$ and implies that the underlying distribution of responses to class 5 is more dispersed than the underlying distribution of responses to class 3.

Interpretation is possible directly in terms of the cut-points. For class 3 the cut-points are located at $(-7.24, \ldots, 10.46)$. For class 5 these cut-points are shifted by -9.27 giving $(-16.51, \ldots, 1.19)$ and then condensed by a factor $\exp(0.56) = 1.76$. This gives the cut-points for class 5 as $(-9.41, \ldots, 0.68)$. The point has been made in §2.3 that such a clustering of cut-points towards the middle of the scale yields larger fitted probabilities at the extreme categories. For class 5 the final cut-point is at 0.68 giving a fitted probability of 0.34 in category 12 for this class. Without taking the scaling effect into account, the cut-point is at 1.19 which gives a fitted probability of 0.23 in category 12 (although to make this comparison properly the model should be fitted without the scaling effect and then the cut-point calculated). Note that the observed proportion in this cell is $13/38 = 0.34$.

The deviance of 27.6 on 36 degrees of freedom suggests that the model fits the data well. The same conclusion is obtained from the χ^2 test of 25.7 on 36 degrees of freedom, although with either statistic the test validity is open to question because of the sparse nature of the data.

After fitting the model with the above code one could subsequently remove the scaling term from the model with

```
$delete s_model$
$use ORDINAL$
```

and to then test for the significance of the scaling term using the change in deviance. This test shows an increase in deviance of 4.4 on 4 degrees of freedom when the term is removed from the model. Thus there is no significant evidence to support the inclusion of the scaling term in the model. Investigation of the parameter estimates for the location effect of `class` shows them to be roughly equally spaced. Hence it may be sufficient to consider `class` as a variate rather than a factor. This may be tested by using

```
$ca class=class-3$
$var class$
$use ORDINAL$
```

and noting an increase in deviance of only 0.9 on 3 degrees of freedom. From this we conclude that there is no significant evidence of nonlinearity in the dependence on class.

This example fits model (4.15) to the data. To make a direct comparison with the results of Kijewski et al (1989) we need to fit the cumulative *probit* model with the same form of scaling term as given by model (4.16). The form of the cumulative logit model with a scaling term that is used by the macros is governed by a scalar argument (the fourth formal argument) to the macro `ORDINAL`. If this scalar is not provided or has value 1 then the form of the model in equation (4.15) is fitted. If the scalar has value 2 then the form of the model in equation (4.16) is used. To fit the cumulative probit model of Kijewski et al (1989) the following commands are used.

```
$link p$
$calc %f=2$
$use ORDINAL obs rows cats %f$
```

However when fitting this form of the model to this particular data the fitting procedure breaks down. This is an example where the user needs to provide initial parameter estimates that are more informative than those calculated automatically by ORDINAL. The vector of intital parameter estimates is given as the fifth formal argument to ORDINAL. In this case using

```
$ass p=-4,-0.1,-0.1,-0.1,-0.1,-0.1,-0.1,-0.1,-0.1,0.1,-0.1,
      -4,-2,0,2,4,0,0,0,0,0$
$use ORDINAL * * * * p$
```

enables ORDINAL to fit the model. The results from the general-purpose macros agree with the results given by Kijewski et al (1989).

4.3.3 The method of independent binomials

In §4.2.4 the method of independent binomials for fitting the cumulative logit model with location terms only is discussed. It is pointed out that a practical advantage of this fitting method is that a standard logistic regression routine is sufficient to implement the method. A further advantage to the method is that it is also possible to fit the cumulative logit model with location and scaling terms by repeated use of the logistic regression routine. To do this a successive relaxation approach to fitting models with non-linear systematic components is used. The term successive relaxation is used because at successive iterations a set of parameters is taken as constant to reduce the systematic component to a linear one.

In a generalized linear model the linear predictor η is a linear combination of explanatory effects which need to be estimated. For example

$$\eta_{ij} = \mathbf{x}_i\alpha + \mathbf{z}_j\beta \tag{4.18}$$

where $\mathbf{x}_i, \mathbf{z}_j$ are design vectors for covariates and α, β are the parameter vectors to be estimated. Suppose now that we have a model which satisfies all the conditions to be termed a generalized linear model except that the systematic component of the model is not a linear combination of explanatory variables. Suppose for example, that the form of this non-linearity is

$$\eta^{\star}_{ijk} = \mathbf{x}_i\alpha \times \mathbf{w}_k\gamma + \mathbf{z}_j\beta \tag{4.19}$$

where $\mathbf{x}_i, \mathbf{z}_j, \alpha, \beta$ are as above, $\mathbf{w}_k$ is another design vector and γ is another parameter vector to be estimated. The redundancy in the parameters α, γ can be removed by taking $\alpha_1 = 1$ for example. In order to estimate the parameters (α,β,γ) the following idea may be used. At the first step the values of α are taken to be equal to 1. This gives a linear predictor and the parameters β and γ may be straightforwardly estimated by a routine for generalized linear models. At the second step, γ is taken to be constant with values given by the estimates obtained in the first step. The vector α is no longer taken to be constant. This gives a linear predictor and estimates can be obtained for α and β. In the third step the vector α is again taken to be constant, this time with values given by the estimates obtained in step 2. Updated estimates for β and γ are obtained. The method "see-saws" between taking α to be constant while obtaining estimates for γ, and taking γ to be constant while obtaining estimates for α. This successive relaxation process is continued until convergence is obtained.

In §4.2.4 there is a discussion of how the method of independent binomials for fitting the cumulative logit model with location terms only may be implemented using a routine for logistic regression. This is possible because the generalized estimating equations in this case reduce to the score equations of an analysis of independent binomial observations. Now consider using a logistic regression routine to fit the cumulative logit model with location and scaling terms (model 4.16). The systematic component of the model has the same form as that in equation (4.19). Since the score equations for a generalized linear model are being used, it is possible to use the successive relaxation approach to fit the model with a scaling term. It is similarly possible to use this approach to fit the form of the model with location and scaling terms given in equation (4.15) although this is not pursued in this thesis.

While successive relaxation leads to satisfactory parameter estimates it must be noted that the standard errors associated with these estimates will be incorrect. At each iteration one set of parameters is held fixed while the other parameters are estimated. This means that the standard errors calculated for the estimated parameters at that iteration will be too small. In contrast, fitting the model with a scaling term by the method of maximum likelihood, implemented in the `ORDINAL` macros of Wolfe (1996), will give the correct standard errors.

Chapter 5

Subject effects in the cumulative logit model

5.1 Introduction

This chapter presents a novel approach to the analysis of longitudinal ordinal responses. The approach is motivated by data from the BT experiments which are described in Chapter 3. In these experiments subjects make repeated responses on a 5-category ordinal scale. Some exploratory analyses are performed on subject's patterns of response across the categories. The frequency of response category use for each experiment is also examined. From these exploratory analyses it is concluded that in addition to a subject-specific location term in the model it is necessary to investigate the significance of another subject-specific term; a subject-specific scaling term. This investigation is done by fitting a hierarchy of cumulative logit models. The cumulative logit model is chosen because the interpretation of a scaling term is simpler in this model than any of the other multinomial models for an ordinal response discussed in Chapter 2. This hierarchy of models is fitted to data from experiment E198 to fully illustrate the approach and then applied to a wide range of BT experiments.

A number of technical issues involved with the fitting of the model hierarchy are discussed. Two methods of fitting the cumulative logit model, maximum-likelihood and independent binomials (both discussed in Chapter 4), are used for applying the hierarchy to the BT data and the results from each are compared. There is also a comparison of two different forms of the cumuative logit model with a scaling term. These two forms are discussed in §2.3 and are introduced by McCullagh (1980) and Kijewski et al (1989). The application of the model hierarchy to ordinal scales with numbers of categories different to 5 is mentioned.

The approach to analysing longitudinal ordinal responses of fitting a hierarchy of models takes the correlation between intra-subject responses into account by including subject-specific explanatory variables in the model. A further consideration, leading to a more familiar modelling approach to repeated response data, is to assume that these subject-specific terms are random effects with a particular distribution. The cumulative logit model with random location effects, discussed in §2.2.3, is applied to the data and the results examined. The cumulative logit model with both random location and scaling effects is considered and the score equations for fitting the model are presented.

5.2 Exploratory data analysis

5.2.1 Differences in subjects' response patterns

Table 5.1 gives the observed response distributions for 4 particular subjects in experiment E198. Comparing the responses for subjects 9 and 10 we can immediately see that subject 10 tends to score higher on the scale than subject 9. This comparison can also be made via the observed cumulative proportions. The observed

Table 5.1: Examples of subject response patterns from experiment E198

Subject	Category				
	1	2	3	4	5
9	2	1	1	4	0
10	0	2	2	1	3
13	0	1	3	4	0
16	2	1	1	1	3

cumulative proportions in each category for subject 9 are $(\frac{2}{8}, \frac{3}{8}, \frac{4}{8}, 1, 1)$ and those for subject 10 are $(0, \frac{2}{8}, \frac{4}{8}, \frac{5}{8}, 1)$. For each of the categories, the observed cumulative proportion for subject 10 is not larger than that for subject 9.

The difference between subjects 13 and 16 is qualitatively different from the difference between subjects 9 and 10. Inspection of the observed responses for the former pair leads to the conclusion that subject 16 tends to spread their responses towards the extremes of the scale, relative to subject 13 who tends to cluster their responses in the central categories. Again this comparison can be made via observed cumulative proportions. For subject 13 these are $(0, \frac{1}{8}, \frac{4}{8}, 1, 1)$ and for subject 16 $(\frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, 1)$. For categories 1 and 2 the observed cumulative proportions for subject 16 are larger than those for subject 13. For category 4 the observed cumulative proportion for subject 16 is smaller than that for subject 13.

An advantage of thinking in terms of observed cumulative proportions is that models exist which relate underlying cumulative probabilities to co-variate effects. One such model is the cumulative logit model which is introduced in Chapter 2. To recap briefly on a motivation for this model, an unobservable continuous response, Y^*, is thought of as underlying the observed ordinal response, Y (with K categories). The ordinal response is observed as $Y = j$ if Y^* is in the interval $(\theta_{j-1}, \theta_j]$ where the θ_j parameters can be thought of as cut-points on the underlying continuum.

The form $Y_i^* = \mathbf{x}_i\beta + \epsilon_i$ is taken as a model for the underlying response. When the logistic distribution is assumed for the errors ϵ_i, this gives the cumulative logit model

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \mathbf{x}_i\beta. \tag{5.1}$$

The location effect parameters in the vector β allow for the cut-points to shift, *en bloc*, up or down the continuum. Hence, a subject-specific location term quantifies differences between subjects scoring higher or lower than each other. This is the kind of difference observed between subjects 9 and 10 in Table 5.1.

A more general form of the cumulative logit model is also defined in Chapter 2. This involves a scaling term in the model. There are two forms of the model with a scaling term described in §2.3. These are

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \frac{\theta_j - \mathbf{x}_i\beta}{\exp(\mathbf{z}_i\tau)} \tag{5.2}$$

which is presented by McCullagh (1980) and

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j \exp(\mathbf{z}_i\lambda) - \mathbf{x}_i\beta \tag{5.3}$$

which is discussed by Kijewski et al (1989). The scaling term quantifies how much responses are spread out (or concentrated) across the ordinal scale. Hence a subject-specific scaling term quantifies the type of difference between subjects observed in subjects 13 and 16 in Table 5.1.

5.2.2 Further motivation for the model

In experiment E198, 9 subjects did not use category 1, 15 subjects used category 1 once, 6 subjects used category 1 twice and 2 subjects used category 1 thrice. This is the frequency distribution for category 1 in experiment E198. The frequency distribution for each category is given in Appendix C for all experiments in Table 3.1. Examination of the marginal totals for each response category in every experiment shows that categories 1 and 5 are the least used in all experiments except F064 and F065.

In this section the frequency distributions of categories presented in Appendix C will be investigated in detail and hypotheses about subject response behaviour will be based upon them. Two points about Appendix C are noted before progressing.

1. For each experiment the frequency distribution of the response categories will depend on the set of conditions used in the experiment. If all the conditions are good, it is expected that the higher categories will be used very frequently. If there are no large differences between the conditions used in an experiment then the dispersion amongst responses is expected to be small.

2. In Chapter 3 the design of these experiments is introduced. The number of times each condition appears in each row in the experiment will influence how many times a subject uses each response category. Ideally the conditions will be orthogonal to rows so that every subject receives each condition the same number of times. This is a feature of the design of all the experiments considered here. However, in the *implementation* of some of these experiments, mistakes were made in the allocation of condition. The experiments not to have conditions orthogonal to rows are conversation experiments E199, E200, E211, E212 and E213. Of these, experiments E200 and E212 have only minor departures from orthogonality. Particular care must be taken for experiments E199, E211 and E213.

Further inspection of the marginal distributions of responses in Appendix C reveals that for conversation experiments E136-E213, category 4 is the modal response category and categories 1 and 5 have roughly equal totals, both less than the other category totals. It is of particular interest to note that the marginal total in category 1 and the marginal total in category 5 are sums of quite different frequency distributions. The responses in category 5 come from a small number of subjects using that category very frequently. The responses in category 1 accumulate from a lot of subjects using that category a small number of times.

As an example consider the marginal distribution of responses from experiment E140 in Appendix C. Category 4 is the mode in the marginal response totals. Categories 1 and 5 have the smallest values of the marginal totals, 18 and 32 respectively. The 18 responses in category 1 arise from 11 subjects using category 1 once, 2 subjects using category 1 twice and 1 subject using it thrice. The 32 responses in category 5 arise from 3 subjects using category 5 once, 1 subject using it twice, 2 subjects using it four times and 3 subjects using it five times or more. Although fewer subjects use category 5 than category 1 (9 to 14) the subjects who use category 5 use it very frequently.

In experiments E136-E213, it is surprising that category 5 is not used more often given that the mode for all these experiments is category 4. What is striking is the large number of subjects (roughly half on average) who do not give a response in category 5. The subjects who do not use category 5 are not all just low scorers; if they were then they would increase the frequency of their category 1 responses. From the tables in Appendix C it can be seen that no subject uses category 1 very frequently for this set of experiments. A hypothesis to explain this is that a lot of subjects are concentrating their responses in the middle of the scale. Another hypothesis is that many subjects use category 4 instead of category 5 in order to keep category 5 'up their sleeve'. The subjects who do use category 5 tend to use it frequently. Of these, some may be high scorers who correspondingly do not use the lower categories very frequently. However, not all of these subjects are just high scorers. Table 5.2 demonstrates that in experiment E200 there is only a small difference in the use of category 1 between those subjects who do not use category 5 and those subjects who do use category 5. From Table 5.2 it can also be observed that there are subjects who do not use either extreme category (4 subjects from 32), which means that they have concentrated their responses in middle categories.

There are subjects who use both categories 1 and 5 (10 subjects from 32), which means that they have spread their responses right across the response scale.

Table 5.2: Use of category 1 according to use of category 5 for E200

Use of cat. 5	Use of category 1			
	0 times	Once	Twice	Thrice
0 times	4	4	6	3
At least once	5	5	5	0

It must be noted regarding point 2 at the beginning of this section that close inspection of the subject response profiles in experiments E199, E211 and E213 shows that in these experiments the pattern of subject behaviour with the extreme categories seems to exist regardless of the effect that lack of orthogonality has on these response profiles. To assess this properly would require a model-based analysis, the proposal of which is the main purpose of this chapter.

The discussion for experiments E136-E213 does not immediately apply to the other experiments. First, consider experiment E216 which is the only other conversation experiment in Table 3.1. The mode for E216 is category 2 and the discussion used earlier is still relevant here but now the situation is reversed: subjects are reluctant to use category 1 even though category 2 is the mode. There still appear to be differences between subjects in their use of the scale. These differences cannot be solely explained by the idea of some subjects scoring lower than others: the idea of subjects bunching responses in the middle of the scale or pushing responses to both extremes must be considered.

The other 4 experiments are listening experiments. The marginal distributions of responses are different from those encountered in the conversation experiments. Experiments F064 and F065 have category 1 as the mode and there is nothing obviously peculiar about the subjects' use of categories in these 2 experiments. Experiments E247 and E264 have fewer responses in the extreme categories than in any of the middle three categories. Note that subjects make many more responses in listening experiments than in conversation experiments. The listening experiment with fewest responses per subject is F065 with 36 whereas the conversation experiments with the largest number of responses per subject are E216 with 18 and E136 with 15. It could be argued that the amount of difference between subjects in the spread of their responses diminishes with an increased ratio of number of responses to number of response categories. For the BT data an informal check is performed in §5.4 to see if this might be the case.

In conclusion, there is a group of experiments (group 1) which exhibit the same pattern of subject behaviour and frequency of category use. It consists of experiments E136, E139, E140, E198, E199, E200, E211, E212, E213. The patterns of subject behaviour and frequency of category use in the other experiments (E216, E247, E264, F064, F065) do not conform to the pattern observed in group 1, nor is there any homogeneity in these other patterns.

A final point to be made is repeated from §3.4.5. A full account of subject variablity must be allowed for before testing for carry-over effects. If this is not done then it is possible to detect a subject-specific effect that has not been allowed for and to mis-interpret it as a carry-over effect. It is possible that the significance of the carry-over parameter matrices in §3.4.4 may be due to the confounded subject-specific scaling term which has not been allowed for in the base model used in the analyses in §3.4.4. This is likely given that group 1 of experiments identified in this section correspond exactly with Group 1 in Figure 3.5.

5.3 Subject-specific fixed effects in the cumulative logit model

The cumulative logit model was discussed in the previous section. Two types of subject-specific effects, location and scale, can be estimated via this model and their interpretation was also discussed in the previous section. The cumulative logit model with a subject-specific location term is

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \theta_j - \alpha_s - \mathrm{x}_i\beta \tag{5.4}$$

where subjects are denoted by s and all other co-variates are denoted by i. Model (5.4) describes differences in location between subjects which are quantified by the location parameters α_s. Given values of the co-variates indexed by i the difference in location between subject 1 and subject 2 is $\alpha_1 - \alpha_2$.

A form of the cumulative logit model that allows for a completely general subject-specific effect is as follows,

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \theta_j + \theta_{js} - \mathrm{x}_i\beta \tag{5.5}$$

which is introduced in §2.2.3. The cut-point by subject interaction term θ_{js} allows as fully as possible for differences between subjects in their use of the response scale. Given values for co-variates indexed by i the difference between two subjects (say subject 1 and 2) is given by $\theta_{j1} - \theta_{j2}$. This difference depends on which cut-point is considered. In model (5.4) the difference between two sets of subjects' cut-points was a linear function of the location effects for those subjects, *i.e.* the difference in their α_s values, and not dependent on the cut-points. Hence model (5.4) can be thought of as a special case of model (5.5).

The cumulative logit model with a subject-specific scaling term of the form introduced by Kijewski et al (1989) is

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \theta_j \exp(\lambda_s) - \alpha_s - \mathrm{x}_i\beta \tag{5.6}$$

and this can also be thought of as a special case of model (5.5). Note that the form of the cumulative logit model with a scaling term (5.2) introduced by McCullagh (1980) is not a special case of model (5.5).

In model (5.6), given values for co-variates indexed by i, there is a set of cut-points associated with each subject. For example taking $\mathbf{x}_i\beta = 0$, the cut-points associated with subject 1 are

$$\theta_1 \exp(\lambda_1) - \alpha_1, \quad \theta_2 \exp(\lambda_1) - \alpha_1, \quad \theta_3 \exp(\lambda_1) - \alpha_1, \quad \theta_4 \exp(\lambda_1) - \alpha_1$$

and the cut-points associated with subject 2 are

$$\theta_1 \exp(\lambda_2) - \alpha_2, \quad \theta_2 \exp(\lambda_2) - \alpha_2, \quad \theta_3 \exp(\lambda_2) - \alpha_2, \quad \theta_4 \exp(\lambda_2) - \alpha_2$$

The differences between these two sets of cut-points are given by

$$\theta_j \left[\exp(\lambda_1) - \exp(\lambda_2)\right] - (\alpha_1 - \alpha_2) \tag{5.7}$$

for each cut-point θ_j. This can be seen as a particular structure for the interaction term of model (5.5). Also since the difference between the α_s parameters is a special case of formula (5.7), the model with a location term only (5.4), is a special case of the model with a scaling term (5.6). This gives the following hierarchy of models, the simplest model coming first,

$$\begin{aligned}
\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) &= \theta_j - \mathbf{x}_i\beta \\
\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) &= \theta_j - \alpha_s - \mathbf{x}_i\beta \\
\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) &= \theta_j \exp(\lambda_s) - \alpha_s - \mathbf{x}_i\beta \\
\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) &= \theta_j + \theta_{js} - \mathbf{x}_i\beta.
\end{aligned}$$

This hierarchy provides a natural way of testing for evidence of subject effects. Initially the model without any subject effects is fitted and the deviance and degrees of freedom noted. Then the model with the subject location term is fitted. The change in deviance and degrees of freedom can be attributed to the subject location term. The model with subject location and subject scaling terms is then fitted. The change in deviance and degrees of freedom can be attributed to the subject scaling term. Finally the model with the full interaction is fitted and the deviance and degrees of freedom noted. The change in deviance and degrees of freedom can be thought of as "residual" subject description. This describes any differences between subjects after location and scale have both been taken into account. All changes in deviance can be compared with an appropriate χ^2 distribution to assess the significance of components of the interaction between cut-points and subjects.

5.4 Analysis using subject-specific fixed effects

The hierarchy of cumulative logit models introduced in the previous section is used to analyse data from conversation experiments E136-E213 and listening experiments E247, E264, F064 and F065. The method of independent binomials described in §4.3.3 is used to fit the model in GLIM. Note that maximum likelihood estimation of models (5.4) and (5.6) is possible in the GLIM4 macros of Wolfe (1996). However, estimation of the full interaction term θ_{js} in model (5.5) is beyond the scope of these macros. Comparison is made in this section of the results from the two methods of fitting and the results are found to be very similar. Results obtained from fitting the hierarchy of models to 12 BT experiments are very heterogenous, so initially discussion will be limited to the analysis of experiment E198.

In experiment E198 the ordinal response is "Opinion score", a subject's rating of a telephone line transmission condition on the 5 category ordered scale given in Figure 3.2. The 'base' cumulative logit model includes the following set of explanatory variables: Condition, Column, End, and the interactions of Condition and Column with End. A full description of these explanatory variables and the experimental

Table 5.3: Deviance table for experiment E198

	Deviance	df
Base model	534.0	991
+ Subject location term	111.0	30
+ Subject scaling term	69.1	31
+ Residual subject description	73.8	62
Model with full subject description	280.1	868

design can be found in Chapter 3. An analysis of deviance table was constructed following the guidelines of §5.3 and is presented in Table 5.3. The base model has residual deviance of 534.0 on 991 degrees of freedom. The apparently large number of degrees of freedom (there are only 256 responses in the experiment) arises from the expansion to independent binary responses used to fit the model. Each 5-category ordinal response expands to 4 independent binary responses, thus for 256 ordinal responses there are 1024 binary responses.

There are 32 subjects in experiment E198, arranged in 16 pairs since it is a conversation experiment. When the subject location term is included in the model (*i.e.* Subject is included as an explanatory factor) the decrease in deviance is 111.0 on 30 degrees of freedom. This is highly significant when tested against the χ^2_{30} distribution. Note that one degree of freedom of the subject location term is used up by estimating End in the base model.

When the subject scaling term is included in the model it is necessary to use a more complex fitting process to fit the model by independent binomials. The successive relaxation approach to fitting models which contain a mixture of multiplicative and additive explanatory terms, described in §4.3.3, is employed. The reduction in deviance is 69.1 on 31 degrees of freedom. This is highly significant when tested against the χ^2_{31} distribution.

Finally a residual deviance for subject description is calculated by fitting the base model with an interaction term between Subject and Cut-point. The deviance for this model is 280.1 on 868 degrees of freedom. The deviance attributable to residual subject description is then calculated as 73.8 on 62 degrees of freedom, which is not significant if tested against the χ^2_{62} distribution.

The analysis of deviance tables for experiments E136-E140, E199-E213, E247, E264, F064 and F065 are collected in Table 5.4. The base model for limited duration experiments (E200, E212) is the same as that for E198. For unlimited duration experiments (E136, E139, E140, E199, E211, E213) there is also an explanatory variable for Picture-sets included in the base model. For the listening experiments (E247, E264, F064, F065) the base model includes the explanatory variables, Condition, Column, Sentence list, Listening level and Voice. There is agreement across the experiments in the significance of the subject location and scaling terms. Similarly there is agreement in the lack of significance of the residual subject description: experiment E140 is the only experiment that has a highly significant amount of subject-specific residual deviance and experiments E211 and F065 have borderline significant amounts of subject-specific residual deviance. Appendix D contains the subject-specific contributions to the residual deviance in experiment E140. Inspection of these indicates that the improvement in fit obtained from the full interaction model over the scaling model is largest for subjects 2, 9, 6, 3 and 4. Indeed the deviance attributed to the residual subject description (91.6 on 46 degrees of freedom) is no longer significant if the contribution of subjects 2 and 9 is subtracted. The subject response profiles in experiment E140 are also presented in Appendix D but it is difficult to deduce from these in what way the subjects with large contributions to the residual deviance differ from the other subjects in the experiment.

Table 5.4: Deviance tables for experiments E136-E140 E199-E213, E247, E264, F064 and F065

		E199	E211		E213		E200	E212
	df	dev.	dev.	df	dev.	df	dev.	dev.
Base model	984	482.4	466.2	982	512.2	991	551.8	569.8
+ Location	30	118.5	101.9	30	134.3	30	150.1	102.1
+ Scaling	31	72.9	97.4	31	88.3	31	60.5	45.4
+ Residual	62	62.4	83.0	62	60.2	62	48.5	66.7
Full model	861	228.6	183.9	859	229.4	868	292.7	355.6

		E136		E139	E140
	df	dev.	df	dev.	dev.
Base model	1725	1055.9	1092	677.2	620.0
+ Location	28	267.0	22	249.3	163.7
+ Scaling	29	95.9	23	131.2	64.2
+ Residual	58	65.7	46	53.4	91.6
Full model	1610	627.3	1001	243.3	300.5

		E247	E264		F064		F065
	df	dev.	dev.	df	dev.	df	dev.
Base model	5717	2909.8	2934	2837	1387.2	1687	809.9
+ Location	23	273.3	217	11	135.7	11	168.1
+ Scaling	23	68.9	121	11	26.9	11	36.0
+ Residual	46	56.1	52	22	16.9	22	35.4
Full model	5625	2511.5	2544	2793	1207.7	1643	570.4

Table 5.5: Number of responses per subject in the BT experiments

Experiment	Responses per subject
E198-E213	8
E139, E140	12
E136	15
F065	36
F064, E247, E264	60

It is possible to examine the results of the subject-specific hierarchy analysis to check whether there is any obvious change in strength of the scaling effect with increasing numbers of responses per subject. This concern is discussed in §5.2.2. The number of responses made by each subject in a range of BT experiments is given in Table 5.5. Simply examining the strength of the scaling effect across these experiments indicates that there is no evidence of a change in the effect with an increase in the number of responses per subject.

In all of the experiments considered in this thesis the ordinal response scale has 5 categories. Using the responses from experiment F100 it is possible to check

Table 5.6: Deviance table for experiment F100

	Deviance	df
Base model	652.5	827
+ Subject location term	181.4	15
+ Subject scaling term	36.5	15
+ Residual subject description	19.7	15
Model with full subject description	414.8	782

whether the scaling effect only exists for response scales with 5 categories. Experiment F100 (discussed briefly in §3.4.1) was performed with the same opinion score response scale as these other experiments, but the conditions used were such that no responses in category 1 were recorded in the experiment. The analysis in this case is effectively reduced to the analysis of a 4-category response. The hierarchy of models was fitted to test for subject effects, the only difference to the analyses in Table 5.4 being that the residual subject description has one less dimension for F100. Testing the deviance results in Table 5.6 against the χ^2_{15} distribution show that both subject location and subject scaling effects are significant and that residual subject description is not significant. So for an analysis of a 4-category response the subject effects are the same as in the analyses of a 5-category response. This comparison should be performed for an experiment where the actual response scale has 4 categories before any firm conclusions are drawn. Note that if the response scale only contains 3 ordered categories then there are only two dimensions to the full interaction of subject and cut-point. If subject location and subject scale are included in a model for a 3-category ordinal response, then there is no residual subject description.

The deviances given in Table 5.4 are calculated using the likelihood function for multinomial data (equation 4.11 in §4.2.4). These deviances are labelled $D_M(y;\hat{\pi}_{IB})$ (see §4.2.4) as they are values of the multinomial deviance function calculated with fitted values from the method of independent binomials. The deviance that is quoted by a binary logistic regression routine comes from the likelihood function for binary data (equation 4.10 in §4.2.4). This deviance is labelled $D_B(y;\hat{\pi}_{IB})$ and it can be seen from Table 5.7 how the value of this deviance is much greater than $D_M(y;\hat{\pi}_{IB})$ for the four experiments included in the table. Also note that the value of the deviance allotted to the subject-specific effects would be distorted if $D_B(y;\hat{\pi}_{IB})$ were used to make the deviance tables collected together in Table 5.4.

Table 5.7: Comparison of deviances in listening experiments

Model	Experiment	Independent binomials		Maximum likelihood	
				Form 1	Form 2
		$D_B(y;\hat{\pi}_{IB})$	$D_M(y;\hat{\pi}_{IB})$	$D_M(y;\hat{\pi}_{ML})$	$D_M(y;\hat{\pi}_{ML})$
Base	E247	3193.6	2910	2907	
	E264	3244.3	2934	2930	
	F064	1545.7	1387	1384	
	F065	900.3	809.9	807.9	
+Subject location	E247	2834.1	2636	2633.1	
	E264	2936.7	2717	2713.3	
	F064	1354.2	1252	1248.2	
	F065	675.3	641.8	639.9	
+Subject scale	E247	2762.8	2567.6	2563.3	2611.6
	E264	2806.6	2596.0	2591.0	2660.0
	F064	1319.7	1224.6	1220.5	1201.9
	F065	636.8	605.8	602.6	607.3
+Cut-point by Subject interaction	E247	2718.7	2511.5	-	
	E264	2764.2	2544.0	-	
	F064	1307.0	1207.7	-	
	F065	604.0	570.4	-	

With the exception of the model with the full interaction between cut-points and subjects, all of the models fitted here by the method of independent binomials can also be fitted by maximum likelihood in GLIM4 using the macros presented in Appendix B. Thus it is possible to compare the values of $D_M(y;\hat{\pi}_{IB})$ with the deviances obtained from maximum likelihood, $D_M(y;\hat{\pi}_{ML})$. This is done for experiments E247, E264, F064 and F065 in Table 5.7 resulting in the conclusion for these data that

$$D_M(y;\hat{\pi}_{IB}) \approx D_M(y;\hat{\pi}_{ML}). \tag{5.8}$$

Note that $D_M(y;\hat{\pi}_{IB})$ is always greater than $D_M(y;\hat{\pi}_{ML})$ as this latter deviance is minimised by maximum-likelihood fitting of the model.

The form of the cumulative logit model with a scaling term that was introduced by Kijewski et al (1989) (model 5.3) is used for the analyses in Table 5.4. This form is used because of its place in the model hierarchy for analysing subject-specific effects discussed in the previous section. This form of the model is labelled Form 1 in Table 5.7. When using the GLIM4 macros (Wolfe, 1996) it is straightforward to change from this form of the model to the alternative form introduced by McCullagh (1980) (model 2.7). This form is labelled Form 2 in Table 5.7. The deviances obtained for the two forms of the model may be compared. For experiments E247, E264, F064 and F065 this comparison shows that Form 1 of the model gives a better fit to the data in 3 experiments. For experiment F064 Form 2 gives a better fit.

It is worthwhile pursuing the comparison in experiment F064 further. As already noted Form 2 of the model with a scaling term gives a better fit to the data in this experiment than Form 1. The values of $D_M(y;\hat{\pi}_{ML})$ for the two forms are 1220.5 and 1201.9 respectively. In Table 5.4 the value of $D_M(y;\hat{\pi}_{IB})$ for the *full* model is 1207.7. This value is greater than the value of $D_M(y;\hat{\pi}_{ML})$ for this model but from the conclusion made in (5.8) this discrepancy would be expected to be small. So in this experiment Form 1 of the cumulative logit model with a scaling term seems to give a better fit to the data than the cumulative logit model with a cut-point by subject interaction term. This emphasises the need to consider both forms of the cumulative logit model with a scaling term.

We may conclude for this set of experiments that the majority of variation due to differences between subjects is contained in the subject location effect and the subject scaling effect, both of which are highly significant. Any remaining subject-specific effects are not significant.

5.5 Subject-specific random effects in the cumulative logit model

The analysis presented in the previous section is one approach to modelling longitudinal ordinal responses. The cumulative logit model is used for analysis and the correlation between intra-subject responses is modelled by the inclusion of a subject-specific location term and a subject-specific scaling term in the model. Two methods are used to fit this model: maximum-likelihood, and generalized estimating equations with an independence working correlation structure (independent binomials). In general, maximum-likelihood estimates are only consistent if the number of parameters in the model remains small as the number of sampling units in the data increases. In the analysis of the previous section the sampling units are the subjects. For each subject there are two subject-specific parameters in the model. Hence, in general, the maximum-likelihood estimates of parameters in the model with subject-specific location and scaling terms are not consistent as the number of subjects increases to infinity. Similarly the independent-binomials estimates will not, in general, be consistent. In the BT data there is a lot of information per subject as is demonstrated by the number of responses per subject displayed in Table 5.5. The amount of information per subject supports the use of the modelling approach with subject-specific location and scaling terms because it means that the total number of responses in the data is much larger than the total number of subjects. In a fixed effect analysis such as this it is effectively the responses that are taken to be the sampling units.

An alternative approach to subject-specific modelling which avoids the possibility of inconsistent estimation is to assume that the subject-specific terms in the model are random effects. Random location effects in the cumulative logit model are discussed in §2.2.3. The cumulative logit model with a subject-specific location random effect can be written as

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \theta_j - \omega_s - \mathrm{x}_i\beta \tag{5.9}$$

where ω_s is the random effect and is assumed to have some distribution, *e.g.* $\omega_s \sim \mathrm{N}(0, \sigma^2_\omega)$. An example of the use of this covariate structure for a binary logistic regression model (a simple case of model (5.9) with $K = 2$) is given by Zeger et al (1988). They note the tendency of the absolute values of the covariate parameters to increase with the variance in the random effect σ^2_ω.

Estimation of the parameters by maximum likelihood requires maximisation of the log-likelihood function

$$\log L = \sum_i^n \log\left\{\int_0^\infty \prod_{j=1}^K \pi_{ijs}^{m_{ijs}}\, \phi(\omega_s)d\omega_s\right\} \tag{5.10}$$

where $\phi(\omega_s)$ represents the normal density function. This can be done by Gaussian quadrature as described by Jansen (1990) and implemented in MIXOR by Hedeker (1993). An alternative approach is to use the Gibbs sampler which is implemented in the software BUGS (Spiegelhalter et al, 1994).

Estimation of the parameters in the cumulative logit model with random effects is also possible in principle using generalized estimating equations. Zeger et al (1988) demonstrate how this is done in the case of $K = 2$, the logistic regression model for binary responses. The use of GEE to fit the model for general K needs to be investigated. Using GEE to fit the model involves the use of a working correlation matrix. This working correlation may be parameterised appropriately to take intra-subject correlation into account. The random effect for subject location then describes the inter-subject differences that are manifested in the higher or lower scoring by subjects on the response scale.

In the BT data significant subject-specific differences have been found which are manifested firstly in the higher or lower scoring by subjects on the response scale and secondly by the clustering or spreading to the extremes of responses by the subjects. A new random effects model must be proposed to cater for these two types of subject difference. The model chosen is the cumulative logit model with form:

$$\log\left(\frac{\gamma_{ijs}}{1-\gamma_{ijs}}\right) = \theta_j \exp(\zeta_s \sigma_\zeta) - \omega_s \sigma_\omega - \mathbf{x}_i \beta \tag{5.11}$$

where ω_s is a random effect describing subject location, ζ_s is a random effect describing subject scaling and these two random effects are assumed to come from two independent standard normal distributions. The amount of random dispersion is measured by σ_ζ and σ_ω.

The parameters in the proposed model (5.11) can be estimated by maximum likelihood. Consider n_s repeated responses for each subject s (from a total of S subjects) such that $\sum_s n_s = n$. The individual cell probabilities for a given response i from subject s are

$$\begin{aligned} P(Y_{is} = j \mid \zeta, \omega, \beta) &= \pi_{ijs} \\ &= F(\theta_j \exp(\zeta\sigma_\zeta) - \omega\sigma_\omega - \mathbf{x}_i\beta) \\ &\quad - F(\theta_{j-1} \exp(\zeta\sigma_\zeta) - \omega\sigma_\omega - \mathbf{x}_i\beta) \end{aligned}$$

where F denotes the cumulative logistic distribution and the subscript s is dropped from ω_s and ζ_s for clarity.

The likelihood function for a given subject s with response vector Y_s is given by

$$l(Y_s|\omega, \zeta, \beta) = \prod_{i=1}^{n_s} \prod_{j=1}^{K} \pi_{ijs}^{m_{ijs}}$$

where

$$m_{ijs} = \begin{cases} 1 & \text{if} \quad Y_{is} = j \\ 0 & \text{otherwise.} \end{cases}$$

The marginal density of Y_s is then expressed as the integral

$$h(Y_s) = \int_\zeta \int_\omega l(Y_s|\omega, \zeta, \beta) g(\omega) g(\zeta) d\omega d\zeta$$

where $g(.)$ represents the standard normal density. The marginal log-likelihood can be written as

$$\log L = \sum_{s=1}^{S} \log h(Y_s)$$

and this can be differentiated with respect to the model parameters and set to zero in order to obtain the maximum-likelihood estimates. The differentiation of the marginal log-likelihood follows the chain rule. For example, differentiation with respect to some particular location parameter β_r is obtained from

$$\frac{\partial \log L}{\partial \beta_r} = \sum_{s=1}^{S} h^{-1}(Y_s) \frac{\partial h(Y_s)}{\partial l(Y_s|\omega, \zeta, \beta)} \quad \frac{\partial l(Y_s|\omega, \zeta, \beta)}{\partial \pi_{ijs}} \quad \frac{\partial \pi_{ijs}}{\partial \beta_r}.$$

For notational simplicity only part of this chain is considered, so for β_r the following derivative is obtained:

$$\frac{\partial h(Y_s)}{\partial \beta_r} = \int_\zeta \int_\omega \sum_{i=1}^{n_s} \sum_{j=1}^{K} m_{ijs} Q_j \, l(Y_s|\omega, \zeta, \beta) g(\omega) g(\zeta) \frac{\partial \eta_{ijs}}{\partial \beta_r} d\omega d\zeta \qquad (5.12)$$

where

$$Q_j = \frac{f(\eta_{ijs}) - f(\eta_{i(j-1)s})}{F(\eta_{ijs}) - F(\eta_{i(j-1)s})},$$

$$\eta_{ijs} = \theta_j \exp(\zeta \sigma_\zeta) - \omega \sigma_\omega - \mathbf{x}_i \beta$$

and

$$\frac{\partial \eta_{ijs}}{\partial \beta_r} = -\mathbf{x}_{ir}.$$

The expression for $\partial h(Y_s)/\partial\sigma_\omega$ differs from expression (5.12) only in that $\partial\eta_{ijs}/\partial\beta_r$ is replaced by

$$\frac{\partial\eta_{ijs}}{\partial\sigma_\omega} = -\omega.$$

For a particular cut-point parameter $\theta_{j'}$ the derivative

$$\frac{\partial h(Y_s)}{\partial\theta_{j'}} = \int_\zeta\int_\omega \sum_{i=1}^{n_s}\sum_{j=1}^{K} m_{ijs} R_{j'}\, l(Y_s|\omega,\zeta,\beta) g(\omega) g(\zeta)\exp(\zeta\sigma_\zeta) d\omega d\zeta$$

is obtained, where

$$R_{j'} = \begin{cases} \frac{f(\eta_{ijs})}{F(\eta_{ijs})-F(\eta_{i(j-1)s})} & \text{if} \quad j' = j \\ \frac{-f(\eta_{i(j-1)s})}{F(\eta_{ijs})-F(\eta_{i(j-1)s})} & \text{if} \quad j' = j-1. \end{cases}$$

Finally, differentiation with respect to the dispersion parameter of the subject-specific random scaling effect yields

$$\frac{\partial h(Y_s)}{\partial\sigma_\zeta} = \int_\zeta\int_\omega \sum_{i=1}^{n_s}\sum_{j=1}^{K} m_{ijs} T_j\, l(Y_s|\omega,\zeta,\beta) g(\omega) g(\zeta)\zeta\exp(\zeta\sigma_\zeta d\omega d\zeta)$$

where

$$T_j = \frac{\theta_j f(\eta_{ijs}) - \theta_{j-1} f(\eta_{i(j-1)s})}{F(\eta_{ijs}) - F(\eta_{i(j-1)s})}.$$

The score equations that are defined by

$$\frac{\partial \log L}{\partial\Phi} = \sum_{s=1}^{S} h^{-1}(Y_s)\frac{\partial h(Y_s)}{\partial\Phi}$$

where $\Phi = (\theta, \beta, \sigma_\omega, \sigma_\zeta)$ have not been implemented in any general-purpose software routine. It should be possible to extend the MIXOR program of Hedeker (1993) to cope with this second type of random effect. Unfortunately time ran out before this could be attempted as a contribution to this thesis. The development of general-purpose software to fit the cumulative logit model with both random location and scale effects would be of practical value. For the BT data analysed in this thesis the significance of subject-specific location and scaling terms has been established. There is an argument for including terms in the model for these effects that are random as opposed to fixed. For the BT data, at least, general-purpose software for fitting random location and scaling terms in a cumulative logit model would be of great use.

5.6 Analysis using subject-specific random effects

A comparison between two types of subject-specific analysis is now made. The first type of analysis includes the subject-specific location term as a fixed effect term in the model (equation 5.4). The second type of analysis considers the subject location term to be a random effect and a variance term for this random effect is included in the model (equation 5.9). Results are presented for a cumulative logit model with a subject location term fitted to data from experiment E198.

To make the comparison of the two types of modelling approach only Condition is included as an explanatory term in the design vector $\mathbf{x}_i$. A full list of the parameter estimates and standard errors are given in Appendix E and the estimates of the Condition effects for both models are also given in Table 5.8. The cut-points in the fixed effects model relate to subject 1. In the random effect model the cut-points relate to an average subject. Hence, it is not sensible to make a direct comparison of the cut-points from the two models. The estimate of the random effect variance term is $\hat{\sigma}_\omega = 1.11$. In the fixed effects model (5.4) the estimates of subject-specific location terms vary between $\hat{\alpha}_{15} = -4.40$ and $\hat{\alpha}_{27} = 1.81$ relative to subject 1. The standard deviation in these individual fixed effects estimates is 1.46.

The deviance for the fixed effects model is 446.1 on 918 degrees of freedom and for the random effect model it is 519.7 on 948 degrees of freedom. The reduction in deviance from a model with no subject location term to a model with either fixed or random subject location can be tested. The reduction in deviance for the fixed effect subject location term is 102.3 on 31 degrees of freedom and the reduction in deviance for the random effect subject location term is 28.6 on 1 degree of freedom. When compared to the relevant χ^2 distributions both of these are highly significant reductions in deviance.

Table 5.8: Parameter estimates and standard errors of Condition effect for two approaches to subject-specific location modelling of experiment E198

	Type of modelling approach			
	Fixed effects		Random effect	
	p.e.	s.e	p.e.	s.e.
Condition 1	0	-	0	-
Condition 2	-0.39	(0.50)	-0.36	(0.75)
Condition 3	-0.64	(0.50)	-0.58	(0.59)
Condition 4	-7.29	(0.74)	-6.67	(1.07)
Condition 5	-0.34	(0.50)	-0.33	(0.90)
Condition 6	-2.94	(0.54)	-2.68	(0.81)
Condition 7	-6.39	(0.69)	-5.89	(0.94)
Condition 8	-9.26	(0.84)	-8.45	(0.98)

The parameter estimates of the condition effects (given in Table 5.8) show that the two approaches to modelling subject location yield similar estimates of the other explanatory variables in the model. The estimates from the random effect model are closer to zero than those from the fixed effect model. Despite this shrinkage towards zero the estimates from the random effect model have larger standard errors than those from the fixed effect model. This seems to indicate that the fixed effect model gives a false sense of accuracy in the estimation of these covariate effects. This might be explained by the fact that in the fixed effect analysis although the sampling units are the subjects, effectively the individual responses are taken as the sampling units, implying more independent information than actually exists. In the random-effect analysis the true situation of the subjects being the sampling units is acknowledged.

5.7 Conclusions

The proposed random effects model with random subject-specific location and scaling terms (5.11) deserves further investigation. It can be seen from the analysis in the previous section how the model with a random subject-specific location effect only (5.9) applies to the BT data. An advantage to using these random effects models as opposed to their fixed effects counterparts is that the theoretical problem of inconsistent estimation discussed in §5.5 is not encountered. As noted in that section, the use of the fixed effect approach is supported by the large number of responses per subject. This is borne out in the similarity in the parameter estimates obtained from the fixed and random effect approaches investigated in §5.6.

The interpretation of the covariate terms (β) is the same in both approaches. However, prediction is different for the two approaches. In a random effects model prediction is made for an 'average' subject. In a fixed effects model prediction is made for particular subjects involved in the experiment.

Finally, a drawback to the random effects approach is the requirement of assuming a distributional shape for each random effect. For the work in this chapter, independent normal distributions are assumed for the two random effects. A sensitivity analysis needs to be performed to check whether the results depend on this choice of distributional shape.

Chapter 6

Comparing different methods of fitting the cumulative logit model

6.1 Introduction

In Chapter 4 several different methods of fitting the cumulative logit model are introduced. This chapter is concerned with comparisons between these different methods of fitting the cumulative logit model. The methods investigated are maximum likelihood, generalized estimating equations (GEE) using the true correlation matrix for multinomial variables, GEE with Clayton's working correlation matrix, GEE with a working correlation matrix based on equal marginal probabilities, and 'independent binomials' (*i.e.*, GEE with an independence working correlation matrix). As noted in §4.2.3 the first two of these are equivalent. One other approach for modelling an ordinal response is considered. This approach involves dichotomising the ordinal scale and modelling the resultant data with a logistic regression model for binomial observations.

The initial comparisons are empirical in nature. These serve to compare different methods in real-life data modelling. Some theoretical comparisons are made by investigating the asymptotic relative efficiency of estimates from different methods. Theoretical comparisons are not included for Clayton's GEE method because the working correlation matrix used in this method is data-dependent. As an alternative, theoretical comparisons are made for GEE with an equal-margins working correlation matrix. This method is similar in spirit to Clayton's GEE method and in many practical analyses the two methods will give similar results.

The comparisons made in this chapter were prompted by practical experience of using the different methods of fitting the cumulative logit model. This experience indicates that there is not much gain to be made by using GEE with either Clayton's or equal-margins working correlation matrices over the method of independent binomials. In most instances of practical data analysis encountered during this work the estimates obtained from independent binomials are similar to those obtained from maximum likelihood. For almost all the instances encountered during this work, when the estimates are compared with their standard errors (by using a t-test say) then, although the exact results are different, the qualitative conclusions drawn from these results are the same for the two fitting methods.

6.2 Empirical comparison

6.2.1 Example 1

Consider the data analysed by Clayton (1992) and presented in Table 6.1. The ordinal reponse variable is the time taken to fall asleep for 239 subjects. The subjects are grouped according to whether they had received an active treatment or a placebo. The data are analysed by fitting the cumulative logit model. Four different methods are used to fit the model: maximum likelihood, Clayton's GEE method, GEE with correlation based on equal marginal probabilities, and independent binomials. The

Table 6.1: Time to falling asleep for 239 subjects

	Time (minutes)			
Treatment	< 20	20–30	30–60	> 60
Active	40	49	19	11
Placebo	31	29	35	25

model is

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \beta_i \tag{6.1}$$

where the parameter β_i denotes the treatment effect. The convention $\beta_1 = 0$ is employed. The maximum likelihood estimates are obtained from GLIM4 using the `ORDINAL` macros (see §4.2.2). The estimates from the three GEE methods are obtained from the `gee` function in S-Plus.

Table 6.2: Parameter estimates and standard errors for treatment effect β_2.

Method of fitting the cumulative logit model							
Maximum likelihood		Clayton's GEE		Equal-margins GEE		Independent binomials	
p.e.	s.e	p.e.	Robust s.e.	p.e.	Robust s.e.	p.e.	Robust s.e.
0.761	(0.238)	0.762	(0.239)	0.770	(0.239)	0.789	(0.240)

The parameter estimates and standard errors for β_2 from the four analyses are given in Table 6.2. In this example the different estimates of the treatment effect parameter obtained from fitting the model by maximum likelihood and by Clayton's GEE method agree to two decimal places, 0.76. The estimate is slightly larger, 0.77, when the model is fitted by equal-margins GEE and larger again, 0.79, when the model is fitted by independent binomials. The standard error of the maximum-likelihood estimate agrees to two decimal places with the robust standard errors of the estimates from all three GEE methods. A test of significance of the treatment effect shows that the parameter is significantly different from zero whichever method of fitting the model is used.

6.2.2 Example 2

The data in this example come from experiment E198 in the BT set of experiments. The same four methods of fitting the cumulative logit model used in Example 1 are employed. The model fitted to these data is

$$\log\left(\frac{\gamma_{ijk}}{1-\gamma_{ijk}}\right) = \theta_j - \beta_i - \mathbf{x}_k\delta. \tag{6.2}$$

where β_i denote the condition contrasts, and the explanatory variables in the design vector $\mathbf{x}_k$ are End and Column. In this example the parameters of interest are β_i. As in the previous example the maximum likelihood estimates are obtained using GLIM4 and the estimates from the three GEE methods are obtained using S-Plus.

The estimates of β_i from the four methods of fitting the cumulative logit model are reported in Table 6.3. From this table it can be seen that the estimates obtained from Clayton's GEE and equal-margins GEE methods are similar to each other. The largest absolute difference between these two sets of parameters is 0.23 for β_7 and the largest relative difference is that for β_5. The standard errors for these two sets of estimates agree closely except for β_7, the parameter with the largest absolute difference in the parameter estimates.

It also seems reasonable to say that the estimates obtained from fitting the model by maximum likelihood and independent binomials agree closely. The largest absolute difference between these two sets of parameter estimates is 0.21 for β_7 and the largest relative difference is that for β_2. The standard errors are slightly different but the difference is not systematic. There is a systematic difference between the

Table 6.3: Parameter estimates and standard errors for Condition contrasts in experiment E198

	Method of fitting the cumulative logit model							
	Maximum likelihood		Clayton's GEE		Equal-margins GEE		Independent binomials	
				Robust		Robust		Robust
	p.e.	s.e	p.e.	s.e.	p.e.	s.e.	p.e.	s.e.
$-\hat{\beta}_1$	0	-	0	-	0	-	0	-
$-\hat{\beta}_2$	0.29	(0.47)	0.20	(0.47)	0.22	(0.48)	0.24	(0.42)
$-\hat{\beta}_3$	0.48	(0.47)	0.20	(0.50)	0.18	(0.51)	0.55	(0.49)
$-\hat{\beta}_4$	5.74	(0.64)	5.51	(0.76)	5.53	(0.78)	5.73	(0.60)
$-\hat{\beta}_5$	0.37	(0.47)	0.36	(0.55)	0.41	(0.55)	0.40	(0.51)
$-\hat{\beta}_6$	2.33	(0.50)	2.32	(0.55)	2.42	(0.57)	2.29	(0.50)
$-\hat{\beta}_7$	5.13	(0.61)	5.73	(0.68)	5.96	(0.77)	4.92	(0.70)
$-\hat{\beta}_8$	7.21	(0.69)	7.12	(0.77)	7.14	(0.79)	7.21	(0.64)

standard errors obtained from maximum likelihood and the standard errors obtained from GEE using either Clayton's or equal-margins working correlation matrices. The standard errors of the maximum-likelihood estimates are smaller than the standard errors of the estimates from either of these GEE methods.

6.3 Asymptotic relative efficiency comparisons

6.3.1 Introduction

Asymptotic relative efficiency (ARE) is defined as follows. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two consistent estimators of θ. The asymptotic relative efficiency of $\hat{\theta}_1$ with respect to

$\hat{\theta}_2$ is the limit

$$\lim_{n \to \infty} \frac{\mathrm{var}(\hat{\theta}_2)}{\mathrm{var}(\hat{\theta}_1)}$$

where n is the sample size. If this limit is less than 1 we say that $\hat{\theta}_1$ is asymptotically less efficient than $\hat{\theta}_2$. If the limit is exactly 1, the two estimators are asymptotically equally efficient to first order approximation.

In this section ARE comparisons are made of the parameter estimates obtained from the different methods of fitting the cumulative logit model that are discussed in §6.1. Note that consistency is a general property of maximum-likelihood estimates under the standard regularity conditions. Note that, again subject to standard regularity conditions, consistency is also a general property of GEE estimates irrespective of the choice of working correlation matrix.

The ARE comparisons will be made for the simple case of a cumulative logit model with one binary covariate. This is

$$\left.\begin{aligned} \log\left(\frac{\gamma_{1j}}{1-\gamma_{1j}}\right) &= \theta_j \\ \log\left(\frac{\gamma_{2j}}{1-\gamma_{2j}}\right) &= \theta_j - \Delta \end{aligned}\right\} \quad j = 1, \ldots, K-1 \tag{6.3}$$

for a K-category ordinal response, where Δ is the covariate effect parameter of interest. It is assumed that the number of responses in group 1 (n_1) is equal to the number of responses in group 2 (n_2) and are denoted $n_1 = n_2 = n$. This model can be thought of as comparing the multinomial distribution of probabilities in the $2 \times K$ contingency table in Table 6.4. To simplify the ARE comparisons it is necessary to make assumptions about the model and in the comparisons made in this chapter it is assumed that the cell probabilities in the first row of the table are equal. That is

$$\pi_{11} = \pi_{12} = \cdots = \pi_{1K} = \frac{1}{K}.$$

This assumption is referred to as 'equal πs in group 1'. It implies that the cut-points

Table 6.4: Cell probabilities in a $2 \times K$ contingency table

Row	Category				
	1	2	3	$\cdots$	K
1	π_{11}	π_{12}	π_{13}	$\cdots$	π_{1K}
2	π_{21}	π_{22}	π_{23}	$\cdots$	π_{2K}

in model (6.3) are

$$\begin{aligned}\theta_j &= \log\left(\frac{\gamma_{1j}}{1-\gamma_{1j}}\right) \\ &= \log\left(\frac{j/K}{1-j/K}\right) \\ &= \log\left(\frac{j}{K-j}\right).\end{aligned}$$

Note, however, that it is not assumed that the cut-points are known in the estimation of model (6.3) and, hence, in the calculation of the asymptotic relative efficiencies for the estimates of Δ.

6.3.2 Variance of maximum-likelihood estimator of Δ

In §4.2.2 the method of fitting the cumulative logit model by maximum likelihood is discussed. The estimate of Δ in model (6.3) obtained from fitting the model by maximum likelihood is denoted $\hat{\Delta}_{ML}$. A standard result from maximum-likelihood theory for the asymptotic variance of maximum-likelihood estimators is

$$\mathrm{avar}(\hat{\beta}_{ML}) = -\left[E\left(\frac{\partial^2 l}{\partial \beta^2}\right)\right]^{-1}$$

where l is the log-likelihood function for the model. The following result for the general cumulative logit model (4.1)

$$-E\left(\frac{\partial^2 l}{\partial \beta^2}\right) = \sum_i n_i \left(\sum_{j=1}^{K-1}\left[\frac{\gamma_{i(j+1)}}{\gamma_{ij}\pi_{i(j+1)}}\right] a_{ij}^2\right) \tag{6.4}$$

is given by Agresti (1984), Appendix B.3, with

$$a_{ij} = \gamma_{ij}(1-\gamma_{ij})x_{ij} - \gamma_{ij}(1-\gamma_{i(j+1)})x_{i(j+1)}.$$

Equation (6.4) can be used to obtain

$$\begin{aligned} \operatorname{avar}(\hat{\Delta}_{ML}) &= \sum_{i=1}^{2}\left(n_i \sum_{j=1}^{K-1} \gamma_{ij}\gamma_{i(j+1)}\pi_{i(j+1)}\right)^{-1} \\ &= \sum_{i=1}^{2}\frac{3}{n_i}\left(1-\sum_{j=1}^{K}\pi_{ij}^3\right)^{-1}. \end{aligned} \tag{6.5}$$

See McCullagh & Nelder (1989), Excercise 5.3 for alternative forms of the summation of probabilities in equation (6.5). In the case of $K = 3$, $n = n_1 = n_2$ and when the equal-πs-in-group-1 assumption is in force, equation (6.5) gives

$$\operatorname{avar}(\hat{\Delta}_{ML}) = \frac{3}{n}\left[\frac{9}{8} + \left(1-\sum_{j=1}^{3}\pi_{2j}^3\right)^{-1}\right]. \tag{6.6}$$

6.3.3 Variance of the independent-binomials estimator of Δ

In §4.2.3 the method of fitting the cumulative logit model by generalized estimating equations (GEE) is discussed. The notion of a working correlation matrix, $R(\alpha)$, described by the unknown parameters α is introduced. The asymptotic variance of GEE estimates is labelled $\operatorname{avar}(\hat{\beta})$. The information sandwich formula

$$\operatorname{acov}(\hat{\beta}) = (D^T V^{-1} D)^{-1} D^T V^{-1} \operatorname{cov}(y^\dagger) V^{-1} D (D^T V^{-1} D)^{-1}, \tag{6.7}$$

where $\operatorname{cov}(y^\dagger)$ is the known form of covariance matrix of the expanded binary responses (see definition 4.4), is used to calculate the asymptotical covariance matrix of the parameter estimates. The diagonal elements of the matrix in equation (6.7) give $\operatorname{avar}(\hat{\beta})$. Note that the matrix $D = AX$ contains the derivatives $\partial\gamma_{ij}/\partial\beta_r$ where $A = \operatorname{diag}\{n_i\gamma_{ij}(1-\gamma_{ij})\}$, X is the model design matrix and

$$V = A^{\frac{1}{2}} R(\alpha) A^{\frac{1}{2}}. \tag{6.8}$$

The method of independent binomials employs the identity matrix as a working correlation matrix. We denote the estimate of Δ in model (6.3) as $\hat{\Delta}_{IB}$ when fitting the model by the method of independent binomials. The asymptotic variance of this estimate is given by the final diagonal element of

$$\text{acov}(\hat{\theta}_{IB}, \hat{\Delta}_{IB}) = (X^T A X)^{-1} X^T \text{ cov}(y^\dagger) X (X^T A X)^{-1}. \tag{6.9}$$

For general K and $n_1 = n_2 = n$, the value of $\text{avar}(\hat{\Delta}_{IB})$ is given by the equation

$$\begin{aligned} \text{avar}(\hat{\Delta}_{IB}) &= n\alpha^2 \left[\sum_{j=1}^{K-1} \left(\gamma_{2j}(K - j - \sum_{k=1}^{K-1} \gamma_{2k}) + \sum_{k=1}^{j-1} \gamma_{2k} \right) \right. \\ &\qquad \left. + \frac{\gamma_{2j}(1-\gamma_{2j})}{\gamma_{1j}(1-\gamma_{1j}) + \gamma_{2j}(1-\gamma_{2j})} s_{jk} \right] \end{aligned} \tag{6.10}$$

where

$$\alpha = \frac{1}{n} \left(\sum_{j=1}^{K-1} \frac{\gamma_{1j}(1-\gamma_{1j})\gamma_{2j}(1-\gamma_{2j})}{\gamma_{1j}(1-\gamma_{1j}) + \gamma_{2j}(1-\gamma_{2j})} \right)^{-1}$$

is the final diagonal element of $(X^T A X)^{-1}$ and

$$\begin{aligned} s_{jk} &= \gamma_{2j}(1-\gamma_{2j}) - 2\left[\gamma_{2j} \left(K - j - \sum_{k=1}^{K-1} \gamma_{2k} \right) + \sum_{k=1}^{j-1} \gamma_{2k} \right] \\ &\quad + 2 \sum_{k=j+1}^{K-1} \frac{\gamma_{2k}(1-\gamma_{2k}) \left[\gamma_{1j}(1-\gamma_{1k}) + \gamma_{2j}(1-\gamma_{2k}) \right]}{\gamma_{1k}(1-\gamma_{1k}) + \gamma_{2k}(1-\gamma_{2k})}. \end{aligned}$$

The asymptotic relative efficiency comparisons made in this chapter are made for responses with 3 categories, ($K = 3$), under the equal-πs-in-group-1 assumption and with equal group sizes $n_1 = n_2 = n$. In this case, the result for the asymptotic variance of the independent-binomials estimator of Δ is

$$\begin{aligned} \text{avar}(\hat{\Delta}_{IB}) &= n\alpha^2 \left[\gamma_{21}(2 - \gamma_{21} - \gamma_{22}) + \frac{\gamma_{21}(1-\gamma_{21})}{\frac{2}{9} + \gamma_{21}(1-\gamma_{21})} s_{1k} \right. \\ &\qquad \left. + \gamma_{22}(1 - \gamma_{21} - \gamma_{22}) + \gamma_{21} + \frac{\gamma_{22}(1-\gamma_{22})}{\frac{2}{9} + \gamma_{22}(1-\gamma_{22})} s_{2k} \right] \end{aligned} \tag{6.11}$$

where

$$\alpha = \frac{1}{n} \left[\frac{\frac{2}{9}\gamma_{21}(1-\gamma_{21})}{\frac{2}{9} + \gamma_{21}(1-\gamma_{21})} + \frac{\frac{2}{9}\gamma_{22}(1-\gamma_{22})}{\frac{2}{9} + \gamma_{22}(1-\gamma_{22})} \right]^{-1}$$

and

$$
\begin{aligned}
s_{1k} &= -\gamma_{21}(1-\gamma_{21}) - 2\left[\gamma_{21}(1-\gamma_{22}) + \frac{\gamma_{22}(1-\gamma_{22})\left[\frac{1}{9}+\gamma_{21}(1-\gamma_{22})\right]}{\frac{2}{9}+\gamma_{22}(1-\gamma_{22})}\right] \\
s_{2k} &= -\gamma_{22}(1-\gamma_{22}) - 2\gamma_{21}(1-\gamma_{22}).
\end{aligned}
$$

6.3.4 The GEE-with-true-correlation estimator of Δ

If the cumulative logit model is fitted by GEE with the true correlation matrix in place of a working correlation matrix then the estimation is equivalent to maximum-likelihood estimation. This is discussed in §4.2.3. The explicit equivalence in the case of $K = 3$ for model (6.3) is now shown. The GEEs are given by

$$X^T A \operatorname{cov}^{-1}(y^\dagger)(y^\dagger - \gamma) = 0$$

which reduce to the simultaneous equations

$$
\begin{pmatrix}
\frac{\gamma_{12}(1-\gamma_{11})}{\gamma_{12}-\gamma_{11}} & \frac{-\gamma_{11}(1-\gamma_{11})}{\gamma_{12}-\gamma_{11}} & \frac{\gamma_{22}(1-\gamma_{21})}{\gamma_{22}-\gamma_{21}} & \frac{-\gamma_{21}(1-\gamma_{21})}{\gamma_{22}-\gamma_{21}} \\
\frac{-\gamma_{12}(1-\gamma_{12})}{\gamma_{12}-\gamma_{11}} & \frac{\gamma_{12}(1-\gamma_{11})}{\gamma_{12}-\gamma_{11}} & \frac{-\gamma_{22}(1-\gamma_{22})}{\gamma_{22}-\gamma_{21}} & \frac{\gamma_{22}(1-\gamma_{21})}{\gamma_{22}-\gamma_{21}} \\
0 & 0 & \gamma_{22} & 1-\gamma_{21}
\end{pmatrix}
\begin{pmatrix}
y^\dagger_{11} - \gamma_{11} \\
y^\dagger_{12} - \gamma_{12} \\
y^\dagger_{21} - \gamma_{21} \\
y^\dagger_{22} - \gamma_{22}
\end{pmatrix}
=
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.
$$

Using the relationship $y_{ij} = y^\dagger_{ij} - y^\dagger_{ij-1}$ the maximum-likelihood equations are

$$
\begin{aligned}
\frac{\partial l}{\partial \theta_1} &= \gamma_{11}(1-\gamma_{11})\left[\frac{y^\dagger_{11}}{\pi_{11}} - \frac{y^\dagger_{12}-y^\dagger_{11}}{\pi_{12}}\right] \\
&\quad +\gamma_{21}(1-\gamma_{21})\left[\frac{y^\dagger_{21}}{\pi_{21}} - \frac{y^\dagger_{22}-y^\dagger_{21}}{\pi_{22}}\right] = 0 \\
\frac{\partial l}{\partial \theta_2} &= \gamma_{12}(1-\gamma_{12})\left[\frac{y^\dagger_{12}-y^\dagger_{11}}{\pi_{12}} - \frac{1-y^\dagger_{12}}{1-\gamma_{12}}\right] \\
&\quad +\gamma_{22}(1-\gamma_{22})\left[\frac{y^\dagger_{22}-y^\dagger_{21}}{\pi_{22}} - \frac{1-y^\dagger_{22}}{1-\gamma_{22}}\right] = 0 \\
\frac{\partial l}{\partial \Delta} &= \gamma_{21}(1-\gamma_{21})\left[\frac{y^\dagger_{21}}{\pi_{21}} - \frac{y^\dagger_{22}-y^\dagger_{21}}{\pi_{22}}\right]
\end{aligned}
$$

$$+\gamma_{22}(1-\gamma_{22})\left[\frac{y_{22}^{\dagger}-y_{21}^{\dagger}}{\pi_{22}}-\frac{1-y_{22}^{\dagger}}{1-\gamma_{22}}\right]=0$$

and it is straightforward to show that these two sets of estimating equations are equal. For example, the third and final equation in both sets reduces to the simple form

$$(1-\gamma_{21})y_{22}^{\dagger}-\gamma_{22}(1-y_{21}^{\dagger})=0.$$

6.3.5 Variance of the GEE-with-equal-margins-correlation estimator of Δ

The method of fitting the cumulative logit model by GEE with a working correlation matrix based on equal marginal probabilities is discussed in §4.2.3. This working correlation matrix is symmetric and block diagonal with the elements in every block being given by

$$\begin{aligned} R_{jj}(\alpha) &= 1 & 1 \le j \le K-1 \\ R_{jk}(\alpha) &= \sqrt{\frac{j(K-k)}{k(K-j)}} & 1 \le j < k \le K-1. \end{aligned}$$

This working correlation matrix is used to obtain a covariance matrix for equal-margin GEE, labelled V_E, from the familiar relationship given in equation (6.8).

The estimate of Δ in model (6.3) obtained from this fitting method is denoted by $\hat{\Delta}_E$ and the asymptotic variance of this estimate is denoted by avar($\hat{\Delta}_E$). The information sandwich formula gives

$$\text{acov}(\hat{\theta}_E,\hat{\Delta}_E)=(X^TAV_E^{-1}AX)^{-1}X^TAV_E^{-1}\,\text{cov}(y^{\dagger})V_E^{-1}AX(X^TAV_E^{-1}AX)^{-1} \quad (6.12)$$

the final diagonal element of which is avar($\hat{\Delta}_E$).

In the case of $K=3$, $n_1=n_2=n$ and under the equal-πs-in-group-1 assumption the components of the information sandwich (equation 6.12) are

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

$$A = n \begin{pmatrix} \frac{2}{9} & 0 & 0 & 0 \\ 0 & \frac{2}{9} & 0 & 0 \\ 0 & 0 & \gamma_{21}(1-\gamma_{21}) & 0 \\ 0 & 0 & 0 & \gamma_{22}(1-\gamma_{22}) \end{pmatrix},$$

$$R^{-1}(\alpha) = \frac{4}{3} \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

and

$$\mathrm{cov}(y^\dagger) = n \begin{pmatrix} \frac{2}{9} & \frac{1}{9} & 0 & 0 \\ \frac{1}{9} & \frac{2}{9} & 0 & 0 \\ 0 & 0 & \gamma_{21}(1-\gamma_{21}) & \gamma_{21}(1-\gamma_{22}) \\ 0 & 0 & \gamma_{21}(1-\gamma_{22}) & \gamma_{22}(1-\gamma_{22}) \end{pmatrix}.$$

The explicit expression for the asymptotic variance of $\hat{\Delta}_E$ in this case is cumbersome and is relegated to Appendix F.

6.3.6 Variance of the dichotomisation estimator of Δ

Consider dichotomising the K-category ordinal response between categories m and $m+1$. The resultant binary observations can then be analysed by logistic regression. The estimate of Δ in model (6.3) obtained from this method of analysis is denoted by $\hat{\Delta}_{LR}$ and the asymptotic variance of this estimate is denoted by avar($\hat{\Delta}_{LR}$). Note that when $K = 2$ this method is identical to maximum likelihood. An expression for this variance is given in Appendix 2 of Armstrong & Sloan (1989) as

$$\text{avar}(\hat{\Delta}_{LR}) = \frac{3}{n}\sum_{i=1}^{2}\left[1 - \left(\sum_{j=1}^{m}\pi_{ij}\right)^3 - \left(\sum_{j=m+1}^{K}\pi_{ij}\right)^3\right]^{-1} \qquad (6.13)$$

which is obtained by considering $\hat{\Delta}_{LR}$ as a special case of $\hat{\Delta}_{ML}$ when $K = 2$ and employing expression (6.5). Note that in Appendix 2 of Armstrong & Sloan (1989) the inverse values are mistakenly printed for the variances, $\text{var}(\hat{\Delta}_{ML})$ and $\text{var}(\hat{\Delta}_{LR})$.

In the case of $K = 3$, $n_1 = n_2 = n$, under the equal-πs-in-group-1 assumption and with a dichotomy made at $m = 1$, the asymptotic variance of $\hat{\Delta}_{LR}$ is

$$\text{avar}(\hat{\Delta}_{LR}) = \frac{3}{n}\left(1.5 - \left[1 - \gamma_{21}^3 - (1-\gamma_{21})^3\right]^{-1}\right). \qquad (6.14)$$

If the dichotomy is made at $m = 2$ the asymptotic variance of $\hat{\Delta}_{LR}$ is

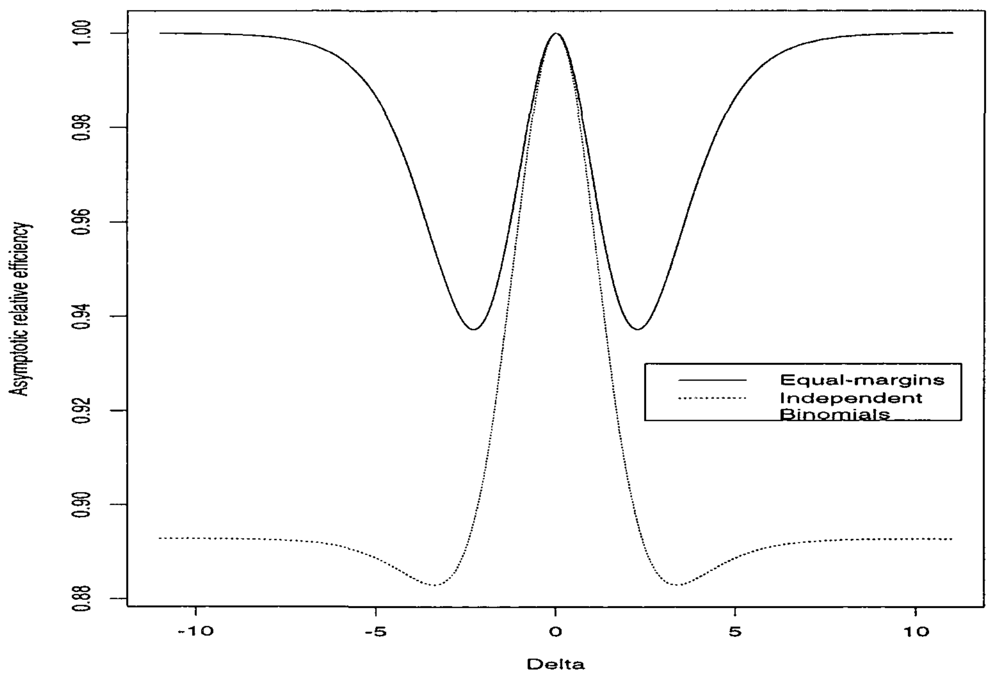
$$\text{avar}(\hat{\Delta}_{LR}) = \frac{3}{n}\left(1.5 - \left[1 - \gamma_{22}^3 - (1-\gamma_{22})^3\right]^{-1}\right). \qquad (6.15)$$

6.3.7 Comparison of GEE methods with Maximum likelihood for a 3-category ordinal response

Figure 6.1 shows the asymptotic relative efficiencies of the estimators of Δ in model (6.3) from two GEE methods of fitting compared with the estimator from maximum likelihood. The two GEE methods are independent binomials and equal-margins GEE. The graph demonstrates that at $\Delta = 0$ both of the GEE methods and maximum likelihood are asymptotically equally efficient to a first order approximation since the value of the ARE is 1 for both GEE estimators.

The asymptotic efficiencies of the GEE estimators relative to the maximum-likelihood estimator decrease as the value of Δ increases or decreases from the origin until a turning point is reached. The ARE value is at a minimum at the turning points, hence inefficiency in the GEE estimates is largest at these points. As the

Figure 6.1: Asymptotic relative efficiency for independent binomials and equal-margins GEE compared with maximum likelihood



value of Δ increases further (decreases further) beyond the turning points the ARE values increase. Hence, the GEE estimators 'recover' efficiency relative to maximum likelihood and this efficiency reaches the same asymptote as $\Delta \to \infty$ or $\Delta \to -\infty$. This asymptote is equal to 1 for the equal-margins GEE estimator and hence this estimator is asymptotically as efficient as the maximum-likelihood estimator as the absolute value of Δ increases. The asymptote is equal to $25/28 = 0.89$ for the independent-binomials estimator.

6.3.8 Comparison of Dichotomisation method with Maximum likelihood for a 3-category ordinal response

Figure 6.2: Asymptotic relative efficiency for dichotomisation and maximum likelihood

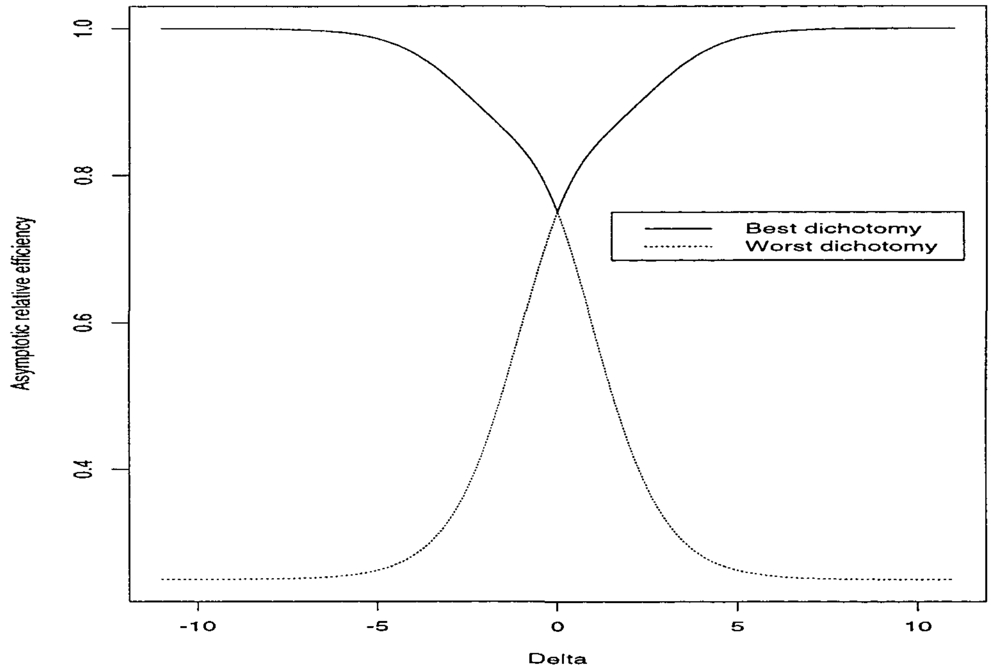


Figure 6.2 shows the asymptotic relative efficiency of the estimator of Δ in model (6.3) from the dichotomisation method of fitting compared with the estimator from maximum likelihood. For Figure 6.2 the dichotomy given by $m = 1$ is chosen for $\Delta \leq 0$ and the dichotomy given by $m = 2$ is chosen for $\Delta \geq 0$ (and is drawn as a solid line). As Armstrong & Sloan (1989) note, the greatest loss of efficiency when using the dichotomisation method will be a the point $\Delta = 0$ where the value of the asymptotic relative efficiency is $\frac{3}{4}$. This is confirmed in Figure 6.2 and it is further noted that the two estimators are asymptotically equally efficient as $\Delta \to \infty$ or $\Delta \to -\infty$. The region of interest in most applied problems is roughly $-5 \leq \Delta \leq 5$. It is worth noting that the ARE is equal to approximately 0.9 at a treament effect of ± 2.5.

Note that if the dichotomy is chosen prior to analysis then the asymptotic efficiency of the dichotomisation estimator relative to the maximum likelihood estimator has an asymptote at 1 in one direction and 0.25 in the other direction. Thus if the dichotomy is chosen prior to analysis the dichotomistaion estimator of Δ could be up to 4 times less efficient than the maximum-likelihood estimator.

6.4 Conclusion

The two examples presented in this chapter provide an empirical comparison of methods of fitting the cumulative logit model. From these examples we may conclude that there are small differences in the estimates obtained from maximum likelihood and independent binomials. In example 2 there are considerable differences in individual parameter estimates obtained from maximum likelihood and either Clayton's GEE or equal-margins GEE.

The asymptotic relative efficiency comparison compares the estimate of a treatment effect obtained from each of two GEE methods with the estimate obtained from maximum likelihood. The comparison indicates that there is a loss of efficiency in parameter estimation when using the GEE methods but that there is a limit to this loss. For the particular case examined where the response has 3 categories, the model has 1 treatment parameter, there are equal group sizes and under the equal-πs-in-group-1 assumption, this loss of efficiency is always less than 12% for the independent binomials estimator and less than 7% for the equal-margins GEE estimator and depends on the size of the treatment effect.

Interestingly, in the region of potentially greatest statistical interest (where the treatment effect is near zero), the loss of efficiency in the independent-binomials estimator compared with the equal-margins GEE estimator is slight. For example the asymptotic relative efficiency between the independent-binomials and equal-margins GEE estimators is approximately 0.95 at a treament effect of ± 2.5. The asymptotic relative efficiency of these two estimators has a minimum of 0.89 which is the asymptote as the treatment effect tends to $\pm\infty$.

Chapter 7

Concluding remarks

The purpose of this chapter is to draw out the main conclusions from the thesis and where appropriate to discuss them in a wider context. Areas are pointed out where further work would be worthwhile. There has been a large amount of published work in recent years on the topic of repeated measures. As the ordinal responses from the BT data investigated in this thesis are longitudinal in nature a discussion of how the approach taken in this thesis to modelling these responses relates to other methods of modelling repeated ordinal responses is needed. Some interesting work, that was published too recently to be included in any detail in this thesis, is briefly discussed.

7.1 Conclusions and suggestions for future work

The tools used to analyse ordinal response data that are examined in this thesis are members of the family of multinomial models defined by the framework in equation (2.2). One of the issues that is discussed in this thesis is a problem of terminology with regard to these models. In §2.5.4 an argument is put forward for reference to these models by use of the probability and link function from which each model is defined. For example, it is suggested that the name cumulative logit model should be used rather than proportional odds model. These models are now well established and a standard terminology needs to be adopted for them. Recently published articles indicate that the scientific community is still divided in the names used to refer to these models. This can be illustrated using the cumulative logit model which in Kenward et al (1994) is called the marginal proportional odds model. Hedeker & Gibbons (1994) refer to it as the logistic regression with the threshold concept, econometricians refer to ordered logit analysis while Miller et al (1993) agree with the name cumulative logit model. Whatever terminology is used, it is necessary to distinguish between different types of terms included in the model. The names 'location terms' and 'scaling terms' are used and as noted in §2.2.3 and §5.5 these terms may represent fixed or random effects.

In Chapter 3 a comparison is performed between the linear model underlying analysis of variance and the continuation-ratio logit model. This comparison is a goodness-of-fit comparison and is quantified using both fitted mean scores and fitted categories. The conclusion is that the multinomial model fits the data better than a linear model. This is not a surprising conclusion to reach given the drawbacks to the linear model, namely the necessity of assigning arbitrary integers to the categories and assuming that the scored observations come from a normal distribution. It is conceivable that for other data the fit of the linear model, as judged by the criteria used in Chapter 3, could be better than the fit of the continuation-ratio logit model. It is also possible that a different choice of arbitrary scores gives a linear model that fits the data better than the continuation-ratio logit model. The results from the linear model and any conclusions drawn from these results are always dependent on the chosen set of scores. The continuation-ratio logit and any other multinomial model for ordinal responses defined by equation (2.2) avoid this dependency.

The continuation-ratio logit model is used in §3.4 for an analysis of the carry-over effect of response. Note that the model includes a subject-specific location term.

By including an interaction between previous score and the threshold parameters, the model is parameterised to allow for the current response to depend in different ways on different immediately preceding responses. This threshold by previous response interaction provides a 'carry-over parameter matrix' which can be tested for significance and investigated for potential patterns of carry-over. When this analysis is applied to the BT data it is found that the carry-over parameter matrix is significant and a pattern emerges across the different experiments considered. However, a cautionary tone is introduced when these results are compared with other exploratory analyses of the same data. These exploratory analyses suggest that there is a subject-specific scaling effect which is not taken account of in the carry-over analysis. It is possible what is being called a carry-over effect of response can alternatively be explained by this subject-specific scaling effect.

These considerations lead to the analyses in Chapter 5 where in addition to a subject-specific location effect, a subject-specific scaling effect is investigated. The need for a subject-specific approach to longitudinal ordinal response data as suggested in Chapter 5 has been recognised at least since Torgerson (1958). The model used to perform these subject-specific analyses is the cumulative logit model. The inclusion of a scaling term in the cumulative logit model is more straightforward than in a continuation-ratio model. This is because the cumulative logit model can be developed by assuming that the ordinal response is a manifestation of an underlying continuous response and that the threshold parameters indicate cut-points on the underlying continuum between different categorical responses. Necessarily the threshold parameters are ordered in this case. However, in a continuation-ratio logit model the threshold parameters are not, in general, ordered. Thus scaling terms in a continuation-ratio model are not intuitively appealing although further research could be done to investigate such a model.

Part of the work for this thesis was the development of a piece of general purpose software to fit cumulative link models with location and scaling terms by maximum likelihood. The software is in the form of macros which can be used with GLIM4 and these have been included in the GLIM Macro Library. The fitting of cumulative link models with a scaling term is not a general feature of any other widely-available software so these macros are a valuable practical tool for analysis. Further work that would make these macros even more useful would be to allow interaction terms between explanatory variables and the cut-point parameters. This would enable the fitting of the carry-over parameter matrix in a cumulative logit model.

The GLIM4 macros are discussed in Chapter 4 along with an alternative method of fitting the model with a scaling term. This alternative method uses independent binomials in a succesive-relaxation scheme. A comparison of the application of the two methods to the BT data is presented in Chapter 5 where it is concluded that while the independent-binomials method cannot give a better fit (in terms of deviance) to the data than maximum-likelihood, for these data the deviance obtained from the former is reasonably close to the deviance obtained for the later. Maximum-likelihood estimation is preferable to independent binomials in that the maximum-likelihood estimates are optimal, but it is concluded that for these data, at least, the independent-binomials method does not give misleading results.

The discussion of scaling terms in this thesis makes it clear that there is more than one possible formulation of the cumulative logit model when a scaling term is included. Two formulations are considered and are loosely referred to as McCullagh's and Kijewski's forms. It is possible to fit both forms of the model with the GLIM4 macros. In the analyses in Chapter 5 Kijewski's form is chosen because of its place in a hierarchy of models. However, the analysis presented in Table 5.7 shows that this form of the model does not necessarily give a better fit to the data. Of 4 experiments analysed, Kijewski's form gives a better fit to the data from three and McCullagh's form gives a better fit to the other. The conclusion is that both forms have a place in the analyst's toolkit. Note that the interpretation of the scaling term is slightly different for the two forms. McCullagh's form of a subject-specific scaling term implies that subjects differ in the dispersion of their underlying response distribution. Kijewski's form of a subject-specific scaling term implies that the dispersion of these underlying distributions are the same for all subjects but that the subjects differ in the way they interpret the scale.

The comparison of maximum-likelihood estimates and independent-binomials estimates is continued in Chapter 6. Here the emphasis is on some theoretical results. The asymptotic variance of the maximum-likelihood estimator of a treatment effect in a simple cumulative logit model is stated. Results are presented for the asymptotic variance of the independent-binomials estimator of the same treatment effect. The asymptotic relative efficiency of the independent-binomials estimator to the maximum-likelihood estimator demonstrates that the loss of efficiency in the former tends to an asymptotic value as the absolute value of the treatment effect increases. It is noted that this asymptote is not the maximum loss of efficiency. In the particular case investigated in Chapter 6 the difference between the maximum loss of efficiency and the asymptotic loss of efficiency is about 1%. The asymptotic loss of efficiency in the independent-binomials estimator tends to 11% as the absolute value of the treatment effect increases. Also of interest is the fact that when there is no treatment effect, the two estimators are asymptotically equally efficient to first order approximations. The broad conclusion is that the loss of efficiency in the independent-binomials estimator might be tolerable if other considerations, such as ease of computation, make it preferable to the maximum-likelihood estimator.

The asymptotic efficiencies of two other estimators relative to the maximum-likelihood estimator are also presented in Chapter 6. Firstly, the estimator from a GEE method using an equal-margins working correlation matrix is found to have a loss of efficiency that is not as great as the independent-binomials estimator. Indeed as the treatment effect increases in absolute value, the former is found to be asymptotically equally efficient to maximum-likelihood. The equal-margins working correlation matrix is block diagonal because for this simple case of the model, the ordinal responses are assumed to be independent. When modelling repeated ordinal responses, some of the off-diagonal blocks can be assigned a correlation structure to represent the correlation between repeated measurements on the same sampling unit. The asymptotic relative efficiency result augurs well for estimates obtained from GEE with a working correlation constructed from the diagonal blocks of the equal-margins correlation matrix with other off-diagonal blocks allowing for intra-unit repeated measures.

The other asymptotic relative efficiency comparison given in Chapter 6 is for an estimator of the treatment effect using a quite different method of analysing the data. This is the dichotomisation method where the ordinal scale is dichotomised by amalgamating the responses above and below a particular cut-point. The resultant data is analysed using binomial logistic regression. This is a technique which has been used historically when binomial logistic regression routines were available in numerous statistical software packages and multinomial modelling methods were not well-developed. The asymptotic relative efficiency comparison shows that if the cut-point is chosen to minimise the variance of the estimtor then the loss in efficiency reaches a maximum of 25% at a zero treatment effect. With the optimal choice of cut-point the estimator is, as the absolute value of the treatment effect increases, asymptotically equally efficient to the maximum-likelihood estimator. However, if a sub-optimal cut-point is chosen for the dichotomisation then the loss of efficiency in the estimator of the treatment effect can be substantial.

7.2 Repeated ordinal responses

The subject-specific analyses in Chapter 5 take the form of fitting a hierarchy of cumulative logit models. Two different methods are employed to fit the models, namely maximum-likelihood and independent-binomials. The hierarchy provides a way of testing the strength of subject-specific location and scaling terms. These terms are initially considered to be fixed effects and when applied to the BT data the conclusion is that both subject-specific location and scaling terms are significant. It is also concluded that there are no further significant subject-specific terms. The analysis with subject-specific location and scaling considered as random effects is also investigated. A subject-specific random location term is found to be significant. The reported precision in estimation of other explanatory terms reduces slightly when the subject-specific location term is taken to be random rather than fixed. This suggests that the fixed effect analysis gives a false sense of accuracy in the estimation of these other explanatory terms. This is partly due to the fact that in the fixed effect analysis the sampling units are effectively taken to be the individual responses. In the random effect analysis the true situation of subjects being the sampling units is considered and there are far fewer subjects than responses. The investigation of the model with a random subject-specific scaling term is developed only insofar as the equations necessary for fitting the model by maximum likelihood are presented. At present these equations have not been implemented. Unfortunately time did not allow the fitting of this model but this remains an important area for further research.

Kenward et al (1994) model repeated ordinal response data in two different ways. They ignore any intra-unit correlation and fit the cumulative logit model using GEE with the true correlation structure for the collapsed binary responses. Secondly, they introduce notation to generalise the Dale model (Dale, 1986) for bivariate ordinal responses to the multivariate case. This approach takes intra-unit correlation into account by increased complexity in the relationship between the model parameters and the joint probabilities that define the likelihood. They find that the GEE approach and the likelihood-based approach give similar results in the case of complete-history data but that there are substantial differences in the case of data missing at random (MAR). The conclusion is that non-likelihood methods (GEE) should be used with caution when data are MAR. This is misleading however as they do not mention the equivalence between the GEE analysis that they use and the maximum-likelihood score equations ignoring intra-unit correlation. That is, GEE in the way that they use it is equivalent to a likelihood-based approach. The results that they find should be summed up by saying that taking intra-unit correlation into account in the model is particularly important when the data are MAR. Work needs to be done to establish which means of taking this intra-unit correlation into account, a more complicated working correlation in GEE or the Dale model as defined by Kenward et al (1994), is better.

Other recent work has investigated the use of GEE to fit the cumulative logit model to repeated measures data when intra-unit correlation is taken into account, *e.g.* Clayton (1992), Miller et al (1993) and Lipsitz et al (1994). The latter paper investigates the performance of what is referred to in this thesis as full GEE (*i.e.* GEE with the true correlation matrix for collapsed binary responses) using a variety of choices for modelling the off-block-diagonal correlation between repeated ordinal responses. They have written a specific SAS macro to implement this analysis. In Chapter 5 of this thesis an analysis of repeated ordinal responses is presented with a novel approach to taking correlation between intra-subject repeated responses into account by including a subject-specific scaling term in the cumulative logit model. The use of a complicated working correlation structure in GEE to analyse this problem needs to be compared with the results from the scaling term approach. It is possible that a parsimonious parameterisation of the working correlation matrix may be found which gives a similarly good fit to the data as found by the scaling-term approach.

Three approaches have been discussed for tackling intra-subject repeated responses and these are:

1. Including location and scaling terms (either fixed or random) in the model.

2. Using the likelihood-based multivariate version of the Dale model.

3. Using GEE with a working correlation matrix designed to account for intra-subject correlation.

The first two of these are examples of a subject-specific approach to the problem of longitudinal data. The estimates of the model parameters are interpreted as the average effect of the covariate on a particular subject (either a particular subject in the data or an average subject from the population of subjects in the data). Zeger et al (1988) distinguish between population-averaged and subject-specific analyses. The third approach listed above, GEE, is an example of a population-averaged model. The parameters from a population-averaged model are interpreted as the average effect of covariates on the population from which the subjects were drawn.

7.3 A Bayesian perspective

The two methods of fitting the cumulative logit model that are investigated in this thesis are maximum-likelihood and GEE. A recent proposal by Chipman & Hamada (1996) to fit cumulative link models with location terms only, is to assume a prior distribution for the location terms and the cut-point parameters and to employ Gibbs sampling to estimate the posterior distributions of these parameters. In maximum-likelihood or GEE fitting, point estimates of the parameters are obtained and in extreme situations these estimates may, albeit justifiably, be infinite. This can cause problems with fitting algorithms used to implement these methods. An advantage to fitting the model by the Gibbs sampling method is that estimation is more stable as infinite parameter estimates are not encountered. However, it is likely that the posterior distribution will depend heavily on the prior distribution for parameters which are located at infinity in maximum-likelihood fitting. The application of Gibbs sampling to the fitting of the cumulative logit model is an interesting development that deserves thorough examination, both into its properties and into comparisons between results from it and from maximum likelihood.

Another reason for further investigation of a Gibbs sampling approach to fitting the cumulative logit model is the extension to deal with random effects. Ten Have & Uttal (1994) demonstrate how Gibbs sampling may be used to integrate a random effect out of a maximum-likelihood fit of a continuation-ratio logit model. The widely and freely available BUGS software (Gilks et al, 1994) will soon offer the capability of fitting the cumulative logit model by use of Gibbs sampling. This should provide a platform for a full investigation of the use of Gibbs sampling to fit this model. It will also allow an investigation into Gibbs sampling as a method of fitting the cumulative logit model with random effects.

References

Agresti, A. (1984) *Analysis of Ordinal Categorical Data.* Wiley, New York

Anderson, J. (1984) Regression and ordered categorical variables. *JRSS*, **B46**, 1-30.

Anderson, J. & Philips, P. (1981) Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics*, **30**, 22-31.

Armstrong, B. & Sloan, M. (1989) Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, **129**, 191–204.

Chipman, H. & Hamada, M. (1996) Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, **38**, 1–10.

Clayton, D. (1992) Repeated ordinal measurements: a generalised estimating equation approach. *MRC Biostatistics Unit technical report*, Cambridge.

Collett, D. (1991) *Modelling Binary Data.* Chapman & Hall, London.

Cox, C. (1995) Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine*, **14**, 1191–1203.

Cox, D.R. (1972) Regression models for life-tables (with discussion). *JRSS*, **B74**, 187-220.

Cox, D.R. & Oakes, D. (1984) *Analysis of Survival Data.* Chapman & Hall, London.

Crouchley, R. (1995) A random-effects model for ordered categorical data. *J. American Statistical Ass.*, **90**, 489–498.

Dale, J.R. (1986) Global cross-ratio models for bivariate, discrte, ordered responses. *Biometrics*, **42**, 909–917.

Ekholm, A. Green, M. & Palmgren, J. (1986) Fitting exponential family nonlinear models in GLIM3.77. *GLIM Newsletter*, **13**, 4–13.

Ekholm, A. & Palmgren, J. (1988) Regression models for an ordinal response are best handled as nonlinear models. *GLIM Newsletter*, **18**, 31–35.

Ekholm, A. & Green, M. (1995) Fitting nonlinear models in GLIM4 using numerical derivatives. *GLIM Newsletter*, **23**, 12–20.

Ezzet, F. & Whitehead, J. (1991) A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine*, **10**, 901–907.

Fisher, R. & Yates, F. (1963) *Statistical Tables.* Oliver & Boyd, Edinburgh.

Francis, B., Green, M. & Payne, C. (Eds) (1993) *The GLIM System: Release 4 Manual.* Oxford University Press, Oxford.

Gilks, W.R., Thomas, A. & Spiegelhalter, D.J. (1994) A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169–178.

Godambe, V.P. & Heyde, C.C. (1987) Quasi-likelihood and optimal estimation. *International Statistical Review*, **55**, 231-244.

Hedeker, D. (1993) MIXOR *Unpublished technical report: Prevention Research Center, School of Public Health.* University of Illinois at Chicago.

Hedeker, D. & Gibbons, R.D. (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–944.

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.

Hutchison, D. (1985) Ordinal variable regression using the McCullagh (proportional odds) model. *GLIM Newsletter*, **9**, 9–17.

Jansen, J. (1990) On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, **39**, 75-84.

Jones, B. & Kenward, M.G. (1989) *Design and Analysis of Cross-over Trials*, Chapman & Hall, London.

Kenward, M.G., Lesaffre, E. & Molenberghs, G. An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, **50**, 945–953.

Kijewski, M.F., Swensson, R.G. & Judy, P.F. (1989) Analysis of rating data from multiple-alternative tasks. *Journal of Mathematical Psychology*, **33**, 403–417.

Läärä, E. & Matthews, J. (1985) The equivalence of two models for ordinal data. *Biometrika*, **72**, 206-207.

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

Lipsitz, S., Kim, K. & Zhao, L. (1994) Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

McCullagh, P. (1978) A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, **65**, 413-415.

McCullagh, P. (1980) Regression models for ordinal data (with discussion). *JRSS*, **B42**, 109-142.

McCullagh, P. (1983) Quasi-likelihood functions. *Ann. Statist.*, **11**, 59–67.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models, 2nd ed.* Chapman & Hall, London.

Miller, M.E., Davis, C.S. & Landis, J.R. (1993) The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics*, **49**, 1033–1044.

Montgomery, D.C. (1984) *Design and Analysis of Experiments, 2nd ed.* Wiley, New York.

Payne, R. *et al* (1993) *GENSTAT 5 Release 3 Reference Manual.* Oxford University Press.

Richards, D.L. (1973) *Telecommunications by speech.* Butterworths, London.

SAS institute, Inc. (1987) *SAS/STAT Users Guide Vol.2 (Version 6, 4th ed.).* Author, Cary, NC.

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1994) *Bayesian Inference using Gibbs Sampling, Version 3.0.* MRC Biostatistics Unit, Cambridge.

Statistical Sciences (1993) *Users Guide, Version 3.2.* Seattle: StatSci, a division of MathSoft, Inc.

Street, A.P. & Street, D.J. (1987) *Combinatorics of Experimental Design.* Oxford University Press.

Ten Have, T. & Uttal, D. (1994) Subject-specific and population-averaged continuation-ratio logit models for multiple discrete time survival profiles. *Applied Statistics*, **43**, 371-384.

Terza, J. (1985) Ordinal probit: A generalization. *Commun. Statist.-Theory. Meth.*, **14(1)**, 1-11.

Thompson, R. & Baker, R.J. (1981) Composite link functions in generalised linear models. *Applied Statistics*, **30**, 125–131.

Torgerson, W.S. (1958) *Theory and Methods of Scaling.* J. Wiley & Sons, New York.

Tutz, G. (1989) Compound regression models for categorical ordinal data. *Biometrical Journal*, **31**, 259-272.

Tutz, G. (1991) Sequential models in categorical regression. *Computational Statistics and Data Analysis*, **11**.

Wang, Y. (1986) Order-dependent parameterization of multinomial distributions. *Scandinavian Journal of Statistics*, **13**, 199-205.

Wedderburn, R.W.M. (1974) Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

Williams, E.J. (1949) Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, **2**, 149–168.

Williams, E.J. (1950) Experimental designs balanced for pairs of residual effects. *Australian Journal of Scientific Research*, **3**, 351–363.

Wolfe, R. (1996) General purpose macros to fit models to an ordinal response. *GLIM Newsletter*, **26**, 20–27.

Wolfe, R. (1996) Macro `ORDINAL`. GLIM4 Macro Library Handbook (*ed.* A.J. Scallan), Numerical Algorithms Group, Oxford.

Zeger, S.L., Liang, K.-Y. & Albert, P.S. (1988) Models for longitudinal data: A generalized estimating approach. *Biometrics*, **44**, 1049–1060.

Appendix A

Tabulations of response in present period against response in previous period for each experiment

A.1 Experiment E136

Table A.1: Previous response against present response (E136)

Previous response	1	2	3	4	5	Total
1	0	5	10	7	4	26
Present 2	5	13	21	21	5	65
response 3	11	14	36	52	8	120
4	6	27	51	77	13	174
5	1	4	9	17	33	65

Table A.2: Previous response against fitted present response (E136)

Previous response	1	2	3	4	5
1	0.39	4.27	11.18	8.01	2.15
Fitted 2	3.89	10.03	19.66	26.14	5.04
present 3	10.24	20.12	35.37	47.92	8.28
response 4	7.84	23.65	48.30	73.59	20.03
5	0.64	4.93	12.49	18.35	27.50

Table A.3: Residuals calculated from [O-E]/sqrt[E] (E136)

Previous response	1	2	3	4	5
1	-0.626	0.351	-0.353	-0.356	1.267
Present 2	0.560	0.939	0.302	-1.005	-0.020
response 3	0.239	-1.365	0.106	0.590	-0.098
4	-0.657	0.689	0.389	0.398	-1.571
5	0.451	-0.418	-0.988	-0.315	1.050

A.2 Experiment E139

Table A.4: Previous response against present response (E139)

Previous response	1	2	3	4	5	Total
1	4	5	3	10	2	24
Present 2	9	14	25	11	1	60
response 3	6	20	34	25	5	90
4	3	13	25	47	6	94
5	1	1	2	5	11	20

Table A.5: Previous response against fitted present response (E139)

Previous response	1	2	3	4	5
1	2.92	5.95	7.18	7.21	0.74
Fitted 2	9.12	12.61	21.74	13.40	2.90
present 3	6.26	20.59	32.21	27.63	4.39
response 4	4.45	13.28	26.35	40.02	9.15
5	0.26	0.58	1.52	9.74	7.82

Table A.6: Residuals calculated from [O-E]/sqrt[E] (E139)

Previous response	1	2	3	4	5
1	0.634	-0.390	-1.560	1.037	1.473
Present 2	-0.039	0.393	0.698	-0.655	-1.116
response 3	-0.104	-0.130	0.315	-0.500	0.289
4	-0.685	-0.077	-0.262	1.103	-1.041
5	1.450	0.561	0.389	-1.518	1.138

A.3 Experiment E140

Table A.7: Previous response against present response (E140)

	Previous response 1	2	3	4	5	Total
1	2	2	6	7	1	18
Present 2	2	8	19	13	8	50
response 3	3	20	20	38	3	84
4	9	14	35	36	10	104
5	1	3	3	10	15	32

Table A.8: Previous response against fitted present response (E140)

	Previous response 1	2	3	4	5
1	1.65	2.29	7.11	6.14	0.79
Fitted 2	3.33	10.63	15.37	17.20	3.48
present 3	4.70	16.47	24.54	31.89	6.56
response 4	6.79	14.81	31.12	37.29	14.32
5	0.52	2.79	4.86	11.48	11.84

Table A.9: Residuals calculated from [O-E]/sqrt[E] (E140)

	Previous response 1	2	3	4	5
1	0.271	-0.194	-0.418	0.346	0.230
Present 2	-0.730	-0.807	0.926	-1.012	2.426
response 3	-0.785	0.870	-0.916	1.082	-1.390
4	0.846	-0.210	0.696	-0.211	-1.142
5	0.667	0.124	-0.845	-0.437	0.917

A.4 Experiment E198

Table A.10: Previous response against present response (E198)

Previous response	1	2	3	4	5	Total
1	3	9	8	7	6	33
Present 2	9	5	14	19	9	56
response 3	6	16	16	14	3	55
4	7	14	19	29	4	73
5	4	7	2	10	16	39

Table A.11: Previous response against fitted present response (E198)

Previous response	1	2	3	4	5
1	3.85	6.45	9.95	8.87	3.88
Fitted 2	7.76	9.88	12.34	17.69	8.60
present 3	7.16	11.31	13.77	16.89	5.88
response 4	8.60	16.90	15.44	21.75	10.08
5	1.62	6.47	7.50	13.80	9.55

Table A.12: Residuals calculated from [O-E]/sqrt[E] (E198)

Previous response	1	2	3	4	5
1	-0.433	1.006	-0.619	-0.627	1.074
Present 2	0.444	-1.552	0.473	0.311	0.135
response 3	-0.433	1.395	0.600	-0.704	-1.188
4	-0.547	-0.705	0.905	1.556	-1.915
5	1.863	0.210	-2.007	-1.023	2.087

A.5 Experiment E199

Table A.13: Previous response against present response (E199)

Previous response	1	2	3	4	5	Total
1	1	2	8	17	6	34
Present 2	3	3	11	22	2	41
response 3	7	15	5	20	4	51
4	16	17	21	33	11	98
5	4	4	2	10	12	32

Table A.14: Previous response against fitted present response (E199)

Previous response	1	2	3	4	5
1	1.60	3.09	10.02	16.37	2.92
Fitted 2	3.55	4.51	8.17	19.97	4.66
present 3	7.34	11.27	8.10	20.28	4.14
response 4	16.05	18.06	18.22	34.63	11.42
5	2.46	4.07	2.49	10.75	11.86

Table A.15: Residuals calculated from [O-E]/sqrt[E] (E199)

Previous response	1	2	3	4	5
1	-0.475	-0.621	-0.637	0.155	1.806
Present 2	-0.294	-0.710	0.989	0.454	-1.232
response 3	-0.125	1.111	-1.090	-0.062	-0.071
4	-0.012	-0.250	0.651	-0.276	-0.125
5	0.982	-0.033	-0.308	-0.229	0.042

A.6 Experiment E200

Table A.16: Previous response against present response (E200)

Previous response	1	2	3	4	5	Total
1	10	6	10	9	5	40
Present 2	6	8	9	15	3	41
response 3	6	11	10	15	6	48
4	12	11	22	37	8	90
5	4	2	4	13	14	37

Table A.17: Previous response against fitted present response (E200)

Previous response	1	2	3	4	5
1	9.61	7.18	10.43	10.74	2.04
Fitted 2	6.79	7.82	8.98	12.32	4.47
present 3	6.85	7.92	11.10	15.95	6.53
response 4	11.53	12.11	18.74	37.93	10.15
5	3.22	2.97	5.74	12.07	12.81

Table A.18: Residuals calculated from [O-E]/sqrt[E] (E200)

Previous response	1	2	3	4	5
1	0.126	-0.440	-0.135	-0.530	2.069
Present 2	-0.303	0.064	0.005	0.765	-0.694
response 3	-0.326	1.094	-0.329	-0.237	-0.209
4	0.139	-0.319	0.753	-0.150	-0.675
5	0.434	-0.565	-0.727	0.266	0.333

A.7 Experiment E211

Table A.19: Previous response against present response (E211)

Previous response	1	2	3	4	5	Total
1	0	5	3	12	2	22
Present 2	4	6	9	17	8	44
response 3	7	11	10	16	4	48
4	8	17	24	57	8	114
5	2	1	1	14	10	28

Table A.20: Previous response against fitted present response (E211)

Previous response	1	2	3	4	5
1	0.87	5.42	4.75	9.25	1.70
Fitted 2	4.29	7.88	6.48	18.37	6.04
present 3	4.69	7.19	10.35	21.00	5.81
response 4	9.37	17.04	22.08	54.52	11.18
5	1.78	2.47	3.34	12.86	7.27

Table A.21: Residuals calculated from [O-E]/sqrt[E] (E211)

Previous response	1	2	3	4	5
1	-0.935	-0.179	-0.804	0.903	0.227
Present 2	-0.138	-0.671	0.989	-0.320	0.797
response 3	1.067	1.421	-0.109	-1.090	-0.750
4	-0.449	-0.009	0.410	0.336	-0.951
5	0.167	-0.937	-1.281	0.318	1.013

A.8 Experiment E212

Table A.22: Previous response against present response (E212)

Previous response	1	2	3	4	5	Total
1	3	4	7	14	3	31
Present 2	6	3	12	14	2	37
response 3	6	13	19	18	7	63
4	8	15	29	36	9	97
5	2	1	1	15	9	28

Table A.23: Previous response against fitted present response (E212)

Previous response	1	2	3	4	5
1	3.29	2.61	9.98	13.18	1.94
Fitted 2	3.79	5.06	10.90	13.93	3.63
present 3	7.20	10.97	18.79	21.31	5.91
response 4	9.51	13.97	23.35	37.18	11.56
5	1.21	3.39	4.98	11.41	6.97

Table A.24: Residuals calculated from [O-E]/sqrt[E] (E212)

Previous response	1	2	3	4	5
1	-0.161	0.861	-0.944	0.226	0.763
Present 2	1.137	-0.915	0.335	0.020	-0.855
response 3	-0.448	0.613	0.048	-0.717	0.447
4	-0.490	0.275	1.169	-0.193	-0.752
5	0.722	-1.299	-1.783	1.064	0.771

A.9 Experiment E213

Table A.25: Previous response against present response (E213)

Previous response	1	2	3	4	5	Total
1	1	8	0	5	6	20
Present 2	3	12	5	21	4	45
response 3	2	5	13	15	8	43
4	8	14	18	47	10	97
5	4	4	9	12	22	51

Table A.26: Previous response against fitted present response (E213)

Previous response	1	2	3	4	5
1	0.89	6.05	1.12	8.44	3.50
Fitted 2	1.98	10.23	7.56	17.15	7.39
present 3	2.37	8.24	9.31	17.44	6.12
response 4	9.75	14.15	18.62	39.99	14.10
5	3.00	4.33	8.39	16.97	18.90

Table A.27: Residuals calculated from [O-E]/sqrt[E] (E213)

Previous response	1	2	3	4	5
1	0.114	0.795	-1.058	-1.185	1.336
Present 2	0.720	0.553	-0.931	0.928	-1.246
response 3	-0.239	-1.127	1.209	-0.584	0.762
4	-0.561	-0.041	-0.144	1.108	-1.091
5	0.575	-0.160	0.210	-1.207	0.713

A.10 Experiment E216

Table A.28: Previous response against present response (E216)

Previous response	1	2	3	4	5	Total
1	34	47	22	15	1	119
Present 2	51	71	60	44	12	238
response 3	18	75	43	33	1	170
4	12	36	46	34	5	133
5	4	6	2	8	4	24

Table A.29: Previous response against fitted present response (E216)

Previous response	1	2	3	4	5
1	28.15	43.97	27.05	18.32	1.50
Fitted 2	50.01	83.35	57.40	41.30	7.12
present 3	26.08	56.09	45.19	35.55	5.84
response 4	13.23	44.88	37.24	31.44	5.82
5	1.53	6.71	6.11	7.39	2.72

Table A.30: Residuals calculated from [O-E]/sqrt[E] (E216)

Previous response	1	2	3	4	5
1	1.102	0.457	-0.972	-0.776	-0.409
Present 2	0.141	-1.353	0.343	0.421	1.830
response 3	-1.583	2.525	-0.327	-0.429	-2.004
4	-0.339	-1.326	1.436	0.457	-0.340
5	2.000	-0.274	-1.662	0.225	0.780

A.11 Experiment E247

Table A.31: Previous response against present response (E247)

Previous response	1	2	3	4	5	Total
1	118	69	25	12	9	233
Present 2	69	122	94	57	20	362
response 3	24	102	102	72	40	340
4	8	45	83	72	70	278
5	13	22	32	72	88	227

Table A.32: Previous response against fitted present response (E247)

Previous response	1	2	3	4	5
1	102.43	73.80	30.06	15.72	10.98
Fitted 2	82.32	122.86	87.07	49.44	19.76
present 3	29.17	96.61	103.12	75.29	38.69
response 4	11.05	48.76	76.60	78.89	65.21
5	7.03	17.96	39.14	65.66	92.36

Table A.33: Residuals calculated from [O-E]/sqrt[E] (E247)

Previous response	1	2	3	4	5
1	1.538	-0.559	-0.924	-0.939	-0.598
Present 2	-1.468	-0.078	0.742	1.076	0.055
response 3	-0.956	0.548	-0.111	-0.379	0.210
4	-0.918	-0.539	0.732	-0.775	0.594
5	2.251	0.953	-1.142	0.782	-0.454

A.12 Experiment E264

Table A.34: Previous response against present response (E264)

Previous response	1	2	3	4	5	Total
1	114	57	19	18	15	223
Present 2	67	92	80	53	23	315
response 3	17	97	101	70	45	330
4	14	44	86	106	75	325
5	7	22	47	80	91	247

Table A.35: Previous response against fitted present response (E264)

Previous response	1	2	3	4	5
1	98.92	60.96	32.66	21.15	9.32
Fitted 2	66.87	92.73	74.94	57.21	22.01
present 3	31.59	87.30	90.27	81.13	45.31
response 4	14.46	54.74	84.46	95.70	74.86
5	7.15	16.27	50.67	71.81	97.50

Table A.36: Residuals calculated from [O-E]/sqrt[E] (E264)

Previous response	1	2	3	4	5
1	1.516	-0.507	-2.390	-0.685	1.859
Present 2	0.015	-0.075	0.584	-0.557	0.211
response 3	-2.596	1.038	1.129	-1.236	-0.047
4	-0.121	-1.452	0.168	1.053	0.017
5	-0.057	1.420	-0.516	0.966	-0.658

A.13 Experiment F064

Table A.37: Previous response against present response (F064)

Previous response	1	2	3	4	5	Total
1	153	42	24	7	5	231
Present 2	45	63	52	13	5	178
response 3	21	48	52	42	6	169
4	8	18	32	23	18	99
5	4	5	10	15	9	43

Table A.38: Previous response against fitted present response (F064)

Previous response	1	2	3	4	5
1	139.24	48.58	31.96	7.11	4.12
Fitted 2	57.22	53.25	45.58	18.10	6.68
present 3	22.71	47.64	50.78	33.87	11.60
response 4	8.82	20.90	28.96	27.75	11.52
5	3.02	5.64	12.72	13.18	9.08

Table A.39: Residuals calculated from [O-E]/sqrt[E] (F064)

Previous response	1	2	3	4	5
1	1.166	-0.944	-1.407	-0.040	0.436
Present 2	-1.615	1.336	0.951	-1.198	-0.651
response 3	-0.358	0.053	0.171	1.397	-1.645
4	-0.274	-0.634	0.564	-0.902	1.907
5	0.565	-0.269	-0.763	0.502	-0.025

A.14 Experiment F065

Table A.40: Previous response against present response (F065)

	Previous response 1	2	3	4	5	Total
1	72	29	8	7	8	124
Present 2	22	43	27	7	4	103
response 3	15	27	37	10	7	96
4	12	3	19	28	7	69
5	2	2	7	15	14	40

Table A.41: Previous response against fitted present response (F065)

	Previous response 1	2	3	4	5
1	73.02	31.15	9.02	5.14	5.69
Fitted 2	23.47	40.01	27.38	7.82	4.62
present 3	12.36	23.80	35.31	16.00	7.44
response 4	9.72	7.63	17.29	25.36	9.65
5	4.44	1.41	9.00	12.68	12.61

Table A.42: Residuals calculated from [O-E]/sqrt[E] (F065)

	Previous response 1	2	3	4	5
1	-0.119	-0.385	-0.339	0.822	0.971
Present 2	-0.302	0.473	-0.072	-0.292	-0.289
response 3	0.752	0.656	0.284	-1.500	-0.161
4	0.732	-1.676	0.411	0.523	-0.852
5	-1.159	0.493	-0.667	0.650	0.391

Appendix B

Listing of the GLIM4 macros for fitting cumulative link models with location and scaling terms

These macros, instructions on how to use them and a GLIM Newsletter article (Wolfe 1996) describing them are all available on the Internet at URL
http://www.maths.soton.ac.uk/rw/r_wolfe.html

```
$sub ORDINAL
!
$pr : '    Author: R. Wolfe'$!
$pr '    Version 1.1 GLIM 4 November 1995'$!
$pr '    Main macros:'$!
$pr '     ORDINAL Fits a cumulative link model to an ordinal response.'$!
$pr '             The following formal arguments must be set:'$!
$pr '          %1   the response variable'$!
$pr '          %2   the rows of the contingency table'$!
$pr '          %3   the categories of the ordinal response'$!
$pr '             The following formal arguments are optional:'$!
$pr '          %4   scalar denoting the form of model for the scale effects'$!
$pr '          %5   vector of intial values for parameter estimates'$!
$pr $!
$pr '             The two lists; '$!
$pr '       model   a list of terms to include as location effects'$!
$pr '     s_model   a list of terms to include as scale effects'$!
$pr '             are available to pass terms into the model.'$!
$pr '    INDI_DAT: A macro that may be used if the data are in'$!
$pr '             individual-level format.'$!
```

```
$pr '             It creates new variables for rows and categories. It'$!
$pr '             expands the response and any variables in model or s_model'$!
$pr '          Formal arguments: %1 individual-level response variable.'$!
$pr '                            %2 name for rows variable.'$!
$pr '                            %3 name for categories variable.'$!
$pr '                            %4 name for expanded response variable.'$!
$pr '    Output: The deviance and degrees of freedom for the model are'$!
$pr '            displayed. Also information about cut-point parameters and'$!
$pr '            interactions (if included) is printed.'$!
$pr '    Example of use:'$!
$pr '            $units 8$'$!
$pr '            $data treat count rows cats$'$!
$pr '            $read'$!
$pr '            1 40 1 1   1 49 1 2   1 19 1 3   1 11 1 4'$!
$pr '            2 31 2 1   2 29 2 2   2 35 2 3   2 25 2 4'$!
$pr '            $factor treat 2$'$!
$pr '            $use ORDINAL count rows cats$'$!
$pr '            $list model=treat$'$!
$pr '            $use ORDINAL$' : $!
!
!---------------------------------------------------------------------------
!
! Important identifiers (all macros).
! T1_ T2_ T3_ : Three temporary macros that are used and reused.
! C_ETA1_ : Used to calculate values for the identifier eta1_.
! C_ETA2_ : Used to calculate values for the identifier eta2_.
! C_TAU_  : Used to calculate values for the identifier tau_.
! C_LP_   : Used in the initialisation to calculate %lp.
! C_L_    : Used to calculate values for the identifiers l1_, l2_, l3_ etc.
! C_DLP_  : Used to calculate values for the identifier dlp_.
!
! Important identifiers (maxj_ and link_ are scalars. Others are all vectors).
! ---------------------------------------------------------------------------
! maxj_ : the number of response categories.
! link_ : indicates which link function is required.
! p_    : the current parameter estimates.
! eta1_ : contribution of cut-points to right-hand side of model.
! eta2_ : contribution of linear terms (from model list) to the same.
! tau_  : contribution of scaling terms (from s_model list) to the same.
! rhs_  : right-hand side of model formula.
! gam_  : fitted values of gamma (cumultaive probabilities).
! pi_   : fitted values of pi (cell probabilities).
```

```
! d1_, d2_, d3_ etc. :  These store the derivatives d(pi)/d(p). They are used
!         to calculate new values for the (internal) model matrices. This is
!         the method of fitting non-linear models described by Ekholm and Green
!         (1995) GLIM Newsletter 'Fitting nonlinear models in GLIM4 using
!         numerical derivatives'. We employ the analytical derivatives for the
!         cumulative link model rather than the general numerical derivatives.
! dgam_, ftau_, dlp_ and l1_, l2_ etc. : Other identifiers used in calculating
!                                        the derivatives.
!
! Other identifiers (scalars).
! -------------------------------
! s_       : Indicates the scaling effect model form.
! len_yv_  : The length of the response variable.
! fit1_    : An indicator to decide whether or not to use macro M_PE_.
! oops_    : Used as an indicator of whether a fault has been encountered.
! wipe_    : Inidcates whether or not to use macro O_DEL_.
!  The following three indicators are used in checking the robustness of the
!  fitting procedure.
! dvlast_  : The deviance at the previous iteration.
! div_     : The change in deviance at the previous iteration.
! osc_     : The number of times the deviance has increased.
!  The following two indicators are used in identifying interactions.
! t_1_     : The position of the first term of the interaction in the list.
! t_2_     : The position of the second term of the interaction in the list.
!
! Other identifiers (vectors).
! -------------------------------
! in_ i_ ind1_ ind2_ : all used in indexing other key vectors.
! j_     : Ordered categories - used for indexing.
! th_    : Cut-points on underlying continuum.
! prop_  : Observed proportions per response category.
!
!
!----------------------------------------------------------------------------
!
$m ORDINAL !
!
! This is the main macro. It is called by the user and it in turn calls the
! other macros. The first thing done is to perform some checks to
!  ensure that the macros don't crash.
!
$ca %z1=(1-%a1)*(1-%a2)*(1-%a3)$!
```

```
$ca %z2=1-%eq(%lin,1)-%eq(%lin,2)*%scf-%eq(%lin,3)-%eq(%lin,6)-%eq(%lin,7)$!
$ca %z3=%gt(%len(model)+%len(s_model),30)$!
$de s_ $nu s_$!
$ar O_SETS_ %4 $sw %a4 O_SETS_$!
$ca s_=s_+%eq(%a4,0) : %z4=%gt(s_,2)+%lt(s_,1)$!
$fault %z1 'The y-variate, ROWS variable and CATS variable must be provided.'$!
$fault %z2 'Incorrect link function - (only g,p and c are available).'$!
$fault %z3 'Too many explanatory variables - 30 is the maximum permissible.'$!
$fault %z4 'Incorrect scaling formulation - (only 1 or 2 are available).'$!
!
! Next various identifiers that are required later in the fitting process are
! defined and calculated.
!
$de fit1_ t_1_ t_2_ oops_ div_ osc_ dvlast_ wipe_ maxj_ len_yv_$!
$nu fit1_ t_1_ t_2_ oops_ div_ osc_ dvlast_ link_ wipe_ maxj_ len_yv_$!
$ca link_=(%lt(%lin,4)+2*%eq(%lin,6)+3*%eq(%lin,7))*(1-%eq(%lin,2)) !
      +link_*%eq(%lin,2)$!
$output 0$!
$ta for %3 by cat_nos$!
$output 6$!
$ca maxj_=cat_nos(%len(cat_nos)) : osc_=0 : len_yv_=%len(%1)$!
$wa off $un len_yv_ $de cut_ p_ in_ i_ i1_ sum_ j_ cat_nos $wa on$!
$yvar %1$!
$va i_ i1_$!
$ca i_=%cu(1) : %z1=maxj_-2 : %z2=7 : %z3=16 : %z4=len_yv_/maxj_$!
$so j_,i_ %3,i_ %2,%3$                ! j_ and i_ are used later in indexing.
$ta the %yv total for %2 into rowsum$  !
$ca sum_=rowsum(%2) $de rowsum$        ! sum_ : total observations per row.
!
! O_METH_ is the user-defined method (used later on). It calculates various
!   identifiers by using the contents of appropriate macros. These macros are
!   now created.
!
$pr (s=T2_) '1' : (s=C_ETA1_) 'p_(1)*' T2_ : (s=C_DLP_) 'l1_'$!
$pr (s=C_LP_) '%lp+p_(1)*d3_*sum_/%exp(%eta)';$!
$pr (s=C_L_) 'l1_(i_)=%1(i_*%eq(j_,1)) : i1_=i_*%eq(j_,2)';$!
$pr (s=C_L_) C_L_ ': l1_(i1_)=-l1_(i_((%cu(1)-1)*%gt(j_,1)))';$!
$ar O_CUTS_ %3 $wh %z1 O_CUTS_$!
$pr (s=C_ETA2_) '' : (s=C_TAU_) '' : (s=PE_INFO) ''$!
$ca %z5=6 : %z6=maxj_-1$!
$pr (s=O_LP_) '1+CUT_' : (s=T1_) '' : (s=DEL_TRM_) 'CUT_ '$!
$ca %z1=%len(model)+%eq(%len(model),-1)$!
```

```
$ar O_TERM_ model %z5 C_ETA2_ OI_ OL_ OM_ $wh %z1 O_TERM_$!
$pr (s=T1_) 'S_'$!
$ca %z5=6 : %z1=%len(s_model)+%eq(%len(s_model),-1)$!
$ar O_TERM_ s_model %z5 C_TAU_ OI_ OL_ OM_ $wh %z1 O_TERM_$!
$ca %z1=2-%a5$!
$ar O_CP1_ %5 : O_CP2_ %3 $sw %z1 O_CP1_ O_CP2_ $tidy O_CP2_$!
$pr :'Checks completed, the model will now be fitted. Please be patient...':$!
!
! The internal macros have all been set up. Now GLIM goes on to fit the model.
!
$ca wipe_=0+wipe_ : %z1=%len(p_)$!
$sw wipe_ O_DEL_$!
$va %z1 in_$!
$ca in_=1 : fit1_=0 : d1_=0 : d2_=0 : %z1=-len_yv_/maxj_$!
$baseline 0 %z1$!
$er p $link 1  $ini O_INI_ $me * O_METH_$!
$us O_METH_ %3 $ti O_METH_ $ca fit1_=1$!
$sc 1.0 $fit #O_LP_-1$!
!
! The model has now been fitted, all unwanted structures are deleted and final
! output is calculated and printed.
!
$ti O_METH_$!
$nu theta_$!
$ca %z1=maxj_-2 : theta_=p_(1)$!
$pr (s=PE_INFO) PE_INFO ; 'Cut-point 1 =     CUT_[1]              = '!
theta_;$!
$wh %z1 O_THETA_$!
$wa off$!
$de T1_ T2_ T3_ C_ETA1_ C_ETA2_ C_TAU_ C_DLP_ v2_ C_L_ O_LP_ d1_ d2_ i_ in_$!
$de d3_ l1_ fit1_ t_1_ t_2_ oops_ div_ osc_ dvlast_ i1_ j_ sum_ len_yv_ maxj_$!
$de wipe_ s_ p_$!
$wa on$!
$pr PE_INFO$!
$$e ! This is the end of macro ORDINAL.
!
!----------------------------------------------------------------------------
!
! A user-defined initialisation (O_INI_) and fitting method (O_METH_) are
! defined to fit the internal model (O_LP_). The macro O_METH_ in turn calls
! the macros M_PE_ (which extracts parameter estimates from the previous fit)
! and M_DERIV_ (which performs some of the calculations necessary for
```

```
! calculating derivates.
!
!
$m O_INI_ $arg C_LP_ d1_ d2_ $ca %lp=0 : %lp=#C_LP_ $de C_LP_$$e!
!
$m (local=dlp_,dgam_,eta1_,eta2_,tau_,rhs_,pi_,gam_,ftau_) O_METH_ !
$sw fit1_ M_PE_$!
$ca eta1_=0 : eta2_=0 : tau_=0 : dlp_=0 : %z2=15-%eq(link_,2)*10$!
$ar C_ETA1_ eta1_ : C_ETA2_ eta2_ : C_TAU_ tau_ : C_L_ dgam_ :  C_DLP_ dlp_$!
$ca eta1_=0+#C_ETA1_ : eta2_=0#C_ETA2_ : tau_=0#C_TAU_$!
$ca rhs_=%if(s_==1,(eta1_-eta2_)/%exp(tau_),eta1_*%exp(tau_)-eta2_) !
  : rhs_=%if(rhs_>%z2,%z2,rhs_) : rhs_=%if(rhs_<-%z2,-%z2,rhs_) !
  : gam_=%eq(link_,1)/(1+%exp(-rhs_))+%eq(link_,2)*%np(rhs_) !
         +%eq(link_,3)*%if(rhs_>2.5,0.999995,1-%exp(-%exp(rhs_))) !
  : dgam_=%eq(link_,1)*gam_*(1-gam_)-%eq(link_,3)*(1-gam_)*%log(1-gam_) !
          +%eq(link_,2)*%exp(-(rhs_**2)/2)/%sqrt(2*%pi)$!
$ca #C_L_ : ftau_=%exp((-s_+3*%eq(s_,2))*tau_) : dlp_=#C_DLP_ !
  : d1_=-dlp_*%if(s_==1,1/%exp(tau_),1) !
  : d2_=l1_*(-p_(1)+%if(s_==1,eta2_,0)) !
  : d3_=dlp_*ftau_ : dlp_=dlp_-l1_  : pi_=gam_ : %z1=maxj_-2$!
$pr (s=T1_) ''$!
$ar M_DERIV_ eta2_ dlp_ ftau_ $wh %z1 M_DERIV_$!
$ca gam_=gam_*%ne(%1,maxj_)+%eq(%1,maxj_)$!
  : pi_(i_)=gam_(i_)-gam_(i_((%cu(1)-1)*%gt(j_,1))) !
  : pi_=%if(pi_<0.000005,0.000005,pi_) !
  : d2_=d2_*ftau_*(s_-3*%eq(s_,2)) !
  : %eta=%log(sum_*pi_) : %z6=2 : %z2=maxj_-1$!
!
! Having calculated updated values for the derivatives the method now proceeds
! to create the matrices for the internal model. One matrix is created for
! every term in the two special lists model and s_model.
!
! M_CUTS_ : creates the matrix for the cut-point parameters.
! MI_ : creates the matrices for interaction terms.
! MM_ : creates the matrix for a matrix term.
! ML_ : creates matrices for all other terms.
!
$ar M_CUTS_ v2_ $wh %z2 M_CUTS_$!
$ca %z6=maxj_-1 : %z1=%z6 : %z2=%len(v2_)$!
$va %z2 cut_$!
$ca cut_=v2_(%gl(len_yv_,%z1)+(%gl(%z1,1)-1)*len_yv_)$de v2_$!
$array cut_ len_yv_,%z1$!
```

```
$ca %z1=%len(model)+%eq(%len(model),-1)$!
$pr (s=T1_) ''$!
$ar O_TERM_ model d1_ * MI_ ML_ MM_   $wh %z1 O_TERM_$!
$ca %z1=%len(s_model)+%eq(%len(s_model),-1)$!
$pr (s=T1_) 'S_'$!
$ar O_TERM_ s_model d2_ * MI_ ML_ MM_ $wh %z1 O_TERM_$!
$$e  ! End of macro O_METH_
!
!-----------------------------------------------------------------------------
!
! The macros in this section are general macros used in setting-up by the
! macro ORDINAL, and in fitting the model by the macro O_METH_.
!
$m O_TERM_ !
! This macro identifies the type of the current term and switches to the
! appropriate macros.
$ca %z2=%len(%1)-%z1+1 : %z1=%z1-1$!
$pr (s=T2_) 'i_' : (s=T3_) *n %1[%z2]$!
$ca %z4=2-%eq(%match(T2_,T3_),1)+%gt(%len(%1[%z2]),len_yv_)$!
$ar %4 %1 %2 %3 : %5 %1 %2 %3 : %6 %1 %2 %3 $sw %z4 %4 %5 %6$$e!
!
$m O_CHK1_ !
! This pair of macros (O_CHK1_ and O_CHK2_) identify which terms make up the
! interaction term. Checks are also performed to ensure that the interaction
! specification is valid.
$ar O_CHK2_ %1 $wh %z8 O_CHK2_$!
$ca %z3=%z3+1 : %z8=1 : oops_=%gt(%z3,%len(%1))$!
$fault oops_ 'Number (a or b) specified in i_a_b is outside range of list.'$$e!
!
$m O_CHK2_ !
$pr (s=T2_) 'i_' *i %z3 '_' *i %z8$!
$ca %z7=%match(T2_,T3_)*2/3 : t_1_=%z3 : t_2_=%z8$!
$ca oops_=%eq(%z7,2)*%eq(%z3,%z8)$!
$fault oops_ 'A term cannot interact with itself.'$!
$ca %z8=(%z8+1)*%ne(%z7,2)*%lt(%z8,%len(%1)) : %z3=%z3*%ne(%z7,2)$!
$sk %z7$$e!
!
$m ML_CONT_ $pr (s=T2_) *n %1[%z2]$$e!
$m ML_FACT_ $pr (s=T2_) '(' *n %1[%z2] '==' *i %z8 ')'$$e!
$m M_NULL_  $pr (s=T2_) '0'$$e!
!
!-----------------------------------------------------------------------------
```

```
!
! The macros in this section are repeatedly used in the fitting procedure by
! the macro O_METH_.
! Note that the macros used to calculate the derivatives d(eta)/d(pi) come
! in four alternative groups.
! The first group (just M_CUTS_) is used for the cut-point parameters.
! The second group (ML_ ML_1_) is used for all terms other than interactions,
!   matrices or the cut-points.
! The third group (MI_ MI_1_ MI_2_ MI_C_F_ MI_F_C_ MI_F_F_) is used for
!   interaction terms.
! The fourth group (MM_ MM_1_) is used for matrix terms.
!
$m M_PE_ !
! This macro perform checks on the deviance to ensure robustness of the
! fitting procedure.
$extract %pe$!
$ca p_(IN_*%cu(1))=%pe(IN_*%cu(IN_)) : oops_=%lt(div_,0)*%gt(%dv,dvlast_)$!
$fault oops_ 'Iterations diverged: try using a better set of intial values'$!
$ca div_=(dvlast_-%dv)*%ne(dvlast_,0) : osc_=osc_+%gt(%dv,dvlast_)$!
$ca oops_=%gt(osc_,2) : dvlast_=%dv$!
$fault oops_ 'Deviance oscillating. Use: dis e :a pe might be infinite'$$e!
!
$m M_DERIV_ !
! This macro calculates the vectors d2_, d3_, d4_ etc.
$ca %z2=maxj_-%z1 : %z3=%z2+2 : %z1=%z1-1$!
$pr (s=T2_) 'd' *i %z3 '_'$!
$ca #T2_=%2*%exp(p_(%z2))*%3$!
$pr (s=T2_) 'l' *i %z2 '_' : (s=T1_) T1_ '-%exp(p_(' *i %z2 '))';$!
$ca %2=%2-#T2_ : d2_=d2_+#T2_*(-p_(1)#T1_+%if(s_==1,%1,0))$$e!
!
$m (local=v1_) M_CUTS_ !
! This is the calculation of d(eta)/d(pi) for
! parameter number %z6 (a cut-point parameter).
$ca %z2=%z2-1 : %z6=%z6+1$!
$pr (s=T1_) 'd' *i %z6 '_'$!
$ca v1_=#T1_*sum_/%exp(%eta)$!
$as %1=%1,v1_$de v1_$$e!
!
$m (local=vv1_) ML_ !
! This macro calculates the internal model matrix for term %1[%z2]
$ca %z5=%lev(%1[%z2]) : %z5=%z5+%eq(%z5,0) : %z4=%z5$!
$ar ML_1_ %1 %2 vv1_$wh %z4 ML_1_$!
```

```
$pr (s=T2_) T1_ *n %1[%z2] '__'$!
$ca %z9=%len(vv1_)$var %z9 #T2_$!
$ca #T2_=vv1_(%gl(len_yv_,%z5)+(%gl(%z5,1)-1)*len_yv_)$!
$array #T2_ len_yv_,%z5$de vv1_$$e!
!
$m (local=v1_) ML_1_ !
! This is the calculation of d(eta)/d(pi) for
! parameter number %z6 from term %1[%z2].
$ca %z4=%z4-1 : %z8=%z5-%z4 : %z9=1+%eq(%ref(%1[%z2]),%z8)+2*%eq(%z5,1)$!
$ar ML_FACT_ %1 : ML_CONT_ %1 $sw %z9 ML_FACT_ M_NULL_ ML_CONT_$!
$ca v1_=#T2_*%2*sum_/%exp(%eta) : %z6=%z6+1$!
$as %3=%3,v1_$de v1_$$e!
!
$m (local=vv1_) MI_ !
! This macro calculates the internal model matrix for interaction term %1[%z2]
$ca %z3=1 : %z8=1$!
$ar O_CHK1_ %1 $wh %z3 O_CHK1_$!
$ca %z5=%lev(%1[t_1_]) : %z5=%z5+%eq(%z5,0) : %z4=%z5$!
$ca %z3=%lev(%1[t_2_]) : %z3=%z3+%eq(%z3,0)$!
$ar MI_1_ %1 %2 vv1_ $wh %z4 MI_1_$!
$pr (s=T2_) T1_ *n %1[%z2] '__';$!
$ca %z5=%z5*%z3 : %z9=%len(vv1_)$!
$va %z9 #T2_$!
$ca #T2_=vv1_(%gl(len_yv_,%z5)+(%gl(%z5,1)-1)*len_yv_)$!
$array #T2_ len_yv_,%z5 $de vv1_$$e!
!
$m MI_1_ !
! This macro provides a second level of looping for macro MI_
$ca %z4=%z4-1 : %z7=%z3$!
$ar MI_2_ %1 %2 %3 $wh %z7 MI_2_$$e!
!
$m (local=v1_) MI_2_ !
! This is the calculation of d(eta)/d(pi) for
! parameter number %z6 from the interaction i_[t_1_]_[t_2_].
$ca %z7=%z7-1 : %z6=%z6+1$!
$ca %z9=1+2*%ne(%ref(%1[t_1_]),%lev(%1[t_1_])-%z4) !
+%ne(%ref(%1[t_2_]),%lev(%1[t_2_])-%z7) !
+4*(%eq(%lev(%1[t_1_]),0)+%eq(%lev(%1[t_2_]),0))$!
$ar MI_C_F_ %1 : MI_F_C_ %1 : MI_F_F_ %1$!
$sw %z9 M_NULL_ M_NULL_ M_NULL_ MI_F_F_ M_NULL_ MI_C_F_ MI_F_C_$!
$ca v1_=#T2_*%2*sum_/%exp(%eta)$!
$as %3=%3,v1_$de v1_$$e!
```

```
!
$m MI_C_F_ ! This macro is exclusively used by MI_
$ca %z9=%lev(%1[t_2_])-%z7$!
$pr (s=T2_) *n %1[t_1_] '*(' *n %1[t_2_] '==' *i %z9 ')'$$e!
!
$m MI_F_C_ ! This macro is exclusively used by MI_
$ca %z8=%lev(%1[t_1_])-%z4$!
$pr (s=T2_) '(' *n %1[t_1_] '==' *i %z8 ')*' *n %1[t_2_]$$e!
!
$m MI_F_F_ ! This macro is exclusively used by MI_
$ca %z9=%lev(%1[t_2_])-%z7 : %z8=%lev(%1[t_1_])-%z4$!
$pr (s=T2_) '(' *n %1[t_1_] '==' *i %z8 ')*(' *n %1[t_2_]'==' *i %z9 ')'$$e!
!
$m (local=vv1_) MM_ !
! This macro calculates the internal model matrix for matrix term %1[%z2]
$ca %z7=%len(%1[%z2]) : %z5=%z7/len_yv_ : %z4=%z5$!
$ar MM_1_ %1 %2 vv1_$wh %z4 MM_1_$!
$pr (s=T2_) T1_ *n %1[%z2] '__'$!
$va %z7 #T2_$!
$ca #T2_=vv1_(%gl(len_yv_,%z5)+(%gl(%z5,1)-1)*len_yv_)$!
$array #T2_ len_yv_,%z5$de vv1_$$e!
!
$m (local=v1_) MM_1_ !
! This is the calculation of d(eta)/d(pi) for
! parameter number %z6 of the matrix term %1[%z2]
$ca %z4=%z4-1 : %z8=%z5-%z4 : %z6=%z6+1$!
$pr (s=T2_) 'new' *i %z6 '_'$!
$ca v1_=#T2_*%2*sum_/%exp(%eta)$!
$as %3=%3,v1_$de v1_$$e!
!
!--------------------------------------------------------------------------
!
! The following set of macros are used once, prior to fitting, to set up the
! internal macros.
! Note that these macros are in four groups.
! The first group (just O_CUTS_) is used for the cut-point parameters.
! The second group (OL_ OL_1_) is used for all terms other than interactions,
!   matrices or the cut-points.
! The third group (OI_ OI_1_ OI_2_ OI_C_F_ OI_F_C_ OI_F_F_) is used for
!   interaction terms.
! The fourth group (OM_ OM_1_) is used for matrix terms.
!
```

```
!
$m O_SETS_ ! Sets the value of the scale identifier s_
$ca s_=%1$$e!
!
$m O_CUTS_ ! Sets up macros C_L_, C_LP_, C_ETA1_, C_DLP_.
$ca %z1=%z1-1 : %z2=%z2-1 : %z3=%z3-1 : %z4=maxj_-%z1 : %z5=%z4-1 : %z6=%z5-1$!
$ca %z7=%z4+1 : %z8=1/3 : %z9=2$!
$ar RESET_C_ * %z2 C_ETA1_ %z8 $sw %z2 RESET_C_$!
$ar RESET_C_ * %z3 C_DLP_ %z9  $sw %z3 RESET_C_$!
$pr (s=C_L_) C_L_ ': l' *i %z5 '_(i_)=%1(i_*%eq(j_,' *i %z5 !
')) : i1_=i_*%eq(j_,' *i %z4 ')';$!
$pr (s=C_L_) C_L_ ': l' *i %z5 '_(i1_)=-l' *i %z5 '_(i_((%cu(1)-1)*%gt(j_,' !
*i %z5 ')))';$!
$pr (s=C_LP_) C_LP_ ': %lp=%lp+p_(' *i %z5 ')*d' *i %z7 '_*sum_/%exp(%eta)';$!
$pr (s=T2_) T2_ '-(' *n %1 '==' *i %z6 ')';$!
$pr (s=C_DLP_) C_DLP_ '+l' *i %z5 '_'$!
$pr (s=C_ETA1_) C_ETA1_ '+%exp(p_(' *i %z5 '))*(' T2_ ')';$$e!
!
$m OL_ ! Used to set up macros O_LP_ and to call OL_1_.
$pr (s=O_LP_) O_LP_ '+' T1_ *n %1[%z2] '__';$!
$pr (s=T2_) T1_ *n %1[%z2] '__'$!
$wa off $de #T2_ $wa on$!
$ca %z3=%lev(%1[%z2]) : %z4=%z3+%eq(%z3,0)$!
$ar OL_1_ %1 %2 %3$!
$wh %z4 OL_1_$$e!
!
$m OL_1_ ! Used to set up macros C_LP_ C_ETA2_ C_TAU_.
$ca %z6=%z6+1 : %2=%2-1 : %z9=1$!
$ar RESET_C_ * %2 %3 %z9 $sw %2 RESET_C_$!
$ca %z4=%z4-1 : %z8=%z3-%z4$!
$ca %z9=%eq(%z3,0)*2+1+%eq(%ref(%1[%z2]),%z8)*%ne(%z3,0)$!
$ar ML_FACT_ %1 : ML_CONT_ %1 $sw %z9 ML_FACT_ M_NULL_ ML_CONT_$!
$pr (s=C_LP_) C_LP_ ':%lp=%lp+' T2_ '*p_(' *i %z6 ')*%1*sum_/%exp(%eta)';$!
$pr (s=%3) %3 '+' T2_ '*p_(' *i %z6 ')';$$e!
!
$m OI_ ! Used to set up macros O_LP_ and call OI_1_.
$ar O_CHK1_ %1$!
$ca %z3=1 : %z8=1$wh %z3 O_CHK1_$!
$pr (s=T3_) 'i_' : (s=T2_) *n %1[t_1_]$!
$ca oops_=%match(T3_,T2_)$!
$fault oops_ 'a,b in i_a_b cannot refer to another interaction term.'$!
$pr (s=T3_) 'i_' : (s=T2_) *n %1[t_2_]$!
```

```
$ca oops_=%match(T3_,T2_)$!
$fault oops_ 'a,b in i_a_b cannot refer to another interaction term.'$!
$pr (s=O_LP_) O_LP_ '+' T1_ *n %1[%z2] '__';$!
$pr (s=T2_) T1_ *n %1[%z2] '__'$!
$wa off $de #T2_ $wa on$!
$ca %z4=%lev(%1[t_1_])+%eq(%lev(%1[t_1_]),0)$!
$ar OI_1_ %1 %2 %3 $wh %z4 OI_1_$$e!
!
$m OI_1_ ! Used to provide another level of looping for OI_2_.
$ca %z4=%z4-1 : %z7=%lev(%1[t_1_])-%z4$!
$ca %z8=%lev(%1[t_2_])+%eq(%lev(%1[t_2_]),0)$!
$ar OI_2_ %1 %2 %3 $wh %z8 OI_2_$$e!
!
$m OI_2_ ! Used to set up macros C_LP_, C_ETA2_, C_TAU_ for interaction terms.
$ca %z6=%z6+1 : %2=%2-1 : %z9=1$!
$ar RESET_C_ * %2 %3 %z9 $sw %2 RESET_C_$!
$ca %z8=%z8-1 : %z3=%lev(%1[t_2_])-%z8$!
$ca %z9=1+2*%ne(%ref(%1[t_1_]),%z7)+%ne(%ref(%1[t_2_]),%z3) !
+4*(%eq(%lev(%1[t_1_]),0)+%eq(%lev(%1[t_2_]),0)) : oops_=%eq(%z9,9)$!
$fault oops_ 'No interaction is possible between two continuous variables'$!
$ar OI_F_F_ %1 : OI_C_F_ %1 : OI_F_C_ %1$!
$sw %z9 M_NULL_ M_NULL_ M_NULL_ OI_F_F_ M_NULL_ OI_C_F_ OI_F_C_$!
$pr (s=C_LP_) C_LP_ ':%lp=%lp+' T2_ '*p_(' *i %z6 ')*%2*sum_/%exp(%eta)';$!
$pr (s=%3) %3 '+' T2_ '*p_(' *i %z6 ')';$$e!
!
$m OI_F_F_ !
$pr (s=T2_) '(' *n %1[t_1_] '==' *i %z7 ')*(' *n %1[t_2_] '=='*i %z3 ')'$!
$pr (s=PE_INFO) PE_INFO 'Parameter ' *i %z6 ' is ' *n %1[t_1_] '[' *i %z7 !
'].' *n %1[t_2_] '[' *i %z3 ']';$$e!
!
$m OI_C_F_ !
$pr (s=T2_) *n %1[t_1_] '*(' *n %1[t_2_] '=='*i %z3 ')'$!
$pr (s=PE_INFO) PE_INFO 'Parameter [' *i %z6 '] is ' *n %1[t_1_] '.' !
*n %1[t_2_] '[' *i %z3 ']';$$e!
!
$m OI_F_C_ !
$pr (s=T2_) '(' *n %1[t_1_] '==' *i %z7 ')*' *n %1[t_2_]$!
$pr (s=PE_INFO) PE_INFO 'Parameter [' *i %z6 '] is ' *n %1[t_1_] '[' *i %z7 !
'].' *n %1[t_2_];$$e!
!
$m OM_ ! Used to set up macro O_LP_ and to call macro OM_1_.
$pr (s=O_LP_) O_LP_ '+' T1_ *n %1[%z2] '__';$!
```

```
$pr (s=T2_) T1_ *n %1[%z2] '__'$!
$wa off $de #T2_ $wa on$!
$ca %z4=%len(%1[%z2]) : %z3=%z4/len_yv_ : %z9=1$!
$ar OM_1_ %1 %2 %3 $wh %z3 OM_1_$$e!
!
$m OM_1_ !  Used to set up macros C_LP_, C_ETA2_, C_TAU_ for matrix terms.
$ca %z6=%z6+1 : %2=%2-1 : %z3=%z3-1 : %z8=%z4/len_yv_-%z3$!
$ar RESET_C_ * %2 %3 %z9$sw %2 RESET_C_$!
$va %z4 index_$!
$ca index_=%gl(%z4/len_yv_,1) : index_=%eq(index_,%z8)$!
$pr (s=T2_) 'new' *i %z6 '_'$!
$pi #T2_ %1[%z2] index_$!
$pr (s=C_LP_) C_LP_ ': %lp=%lp+' T2_ '*p_(' *i %z6 ')*%1*sum_/%exp(%eta)';$!
$pr (s=%3) %3 '+' T2_ '*p_(' *i %z6 ')';$$e!
!
!------------------------------------------------------------------------------
!
! Other macros used by the macro ORDINAL before fitting takes place.
!
$m O_CP1_ !
! This macro sets p_ if the user has provided an initial value vector
$ca p_=%1 : oops_=1-%eq(%z6,%len(%1))$!
$fault oops_ 'Initial values vector is incorrect length.'$$e!
!
$m (local=ind1_,ind2_,th_,prop_) O_CP2_ !
! This macro calculates the initial values (p_) if they have not been supplied
! by the user. All the values of p_ are set to zero except the cut-points;
! these are determined from the observed totals in each response category.
$ca %z2=maxj_-1 : %z3=maxj_-2$!
$va %z2 ind1_ th_ : %z3 ind2_ : %z6 p_$!
$ta the %yv total for %1 into prop_$!
$ca prop_=%cu(prop_) : %z3=prop_(%len(prop_)) : ind1_=%cu(1)$!
$ca th_(ind1_)=%eq(link_,1)*%log(prop_(ind1_)/(%z3-prop_(ind1_)))+%eq(link_,2)!
*%nd(prop_(ind1_)/%z3) +%eq(link_,3)*%log(-%log(1-prop_(ind1_)/%z3))$!
$ca ind2_=%cu(1)+1 : p_(1)=th_(1) : p_(ind2_)=%log(th_(ind2_)-th_(ind2_-1))$
$$e!
!
$m O_DEL_ !
! This is used to remove macros after they are finished with.
! The advantage of this is to maximise the number of identifiers available.
! The disadvantage is that the macros need to be reloaded after use.
$de O_CUTS_ OL_ OL_1_ OI_ OI_1_ OI_2_ OM_ OM_1_ WIPE_MAC INDI_DAT INDI_SW_$!
```

```
$de OI_F_F_ OI_C_F_ OI_F_C O_CP1_ O_CP2_ RESET_C_ INDI_WT_ IND_MAT_ IND_TRM_$!
$de INDI_SC_ IND_FAC_ O_SETS_$!
$e!
!
$m RESET_C_ !
! This splits up incoherently long commands written in macros for internal use.
$ca %2=6*%4 $pr (s=%3) %3 ': %1=%1'$$e!
!
$m O_THETA_ !
! This macro calculates and prints the values for the underlying cut-points
! from the functions used in fitting the model.
$ca %z2=maxj_-%z1 : %z3=%z2-1 : %z1=%z1-1 : %z4=%z2+2$!
$ca theta_=%exp(p_(%z2))+theta_$!
$pr (s=PE_INFO) PE_INFO 'Cut-point ' *i %z2 ' = exp(CUT_[' *i %z2 !
']) + cut-point ' *i %z3 ' = ' theta_;$!
$pr (s=T1_) 'l' *i %z2 '_' : (s=T1_) T1_ ' d' *i %z4 '_'$!
$de #T1_$$e!
!
!------------------------------------------------------------------------------
!
! Macros available to the user
!
$m WIPE_MAC $nu wipe_ $ca wipe_=1$$e!    This sets the indicator to use O_DEL_
!
$m INDI_DAT !
! These macros expand individual level data to the form required by ORDINAL.
$ca %z1=1-%a1 : %z2=1-%a2 : %z3=1-%a3 : %z4=1-%a4$!
$fault %z1 'The response variable must be provided.'$!
$fault %z2 'A name for the ROWS variable must be provided.'$!
$fault %z3 'A name for the CATS variable must be provided.'$!
$fault %z4 'A name for the expanded response variable must be provided.'$!
$nu maxy_ newlen_$!
$ta the %1 largest into maxy_$!
$ca %z1=%len(%1) : newlen_=maxy_*%z1$!
$sl newlen_$!
$ca %3=%gl(maxy_,1) : %2=%gl(%z1,maxy_) : %4=%eq(%1(%2),%3)$!
$ar INDI_WT_ %2 $sw %pwf INDI_WT_$!
$ca %z2=%len(model)+%eq(%len(model),-1)$!
$ca %z8=%len(s_model)+%eq(%len(s_model),-1)$!
$list s_model=s_model $pr (s=T1_) s_model$!
$wa off $list t_list $de s_model $wa on$!
$ar INDI_SW_ model IND_TRM_ IND_MAT_ %2 $wh %z2 INDI_SW_$!
```

```
$wa off $list model=t_list $de t_list $wa on$!
$ca %z2=%ne(%z8,0)$!
$sw %z2 INDI_SC_$!
$ca %z2=%z8$
$ar INDI_SW_ s_model IND_TRM_ IND_MAT_ %2 $wh %z2 INDI_SW_$!
$wa off $list s_model=t_list $de t_list maxy_ newlen_$!
$wa on$$e!
!
$m INDI_WT_ $ca wt_=%pw(%1) $de %pw $we wt_ $$e!   Re-sets the weight vector.
!
$m INDI_SW_ !   Switches to the appropriate macro depending on the term type.
$pr (s=T2_) 'i_' : (s=T3_) *n %1[1]$!
$ca %z2=%z2-1 : %z9=%len(%1[1])$!
$list %1=%1-#T3_$!
$ca %z4=(1-%eq(%match(T2_,T3_),1)+%gt(%z9,%z1))*%lt(%z9,newlen_)$!
$ar %2 %4 $sw %z4 %2 %3$!
$list t_list=t_list,#T3_$$e!
!
$m INDI_SC_ $list s_model=#T1_ : t_list$$e!
!
$m (local=term_) IND_TRM_ ! Used if term is neither interaction nor matrix.
$ca term_=#T3_(%1) : %z5=%lev(#T3_) : %z6=%ref(#T3_)$!
$de #T3_$!
$ca #T3_=term_ : %z7=%ne(%z5,0)$!
$sw %z7 IND_FAC_$!
$ti $$e!
!
$m IND_FAC_ $factor #T3_ %z5(%z6)$$e!    Used by IND_TRM_ if term is a factor.
!
$m (local=ind1_) IND_MAT_! Used if term is a matrix.
$ca %z5=%z9*maxy_$
$va %z5 fact_$
$ca fact_=#T3_(%gl(%z9/%z1,1)+(%z9/%z1)*(%gl(%z9,%z9/%z1*maxy_)-1))$!
$de #T3_$!
$ca #T3_=fact_ : %z5=%z1*maxy_ : %z6=%z9/%z1$!
$array #T3_ %z5,%z6$!
$ti $$e
!
!---------------------------------------------------------------------------
!
$RETURN
```

Appendix C

Frequency distributions for all five categories of the response scale

C.1 Conversation experiments (9 or 10 periods)

Note that for these experiments the data included in the tables come from columns 1-8 of the experiment.

Table C.1: E198

		Frequency							
		0	1	2	3	4	5	6	7	Times used
	1	9	15	6	2	-	-	-	-	33
	2	-	14	12	6	-	-	-	-	56
Category	3	4	11	10	4	3	-	-	-	55
	4	4	5	9	6	8	-	-	-	73
	5	16	2	6	7	1	-	-	-	39

Table C.2: E199

	Frequency 0	1	2	3	4	5	6	7	Times used
1	8	14	10	-	-	-	-	-	34
2	5	16	8	3	-	-	-	-	41
Category 3	6	12	5	7	2	-	-	-	51
4	1	2	7	10	8	4	-	-	98
5	19	4	4	1	3	1	-	-	32

Table C.3: E200

	Frequency 0	1	2	3	4	5	6	7	Times used
1	9	9	11	3	-	-	-	-	40
2	7	12	10	3	-	-	-	-	41
Category 3	5	11	11	5	-	-	-	-	48
4	2	3	6	12	7	1	1	-	90
5	17	5	4	1	4	1	-	-	37

Table C.4: E211

	Frequency 0	1	2	3	4	5	6	7	Times used
1	14	14	4	-	-	-	-	-	22
2	6	11	12	3	-	-	-	-	44
Category 3	5	11	13	1	2	-	-	-	48
4	-	5	1	6	12	7	1	-	114
5	21	3	2	3	3	-	-	-	28

Table C.5: E212

	Frequency								
	0	1	2	3	4	5	6	7	Times used
1	13	10	6	3	-	-	-	-	31
2	5	17	10	-	-	-	-	-	37
Category 3	4	8	10	7	2	-	1	-	63
4	1	2	10	7	7	4	1	-	97
5	14	12	5	-	-	-	1	-	28

Table C.6: E213

	Frequency								
	0	1	2	3	4	5	6	7	Times used
1	15	14	3	-	-	-	-	-	20
2	4	14	11	3	-	-	-	-	45
3	8	11	8	4	1	-	-	-	43
4	3	3	6	7	5	7	1	-	97
5	12	5	6	5	2	1	1	-	51

C.2 Conversation experiments (13 periods)

Note that for these experiments the data included in the tables come from columns 1-12 of the experiment.

Table C.7: E139

		Frequency									
		0	1	2	3	4	5	6	7	>7	Times used
	1	13	3	4	3	1	-	-	-	-	24
	2	5	4	3	5	3	1	3	-	-	60
Category	3	2	2	4	6	-	2	5	2	1	90
	4	1	5	2	5	3	3	2	-	3	94
	5	17	4	1	-	-	1	-	-	1	20

Table C.8: E140

		Frequency									
		0	1	2	3	4	5	6	7	8	Times used
	1	10	11	2	1	-	-	-	-	-	18
	2	2	8	5	7	1	-	-	1	-	50
Category	3	3	2	2	3	7	4	1	1	1	84
	4	1	1	-	7	4	5	3	1	2	104
	5	15	3	1	-	2	1	1	-	1	32

C.3 Conversation experiment (16 periods)

Note that for this experiment the data included in the tables come from columns 1-15 of the experiment.

Table C.9: E136

	Frequency of response category														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Times used
1	13	10	5	2	-	-	-	-	-	-	-	-	-	-	26
2	4	10	5	3	5	2	1	-	-	-	-	-	-	-	65
Category 3	1	4	3	4	8	2	3	3	1	-	1	-	-	-	120
4	1	1	-	5	3	4	3	7	1	2	1	1	1	-	174
5	13	8	2	-	2	1	1	-	1	-	-	-	1	1	65

C.4 Conversation experiment (20 periods)

Note that for this experiment the data included in the tables come from columns 1-9 and 11-19 of the experiment.

Table C.10: E216

	Frequency of response category											
	0	1	2	3	4	5	6	7	8	9	10	Times used
1	4	2	7	7	9	2	3	1	1	-	-	116
2	-	-	2	2	3	2	9	6	11	1	-	225
Category 3	-	1	5	6	6	6	10	1	1	-	-	158
4	-	6	3	12	5	5	3	1	1	-	-	126
5	22	9	3	2	-	-	-	-	-	-	-	21

C.5 Listening experiments

Note that for these experiments the data included in the tables come from columns 1-12 of the experiment.

Table C.11: E247 (5 listening levels)

		Frequency of response category					
		0–4	5–9	10–14	15–19	>20	Times used
	1	3	10	5	5	1	233
	2	-	1	14	5	4	362
Category	3	-	2	9	13	-	340
	4	-	8	12	4	-	278
	5	2	12	9	1	-	227

Table C.12: E264 (5 listening levels)

		Frequency of response category					
		0–4	5–9	10–14	15–19	20–24	Times used
	1	4	9	6	5	-	223
	2	-	3	13	8	-	315
Category	3	-	1	14	7	2	330
	4	-	4	10	8	2	325
	5	2	11	8	2	1	247

Table C.13: F064 (5 listening levels)

		Frequency of response category					
		0–4	5–9	10–14	15–19	20–24	Times used
	1	-	-	2	5	5	231
	2	-	2	2	7	1	178
Category	3	-	1	6	5	-	169
	4	-	9	2	1	-	99
	5	9	2	1	-	-	43

Table C.14: F065 (3 listening levels)

	Frequency of response category						
	0–2	3–5	6–8	9–11	12–14	>14	Times used
1	-	2	2	4	-	4	124
2	-	2	5	4	1	-	103
Category 3	-	1	5	5	1	-	96
4	1	5	4	2	-	-	69
5	4	6	1	1	-	-	40

Appendix D

Further subject-specific analysis of experiment E140

Table D.1: Subject response profiles in E140

	Response				
	1	2	3	4	5
Subject					
1	1	2	5	4	0
2	0	7	4	0	1
3	0	3	0	1	8
4	1	1	2	8	0
5	1	1	1	5	4
6	2	1	1	8	0
7	1	4	4	3	0
8	0	2	4	6	0
9	0	3	4	3	2
10	0	0	0	7	5
11	1	3	4	4	0
12	2	1	5	4	0
13	0	0	3	3	6
14	0	2	7	3	0
15	1	2	3	5	1
16	1	2	0	5	4
17	1	1	6	3	1
18	0	3	4	5	0
19	1	3	2	6	0
20	0	1	8	3	0
21	1	1	5	5	0
22	3	3	3	3	0
23	0	1	5	6	0
24	1	3	4	4	0

Table D.2: Subject-specific contributions to deviance in E140

Subject	Scaling model	Interaction model	Residual description
1	9.29	4.45	-4.84
2	27.35	7.48	-19.87
3	18.11	8.18	-9.93
4	12.55	4.98	-7.57
5	11.59	8.21	-3.38
6	19.24	8.67	-10.57
7	19.06	23.90	+4.84
8	7.32	5.22	-2.10
9	25.62	11.67	-13.96
10	13.49	14.00	+0.51
11	13.14	10.14	-3.00
12	14.32	11.92	-2.40
13	23.39	20.27	-3.12
14	9.00	7.14	-1.86
15	12.52	11.85	-0.67
16	21.64	23.32	+1.68
17	18.27	13.18	-5.09
18	8.53	8.77	+0.24
19	16.13	13.84	-2.29
20	22.09	22.39	+0.30
21	10.88	8.28	-2.60
22	25.75	22.48	-3.27
23	22.34	21.53	-0.81
24	10.47	8.60	-1.87
Total	392.09	300.47	-91.63

Appendix E

Results of fixed effects and random effect modelling of subject location for experiment E198

Table E.1: Fixed effects results

$$\begin{aligned} \text{Deviance} &= 446.1 \\ \text{Degrees of freedom} &= 918 \end{aligned}$$

Term	Parameter estimate	Standard error
Cut-point 1	-10.00	(1.07)
Cut-point 2	-6.36	-
Cut-point 3	-3.35	-
Cut-point 4	-0.41	-
Condition 1	0	-
Condition 2	-0.39	(0.50)
Condition 3	-0.64	(0.50)
Condition 4	-7.29	(0.74)
Condition 5	-0.34	(0.50)
Condition 6	-2.94	(0.54)

Table E.1: Fixed effects results *continued*

Term	Parameter estimate	Standard error
Condition 7	-6.39	(0.69)
Condition 8	-9.26	(0.84)
Subject 1	0	-
Subject 2	0.76	(1.02)
Subject 3	-1.66	(1.02)
Subject 4	-1.90	(1.02)
Subject 5	-0.30	(0.99)
Subject 6	-1.29	(1.01)
Subject 7	-1.33	(1.01)
Subject 8	-0.03	(0.99)
Subject 9	-2.07	(1.02)
Subject 10	-0.02	(0.99)
Subject 11	-2.28	(1.02)
Subject 12	-0.85	(1.00)
Subject 13	-0.63	(0.99)
Subject 14	-3.40	(1.04)
Subject 15	-4.40	(1.10)
Subject 16	-0.72	(0.99)
Subject 17	-1.68	(1.02)
Subject 18	-0.44	(0.99)
Subject 19	0.42	(1.00)
Subject 20	-3.80	(1.06)
Subject 21	-1.69	(1.02)
Subject 22	-2.58	(1.02)
Subject 23	-2.52	(1.02)
Subject 24	-1.22	(1.01)
Subject 25	-0.83	(1.00)
Subject 26	-0.43	(0.99)
Subject 27	1.81	(1.09)
Subject 28	-2.12	(1.02)
Subject 29	1.29	(1.05)
Subject 30	-1.81	(1.02)
Subject 31	-3.39	(1.04)
Subject 32	-2.99	(1.02)

Table E.2: Random effects results

$$\begin{aligned} \text{Deviance} &= 519.7 \\ \text{Degrees of freedom} &= 948 \end{aligned}$$

Term	Parameter estimate	Standard error
Cut-point 1	-7.95	(1.13)
Cut-point 2	-4.64	-
Cut-point 3	-1.84	-
Cut-point 4	0.80	-
Condition 1	0	-
Condition 2	-0.36	(0.75)
Condition 3	-0.58	(0.59)
Condition 4	-6.67	(1.07)
Condition 5	-0.33	(0.90)
Condition 6	-2.68	(0.81)
Condition 7	-5.89	(0.94)
Condition 8	-8.45	(0.98)
Variance term	1.11	(0.34)

Calculation of the intracluster correlation

$$\begin{aligned} \text{residual variance} &= \tfrac{\pi^2}{3} \text{ (assumed for the logistic distribution)} \\ \text{cluster variance} &= (1.110)^2 = 1.232 \\ \text{intracluster correlation} &= \tfrac{1.232}{1.232+(\pi^2/3)} = 0.272 \end{aligned}$$

Appendix F

Variance of the estimator of Δ when the fitting method is GEE with an equal-margins working correlation matrix

The method of fitting the cumulative logit model by GEE with a working correlation based on equal marginal probabilities is discussed in §4.2.3. The calculation of the asymptotic variance of the estimate of a treatment effect under this fitting method is described in §6.3.5.

The explicit expression for the asymptotic variance of $\hat{\Delta}_E$ for the case of $K = 3$, $n = n_1 = n_2$ and when the equal πs in group 1 assumption is in force, is

$$\hat{\Delta}_E = \frac{Z_1/(d^2) - 2Z_2/d + Z_3}{\left(\sum_{j=1}^{2} \gamma_{2j}(1-\gamma_{2j}) - \sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})} - Z_4/d\right)^2}$$

where

$$Z_1 = G_1^2 \left[\frac{1}{6} + \gamma_{21}(1-\gamma_{21})\left(\frac{5}{4} - \sqrt{\frac{\gamma_{21}(1-\gamma_{22})}{\gamma_{22}(1-\gamma_{21})}}\right)\right]$$

$$+G_2^2\left[\frac{1}{6}+\gamma_{22}(1-\gamma_{22})\left(\frac{5}{4}-\sqrt{\frac{\gamma_{21}(1-\gamma_{22})}{\gamma_{22}(1-\gamma_{21})}}\right)\right]$$
$$+2G_1G_2\left[-\frac{1}{12}+\frac{5}{4}\gamma_{21}(1-\gamma_{22})-\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})}\right],$$

$$\begin{aligned} Z_2 &= [G_1\gamma_{21}(1-\gamma_{21})+G_2\gamma_{22}(1-\gamma_{22})]\left(\frac{5}{4}-\sqrt{\frac{\gamma_{21}(1-\gamma_{22})}{\gamma_{22}(1-\gamma_{21})}}\right) \\ &\quad (G_1+G_2)\left[\frac{5}{4}\gamma_{21}(1-\gamma_{22})-\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})}\right], \end{aligned}$$

$$\begin{aligned} Z_3 &= [\gamma_{21}(1-\gamma_{21})+\gamma_{22}(1-\gamma_{22})]\left[\frac{5}{4}-\sqrt{\frac{\gamma_{21}(1-\gamma_{22})}{\gamma_{22}(1-\gamma_{21})}}\right] \\ &\quad +\frac{5}{2}\gamma_{21}(1-\gamma_{22})-2\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})}, \end{aligned}$$

$$Z_4 = G_1\gamma_{21}(1-\gamma_{21})+G_2\gamma_{22}(1-\gamma_{22})-\frac{1}{2}(G_1+G_2)\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})},$$

$$\begin{aligned} d &= \frac{1}{27}+\frac{3}{4}\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})+\frac{2}{9}[\gamma_{21}(1-\gamma_{21})+\gamma_{22}(1-\gamma_{22})] \\ &\quad -\frac{1}{9}\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})}, \end{aligned}$$

$$\begin{aligned} G_1 &= \frac{2}{9}\gamma_{21}(1-\gamma_{21})+\frac{1}{9}\gamma_{22}(1-\gamma_{22})+\frac{3}{4}\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22}) \\ &\quad -\frac{1}{6}\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})} \end{aligned}$$

and

$$\begin{aligned} G_2 &= \frac{1}{9}\gamma_{21}(1-\gamma_{21})+\frac{2}{9}\gamma_{22}(1-\gamma_{22})+\frac{3}{4}\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22}) \\ &\quad -\frac{1}{6}\sqrt{\gamma_{21}\gamma_{22}(1-\gamma_{21})(1-\gamma_{22})}. \end{aligned}$$